MEDINFO 2007

Studies in Health Technology and Informatics

This book series was started in 1990 to promote research conducted under the auspices of the EC programmes' Advanced Informatics in Medicine (AIM) and Biomedical and Health Research (BHR) bioengineering branch. A driving aspect of international health informatics is that telecommunication technology, rehabilitative technology, intelligent home technology and many other components are moving together and form one integrated world of information and communication media. The complete series has been accepted in Medline. Volumes from 2005 onwards are available online.

Volume 129

Recently published in this series

ISSN 0926-9630

MEDINFO 2007

Proceedings of the 12th World Congress on Health (Medical) Informatics

Building Sustainable Health Systems

Part 1

Edited by

Klaus A. Kuhn

University Medical Center, Technische Universität München, Germany

James R. Warren

Department of Computer Science, University of Auckland, New Zealand

and

Tze-Yun Leong

School of Computing, National University of Singapore, Singapore

IOS Press

Amsterdam • Berlin • Oxford • Tokyo • Washington, DC

ISBN 978-1-58603-774-1

Publisher
IOS Press
Nieuwe Hemweg 6B
1013 BG Amsterdam
Netherlands
fax: +31 20 687 0019
e-mail: order@iospress.nl

Distributor in the UK and Ireland
Gazelle Books Services Ltd.
White Cross Mills
Hightown
Lancaster LA1 4XS
United Kingdom
fax: +44 1524 63232
e-mail: sales@gazellebooks.co.uk

Distributor in the USA and Canada
IOS Press, Inc.
4502 Rachael Manor Drive
Fairfax, VA 22032
USA
fax: +1 703 323 3668
e-mail: iosbooks@iospress.com

PRINTED IN THE NETHERLANDS

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Table of Contents

Part 2

Chapter 5. Decision Support and Workflow 827

EDITORIAL

Playing an active role in the 12th World Congress on Medical Informatics has been an honour and a privilege for the Editorial Committee. We closely cooperated with our colleagues and friends of the Scientific Program Committee (SPC) and the Organizing Committee (OC).

Out of 638 submissions, 292 high quality papers were selected for publication in these proceedings by the SPC. Out of 218 poster submissions, 21 posters were nominated for a best poster award; they are included as 1 page contributions in these proceedings. While high quality contributions have been selected by the SPC, the Editorial Committee has focussed on providing proceedings with a high quality of format and presentation which made additional editing necessary. The proceedings are also appearing on CD-ROM. All accepted poster contributions are published on an additional CD-ROM.

The MEDINFO 2007 Proceedings present an excellent overview of a dynamic and quickly growing field, demonstrating methodical and practical progress from around the world. Information science and, specifically, (Bio-) Medical Informatics have become core pillars of foundational and clinical research, of medical care, and of prevention. MEDINFO 2007 in Brisbane will be the leading conference of the year, bringing a worldwide community together.

We have organized the proceedings into twelve chapters, covering topics such as eHealth, Decision Support, Improving Quality, Usability, Sustainability, Genomics, Biomedical Image and Signal Processing, and Education and Training. Within each chapter, the articles are organized according to the conference sessions; the session titles are shown in the table of contents.

The assistance of HISA has been invaluable in all editorial steps, including communication with authors, language editing, and formatting of manuscripts. We owe specific thanks to Joan Edgecumbe, Dale Proposch and Tom Morgan.

The Editorial Committee

Klaus A. Kuhn, Germany

James R. Warren, New Zealand

Tze Yun Leong, Singapore

Disclaimer

While the Medinfo 2007 Editorial Committee has made every effort to ensure that all accepted contributions are published in these two volumes of the MEDINFO 2007 Proceedings, it reserves the right to:

- Edit/alter one or more parts of an original contribution, including its title, author and affiliation listings, as it sees fit in order to comply with publications requirements.

- Withhold the publication of a contribution due to one or more of the following circumstances:
  - failure to meet the final deadline for submission of all original/revised contributions;
  - failure to revise the original contribution in accordance with the instructions of the Scientific Programme Committee, including the advice of the Editorial Committee.

The Editorial Committee is not responsible for the alteration and/or omission of part or parts of an original contribution in cases where a contribution has not fully complied with the submission guidelines and has to be partially or fully re-written and/or reformatted for the final publication.

Preface from the Scientific Program Co-Chairs

The theme of Medinfo2007 is "Building Sustainable Health Systems". Particular foci are health challenges for the developing and developed world, the social and political context of health care, the knowledge that is needed for safe and effective health care, and the difficult task of building and maintaining complex health information systems. Sustainable health information systems are those that can meet today's needs without compromising the needs of future generations. It is a challenge and an opportunity to develop and implement systems that can be improved – not replaced. To achieve this and to build a global knowledge society, we need to seek increased scientific and technological co-operation, and we need to facilitate access to and use of high-quality knowledge and information.

We are pleased to report that more than 900 submissions were made to Medinfo2007. This includes full papers, panels, posters, and workshops. Submissions were made from every region of the world and from many countries, including, among others, Argentina, Australia, Belgium, Brazil, Canada, China, Finland, France, Germany, Great Britain, India, Israel, Italy, Japan, New Zealand, Portugal, Russia, Singapore, Switzerland, and the United States, making this a truly international conference.

The majority of submissions (638) were made in the full paper category. Papers were refereed by members of the biomedical informatics community and final decisions were made by the SPC members at a three day meeting in Chamonix, France in late January 2007. Of the 638 papers submitted, 292 were accepted for presentation and publication (a 46% acceptance rate). 260 were not accepted for publication, but their authors were offered the possibility of presenting their work in the Medinfo2007 poster session, and they were also given the opportunity to create a set of slides that is planned to be part of a continuous slide show throughout the meeting. 86 paper submissions were rejected (13%).

The contributions to Medinfo2007 reflect the breadth and depth of the field of biomedical and health informatics. Papers cover topics in health information systems, knowledge and data management, education, standards, consumer health and human factors, emerging technologies, sustainability, organizational and economic issues, genomics, and image and signal processing.

To recognize the truly outstanding contributions to the conference, there will be "Best Paper" and "Best Poster" awards at the meeting. A student paper competition will also be held. All decisions will be made by an international jury at the meeting itself, and the winners will be announced during the closing session of the Congress.

On behalf of the members of the Medinfo2007 program committee (listed below), we wish to thank all those who contributed to Medinfo2007 by sending in their paper, poster, panel, and workshop contributions. We give special thanks to all those who carefully reviewed the many excellent submissions. We expect that the papers included in this volume will be of great interest to anyone engaged in biomedical and health informatics research and application.

Alexa T. McCray, PhD
Harvard Medical School
Boston, Massachusetts, USA

Heimar Marin, RN, PhD
Federal University of São Paulo
São Paulo, Brazil

Medinfo 2007 Scientific Program Committee

SPC Chairs

Alexa McCray (Co-Chair, US)
Heimar Marin (Co-Chair, Brazil)

SPC Members

Europe

Elske Ammenwerth (Austria)
Ricardo Bellazi (Italy)
Jan van Bemmel (Netherlands)
Patrice Degoulet (France)
Nicos Maglaveras (Greece)
George Mihalas (Romania)

Americas

Dominik Aronsky (US)
Marion Ball (US)
Alex Jadad (Canada)
Alvaro Margolis (Uruguay)
Charles Safran (US)

Asia-Pacific

John Bennett (Tutorial chair, Australia)
HM Goh (Malaysia)
Kazuhiko Ohe (Japan)
Hyeoun-Ae Park (Korea)

Africa

Ben Bondo Monga (Congo)

Middle East

Yuval Shahar (Israel)
Ahmed Al Barrak (Saudi Arabia)

In memory of

Dr Branko Cesnik

Australia

It is with sadness that as we complete this foreword we acknowledge the passing of Branko Cesnik Melbourne on Sunday, 10, June, 2007 after a courageous battle with cancer.

Branko was one of the founders of the Health Informatics Society of Australia, the Asia Pacific Association of Medical Informatics and the Australian College of Health Informatics.

Branko was also a member of the Board of the International Medical Informatics Association (IMIA) as the Vice-President of Membership from 1999 to 2003. He was the lead editor of the Medinfo '98: 9th World Congress Medical Informatics Proceedings (Korea). In addition he served IMIA through a variety of other responsibilities, such as being a member of a Medinfo Scientific Program Committee.

Those who knew him as a colleague will remember his knowledge, level-headedness, spirit of cooperation and sharing, and above all getting things done—no matter what was required or the amount of time required. Those who knew him as a friend will remember his panache, his spirit and his zest for living life on the edge and to the fullest. That zest and spirit accompanied him throughout all of the latter days of his struggle.

His great leadership and vision were instrumental in laying the foundation for health informatics developments “Down Under” that has this year occasioned MEDINFO 2007 to come to Australia.

This page intentionally left blank

Chapter 1.

eHealth

This page intentionally left blank

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Linkcare – Enabling Continuity of Care for the Chronically Ill across Levels and Profession

Marius Mikalsen[a], Ståle Walderhaug[a+b], Per Håkon Meland[a], Ole Martin Winnem[a]

[a] *SINTEF ICT, Norway*
[b] *Medical Informatics and Telemedicine group, Dept for Computer Science, University of Tromsø, Tromsø, Norway*

Abstract

Chronic diseases are increasing rapidly and this phenomenon is becoming a major burden to the health delivery system around the world. A new health care paradigm with focus on chronic treatment and care will actualize the need for interoperable standards based services due to the complexity of care where different health levels and professions are involved. Given the complexity of the domain, we argue the need for a systematic and formal approach to the development of interoperable information systems if there shall be any real support of the cooperating actors. We describe our work on technical interoperability done in the Linkcare project addressing new models of care and technology to support them in the domain of the chronically ill using concrete results from an architecture built using the MAFIIA architecture framework and the UML 2.0 profile for software services, and argue that building formal architectural descriptions on the basis of shared interface descriptions and profiles are an important part of achieving continuity of care based on sustainable health systems.

Keywords:

medical informatics, information systems, systems integration, continuity of patient care

Introduction

The number of patients suffering from chronic diseases has increased rapidly over the recent decades. The disease burden is changing from acute to chronic care, 35.000.000 people died from chronic diseases in 2005, and 60% of all deaths are due to chronic diseases [1].

The new conditions require that we rethink traditional models of care. One prominent aspect of new care models used for chronic conditions is the cooperation of several different stakeholders in the care process [2]. Stakeholders include the patient himself, his relatives, and caregivers.

Additionally, patients are no longer passive consumers of health services, but are instead demanding more control over their own treatment, together with increased responsiveness and improved quality of care services provided by the involved healthcare institutions. Today, healthcare systems are expected to maintain the continuity of care, shared care, and the empowerment of patients in the management process [3].

The new models of care, including increasingly cooperating stakeholders and empowered users, would benefit from interoperable technology to support continuity of care, e.g., shared access to electronic patient records (EPRs). The situation today is that most stakeholders in the care model have their own proprietary systems, with no or limited possibilities for exchanging information with the outside world in a standardized fashion. Development of interoperable healthcare services would make for more efficient work processes and constitute considerable savings. Walker et.al. [4] argue that the value of health care information exchange and interoperability (HIEI) in the United States alone can yield a net value of $77.8 billion per year once fully implemented.

There are obstacles to interoperability though. Iakovidis lists reasons such as organizational and cultural matters, technological gap between healthcare professionals and information science experts, and legal requirements as to why integrating services in healthcare is a challenge [5].

This article presents Linkcare[1], a European Union project, addressing new models of care for the chronically ill and the technology to support them. Three standalone systems in three different countries (Norway, Cyprus and Spain) supporting different aspects of care of the chronically ill provided the outset for the Linkcare project. This is a relevant scenario as cooperation between different stakeholders using different systems is what characterizes care for the chronically ill.

There is a need for a new way of designing information services in healthcare to ensure an interoperability level that supports coordinated management of care.

In this paper we describe our approach to achieve interoperability between the above mentioned systems. We define interoperability as [9]: *"A system's ability to exchange information with other systems and to mutually use the exchanged information".* This ability is fundamental in order to allow stakeholders in the continuity of care process to provide high quality care to the chronically ill. We

1 Linkcare project (eTen, Grant Agreement Number C517435)

will show how we applied the MAFIIA architecture framework [9] together with the UML 2.0 profile for software services [13] to build an architecture enabling sustainable interoperability between systems in this domain.

This article is organized as follows. First, in the Methods section, we describe the methods used to build the architecture, namely the MAFIIA architecture framework, the UML 2.0 profile for software services and the Linkcare process. Second, the Results section presents the results from applying the methodology as a set of MAFIIA viewpoints. In the Discussion section we present the experiences gathered in our work, before making our concluding remarks.

Methods

This section details the background information relevant to the research presented in this paper. We cover the MAFIIA architecture framework, and the IBM UML 2.0 Profile for Software Services, and explain how these were applied by describing the Linkcare process.

MAFIIA

The service architecture that is presented in this document is created using MAFIIA/H (Modelbased Architecture description Framework for Information Integration Abstraction/Healthcare) [9]. MAFIIA is an architectural description framework for software intensive systems with special a focus on Information Integration Systems (IIS).

An architectural description created using MAFIIA consists of a set of views. Viewpoints are used to create a view. The view consists of one or more models that describe and present different aspects related to structure and behaviour for a target system. Five different viewpoints are defined in MAFIIA. They are; *i*) The Context Viewpoint which describes all aspects of the Target System's environment, which is of importance to be able to document all the interfaces between the Target System and its environment, and what the Target System is intended to do in its environment. *ii*) The Requirements Viewpoint, which document all specific requirements related to the Target System. *iii*) The Component Viewpoint which identifies and documents specific physical or logical components. *iv*) The Distribution Viewpoint which describe the logical distribution of software and hardware components. The distribution view shows if some components cannot be separated and if any must be separated. *v*) The Realisation Viewpoint which describes any constraints on how the target system's components should be implemented and deployed into its environment.

In addition to the viewpoints, a MAFIIA description includes description of concerns, system assets and reference architectures. Concerns are related to the documentation of the functional aspects of the target system and its environment. A concern is visible and treated in relation to any view. Concerns are related to functionality. System assets are sources of information that can be used when developing the architecture descriptions. The reference architecture is a high-level, generic architecture which is used as the basis for development of concrete system architectures, and to compare architectures of existing systems to each other.

UML 2.0 Profile for software services

The UML2.0 Profile for Software services is a profile for UML 2.0 which allows for the modeling of services and service oriented architecture solutions. A UML profile extends the expressiveness of UML with domain specific knowledge, in this case software service specification. The use of a service oriented architecture approach in Linkcare followed from the fact that the original system platforms already were build for the web and web services.

We present a subset of the profile, intended to explain the results and discussion sections. The complete profile is explained by Johnston [13], and the following are relevant excerpts. A **Service Partition** represents some logical or physical boundary of the system. Service partitions are used to represent the web, business, and date tiers of a traditional n-tier application. Any owned parts of a partition shall be a **Service Provider.** A Service Provider is a software element that provides one or more services. A service provider has a property that captures information about its location, but the meaning of this is implementation independent. A **Service Gateway** is only available for use on partitions and not directly on service providers. Gateways are used on partitions to expose services. A **Service** is the model element that provides the end-point for service interaction (in web-service terminology). A **Service Consumer** is any component (includes services) that consumes services. A **Service Channel** represents the communication path between two services. Interaction may occur over a channel, but the channel does not represent any particular interaction. In web services, each service denotes the bindings associated with it, and in the modeling profile, you denote binding either on the communication between services or between a service and a consumer.

The Linkcare process

The guiding principles of the Linkcare architectural process was the following. First, it was essential to preserve the heritage application, that is the original systems that the Linkcare services were built on, and minimal change should be introduced upon these systems. Second, based on a market survey performed in the project, the services created should meet concrete needs in the market. Third, it should be easy to compose new services compliant to the architecture.

We applied the MAFIIA workflow, starting with definition of system concerns and assets, before specifying the five architectural views. First, we used the Linkcare heritage systems as the foundation for capturing the core business processes (functionality) that must be described in the architecture and identified all relevant stakeholders in relation to the Linkcare services (described in the context viewpoint). Second, from the use cases, architectural requirements were defined (requirements model). Third, in the component viewpoint, the resulting system informa-

tion model was described, together with models that show how the Linkcare services are composed, the collaboration of services, and the interfaces of the services. Fourth, in the distribution viewpoint the logical distribution of Linkcare components were modelled, and in the realisation viewpoint we investigated the platforms on which the Linkcare systems were running to identity constraints influencing the realisation of the architecture (such as technical platforms and development processes).

To enable information sharing and version management in the process a CSCW tool named eRoom was used. The modelling was done using Rational Software Architect, and the complete architecture was written in Microsoft Word. Three project meetings, and one two day workshop were used to cooperate and decide on the architecture.

Results – viewpoints of the architecture

In the following we introduce core results from the MAFIIA viewpoints modeled using the UML 2.0 profile for software services.

Concerns

One of the main concerns was identified to be interoperability as defined by MAFIIA. Another important concern was security, but it is outside the scope of this paper to address this.

System assets

Relevant standards, a project dictionary, and software profiles and patterns were identified as important assets. We applied Chari et al's [8] framework for weighing the scope and usability of a set of selected standards, and found that CEN ENV 13940 (Health Informatics System of Concepts to support continuity of Care) was the most appropriate standard. The dictionary, an agreed upon reference list of concepts, was created based upon scenario descriptions and story boards from the three heritage systems, and, as far as possible, mapped to the terminology of the chosen standard. The profile chosen was the UML 2.0 profile for software services, and web service patterns such as the publish-subscriber, asynchronous query, command façade, and message bus pattern were used to guide the design.

Reference architecture

The MAFIIA reference architecture divides the target system (in our case, the Linkcare services) into a set of logical tiers, and defines how the system interfaces with the environment. A tier is a logical partitioning of a system where each tier has a unique responsibility. Each tier is loosely coupled to the tiers directly above and below. All components in the Linkcare architecture were linked to the reference architecture in the component viewpoint.

Context viewpoint

The context viewpoint consists of the business aspects model, the environment systems model, and the business to system mapping model.

The business aspect models documents any business related concern that increases the understanding of what problems the target system solves, or what functionality it implements (e.g. stakeholders). Incorporating the results of the Linkcare process (market analysis and studying the Linkcare systems), we identified 17 core stakeholders modeled in a UML use case model/actor hierarchy accompanied by text outlining each stakeholder.

The Environment Systems model documents other technical systems (environment systems) that are involved in the implementation of the business models, or influences the operation of the target system. We identified seven services that constituted the Linkcare services (the target system). These services are modeled as a Service Partition and made available for other services (Service Consumers) through Service Gateways. This was done in order use the gateway construct to control the access to services as the partition in Norway differs from the Spanish partition as different services are provided.

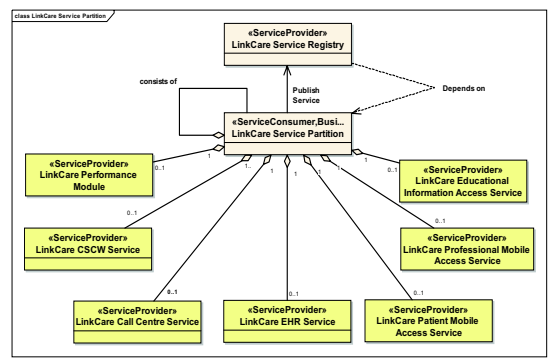


Figure 1 - The Linkcare service partition

Applying the service profile in this way implies that all Linkcare services can be accessed through well-defined interfaces (the gateways). We designed a Linkcare Service Registry that controls the location and availability of the Linkcare services (using the concept of location in the Service provider). This structure enables new Linkcare services to be added, existing services to be modified and relocated without notifying its clients and communication (Service Channels) between Linkcare service partitions can be managed through the gateways.

We used UML component diagrams to stereotype environment systems either as Service Providers or Service Consumers according to the software services profile. UML Sequence diagrams were used to detail the interaction between the Linkcare services and the environment systems.

Requirements viewpoint

Having a clear understanding of what were to be solved following the context viewpoint, we gathered requirements for the target system We organized the requirements into generic requirements for all the services, requirements for each service, and generic interoperability requirements.

Component viewpoint

The component viewpoint consists of the system information model, the system decomposition model, the system collaboration model and the component and interface specification model.

The system information model describes the most important information elements in the Linkcare system. As in the dictionary, the elements in this model were mapped to the CEN standard as far as possible. The model was built using a UML class diagram and the elements were additionally explained using text.

The system decomposition model describes how the target system is divided into different subsystems and components. The Linkcare services were modeled as being part of a Service Partition consisting of one to many Service Providers (the services). The Service Partition publishes services to the Linkcare Service Registry. Interaction with the service registry was detailed in a UML sequence diagram. A UML component diagram was used to stereotype the components to the MAFIIA reference architecture.

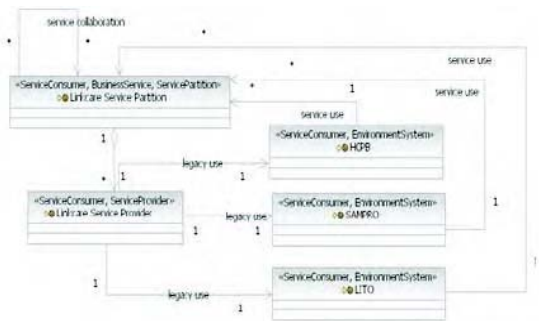


Figure 2 - Mapping to MAFIIA Reference architecture

A UML composite structure diagram were used to model the Service Gateways on a partition, and a UML collaboration diagram were used to model the Service Channels on the gateways.

In the system collaboration model we defined the main interactions between the components in the architecture using a component diagram showing use relations. This was elaborated with text explaining the collaborations, and detailed using UML communication diagrams showing the intra-service communication.

In the component and interface specification model each interface has a specification that shows which methods are available and the input/output parameters of these methods.

Distribution viewpoint

In the distribution viewpoint we modeled the logical distribution of services focusing on what functionality could be distributed together and what should be kept separate (e.g. for business or security reasons). We used a UML deployment diagram and identified four nodes (a node is something that can host software).

The original Linkcare system instances are kept on separate application servers. This is because these are separate systems, maintained by separate organizations in separate countries. The services will cooperate through the Linkcare services, which will help to keep the core business functionality separated, and avoiding building monolithic systems.

Some services, such as the Linkcare patient mobile access service and the Linkcare patient mobile access service are services that typically require rich clients and introduce special requirements on services. Examples of which are extended session logic to handle session interruptions (e.g. when transferring measurement data) and extended security mechanisms.

The Linkcare EHR service is also kept separated. This is because sensitive data typically is transferred using this service, and extra security mechanisms (such as firewalls, log on mechanisms etc) is needed.

The Linkcare service registry service will receive a higher amount of service requests (service lookup) than other services. In order for this not to influence other services, this service is also kept separate.

Realization viewpoint

In the realization viewpoint we considered the technical platforms that the original Linkcare systems were running on. The Norwegian and Cypriot system ran on .NET technology, while the Spanish system is based on Oracle solutions. In the realization view we outlined how web services (one way of implementing services) was to be done on the different platforms. We were focusing on, service description, service implementation, service publishing, discovery and binding, and service invocation and execution.

Discussion

The reason we chose to use MAFIIA in this domain is because it is a proven and documented framework for creating information integration architecture descriptions. An overview of other projects in which MAFIIA is used is provided by Walderhaug et al [9]. In Linkcare we use MAFIIA to ensure that the architectural description created for Linkcare is inline with requirements to architectural descriptions for interoperable systems, and contains the necessary viewpoints. Using the viewpoints and the guidelines that MAFIIA provides, allows us to "think right" when designing the system and ensures that all relevant information, such as important stakeholder, what standards are applied, decomposition of the system and the systems environment and how the system relates to the environment, is sufficiently documented. We believe that the lack of such documentation can considerably hamper the development of interoperable and sustainable systems for continuity of care.

In addition to enforce the presence of essential documentation, the use of architecture frameworks also ensure that the architecture is described on a sufficient level of abstraction, making it more resilient to change in the

domain it is addressing. It is important to note that abstract does not imply informal. On the contrary, MAFIIA suggests using UML as the formalism when describing models. We are inline with Fowler 10], in that there are three ways of using UML; as sketch, as blueprint and as programming languages. We have used UML as blueprints to describe interfaces, and let the implementation of components be hidden in the architecture. This is inline with other current standardisation initiatives see e.g. the Healthcare Service Specification Project [11], a joint effort between the OMG and HL7.

We have used design patterns and the UML 2.0 profile for software services when modelling the architecture. Patterns describe a well proven solution to a recurring design problem. By applying agreed upon patterns (such as the publish-subscriber pattern) we are one step closer to having interoperable systems. A UML profile can be used to further specialize the architecture design. The UML profile defines stereotypes, tagged values and constraints that can be assigned to modelling elements in the design process [13]. We have designed our architecture using the UML 2.0 profile for software services, modelling our services as service partitions with stereotypes like service gateways, communication over service channels, and the notion of service providers and consumers. We experienced several benefits of using this approach. First, since the MAFIIA framework was already assuming the use of UML, applying a UML profile in the framework was straightforward. Second, the profile incorporated software service design terms, allowing us to unambiguously represent the system artefacts in a formal language. Third, the UML Profile assists in communicating concepts in the project. A Service Provider for instance, is a well defined concept and elements stereotyping this concept are bound by the constraints as defined in the profile. We used the profile in the tool it was created for, which is Rational Software architect, so a potential drawback of the profile is that it is not automatically applicable in other tools.

Conclusion and future work

In this paper we have argued for a systematic and formal approach to the development of interoperable information systems using architecture frameworks such as MAFIIA that en-able formal software interface descriptions. We have shown how we applied a UML profile providing us support in design-ing our architecture inline with best practice. Healthcare enter-prises could impose vendors to apply the same design patterns in their system design, thus preparing the ground for interoper-able healthcare enterprises. Consequently, we are currently continuing our work on profiles by creating domains specific healthcare profiles that incorporate information from standards (such as CEN) in support of developing interoperable healthcare systems (see e.g., [12]).

References

[1] Beaglehole R. (ed.). Preventing chronic diseases: a vital investment: WHO global report. World Health Organisation, ISBN: 92 4 156300 1. 2005.

[2] Winnem OM, Walderhaug S. Distributed, role based, guideline based decision support, Proceedings of E-he@lth in Common Europe, Springer pp. 101-109, ISBN 83-915141-2-9. 2002.

[3] Tattersall R. The expert patient: a new approach to chronic disease management for the twenty-first century. Clin Med 2002; 2(3): 227-229

[4] Walker J, Pan E, Johnston D, Adler-Milstein J,. Bates DW, Middleton B. The value of health care information exchange and interoperability. Health Affairs Web Exclusive 2005 (Jan.19): W5-10 W5-18.

[5] Iakovidis I. Towards a Health Telematics Infrastructure in the European Union, In Information technology strategies from US and the European union: transferring research to practice for healthcare improvement, Amsterdam, IOS Press, 2000.

[6] Patient Mobility in the European Union - Learning from experience: World Health Organization 2006, 2006.

[7] MDA Guide Version 1.0.1 http://www.omg.org/cgi-bin/doc?omg/03-06-01

[8] Chari K, Seshadri S. Demystifying Integration. Communications of the ACM, vol. 47, pp. 58-63, 2004.

[9] Walderhaug S, Stav, E, Tomassen SL, Røstad L, Moe NB MAFIIA - an Architectural Description Framework: Experience from the Health Care Domain. In: Konstantas D, Bourrières JP, Léonard M, Boudjlida N. (eds.): Interoperability of Enterprise Software and Applications. Springer, 2005; 43-54

[10] Fowler M. UML Distilled, 3 ed: Addison Wesley, 2004.

[11] The Healthcare Service Specification Project, http://hssp.wikispaces.com/

[12] Walderhaug S, Mikalsen M, Hartvigsen G, Aagedal J. Improving Systems Interoperability with Model-driven Software Development for Healthcare. In Proceedings of Medinfo 2007.

[13] Johnston S. UML 2.0 Profile for Software Services. http://www-128.ibm.com/developerworks/rational/library/05/419_soa/

Address for correspondence

Marius Mikalsen,
SINTEF ICT, NO-7465 Trondheim, Norway.
marius.mikalsen@sintef.no

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Sustainable Ubiquitous Home Health Care – Architectural Considerations and First Practical Experiences

Michael Marschollek, Klaus-H. Wolf, Oliver-J. Bott, Mirko Geisler, Maik Plischke, Wolfram Ludwig, Andreas Hornberger, Reinhold Haux

Institute for Medical Informatics, Technical University Carolo-Wilhelmina of Braunschweig, Germany

Abstract

Despite the abundance of past home care projects and the maturity of the technologies used, there is no widespread dissemination as yet. The absence of accepted standards and thus interoperability and the inadequate integration into transinstitutional health information systems (tHIS) are perceived as key factors. Based on the respective literature and previous experiences in home care projects we propose an architectural model for home care as part of a transinstitutional health information system using the HL7 clinical document architecture (CDA) as well as the HL7 Arden Syntax for Medical Logic Systems. In two short case studies we describe the practical realization of the architecture as well as first experiences. Our work can be regarded as a first step towards an interoperable – and in our view sustainable – home care architecture based on a prominent document standard from the health information system domain.

Keywords:

Home care, information system architecture, standards, sensors.

Introduction

Telemedicine in general and telemedical systems for home care have been a key interest in medicine and medical informatics in the last two decades. Many research projects and funding programs have been conducted and some of the solutions developed have been successfully introduced in medical care and are commercially available. On the one hand there is a considerable advancement and maturity of the underlying technologies and methods, e.g. computer systems, ubiquitous computing, sensor systems, signal analysis, networks and health information systems, and on the other hand there is first proof that specific patient groups stand to benefit from tele home care [1]. Despite this there is no widespread use of these technologies as yet. The reasons for this lack of sustainability are manifold and it is often argued that, apart from technological barriers, evidence for cost benefit is sparse for tele home care systems [2], that profound changes in the work processes of care providers are required and that physicians lack the resources to interpret huge amounts of sensor data recorded in home settings [3]. In particular, the lack of utilization of standards such as the HL7 Clinical Document Architecture (CDA) [4] or prEN 13606 [5] and the inadequate integration into regional health information system infrastructures, if existent, are regarded as key factors [6, 7]. Only few examples of the use of standards in home care can be found in literature. Van der Linden et al. present PropeR, a re-usable, modular EHR system based on open source components such as the openEHR standard, and report on its use for multidisciplinary care of stroke patients [8, 9], but make no statement on the representation of sensor data. In [10] the processes in home care are described in detail and a framework for home care cooperation is proposed. The authors use the XML standard for messaging between different health information systems.

Concentrating on the use of standards from the health information system domain, we aim:

- to elucidate the requirements for sustainability in home care systems (aim 1),
- to propose a functional, adaptive and modular (and in our view sustainable) architectural model (aim 2), and
- to report on two current projects of the authors and their status of implementation (aim 3).

The research questions addressed are the following:

Q1: What are the basic requirements for sustainable home care architectures with respect to the usage of standards?

Q1.1: What are the categories of data processed in home care, what are the basic paths of information flow, and who are the participants?

Q1.2: What are suitable forms of medical data representation for further use in transinstitutional health information systems?

Q2: What is an architectural model suitable to meet these requirements and how can it be implemented?

Materials and methods

Based on our previous experiences in building home care systems and with regard to the respective literature we summarize the basic architectural requirements for a sustainable, standards-based home care system (first section in *Results*). We then propose an architectural model along

with the strategy for its practical realization in the following section, employing the *Three-layer Graph-based meta model (3LGM²)*-methodology for statical modeling of health information systems [11, 12]. In the final two sections in *Results* we report on two case studies using our architectural model and present their current status of implementation. In the *Discussion* we balance advantages and disadvantages of our architecture and finally conclude with an outlook on our future work.

Results

Requirements for sustainable home care architectures

In home care, mostly two different categories of data are gathered. Firstly, there are sensor data that are inherently a very heterogenous group, with sensors that measure just one parameter once a day, e.g. a sphygomanometer for blood pressure measurements, to multi-sensor devices that continuously record multiple channels and transfer a data stream to a computer system, e.g. a 12-lead ECG. Different persons with individually different diseases or functional disabilities will need customized sensor device compilations that fit their situation best. Therefore there is an urgent need for device interoperability standards that allow to build self-configurating 'plug-and-play' sensor sets with known features [13]. The ISO/IEEE 11073 medical device standard is a candidate [14], though at present hardly used outside intensive care units. Apart from medical sensors, context sensors that provide valuable information for the interpretation of the medical data also should be considered, e.g. accelerometers for the classification of activities of daily life (ADL) [15].

A second category of data are data actively provided by the persons themselves, from relatives and care givers. These may either be structured, e.g. in a standardized questionnaire for well-being, and therefore can be represented with reference to code systems such as SNOMED CT [16], or unstructured such as e.g. in free text in emails or spoken comments. The semantic integration of the latter into existing health information systems remains a largely unsolved problem. The addressees of home care not only are a source but also a target of information, because feedback is a necessary precondition for promoting self-management and patient empowerment.

When multiple sensors are used in home care, huge amounts of data will be recorded every day. There is an irrevocable need for intelligent *processing* of these data, which may be done already within the sensor device, on an additional mobile device worn by the patient, or on a computer at the patient's home, probably all three in form of cascading preprocessing procedure. Data fusion is necessary both on technical and content level, and conclusions should be drawn considering all available data from a patient, including those stored in her or his electronic health record (EHR). As there are different groups of recipients of information – the patients, their care givers and their physicians – with different informational needs and levels of 'health literacy', there should also be different, configurable forms of data aggregation and presentation to fit user needs.

In order to achieve semantic interoperability the heterogenous data gathered from multiple sources have to be represented in a consistent form, so that they can be used in transinstitutional health information systems. A prominent standard for clinical documents is the HL7 CDA, which is designed primarily for clinical documents such as discharge letters or diagnostic findings, and not for huge amounts of sensor readings.

An architectural model for standards-based home care

We propose an architectural model for a sensor-based home care system that makes use of the HL7 CDA (version 2), for data representation and includes a decision support infrastructure based on the HL7 Arden Syntax for Medical Logic Systems and its Medical Logic Modules (MLMs) [17].

Figure 1 shows the general component model of our architecture. The patient or respectively the person is the primary source of data in the home care environment, but relatives and care givers may also provide valuable data on her or him, e.g. in the form of questionnaires or observations. Thus these are also part of the architecture. All data gathered are transformed into CDA documents and then transferred to a home server with a personal electronic health record (pEHR) system. Attached to this system is a decision support system that automatically retrieves stored CDA documents, analyzes them and in turn produces synopses or reports in CDA format or as messages, e.g. email alarms to a health care provider. The processing of the data takes place in two stages: firstly, the huge amounts of sensor data are preprocessed by employing temporal abstraction methods, intelligent filters and machine-learned classifiers to aggregated data units in CDA format. Secondly, the data are analyzed along with other CDA documents on the patient from 'external EHR' systems, i.e. other components of the transinstitutional health information system (cmp. Figure 1), with predefined MLMs. The MLMs contain medical knowledge coded in Arden Syntax to extract medically relevant information that in turn can be used e.g. to trigger alarms. The model architecture also contains two loops of feedback (cmp. Figure 1, dotted arrow) to the patient: an inner loop with machine-generated reports made available via an interface, e.g. the home TV set, and an outer loop, where the feedback is posted by a health professional, e.g. a General Practitioner, on her or his institutional system and then transmitted via the transinstitutional communication infrastructure to the personal EHR at home and then sent to the interface.

Figure 2 shows the 3LGM² model of the architecture described, with all entities, functions and relationships subdivided into three distinct architectural layers: the domain layer and the logical and physical tool layer.

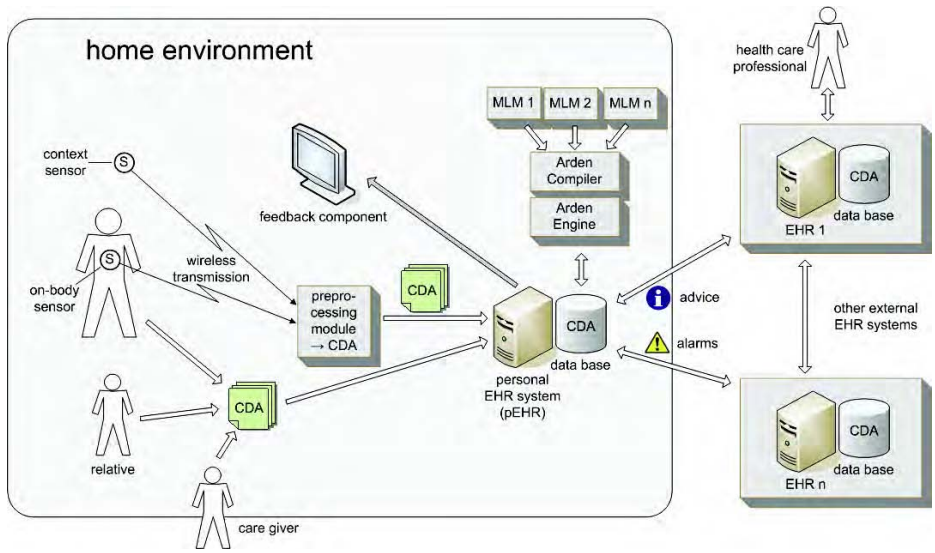


Figure 1 - Component model of the proposed architecture based on the HL7 clinical document architecture standard (CDA). pEHR = personal electronic health record; MLM = Medical Logic Module

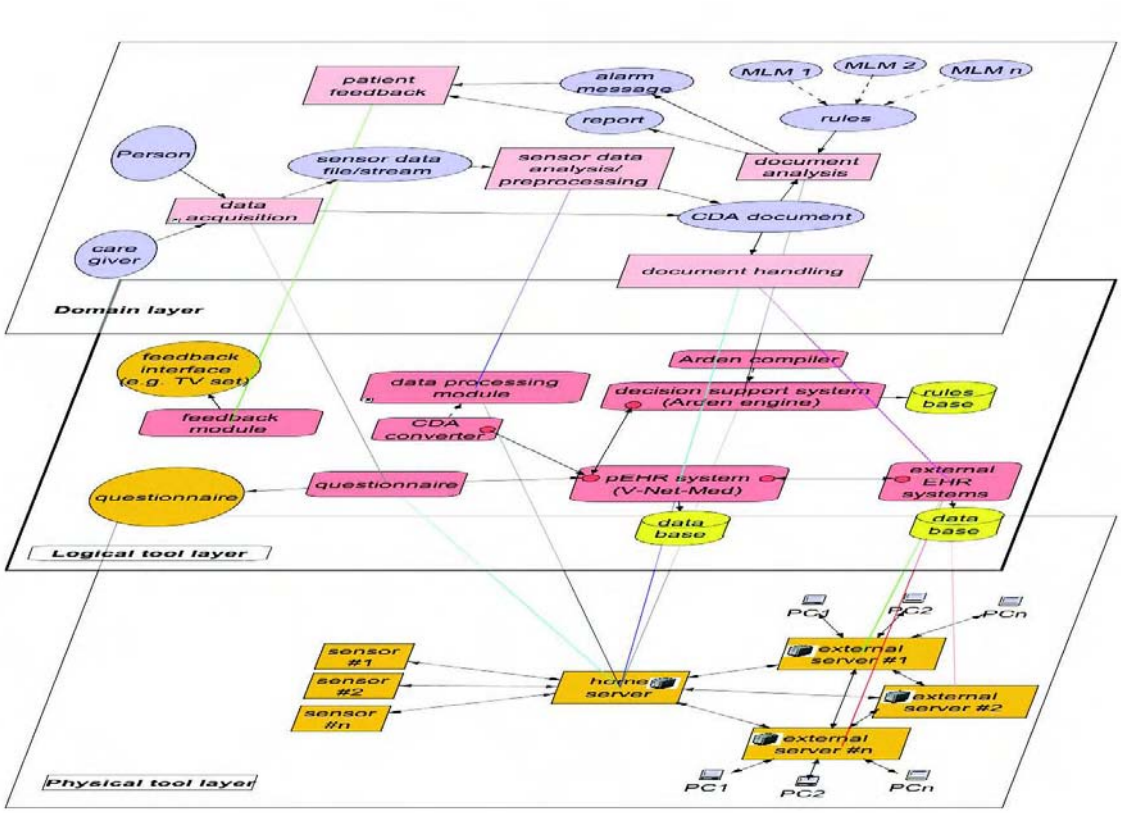


Figure 2 - Three-layer graph-based meta model (3LGM²) of the proposed home care information system. pEHR = personal electronic health record; CDA = clinical document architecture; MLM = Medical Logic Module

Case study 1: Laboratory Installation

In our health-enabling technologies lab, a complete apartment equipped with different sensor technologies, we have installed a prototype system with a personal EHR system developed in our institute [18]. We experiment with different wireless as well as cabled sensors, e.g. triaxial accelerometers, that transfer real-time data to the system which then are analyzed and aggregated into CDA level 3 documents (cmp. Figure 1). In addition to the sensor data, data on the persons' well-being are recorded with computerized questionnaires and stored subsequently, currently in CDA level 2 because we use a non-standardized questionnaire that is not listed in a coding system. We also use an Arden compiler that was constructed to retrieve CDA documents from the EHR system, to analyze them and to generate email messages, e.g. to a health care professional. This compiler was developed and tested in a home care project providing telemedical support for two patients having received a heart-and-lung transplantation [19]. A feedback component is included in the system in the form of a digital TV set with a media portal software. It features a "health" plugin currently showing a frequently updated graph summarizing a person's activity level that is recorded with a wearable wireless accelerometer unit.

Case study 2: The LASS project

The *LASS* project – a German acronym for self-sufficient and self-determined life in old age – aims at assessing the use of health-enabling technologies (e.g. sensor systems for fall detection) to improve both the quality of life and the efficiency of care for elderly people in their home environment. As part of this project, we currently establish an architectural infrastructure as decribed above in Figures 1 and 2.

In a case study [20] we recorded annotated data from five elderly persons over a period of two weeks with a multi-sensor device (*Sensewear Pro 2, Bodymedia, Pittburgh*) and – employing methods of our previous research on this topic [15] – constructed machine-learned classifier algorithms for the classification of activities of daily life and body position that are used by our data preprocessing module. The data from the multi-sensor device are subsequently transformed into a CDA level 3 document that, once stored in the personal health record, can be analyzed by the Arden engine described in the previous section. The specific aim is to identify elderly peoples' activity profiles that provide valuable information on the persons' behavior as well as context information for other medical data recorded in home care. We also installed specially prepared, easy-to-use PC systems in four of the elderly persons' homes and tested daily questionnaires on well-being, the acceptance of which was good. The answers were stored in CDA level 2 documents as stated above.

Discussion

The importance of using standards for information interchange in the health care domain has often been stated and has partly been realized in institutional health information systems. For transinstitutional health information systems – and home care can be considered as part of these – they are vital in order to achieve interoperability and thus sustainability, in this context denoting that such a system meets the requirements for future use. Yet in the past often proprietary document 'standards' or formats were used in this field and still are. In addition to this, no generally approved form of representation for sensor data exists.

Our approach aims at bridging the gap between multiple data types encountered in home care by persistently using a prominent clinical document standard – the HL7 CDA. An obvious advantage of this approach is that such documents can easily be used within existing transinstitutional health information systems. Furthermore their semantics are clear and thus they can be used along with other clinical documents by decision support systems, as demonstrated in our architecture. The installation of a personal in-home EHR system with a decision support component has – apart from reducing data transfer costs – the benefit that decisions can be made in a timely manner and that the data recorded are kept in a private area within a secure system. So the patient may decide which information to pass on. A security solution based on a ticket-based electronic consent system for granting scalable access to the information stored in the personal EHR system has been developed by some of the authors and is part of the system used [18].

Limitations

The HL7 CDA standard is not primarily designed for the representation of sensor data, and the XML files tend to grow large. Therefore intelligent preprocessing is needed, bearing an inherent but unavoidable loss of information. Parts of our prototype architecture have only been tested in our health-enabling technologies lab so far. Further evaluation is necessary.

Apart from this, there is still a lack of analyses regarding the costs and – measurable or perceived – benefits of such a home care system, both for the patients respectively users and the health care professionals. Some studies imply a reduction of costs as the duration of hospital stays is reduced [1], other studies show an increased overall cost as the technology for monitoring and feedback is expensive [21]. Our future work will include a cost-benefit evaluation of the proposed system.

Conclusion

We have proposed an architectural model for a sensor-based home care system that makes use of two prominent standards from the domain of health information systems, the HL7 CDA (version 2) for semantically accessible data representation and the HL7 Arden Syntax for Medical Logic Systems for medical knowledge representation and decision support. We believe that the use of standards is an important precondition for building an interoperable and therefore sustainable home care system [22].

Our future work will be directed towards the evaluation of all parts of our prototype system in a field study outside lab

conditions. Furthermore we will develop and implement further preprocessing modules and transformation components for several sensor devices as well as MLMs with the help of our medical partners. Besides this, we will integrate care givers into our system, both laypersons and professionals.

Acknowledgments

The authors thank the following persons for their support in conducting case study 2: Ines Schapiewski and Elin Schriever (German Red Cross Braunschweig), Hartmut Dybowski and Rainer Schubert (City of Braunschweig) and Hubertus Meyer zu Schwabedissen (Medical Center Braunschweig).

References

[1] Cleland JG, Louis AA, Rigby AS, Janssens U, Balk AH. Noninvasive home telemonitoring for patients with heart failure at high risk of recurrent admission and death: the Trans-European Network-Home-Care Management System (TEN-HMS) study. J Am Coll Cardiol 2005;45(10): 1654-64.

[2] Whitten PS, Mair FS, Haycox A, May CR, Williams TL, Hellmich S. Systematic review of cost effectiveness studies of telemedicine interventions. Bmj 2002;324(7351):1434-7.

[3] Lamothe L, Fortin JP, Labbe F, Gagnon MP, Messikh D. Impacts of telehomecare on patients, providers, and organizations. Telemed J E Health 2006;12(3):363-9.

[4] Dolin RH, Alschuler L, Boyer S, Beebe C, Behlen FM, Biron PV, et al. HL7 Clinical Document Architecture, Release 2. J Am Med Inform Assoc 2006;13(1):30-9.

[5] CEN/TC251. prEN 13606-1: Health informatics — Electronic health record communication — Part 1: Reference model (CEN FV and ISO DIS Draft). 2006 [cited 21.09.2006]; Available from: http://www.centc251.org/WGI/N-documents/WG1_N06-16a_votingreport_ENQ_13606-4dispositionv1.2.doc

[6] Bott OJ. Health information systems: between shared care and body area networks. Methods Inf Med 2006;45 Suppl 1:53-4.

[7] Koch S. Meeting the Challenges - the Role of Medical Informatics in an Ageing Society. Stud Health Technol Inform 2006;124:25-31.

[8] van der Linden H, Grimson J, Tange H, Talmon J, Hasman A. Archetypes: the PropeR way. Medinfo 2004;11 (Pt 2):1110-4.

[9] van der Linden H, Boers G, Tange H, Talmon J, Hasman A. PropeR: a multi disciplinary EPR system. Int J Med Inform 2003;70(2-3):149-60.

[10] Bricon-Souf N, Anceaux F, Bennani N, Dufresne E, Watbled L. A distributed coordination platform for home care: analysis, framework and prototype. Int J Med Inform 2005;74(10):809-25.

[11] Winter A, Brigl B, Wendt T. Modeling hospital information systems. Part 1: The revised three-layer graph-based meta model 3LGM2. Methods Inf Med 2003;42(5):544-51.

[12] Wendt T, Haber A, Brigl B, Winter A. Modeling Hospital Information Systems (Part 2): using the 3LGM2 tool for modeling patient record management. Methods Inf Med 2004;43(3):256-67.

[13] Zywietz C. Communication and interoperability for serial comparison in continuous health care--the new challenges. Stud Health Technol Inform 2004;108:172-80.

[14] Yao J, Warren S. Applying the ISO/IEEE 11073 standards to wearable home health monitoring systems. J Clin Monit Comput 2005;19(6):427-36.

[15] Marschollek M, Wolf KH, Plischke M, Haux R. Classification of activities of daily life from long-term realistic multi-sensor data. In: IEEE Health Pervasive Systems (HPS06); 2006 June 29th; Lyon: IEEE; 2006. p. 30-4.

[16] College of American Pathologists. SNOMED CT. [Webpage] [cited 2006 Nov 16th]; Available from: http://www.snomed.org/snomedct/index.html

[17] Health Level Seven (HL7). HL7 Standards. [Webpage] [cited 2006 Nov 11th]; Available from: http://www.hl7.org/

[18] Bergmann J, Bott OJ, Pretschner DP, Haux R. An e-consent-based shared EHR system architecture for integrated healthcare networks. Int J Med Inform 2006;76(2-3):130-6.

[19] Geisler M, Bott OJ, Tegtbur U, Bergmann J, Pretschner DP. Zur Integration einer auf der Arden-Syntax basierenden Entscheidungsunterstützungskomponente in eine Telematikplattform. Beispielhafte Umsetzung im Rahmen der telemedizinischen Nachbetreuung von Patienten mit Herz-/Kreislauferkrankungen. german medical science [Meeting Abstract] [cited 2006 Nov 11th]; Available from: http://www.egms.de/de/meetings/gmds2006/06gmds281.shtml

[20] Marschollek M, Ludwig W, Schapiewski I, Schriever E, Schubert R, Dybowski H, et al. Multimodal home monitoring of elderly people – first results from the LASS study. In: IEEE First International Workshop on Smart Homes for Tele-Health (SmarTel07); 2007; Niagara Falls, Canada (May 21-23): IEEE; 2007.

[21] Bott OJ, Hoffmann I, Bergmann J, Kosche P, von Ahn C, Mattfeld DC, et al. Simulation based cost-benefit analysis of a telemedical system for closed-loop insulin pump therapy of diabetes. Stud Health Technol Inform 2006;124:435-40.

[22] Haux R. Individualization, globalization and health - about sustainable information technologies and the aim of medical informatics. Int J Med Inform 2006;75(12):795-808.

Address for correspondence

Dr. Michael Marschollek, MD MSc
Institute for Medical Informatics
Technical University Carolo-Wilhelmina
Muehlenpfordtstrasse 23
D-38106 Braunschweig, Germany
+49 531 391 9504;
m.marschollek@mi.tu-bs.de

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Do Physicians Take Action on High Risk Family History Information Provided by Patients Outside of a Clinic Visit?

Lynn A. Volk, MHS[a], Maria Staroselsky[a], Lisa P. Newmark[a], Hannah Pham[a], Alexis Tumolo[a,b], Deborah H. Williams, MHA[a,b], Ruslana Tsurikova[a,b], Jeffrey Schnipper, MD, MPH[b,d] Jonathan Wald, MD[c,d], David W. Bates, MD, MSc[a,b,d]

[a]*Clinical and Quality Analysis, Information Systems, Partners HealthCare System, Wellesley, MA, USA*
[b]*Division of General Medicine, Department of Medicine, Brigham & Women's Hospital, Boston, MA, USA*
[c]*Clinical Information Research & Development, Information Systems, Partners HealthCare System, Wellesley, MA, USA*
[d]*Harvard Medical School, Boston, MA, USA*

Abstract

Clinically relevant family history information is frequently missing or not readily available in electronic health records. Improving the availability of family history information is important for optimum care of many patients. Family history information on five conditions was collected in a survey from 163 primary care patients. Overall, 53% of patients had no family history information in the electronic health record (EHR) either on the patient's problem list or within a templated family history note. New information provided by patients resulted in an increase in the patient's risk level for 32% of patients with a positive family history of breast cancer, 40% for coronary artery disease, 50% for colon cancer, 74% for diabetes, and 95% each for osteoporosis and glaucoma. Informing physicians of new family history information outside of a clinic visit through an electronic clinical message and note in the EHR was not sufficient to achieve recommended follow-up care. Better tools need to be developed to facilitate the collection of family history information and to support clinical decision-making and action.

Keywords:

family history, electronic medical records, clinical decision-making

Introduction

Family history data often goes unrecorded or is available only as uncoded free text in the electronic health record (EHR).[1,2,3] Patient provided family history information may provide incremental value if one can learn of increased risks that may influence care management.[1,4]

Although sophisticated family history assessment is most expertly done by those trained in this area, much can be learned about a patient´s risk from whether a relative with a condition is a primary or secondary relative, how many relatives have the condition, and the approximate age at which they were diagnosed.[5,6] Clinical decision support can aid physicians in identifying patients "at risk" due to family history and in determining appropriate follow-up and monitoring.[7,8]

In this study, we assessed the availability of coded or structured family history information in EHRs and the extent to which solicited input from patients would identify those at higher risk as compared to risk assessed using information readily available in their EHR alone. We also assessed whether physicians took action to change care management based on new family history information provided by a patient outside of a clinic visit.

Methods

Study setting

Partners HealthCare System is an integrated healthcare delivery system in the Boston, Massachusetts, United States (U.S.) area that includes two major academic medical centers, several community hospitals, and an extended physician network. A comprehensive EHR is in use at the majority of ambulatory care practices. This EHR includes the ability to capture provider notes, results of lab and diagnostic tests, problem lists, medication lists, and automatically generates patient-specific reminders for health maintenance items. However, at the time of this study, input of structured or coded family history information was not supported.

Patients who choose to enroll are able to electronically communicate with their physician practice through a secure web-based patient portal called *Patient Gateway.* At the time of the study, features of *Patient Gateway* included patients being able to view their EHR medication and allergy lists, and request prescription renewals, appointments, and referrals.

Recruitment

Patients were recruited from a single Partners-affiliated ambulatory primary care practice in the suburbs of Boston. A packet was sent through postal mail with a cover letter, consent form, opt-out card, privacy notice, and a return envelope to send back the consent form.

Once consent was received, a paper-based survey that asked about medications, health maintenance status, and family history was populated with patient–specific data and sent to consenting patients. A two-step process was required to distribute the survey because of the specific medical information included that could not be generated without patient consent. The Partners Human Research Committee approved the study and the survey was administered from November 2003 – Feburary 2004.

Up to two reminders were sent to non-responders to solicit consent forms and also to complete the survey once they had consented. Survey data on medications and health maintenance are reported elsewhere.[9,10]

Surveys

Patients were asked to provide family history information in a structured format that required specifying the number of family members by degree of relationship and age category that had been diagnosed with a condition. In addition to requesting information on six specific conditions (glaucoma, osteoporosis, coronary artery disease, diabetes mellitus, breast cancer, and colon cancer), patients could complete the form for other conditions for which they had a positive family history.

Patients were then instructed to return the surveys to researchers who provided a copy of the survey back to patients. Patients were informed that information provided in their surveys would be shared with their primary care physicians and may become part of their medical record.

Surveys were tracked and analyzed using Microsoft Access. Once a survey was received, researchers created a note within the patient's electronic health record with Partners Health Information Services approval. This note reported on the information provided by the patient and clearly identified the source of the information as obtained from the patient through a research survey. Physicians were sent a clinical message within 90 days through the EHR messaging system that informed them of the new note in the patient's record.

Analysis

Patient provided family history information was compared with information readily available in the EHR. Readily available was defined as family history information recorded on a patient's problem list or within a note template containing a family and social history section.

To evaluate physician action in response to the family history note, electronic health records were reviewed for visit, medication, and procedure information consistent with recommendations made in these clinical areas by local experts, based on guidelines from the Risk Management Foundation of the Harvard Medical Institutions, Inc., the U.S. Preventive Services Task Force, American Heart Association, National Cholesterol Education Program, American Diabetes Association, and National Osteoporosis Foundation.[11-19]

P-values were calculated for comparisons between respondents and non-respondents using a chi-square test for proportions and Wilcoxon rank-sum test for continuous variables.

Results

Among 1098 patients solicited, 189 patients consented and a family history survey was completed by 163 patients. Respondents and non-respondents were similar in the proportion that were female (66% and 67%, respectively), and their average number of medical problems (4.8 and 4.4, respectively) and medications (3.5 for both) listed in the EHR. Respondents were significantly older than non-respondents on average (50.6 years and 46.8 years, p=0.001).

New information identifying higher risk profile

Fifty-three percent (53%) of the respondents had no family history information in the electronic health record either on the patient´s problem list or within a templated family history note.

For glaucoma, osteoporosis, and colon cancer over 90% of the patients determined to have a positive family history were identified based solely on new information provided by the patients in the survey. (Table 1) A family history of diabetes, breast cancer, or coronary artery disease was more frequently available within the EHR, although more than 80% of patients with a positive family history for these conditions were still identified through the survey alone.

Table 1 - Sources of positive family history information and changes in patient risk level

Condition	Source of Positive Family History		Change in Risk Level with Info from Survey Alone		
	EHR and/or Survey N(%)	Survey Alone N(%)	Reduced N(%)	Did Not Change N(%)	Increased N(%)
Glaucoma	20	19 (95%)	0 (0%)	1 (5.3%)	18 (94.7%)
Osteoporosis	22	20 (91%)	0 (0%)	1 (5.0%)	19 (95.0%)
Diabetes	51	42 (82%)	1 (2.4%)	10 (23.8%)	31 (73.8%)
Breast Cancer	47	40 (85%)	0 (0%)	27 (67.5%)	13 (32.5%)
Colon Cancer	68	66 (97%)	0 (0%)	33 (50.0%)	33 (50.0%)
Coronary Artery Disease	82	67 (82%)	0 (0%)	40 (59.7%)	27 (40.3%)

Table 2 - Recommended actions occurring following clinical message indicating new family history information

Condition & Recommendations (N=total patients with increased risk level due to family history information from survey alone)	Patients Meeting Criteria for Physician Action N	Patients for Whom Recommended Action Occurred After Clinical Message was Sent N (%)
Glaucoma (N=18)		
visit w/in specified time	18	14 (78%)
consider IOP	16	1 (6%)
Osteoporosis (N=19)		
visit w/in specified time	13	6 (46%)
DEXA scan	12	5 (42%)
Diabetes (N=31)		
visit w/in specified time	31	23 (74%)
Consider fasting glucose or HbA1c	31	26 (84%)
Breast Cancer (N=13)		
visit w/in specified time	13	5 (38%)
clinical breast exam	13	3 (23%)
mammogram	13	8 (62%)
referral to genetic counselor*	13	0 (0%)
Colon Cancer (N=33)		
visit w/in specified time	32	13 (41%)
colonoscopy	9	2 (22%)
genetics referral for HNPCC*	1	0 (0%)
Coronary Artery Disease (N=27)		
visit w/in specified time	27	13 (48%)
lipid profile	27	19 (70%)
cholesterol lowering med	11	3 (27%)
blood pressure lowering med	7	3 (43%)
aspirin advice/use	27	5 (18%)
smoking cessation counseling	0	n/a**
exercise/nutrition counseling/referral	22	8 (36%)
consider plasma homocysteine or C-reactive protein	27	2 (7%)

*genetic tests may not have been documented in accessible EHR data
**n/a = not applicable

New information provided by patients led to an increased risk level for 50% of patients with a positive family history of colon cancer, and the majority of patients with a positive family history for diabetes (74%), osteoporosis (95%), and glaucoma (95%). Increases in risk were found for 32% of patients with a positive family history of breast cancer and 40% with coronary artery disease.

Physician action in response to clinical messages

Overall, of the 21 actions physicians should have taken across all conditions in response to new family history information, 16 of the actions were carried out less than 50% of the time. (Table 2) However, for the smoking cessation counseling action, no patients met the criteria of being a current smoker so no action was required. Although patients at higher risk of glaucoma had visits with their physicians more than three quarters of the time after clinical messages were sent, only 1 out of 16 patients (6%) had a documented intraocular pressure measurement (IOP). For osteoporosis, fewer had visits (46%) than did the glaucoma patients, but 42% did receive a dual x-ray absorptiometry (DEXA) scan within the specified time after the clinical message.

In follow-up to new information regarding a patient's family history of colon cancer, only 2 out of 9 (22%) patients meeting the criteria to have a colonoscopy received one. However, 15 of the 33 patients at higher risk for colon cancer had already had a colonoscopy sufficiently recently that they did not meet criteria to have one within a year of

the clinical message. This finding may be reflective of compliance with general screening guidelines for patients regardless of risk level associated with family history. The same may be true of diabetes and breast cancer actions; 84% of the patients meeting the criteria for fasting glucose or HbA1c tests and 62% requiring mammograms had received them within the time specified by the guidelines.

Although referrals for genetic testing related to breast or colon cancer were not documented for any of those patients meeting the criteria, these may have occurred but not been documented in the generally accessible EHR data.

For coronary artery disease (CAD), 70% of patients with a positive family history had a lipid profile done after the clinical message was sent which may also reflect general screening. However, only 27% of patients meeting the criteria for cholesterol lowering medications and 43% of those for blood pressure medications received these prescriptions after the clinical message. Only 36% of those patients with a positive family history for CAD and an LDL >130 and/or body mass index (BMI) of >25 had documented exercise and nutrition counseling or referrals.

Discussion

Accurate and current family history information is becoming increasingly important with the advent of genomics and the promise of personalized medicine. Typically, patients are asked about their family history at their initial visits with their primary care physicians, but this information is generally recorded as free text in notes, and updating this information is done sporadically or not at all. When the information is captured in unstructured format, it is inaccessible to automated clinical decision support. As a result, physicians are missing opportunities for providing optimum screening and prevention for those patients at higher risk for certain conditions based on their family history.

In this study, we found that significant family history information indicating that patients were at an increased risk for certain conditions was frequently missing from the EHR, at least as structured or coded information, or within a templated note. This was particularly true of glaucoma, osteoporosis, and diabetes. Coronary artery disease, breast and colon cancer family history information were more readily available, perhaps due to greater prevention emphasis, patient knowledge of risks, and focus of physicians on soliciting this information. More structured tools to support the electronic capture of this information could help to guide comprehensive data collection for all conditions in which there are known care implications when an individual is at risk due to family history. Decision support that uses this information would also ensure the physicians take appropriate actions when patients have positive family histories.

A recent study reported that, whereas 96% of patients understand the importance of knowing one's family medical history, only 30% have collected this information.[20] Recent U.S. national efforts, such as those by the Centers for Disease Control (CDC) and U.S. Surgeon General's Family History Initiative, have been encouraging patients to complete their family histories and discuss them with their physicians.[21] As personal health records gain momentum and are designed to include family history, the numbers of patients with more comprehensive electronic repositories of this important information will increase.

To get the greatest value from this patient data, however, it will be important to create effective processes to facilitate the transfer and use of this information without undue burden on busy primary care physicians. Sending an electronic clinical message outside of a patient visit to alert physicians of new family history information was investigated in this study as a proxy for patients completing data online and then submitting it to their physicians through a patient portal, such as *Patient Gateway.* Our results show that simply sending a message outside of a visit is not sufficient for achieving appropriate screening rates for patients at increased risk due to family history.

The lack of physician follow-up may be due, in part, to the information being presented outside of a visit when a physician would not be able to discuss it with a patient and when action would be outside the normal physician workflow. In addition, no clinical decision support was provided so that physicians could evaluate the importance of the information and know what action was required. Tools that support the electronic capture of coded data by patients, deliver this information to physicians around visit time, support provider review and documentation in the electronic health record, and offer clinical decision support integrated with electronic ordering to guide appropriate follow-up are important for closing this gap in patient care.

Limitations

The response rate was low, probably in part because a very large packet was sent and a two-step enrollment process was required. The responders were somewhat older than non-responders and may have differed in other ways not measured. In addition, the study was uncontrolled. Whether any recommended actions occurred between a patient's receipt of the survey and when a clinical message was sent were not specifically assessed.

Conclusions

Clinically relevant family history information is frequently missing or not readily available in electronic health records. For all diagnoses of interest, patients often provided family history data that might influence use of screening tests, especially for glaucoma, osteoporosis, and diabetes. With the increased emphasis on family history as the field of genomics becomes more sophisticated, these results demonstrate the potential benefits of soliciting information from patients in a structured format outside of visits. However, sending clinical messages to physicians informing them of new family history information outside of a visit is not sufficient to ensure appropriate screening occurs. Better tools need to be developed to facilitate the flow of family history information from patients to their

primary care physicians and to support clinical decision-making and action.

Acknowledgements

The authors would like to thank Tiffany Wang, Margaret Lippincott, Virginia Carlson, and Tony Yu for their assistance with this study. This work was supported, in part, by the Markle Foundation, New York, USA which is not responsible for the contents of this paper.

References

[1] Murff HJ, Byrne D, Syngal S. Cancer risk assessment: quality and impact of the family history interview. Am J Prev Med 2004; 27: 239-45.

[2] Rich EC, Burke W, Heaton CJ, Haga S, Pinsky L, Short MP, Acheson L. Reconsidering the family history in primary care. JGIM 2004; 19: 273-80.

[3] Guttmacher AE, Collins FS, Carmona, RH. The family history – more important than ever. NEJM 2004; 351: 2333-36.

[4] Yoon P, Scheuner MT, Peterson-Oehlke K, Gwinn M, Faucett A, Khoury MJ. Can family history be used as a tool for public health and preventive medicine? Genet Med 2002; 4: 304-10.

[5] Murff HJ, Spigel DR, Syngal S. Does this patient have a family history of cancer? An evidence-based analysis of the accuracy of family cancer history. JAMA 2004; 292: 1480-89.

[6] Qureshi N, Bethea J, Modell B, Brennan P, Papageorgiou, A, Raeburn S, Hapgood R, Modell, M. Collecting genetic information in primary care: evaluating a new family history tool. Family Practice 2005; 22(6): 663-69.

[7] Emery J, Walton R, Murphy M, Austoker J, Yudkin P, Chapman C, Coulson A, Glasspool D, Fox J. Computer support for interpreting family histories of breast and ovarian cancer in primary care: comparative study with simulated cases. BMJ 2000; 321: 28-32.

[8] Kawamoto K, Houlihan CA, Balas EA, Lobach DF. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. BMJ 2005; 330(7494): 765

[9] Staroselsky, M, Volk LA, Tsurikova R, Pizziferri L, Lippincott M, Wald J, Bates DW. Improving electronic health record (EHR) accuracy and increasing compliance with health maintenance clinical guidelines through patient access and input. Int. J. Med. Inform. 2006; 75: 693-700.

[10] Staroselsky M, Volk LA, Tsurikova R, Newmark LP, Lippincott M, Litvak I, Kittler A, Wang T, Wald J, Bates DW. An effort to improve electronic health record medication list accuracy between visits: Patients' and physicians' response. Int. J. Med. Inform. 2007 (in press)

[11] Risk Management Foundation of the Harvard Medical Institutions. RMF Breast Care Management Algorithm. Cambridge, MA, 2003. http://www.rmf.harvard.edu/files/documents/BCA2003.pdf [cited 2006 Dec 1].

[12] Risk Management Foundation of the Harvard Medical Institutions. RMF Colorectal Cancer Screening Algorithm. Cambridge, MA, 2004.

[13] U.S. Preventive Services Task Force. Aspirin for the primary prevention of cardiovascular events: recommendations and rationale. Ann Intern Med 2002;136:157-60.

[14] U.S. Preventive Services Task Force. Guide to Clinical Preventive Services. Report of the United States Preventive Services Task Force. 3rd ed. International Medical Publishing, 2002. http://www.ahrq.gov/clinic/gcpspu.htm [cited 2006 Dec 1].

[15] U.S. Preventive Services Task Force. Screening for Glaucoma. http://www.ahrq.gov/clinic/uspstf/uspsglau.htm [cited 2006 Dec 1]

[16] National Osteoporosis Foundation. Physician's Guide to Prevention and Treatment of Osteoporosis. http://www.nof.org/physguide/index.asp [cited 2006 Dec 1]

[17] Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III) final report. Circulation. Dec 17 2002; 106(25):3143-3421.

[18] Pearson TA, Blair SN, Daniels SR, Eckel RH, Fair JM, Fortmann, SP, Franklin BA, Goldstein LB, Greenland P, Grundy SM, Hong Y, Miller NH, Lauer RM, Ockene IS, Sacco RL, Sallis JF, Smith SC, Stone NJ, Taubert KA. AHA Guidelines for Primary Prevention of Cardiovascular Disease and Stroke: 2002 Update. American Heart Association Science Advisory and Coordinating Committee. Circulation. 2002; 106(3):338-391.

[19] American Diabetes Association. Standards of Medical Care in Diabetes. Diabetes Care 2006; 29:S4-S42.

[20] Yoon PW, Scheuner MT, Gwinn M, Khoury MJ. Awareness of family health history as a risk factor for disease-United States, 2004. MMWR Morb Mortal Wkly Rep 2004; 53:1044-7.

[21] United States Department of Health & Human Services. U.S. Surgeon General's Family History Initiative. Available from: http://www.hhs.gov/familyhistory. 2006 Nov 15 [cited 2006 Dec 1].

Address for correspondence

Lynn Volk, Partners HealthCare System – Information Systems, 93 Worcester Street, Wellesley, MA 02481 USA, 781-416-8561, lavolk@partners.org

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Sharing Electronic Laboratory Results in a Patient Portal – a Feasibility Pilot

Jonathan S. Wald, MD MPH[a,c], Karen Burk, MA[a], Kate Gardner[a], Raisa Feygin[a], Elizabeth Nelson[a], Marianna Epstein[a], Eric G. Poon, MD MPH[a,b,c], Blackford Middleton, MD MPH MSc[a,b,c]

[a]Partners Information Systems, Partners HealthCare, Boston, MA, USA
[b]Brigham and Women's Hospital, Boston, MA, USA
[c]Harvard Medical School, Boston, MA, USA

Abstract

Surveys of patients consistently demonstrate a very strong interest in having secure electronic access to their own laboratory test results. In recent years, a number of health care providers and lab service centers have offered this capability, which now extends to millions of patients in the United States. Yet, little has been published on the methods of making lab results available. This case report identifies the objectives, methods, and results of a feasibility pilot conducted at Partners Healthcare from May to September, 2006. A candidate set of results were identified, approved for release, programmed into Patient Gateway, Partners' secure patient portal, and studied. Patient and practice feedback was positive. No noticeable rise in patient concerns was observed by practice staff or through patient surveys. One-half of patients who viewed results accessed reference information linked to a result. Organizational and practice-level issues necessary to support continued rollout are described.

Keywords:

Personal health records, diagnostic tests, communication, continuity of patient care, computerized medical records systems.

Introduction

In May, 2006, Partners HealthCare (Boston, MA, USA) began a feasibility pilot to introduce online laboratory results to patients in two primary care practices, via an established, self-developed secure internet portal, *Patient Gateway.* This paper aims to describe the rationale, design, and results of the four-month pilot to share our learning.

Partners' approach was informed by other organizations who have shared online results, notably Group Health Cooperative (Seattle, WA, USA) and CareGroup (Boston, MA, USA). The project context included: a) the desire for a short development time frame, b) the need to coordinate with provider workflow, c) the need to align with practice and organizational policies, and d) the need to address clinical information systems environment dependencies. Releasing laboratory results non-electronically will continue. Practical ways to address these challenges, while somewhat unique for each organization, are the focus of this paper. Since September, 2006, the new features are being deployed among primary and specialty care clinics as a result of the success of the pilot.

Materials and methods

Partners HealthCare, a large integrated delivery network in Eastern Massachusetts comprised of two teaching hospitals, three community hospitals, four specialty hospitals, and a large number of community physicians in primary and specialty care, offers providers an enterprise ambulatory electronic record, the LMR (Longitudinal Medical Record), a patient portal extension, *Patient Gateway* (PG) [1], the CDR (Clinical Data Repository) that stores laboratory results, and other related clinical systems components.

LMR (used by over 3,500 physicians) is interfaced with PG (used by over 30,000 patients since its launch in 2002). PG enables authorized patients to view real-time chart information abstracted from the LMR – currently medications, allergies, and future/past appointments. Laboratory test results are available to providers and staff who use LMR through a simple viewer. Results pending physician action across multiple patients are displayed through an advanced viewer, the Results Manager (RM) [2].

Pilot objectives and metrics

The primary pilot objective was to offer patients secure online access to their own lab results, striking a balance between the goals of "clinical transparency" (avoid withholding information) and "clinical sensitivity" (avoid violating patient and physician communication preferences). A second objective was to understand what technical, workflow, and organizational challenges must be addressed in order to successfully scale the new capability to all practices and patients. Pilot success would be measured by usage, spontaneous and solicited feedback from patients and providers, and brief surveys of patients about their experience.

Results feature design

The feasibility pilot introduced two new features – patient access to browse laboratory results in PG ("Patient Results"), and provider ability to take a Results Letter she

had created using RM in LMR, and instantly send it electronically to the patient, via PG ("Online Results Letter").

Expert panel. An expert panel of 7 clinicians was formed to identify the laboratory results to release and their timing rules (immediate release, or two business day embargo). The expert panel identified approximately 50 different results (chemistry, hematology, drug levels, endocrine, etc.). Since each clinical result (e.g. potassium) may correspond to multiple CDR items (e.g. serum potassium, blood gas potassium, serum potassium from another lab, etc), a total of 170 unique CDR items were included in the pilot. CDR results from all dates (matching a CDR code) were displayed.

Release Framework. Each CDR result was assigned a release category of "release now" or "release in two business days". The immediate release assignment meant that once the result was available in the CDR (to the physician) it was also available to the patient. The embargoed release meant that the result was "held" from display in PG for two business days once it was available for review by the physician. The embargo clock began when the result became available (not pending). Release timing (immediate or 2 day embargo) was not predicated on whether the result was in reference or out of reference range. Software changes in the CDR system would be required to accomplish this, and were out of scope for the pilot.

PG Patient Results design. In the patient portal, a new menu item under the "Health Record" menu was introduced: "Results". This opened a "Results Summary" page with a table of results (Figure 1). Each row in the table lists a single result "name" (e.g. CBC – Complete Blood Count) and most recent value, using the CDR display name familiar to physicians. The name is hyperlinked to a "history" page (e.g. CBC results from all dates). On the Results Summary page, the "date" is hyperlinked to a "Specimen" page (e.g. CBC, PLT – Platelets, etc. for a given date). Sorting by column is supported.

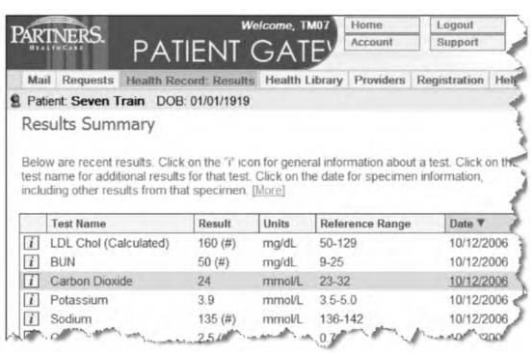


Figure 1 – Results Summary page in Patient Gateway

There is no patient notification of a new result. A "pending" result was displayed as "pending" to indicate it was not yet available (to providers). An embargoed result displayed as "held" during the embargo period, to indicate it was available for review by the provider, but not available yet in PG to the patient. Once displayed, each result included the test name, result value, "#" character if the value is out of reference range, units, reference range, and an "information button" with a hyperlink to general reference information about the result that opens in a separate window and is not maintained by PG.

General reference information in context was a design requirement. Lab Tests Online® (LTO), www.labtestsonline.org, a public resource for patients and caregivers with professionally reviewed articles, was selected for use during the pilot because of the breadth and quality of its content, and the ability of an outside site (like PG) to directly address a search query for a specific item.

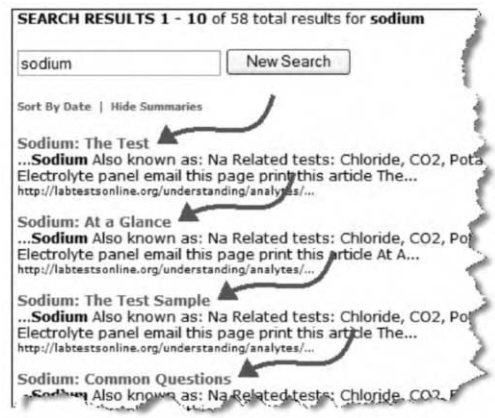


Figure 2 – Lab Tests Online® search results for Sodium

To minimize maintenance and gain sensitivity of the information retrieved given a variety of user interests, the information hyperlink avoids tight coupling to a single LTO page – but rather – executes a search query using the name of the result, bringing the user to the "Search Results" page in LTO (Figure 2). Each link works this way. If LTO were to modify their syntax for search queries, every PG link would break, making it easier to detect the problem, and simple to correct and maintain across all 170 CDR results items.

LMR Physician Results Letter design. The new LMR feature for sending a Results Letter was essentially a button called "Send to PG" placed on the RM screen to allow the provider to take action – to make a Results Letter available to the patient via PG. When this happens, the patient is notified that a message from their doctor is waiting for them in PG. Multiple patient Results Letters can be "sent" as a batch to individual patients, just as batch printing is supported. In addition, the Results Letter display in LMR was enhanced to indicate when the patient had "read" the document (the display includes "first opened" and "last opened"). Before and during this pilot, providers had the ability to type or copy/paste results from LMR into a PG message and send that message to the patient. Both methods were voluntarily used during the pilot by physicians to provide interpretation of results to the patient.

Configuration tool. A setup tool was created to activate the feature for each practice (default: all patients belonging

to that practice have the new features). In addition, a tool was created for activating the feature for a discrete list of patients to support limited deployment during the pilot. Auditing tools for pilot analysis recorded whether a patient ever looked at the new Results page in PG, or whether a patient ever received a Results Letter from a provider. Patient web surveys were sent to participating patients using Perseus SurveySolutions® software. Survey results were submitted anonymously.

Participant recruitment

Two primary care practices volunteered to offer the new features to their patients and physicians. Practice A activated the features for 2 (of 8) physicians and all of those physician's patients. Practice B activated the features for all 8 of their physicians, and all their patients. Patient and physician use of the new features was completely voluntary. Physicians attended a single brief training session to familiarize them with the results that patients could see, and with the new button in LMR to send a Results Letter online to PG.

Patients were notified of the new feature and invited to try it in two ways. A PG message was sent to each eligible patient (within the secure PG portal), which also triggers a notification message to the patient's preferred email address (e.g. gmail, yahoo, etc). Upon signing in to PG, a Home page message announced the new features. This page displayed each time a user signed in.

Pilot physicians and staff were contacted about their experience 8 weeks after the pilot began; after twelve weeks, a brief 10-item web-based survey was administered to patients who had used the new features to assess usability and impact.

Spontaneous "feedback button" information from users was also collected.

Results

Demographics and usage for PG results

The pilot users included a total of 10 physicians in 2 practices who offered the PG Results feature to 3,583 patients with a PG account (mean age 42 years, 49% female) as shown in Table 1. 2,417 (67%) of those accounts had been activated (included an external email address for patient notification). Ten percent of email notification messages were returned "undelivered". Over the next 8 weeks, 842 patients signed in to PG, and 594 (71%) viewed the Results page. Patients in practice A who viewed PG Results also viewed LTO reference information in 50% of cases. Spontaneous feedback from patients was uniformly positive. Suggestions for enhancements included: showing more results; showing radiology and cardiology reports; adding a graphing feature; and some usability improvements.

The survey results (Table 2) were strongly positive. They were collected anonymously (an IRB waiver was obtained since a survey response was considered an implied consent), so demographics are not available. Among 128 respondents, 89% felt the feature was "easy to use", 85% felt results were clearly presented, 68% felt they would recommend the site (with the new feature) to others, and 64% found the LTO reference information "very helpful". Three-quarters of respondents said they viewed the reference information. One in six respondents (17%) said they had questions for their provider as a result of seeing their results information. 31% wished for additional results, and 23% felt improvements would make the feature even more valuable.

LMR results letter "sent to PG"

In the first 6 weeks of the pilot, 5 providers sent 121 Results letters to 107 patients. A total of 65 patients (61% of 107) opened 70 letters (58% of 107). Fifty letters (41%)

Table 1 – Results: Patient Demographics, Account Activation, Results Usage, and Survey Completion Rate

Practice	Pilot MDs	PG Accounts	Pt Mean Age %F	Account Activated	Signed in	Viewed Results	Survey Invited	Survey Complete
A	2	140	49 (62% F)	113 (81%)	64	54 (84%)	180	15 (8%)
B	8	3443	42 (48% F)	2304 (67%)	778	540 (69%)	594	113 (19%)
Total	10	3583	42 (49% F)	2417 (67%)	842	594 (71%)	774	128 (17%)

Table 2 – Results of Patient Survey: PG Results feature

PG Results: Patient Survey Question	Strongly Agree	Agree	Neither	Disagree	Strongly Disagree	Uncertain	Not Answered	N
1. Results easy to use?	33%	56%	3%	4%	2%	1%	1%	128
2. Results clearly presented?	30%	55%	6%	3%	3%	2%	1%	128
3. Would recommend to others?	29%	39%	24%	2%	2%	3%	1%	128
5. Additional info: Very Helpful?	27%	37%	12%	2%	1%	2%	20%	128
Patient Survey Question		**Yes**		**No**		**Uncertain**	**Not Answered**	**N**
4. Viewed additional information?		75%		18%		6%	1%	128
6. Questions for your provider?		17%		72%		7%	4%	128
7. Want additional results?		31%		30%		34%	4%	128
9. Could it be more valuable?		23%		38%		31%	7%	128

were opened in the first week, 60 (50%) by the end of the second week, and 70 (58%) by the 26th day.

Fifteen (19%) of 80 patients who viewed the Results Letter sent by their physician responded to a brief survey. 13 (87%) felt it was easy to use; 10 (67%) felt they would recommend it to others; 10 (67%) preferred receiving the letter "Online only" and 4 (27%) liked receiving it "both online and printed"; a few suggested offering a "tabular" format with export to a spreadsheet. Ad hoc physician feedback was positive; and one physician asked if he could be notified when an online letter was *not* opened by the patient, so clinical staff could escalate if needed.

Spontaneous and solicited feedback from pilot physicians was positive. Some physicians reported that they no longer send hard copy Results Letters via postal mail, while others continue to do so. Physicians reported *no* increase in messages from their patients about clinical inconsequential results, such as an abnormal MCHC (mean corpuscular hemoglobin concentration). They reported *no* extra time expenditures as a result of the new features.

Discussion

What have we learned from this pilot? Patients and their providers have long communicated about laboratory results. Patient access to their medical chart has been legally mandated for decades. Why is patient *electronic* access via a secure portal so important?

Poor communication and disrupted follow-up of abnormal results are a problem in emergency rooms [3] and ambulatory care [4,5]. Furthermore, patients during emergency, primary or specialty care are all too often unaccompanied [6,7,8] by the clinical information from their previous care, leading to poor quality. This can also be costly, especially when errors in test follow-up and interpretation are contributors to malpractice claims [9], as they commonly are.

Physicians play a role in communication breakdowns – usually despite their best efforts and due in large part to a lack of adequate "systems" in place to support error-free practice [10]. When asked directly, many physicians freely acknowledge the need to improve their processes for results management, voicing worries about delays and results-tracking failures [11].

Can communication failures be mitigated or avoided entirely through process understanding and with the help of information transparency? Many patients think so – they want access to their own clinical information, including laboratory results [12,13]. Patient safety experts also believe changes will help, such as explicit criteria for communication of abnormal results, test-tracking systems for ordering providers, and the use of information technologies [14]. Many physicians, recognizing that their usual methods of "surveillance" for clinically important information can be fragile at times, believe that patients [15] and automated systems [16] can help.

So, if patients, providers, and safety experts support transparency, what makes it so challenging to share laboratory results electronically using a secure patient portal? Our pilot experience identified the following challenges:

Clinical sensitivity. Individual physicians resist changes that they feel will create anxiety among their patients, such as an abnormal test result without proper interpretation and delivered at an appropriate time in an appropriate way. When a patient views a result online for the first time, is it empowering? Is it an alarming indicator of a new, serious problem? Will it reinforce the confidence placed in the physician, or create friction?

Secure messaging, Results Letters, LTO, and other tools that enhance *interpretation* of results are important. Any communication approach (e.g. telephone, face-to-face, etc) can lack sensitivity, not just electronic information sharing. Also, if a less permissive provider denies their patients electronic access to results, the patients can still access them from another.

Physician workflow. Results management practices by physicians are extremely varied, even for the same physician. Physicians are notified of results by phone, pager, paper, fax, or computer. The timing of their notification varies from immediate and predictable, to delayed and unpredictable. Some physicians have staff to help them manage results, while others do not. Those who practice part-time or cover for others may not have access to results that are ready for review.

Systems to improve results management must address the larger context challenges, not just the technology. An important reason to offer patients direct electronic access to their results is that it prevents the unintended isolation of information from the patient, particularly when overworked physicians and staff are unable to detect and correct notification errors.

Individual fears. The fears of individual physicians and organizational leaders can cause inaction. Despite years of experience at Partners and elsewhere, uninitiated physicians are convinced an "avalanche" of electronic messages, worried patient telephone calls, malpractice risk, and un-reimbursed tasks will befall them with the activation of online lab results.

Education and leadership are needed to support medical professionals as secure patient portals are introduced. Vigorous efforts are needed to address the fears that are unsupported by data, since they are emotionally based and may be refractory to modification through data alone. Direct physician experience is often the most effective antidote for fear.

When concerns *are* supported by data, they should not be ignored. An example from our pilot experience is illustrative. Two business days from having their tests, patients have access to up to 50 different results, whether their physician has reviewed them or not. Pilot physicians are comfortable with this since results are reviewed and interpreted quickly in their practices. Different practice physicians (beyond the pilot) have questioned whether more delay is needed than two business days, since part-time physicians may not see test results for a week or

longer. Many physicians "batch" their review of results – waiting 3 weeks or longer to write a results letter to a patient, once all results are back (e.g. pap smears take 3 weeks). Work to address these practice variations is underway.

Engaging the patient. In many ways, we are exquisitely attuned to our patients. Yet, in the jumble of priorities, time pressures, reimbursement challenges, risk management, regulations, privacy concerns, budget priorities, and strong practitioner autonomy, the voice of the patient can be weak.

It is critically important to engage patients directly in change efforts and to value their perspective. Patients are invited on the expert panel that recommends additional chart information to share, and patient-employees are participating on another key project. The direct involvement of our patients is key.

Conclusion

Partners' feasibility pilot for sharing a limited set of laboratory results with patients in primary care was successful as measured by usage, patient satisfaction, and physician-reported experience. Now that pilot has ended, the features are being activated for additional practices beyond the pilot.

Our work identified a number of issues and challenges – some with the technology, and others due to workflow and organizational issues. Having gained experience in sharing laboratory results electronically, we plan to leverage the expert panel, lessons learned in the pilot, and patient enthusiasm to accelerate our progress toward medical information transparency.

We believe that electronic patient access to their own results and other medical chart information, if provided with appropriate privacy safeguards and clinical sensitivity, offers a powerful remedy to communication breakdowns. Confidentiality – keeping clinical information isolated from others who do not need or have authority to access it – is critically important. But clinical information isolated from the *patient* can be a critical mistake – unhelpful, inappropriate, and potentially quite dangerous.

We look forward to the challenges ahead of accelerating information transparency for patients while at the same time strengthening the physician-patient relationship, built upon caring, communication, and trust.

Acknowledgments

We are grateful to Ted Eytan MD, Group Health Cooperative, Seattle, WA, USA, who provided useful guidance throughout this pilot.

References

[1] Wald JS, Pedraza LA, Reilly CA, Murphy ME, Kuperman GJ. Requirements development for a patient computing system. Proc AMIA Symp, 2001:731-5.

[2] Poon EG, Wang SJ, Gandhi TK, Bates DW, Kuperman GJ. Design and Implementation of a Comprehensive Outpatient Results Manager. Journal of Biomedical Informatics 2003;36:80-91.

[3] Meza JP, Webster DS. Patient preferences for laboratory test results notification. Am J Manag Care 2000; 6(12):1297-300.

[4] Kern LM, Callahan MA, Brillon DJ, Vargas M, and Mushlin AI. Glucose testing and insufficient follow-up of abnormal results: a cohort study. BMC Health Serv Res, 2006; 6:87.

[5] Rojas M, Mandelblat J, Cagney K, Kerner J, Freeman H. Barriers to follow-up of abnormal screening mammograms among low-income minority women. Cancer Control Center of Harlem. Ethn Health 1996; 1(3):221-8.

[6] Burstin H. "Crossing the Quality Chasm" in emergency medicine. Acad Emerg Med, 2002; 9(11):1074-7.

[7] Gandhi TK, Sittig DF, Franklin M, Sussman AJ, Fairchild DG, Bates DW. Communication breakdown in the outpatient referral process. J Gen Intern Med 2000; 15(9):626-31.

[8] Lauteslager M, Brouwer HJ, Mohrs J, Bindels PJ, Grundmeijer HG. The patient as a source to improve the medical record. Fam Pract 2002; 19:167–71.

[9] Gandhi TK, Kachalia A, Thomas EJ, Puopolo AL, Yoon C, Brennan TA, Studdert DM. Missed and delayed diagnoses in the ambulatory setting: a study of closed malpractice claims. Ann Intern Med 2006; 145(7):488-96.

[10] Institute of Medicine. Crossing the Quality Chasm: a new health system for the 21st century. Washington, DC: National Academies Press, 2001.

[11] Poon EG, Gandhi TK, Sequist TD, Murff HJ, Karson AS, Bates DW. I wish I had seen this test result earlier! Dissatisfaction with test result management systems in primary care. Arch Intern Med 2004; 164(20):2223-8.

[12] Connecting for Health. The personal health working group final report. Markle Foundation; 2003 Jul 1. Available from: http://www.connectingforhealth.org/resources/final_phwg_report1.pdf. Accessed December 2, 2006.

[13] Harris Interactive. Two in Five Adults Keep Personal or Family Health Records and Almost Everybody Thinks This is a Good Idea. August 2004; 4(13). Available from: http://www.harrisinteractive.com/news/newsletters/healthnews/HI_HealthCareNews2004Vol4_Iss13.pdf

[14] Gandhi TK. Fumbled handoffs: one dropped ball after another. Ann Intern Med 2005; 142(5):352-8.

[15] Smith PC, Araya-Guerra R, Bublitz C, Parnes B, Dickinson LM, Van Vorst R, Westfall JM, and Pace WD. Missing clinical information during primary care visits. JAMA 2005; 293:565-71.

[16] Murff HJ, Gandhi TK, Karson AK, Mort EA, Poon EG, Wang SJ, Fairchild DG, and Bates DW. Primary care physician attitudes concerning follow-up of abnormal test results and ambulatory decision support systems. Int J Med Inform 2003; 71(2-3):137-49.

Address for correspondence

Jonathan Wald, MD, MPH, e-mail: jwald@partners.org

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

An Evolving Systems-Based Methodology For Healthcare Planning

Jon Warwick and Gary Bell

London South Bank University, London, UK.

Abstract

Healthcare planning seems beset with problems at all hierarchical levels. These are caused by the 'soft' nature of many of the issues present in healthcare planning and the high levels of complexity inherent in healthcare services. There has, in recent years, been a move to utilize systems thinking ideas in an effort to gain a better understanding of the forces at work within the healthcare environment and these have had some success. This paper argues that systems-based methodologies can be further enhanced by metrication and modeling which assist in exploring the changed emergent behavior of a system resulting from management intervention. The paper describes the Holon Framework as an evolving systems-based approach that has been used to help clients understand complex systems (in the education domain) that would have application in the analysis of healthcare problems.

Keywords:

Healthcare Systems, Systems Thinking

Introduction

In recent years there has been a growing acceptance among those working in the field of operational research that effective planning, conflict resolution, management and the control of change have become more and more difficult in domains that display 'soft' characteristics. Indeed, the emphasis in such situations often changes from one of 'achieving goals', to one of 'understanding and exploring' the problem situation and a number of methods have evolved to help decision makers better understand the web of complex issues with which they are faced [1].

Within the context of health informatics, and indeed of healthcare planning in general, there seems to be agreement that the issues faced by practitioners, decision makers and those involved in establishing policy are in a domain that could be described as [2]:

"… more information intense, more organizationally complex, more turbulent, more paradoxical, more subjective, and more emotionally charged than other industries and businesses."

The characteristics of this description are typically those of issues situated in the 'soft' domain (as described in [1]) and in response there has been some interest in using 'soft' approaches to help decision makers in the healthcare environment understand the issues they are faced with and resolve problems. To date, these approaches seem to have been largely restricted, quite understandably, to applications of systems theory and in particular soft systems methodology (SSM) either in an extended form (see for example [3, 4]) or in combination with other techniques [5]. Other more traditional operational research methods have also been used [6].

In this paper we describe the Holon Framework which is itself an interdisciplinary approach derived from the systems movement and software engineering. It has had some success in teasing out key issues within 'soft' domains and, we argue, could provide those engaged in healthcare planning with an approach for better understanding the issues they are faced with and a means for better controlling change processes in what is often a turbulent environment.

Organizational complexity and 'messes'

All planning processes have limitations. The extent to which these limitations restrict the effectiveness of the resultant plans will, to a certain extent, depend on the nature of the organization concerned and the organizational culture. In terms of healthcare planning, a number of authors have tried to produce a taxonomy of healthcare issues and one such is that of Braithwaite, Hindle, Iedema and Westbrook [3] who characterize four deep-seated problems in healthcare and describe their occurrence in terms of a quasi hierarchical domains. The first domain is that of high level policy formulation and its implementation. Policy formulation is often highly politicized and contested and even what seems to be relatively clear policy is sometimes interpreted and actioned inconsistently. The second domain is the organizational or institutional domain in which the complex nature of interacting, collaborating and co-operating healthcare organizations which make up a national healthcare service can lead to fragmentation of policy, conflicting interpretations and goals, poor communication and a perceived lack of resources. The third domain relates to the service level where clinicians and patients interact. Issues here relate to resources and workloads, unachievable patient expectations and lack of communication between clinical groups. The fourth domain is at the individual level of one-to-one interaction between health professional and patient. Here the issues relate often to the difficult choices that must be made, bal-

ances between organizational and individual work duties, low morale, work loads etc. In terms of producing effective decision making processes we can add to this list the potential problems relating to achieving 'buy-in' both from clinical and non-clinical staff and from senior management, the difficulty of embracing creativity and intuition within the process and the desire for plans which are sufficiently detailed so as to give clear direction but also remain flexible in response to environmental changes. Classic texts such as Mintzberg [7] document many of these issues.

We believe, as do others working in the field of systems theory, that resolving these issues centres around the ability to understand the nature and culture of the organization and to be able to capture some of the complex behavior and beliefs of actors within the organization. This belief is, of course, not new. Ackoff [8] contends that managers are not confronted with problems that are independent of each other, but with situations that consist of dynamic, transient and complex problems that interact with each other. He calls such situations 'messes'. Furthermore, he states [8]:

"Messes are systems of problems, the sum of the optimal solutions to each component problem taken separately is not an optimal solution to the mess. The behavior of a mess depends more on how the solutions to its parts interact than on how they act independently of each other."

The taxonomy of healthcare issues of Braithwaite, Hindle, Iedema and Westbrook briefly described above, captures some of the key problems that must be addressed if complex healthcare issues are to be addressed and which encourage the use of systems based enquiry methods. Firstly, the four domains at which the issues have been identified illustrate the range of 'voices' that must be heard if a full understanding of the problem situation is to be gained. Secondly, these 'voices' speak with very different volume in terms of their political and organizational power. Thirdly, the range of issues addressed through the four domains is huge and includes at one extreme high level policy development and at the other the one-to-one patient-clinician interaction. Fourthly, there is a key issue of complexity which is usually thought to be addressed through the application of systems thinking ideas but is often not clearly understood by stakeholders within the context of the problem. We shall return to the issue of complexity shortly.

Much work has been carried out in recent years in developing problem structuring methods that aim to provide tools for thinking about these messes [1]. Systems thinking has its advocates in healthcare applications [9] and one approach derived from systems thinking is Checkland's Soft Systems Methodology [10]. Soft Systems Methodology advocates two streams of enquiry which explore the facts and logic of the situation from the perspectives of those involved (logic-based enquiry) and also the myths and meanings which constitute what we (as individuals) mean by an organization and organizational culture (cultural enquiry). Cultural enquiry will include roles, norms, and values as well as political and other power related relationships and control processes. Note here that the phrase 'myths and meanings' encompasses a wide range of descriptors and is used to contrast with 'facts and logic' which make up the complementary stream of enquiry.

SSM refers to the importance of enabling change that is both systematically desirable and culturally feasible. Three questions arise at this point. Firstly, how will we know if the planned changes are culturally feasible? In other words, how do we uncover and explore the organizational culture that we are working within? Secondly, it may be the case that the changes are not culturally feasible within the current dominant culture so how might we perhaps influence change in the organization's culture? Thirdly, how can we ensure that the systematically desirable and culturally feasible change is enacted? In other words, how can we control the change process once changes have been identified?

The first two questions we have addressed in other work describes later in this paper. The third is key and we shall concentrate on this since the world of systems theory, information systems and, indeed, IT in general is littered with stories of system changes, organizational restructuring and IT interventions that have failed at the implementation stage [11]. UK healthcare services are no exception to this with new IT systems failing to produce the expected benefits and massive organizational change generating unexpected consequences [12]. Part of the explanation for this can be related, no doubt, to poor implementation procedures or IT system specification, but there are also issues related to complexity which need to be addressed. For example, in relation to the implementation of IT systems within healthcare, it has been noted that [5]:

"Given the complexity of the context, health informatics cannot simply focus on technical or information systems aspects alone. It has to take account of their relationship with clinical and managerial processes and practices as well as deal with multiple stakeholders and organizational cultures and accompanying politics."

We would argue that this statement does not, potentially, go far enough. It may well be the case that there are complex formal managerial and organizational relationships that need to be considered, but complex systems may well have a host of *informal* processes and relationships that are hidden from key stakeholders and organizational charts. The influence of informal processes and relationships may only become apparent once organizational changes are made.

Brookfield and Smith [13] argue that there is an inherent weakness in the management maxim that 'if you can measure it, you can manage it'. Specifically the weakness is concerned with the measurement techniques used which often assume linearity of relationships and a reliance on *a priori* data as a predictor of future performance. If we couple with this a recognition that we may well only have a partial understanding of the effects of system intervention (what Simon referred to as bounded rationality [14]) then predicting how a system may react to structural or environ-

mental change and controlling the change process itself becomes difficult.

System complexity results in only a partial understanding of the true dynamics of the system [13]. Important here are macro and micro system properties and the notion of 'downward causation'. Downward causation is the process through which a system's micro components adapt to macro level intervention and this adaptation can, potentially, be very unpredictable. Thus the effect of macro level managerial intervention could be unpredicted micro level changes the emergent properties of which may subsequently influence the properties of the wider system. Brookfield and Smith argue that there is:

" ... a degree of uncertainty associated with intervention outcomes from a managerial perspective because the performance metrics of models of intervention (their motives, logic, organizational scope, timescales, and implementation) cannot capture easily, if at all, emergent system responses."

Many public bodies in the UK (and this applies particularly to health and education) are subject to high levels of government scrutiny which involves target setting and the measurement of 'quality standards'. Clearly these issues relating to our ability to measure and predict system change are crucial in understanding how systems will respond to management intervention. It is difficult to predict system responses to change however systematically desirable and culturally feasible they might seem to be. Thus high level policy formulation in the earlier taxonomy may have unpredicted effects at the lower levels relating to clinical groups and individual personal transactions. Brookfield and Smith go on to contextualize their argument within the UK healthcare environment and use the introduction of Payment by Results [15] as part of the UK healthcare system as illustration.

The Holon Framework

The Holon Framework [16] takes from the systems movement the notions of systematic wholeness and systematic analysis and combines aspects of SSM with the Goal/Question/Metric (GQM) method [17]. The GQM method is an integral part of a goal-oriented measurement initiative [18] which aims at the continuous improvement of software processes and products and which we now adapt for use with SSM in 'soft' contexts. The GQM method is a mechanism for defining and interpreting a set of operational goals using measurement which address the specific needs of a project and organization. Thus, within its process model the Holon Framework has two modes of working in that the 'soft' part relates to improvement by addressing questions associated with 'the what', 'the where' and 'the who' while the 'hard' part relates to control by addressing 'the how', 'the why' and 'the when'. Checkland argues that researchers who apply systems concepts to investigate social situations face difficulties because these situations are never clearly defined. He prefers to use the word 'Holon' rather than 'system' as it highlights a distinctive approach to investigating such situations. We consider a Holon to be an abstract representation of a social situation that captures all problems. It is used as a framework to discover relevant issues from stakeholders' points of view; these are organized in a layered structure.

The Holon Framework combines soft elements (Framing, Enquiry, Visioning) and hard elements (Metrication and Modeling). It addresses 'the who', 'the what', and 'the where' type questions for the current state S_0, and generates a vision of a desired state S_1. Additionally, this produces a relevant metrics programme, and the collected metrics can be used as dynamic behavior patterns. It is then possible (using modeling techniques such as system dynamics) to tackle 'the how', 'the why' and 'the when' type questions. The metrication and modeling stages of the framework we see as crucial in application to complex systems.

Table 1 - Aims of Holon Framework Stages

Stages	**Stage Aims**
Framing	This stage has a number of objectives among which are that the stakeholders are identified and become familiar with the framework and that the investigators gain a broad understanding of the situation so that relevant holons (and sub-holons) can be identified and labelled.
Enquiry	This stage aims to identify the problems as perceived by the stakeholders.
Visioning	This stage attempts to collate various problems into themes to be addressed. These can be linked with a sub-holon hierarchical level.
Metrication	This stage analyses the themes and links the emergent problems with the appropriate hierarchical level. Metrics are generated to characterize specific problems.
Mathematical Modeling	This stage aims to analyse the data further using appropriate modeling techniques – for example a system dynamics model might be used to explain the situation of concern.
Action	This stage aims to facilitate change having achieved understanding of the area of concern

While it is essential to generate a vision of future activity and to use the associated metrics to measure our progress towards the vision once action has been undertaken, it is

also crucial to understand, through the modeling stage, how the system is evolving and the emerging dynamics that are being exhibited via the data collection process. Morecroft argues [19] that dysfunctional behavior can be prevented by investigating the systemic consequences of various decisions so that there is advantage to be gained by turning to traditional modeling techniques such as system dynamics. Thus as the system moves, perhaps, to a new equilibrium point as a result of the management intervention we can observe and understand this process better through the modeling and metrication stages.

The most important traits of this framework are:

- it provides an holistic view of a situation;
- the use of a soft methodology to enable the capture of the stakeholders' point of view;
- the researcher's role is that of a facilitator;
- the monitoring and controlling of the effects of bounded rationality and seeking to uncover new emergent system behavior;
- an enhanced understanding of the problem situation by the client group and the development of a desirable and feasible vision;
- the creation of a relevant metrics programme;
- production of the 'best solution' to achieve the vision given cost and other environmental constraints;
- the continuous use of system dynamics models to examine various 'what-if' scenarios and enhance understanding of the effects of macro management intervention.

The Framing and Enquiry stages are means of exploring issues, drawing out themes, boundaries and experiences that the stakeholders feel are important within the problem context. These first two stages encourage a thorough examination of the current state, S_0, resulting in its definition. Next we move to Visioning in which the client group explore a vision of the future that they feel is achievable and desirable. The vision will be expressed in terms of the holon structure used throughout the enquiry and may be expressed formally in terms of root definitions. It is important though that the discussion of S_0 and the vision, S_1, are linked through issues and problems. The stakeholder group should identify the critical issues and problems which require resolution if movement towards the vision is to be achieved. The issues and problems will generate goals, questions and metrics. The Metrication stage allows the stakeholders to learn more about the problems and issues in S_0, and the subsequent Metrics Collection stage enables them to measure their progress towards S_1. This is followed by the Action stage in which modeling is undertaken to clarify the processes which can effect movement from S_0 to S_1.

We have applied the emerging Holon Framework to situations characterized as 'messes' and have been able to demonstrate the value of the approach [20]. In addition, previous work with the Holon Framework with regard to the first two questions posed earlier has encouraged the view that it is possible to capture aspects of organizational culture and perhaps to influence changes in culture by using the framework [21]. We have also been able to demonstrate the organizational learning opportunities that are uncovered by use of the framework within the domain of higher education [22] a domain which has many of the same characteristics as the healthcare domain and, indeed, many other complex human activity systems.

Conclusions

The healthcare domain is characterized by many of the same problems that afflict other complex systems in the management and business domains. What is different, however, is the critical nature of many of the decisions that need to be made whether these are part of the care process itself, or related to improving delivery, care and management of services. This emphasizes the need to monitor and control the effects any of intervention directly resulting in system change. There are many interested parties (representing the public, health professionals, politicians and commerce) and, as stated in [9], there needs to be a:

"... judicious consideration and balancing of cultural, managerial, clinical, technical, legal, political and economic issues."

Soft Systems Methodology has been used to some effect to help decision makers understand some of the issues, but we would argue that it has some shortcomings in application to healthcare services that the Holon Framework can help to overcome. In particular we feel that:

- The Holon Framework captures and structures different perspectives of the problem situation in much the same way as SSM would do. To that extent, it has the capacity to allow participant to explore both the systemic issues and the cultural (indeed personal) issues that reflect their actions within the system;
- The framework can also allow exploration of organizational culture and provides a framework within which organizational learning can take place;
- The framework allows the development of a vision and, furthermore, the identification of metrics that will measure progress from the current position towards the vision;
- Metric data collection provides reference mode behavior which can be used as a basis for system dynamics modeling so that underlying dynamic effects within the system can be uncovered. This provides a means of measurement of progress towards the vision and also a window on the system through which unexpected properties emerging from the system may be observed and investigated. These unexpected changes bought about as result of management action will result in learning taking place as the client group seek to understand the nature of this difference between expected and observed outcomes.

The modeling and metrication stages provide stakeholders with the means to uncover system changes at the micro level as they occur and begin to produce emergent effects. Stakeholders are not bound to traditional linear models or a

priori data in trying to assess system response to any macro managerial intervention, but can use the emerging metrics to confirm beliefs or to uncover variance from expectation which might lead to significant learning opportunities regarding their own beliefs and actions, or perhaps those of others.

One strength of the Holon approach lies in discovering the complex interrelationships among the issues at play within the 'mess', and this may be explored using, for example, system dynamics. As an example from one education case study [20] admitting poorly qualified students increases the pressure on student support resources, decreases course quality, increases the burden on academics and yields an increased withdrawal rate among students. This results in the need to find additional students the following year. Dealing with poorer students reduces the academic productivity of staff, which reduces the reputation of the institution and the course. A spiraling negative loop structure begins to dominate and management intervention is required coupled with appropriate resource allocation to reduce the dominance of this loop. Such actions can be explored systematically and dynamically using appropriate modeling methods within the Holon Framework.

We argue, therefore, that the Holon Framework would have considerable advantages over other systems approaches in application to the healthcare sector both in helping to understand and to deliver system change.

References

[1] Rosenhead J, Mingers J, editors. Rational analysis for a problematic world revisited. England: John Wiley and Sons; 2001.

[2] Plummer A. Information systems methodology for building theory in health informatics: the argument for a structured approach to case study research. Proceedings of the 34th Hawaii International Conference on System Sciences; 2001; Hawaii, USA.

[3] Braithwaite J, Hindle D, Iedema R, Westbrook J. Introducing Soft Systems Methodology Plus (SSM+): why we need it and what it can contribute. Australian Health Review. 2002; 25 (2) : 191-198.

[4] Kalim K, Carson E, Cramp D. A soft systems methodology analysis of the NSF for diabetes. Proceedings of the 6th International Conference of Strategic Issues in Health Care Management; 2004; University of St. Andrews, UK.

[5] Atkinson C, Eldabi T, Paul R, Pouloudi A. Integrated approaches to health informatics research and development. Logistics Information Management. 2002; 15 (2): 138-152.

[6] Davis R, Roderick P. Planning resources for renal services throughout UK using simulation. European Journal of Operational Research. 1998; 105 (2): 285-295.

[7] Mintzberg H. The rise and fall of strategic planning. New York: The Free Press; 1994.

[8] Ackoff RL. The future of operational research is past. Journal of the Operational Research Society. 1979; 30 (2): 93-104.

[9] Fahey D, Carson E, Cramp D, Gray J. Information communication technology in public health: the role of systems modeling. Health Informatics Journal. 2003; 9 (3): 163-181.

[10] Checkland PB. Systems thinking, systems practice. England: John Wiley and Sons; 1981.

[11] King J. Survey shows common IT woes. Computerworld. 2003, June 23, [Available on the internet, cited 2006 November 3]. Available from: http://www.computerworld.com/managementtopics/management/story/0,10801,82404,00.html

[12] CBR Staff. Updated – NHS IT system crisis: the story so far. Online Computer Business Review. 2006 August 30. [cited 2006 November 3]. Available from: http://www.cbronline.com/article_cbr.asp?guid=35AC0F09-6C33-4D0E-AC2C-D912E2AA6042

[13] Brookfield D, Smith D. Managerial intervention and instability in healthcare organizations: the role of complexity in explaining the scope of effective management. [paper on the internet], The University of Liverpool Management School, Research Paper Series No. 2006/27, [cited 2006 October 27]. Available from: http://www.liv.ac.uk/managementschool/working%20papers/wp200627.pdf

[14] Simon HA. Rationality and decision making, In Models of Man. New York: John Wiley and Sons; 1957.

[15] NHS. Delivering the NHS plan. [monograph on the internet]. UK; HMSO: 2002 [cited 2006 October 27]. Available from: http://www.nhshistory.net/deliveringthenhsplan.pdf

[16] Bell GA, Warwick JP. Towards establishing the use of holons as an enquiry method. International Transactions in Operational Research. In press 2007.

[17] Basili VR, Caldiera G, Rombach DH. GQM paradigm. In: Marciniak JJ, editor. Encyclopedia of software engineering. New York: John Wiley and Sons; 1994.

[18] Boehm BW, Rombach HD, Zelkowitz MV, editors. Foundations of empirical software engineering: the legacy of Victor R Basili. Berlin Heidelberg,: Springer; 2005.

[19] Morecroft JD. System Dynamics: Portraying bounded rationality. Omega. 1983; 11 (2): 131-142.

[20] Bell GA, Cooper MA, Warwick JP, Kennedy MS. The Holon Framework: process improvement and control for higher education. In: Groccia JE, Miller JE, editors. On becoming a productive university. Bolton: Anker Publishing Company; 2005.

[21] Warwick J, Bell G, Kennedy M. Exploring myths and meanings in higher education planning. In: Williams D, Baryamureeba V, editors. Measuring computing research excellence and vitality, Kampala: Fountain; 2005. p. 143-158.

[22] Warwick J, Bell G. Uncovering opportunities for learning in higher education planning. Proceedings of the World Multi-Conference on Systemics, Cybernetics and Informatics; 2006; Orlando, Florida, USA.

Address for correspondence

Jon Warwick, Faculty of Business, Computing and Information Systems, London South Bank University, Borough Road, London SE1 0AA, United Kingdom.

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Model-Centric Approaches for the Development of Health Information Systems

Mika Tuomainen[a], Juha Mykkänen[a], Heli Luostarinen[a], Assi Pöyhölä[b], Esa Paakkanen[a]

[a] *HIS R&D Unit, IT Services Centre, University of Kuopio, Finland*
[b] *Shiftec, Department of Health Policy and Management, University of Kuopio, Finland*

Abstract

Modeling is used increasingly in healthcare to increase shared knowledge, to improve the processes, and to document the requirements of the solutions related to health information systems (HIS). There are numerous modeling approaches which aim to support these aims, but a careful assessment of their strengths, weaknesses and deficiencies is needed. In this paper, we compare three model-centric approaches in the context of HIS development: the Model-Driven Architecture, Business Process Modeling with BPMN and BPEL and the HL7 Development Framework. The comparison reveals that all these approaches are viable candidates for the development of HIS. However, they have distinct strengths and abstraction levels, they require local and project-specific adaptation and offer varying levels of automation. In addition, illustration of the solutions to the end users must be improved.

Keywords:

health information systems, modeling, information systems development

Introduction

The understanding of healthcare activities is central for the development and integration of health information systems (HIS). The specification and improvement of healthcare processes is pursued in relation to complex activities which involve lots of communication in the changing healthcare environment. To support these needs, health care actors are increasingly applying methods and best practices of business process development and enterprise modeling for analyzing care activities and clinical routines, building patient-centric processes and improving the knowledge and understanding of health care organizations [1, 2]. *Models* are used to illustrate, identify improvements or simulate organizational processes and individual activities in healthcare. Modeling also supports the specification of requirements and creates a basis for the implementation or integration of health information systems.

Modeling is used widely in software engineering and information systems development projects. Models are produced to aid the communication between users, healthcare managers, architects, designers and developers. Models add clarity to the application domain and the design, providing formalization and abstraction to the understanding of problems and solutions. Good models support complexity, cope with changes in the requirements, are economically implementable and promote information and knowledge longevity [3]. Models contain observations, requirements and assumptions about the current or prospective state and solutions of a specific domain. Numerous approaches, methods, notations and tools have been suggested and used in modeling. However, an analysis of different modeling approaches is needed for identifying a suitable approach for given development needs of HIS. It is also necessary to identify which aspects are covered by different modeling approaches and which features of different modeling approaches support the specific needs of healthcare.

In this study, we compare three model-centric approaches in relation to the development and integration of health information systems. The aim is to identify the strengths and weaknesses of different types of approaches and to bridge the gap between the healthcare knowledge and software development. The comparison is based on the use of a conceptual framework, the literature, and our experience of applying these approaches in service-oriented HIS development and integration.

Materials and methods

In this section we introduce three model-centric development approaches and a framework for the analysis and comparison of these approaches. For this analysis, we have selected three very different model-centric approaches which have been applied varyingly in our projects and which can be applied in the development and integration of health information systems.

An analysis framework for modeling approaches

To provide a systematic approach for this study, we are using a structured framework for assessing different modeling approaches. The *main purpose* of each approach is a viable starting point for such an assessment. In addition, we especially consider the support provided by each approach in different phases of an *information systems development (ISD) value chain* [4]. This chain initiates from the *understanding of the target domain*. It then proceeds to the *specification of goals and requirements* and to the *design* and *development* of solutions. The solutions are then *used in the healthcare environment*. To assess the completeness of the approach, we consider distinct aspects of information systems [5]: *structure* (which concepts and entities are included), *function* (which tasks are performed), and *behavior* (when and using what kind of

interactions the tasks are performed). Each aspect has different features in different phases of the ISD chain. We argue that a modeling chain which supports the traceability of these aspects in different phases reduces the distance between healthcare knowledge and software development [4].

Other aspects which improve the traceability should also be considered in relation to HIS [4]. Due to the information-intensive nature of healthcare, special emphasis should be paid to *semantic elements and entities* in different phases to achieve a shared understanding of complex health information and processes. In addition, the *illustration of solutions to the end users* already in the early phase of the process is beneficial. Furthermore, the support for *accurate, consistent, atomic and unambiguous documentation of the requirements* promotes the utilization of tacit knowledge and shared understanding.

In addition, we consider how *definitive* the approach is, or how many local and project-specific extensions are needed or allowed, and if there are guidelines to produce the results on a specific *abstraction level*. Furthermore, the *visibility of end users or process participants* in the models is considered. Traceability and productivity can also be improved by providing *automatic transformations or generation* of implementations from the models. Finally, the specific usage contexts and the *dissemination* situation of the approach are considered.

Model-Driven Architecture (MDA)

The *Model-Driven Architecture (MDA)* by the Object Management Group (OMG) supports software development through modeling techniques such as Unified Modeling Language (UML) [6]. Its abstraction levels separate logical and technological models: *computation independent* (CIM), *platform independent* (PIM) and *platform specific* (PSM). Three primary goals of the MDA are portability, interoperability and reusability through architectural separation of concerns.

A *computation independent model (CIM)* shows the system in its operation environment. The structure of the system, however, remains hidden or undetermined. The domain model of CIM helps in presenting exactly what the system is expected to do. A CIM often consists of several UML models, some providing more detail than others, or focusing on particular concerns. The primary user of the CIM is the domain practitioner, and it aims to bridge the gap between domain and IT experts. CIM is used as an aid to understanding the problem, but also as a source of a shared vocabulary for use in other models.

A *platform independent model (PIM)* describes the system without showing aspects which are specific to the platform, or the technology that is used to realize the system. A PIM might consist of various enterprise, information and computational models. A PIM is suited to one or more architectural styles. Concepts such as technology-neutral virtual machines can be used for platform independence in PIMs [6].

A *platform specific model (PSM)* is produced by transforming a PIM. It combines the specifications in the PIM with details which specify how the system uses a particular type of platform to provide an implementation [6]. Transformations between PIMs and PSMs can also be repeated: each new platform provides new features to the implementation. Thus there can actually be several levels of PIM and PSM descriptions and several definitions of a platform even within one system.

In an MDA specification of a system, the requirements in the CIM should be traceable to the PIM and PSM constructs that implement them, and vice versa. An MDA tool might transform a PIM directly to deployable code, without producing a PSM that would be visible to the user. This requires models on a very detailed level or many tool-specific assumptions. In practice, the application of MDA requires specific interpretations which are not provided in the MDA specifications. For this study, we have used the approach of an MDA toolkit [7] which defines the use of UML models on CIM, PIM and PSM levels and the phases of the software development lifecycle.

Process Modeling with BPMN and BPEL

Business Process Modeling Notation (BPMN) [8] is a graphical notation for process modeling. It aims to be understandable by different users from business analysts to developers and people who manage and monitor business processes. *Business Process Execution Language for Web Services (BPEL)* [9] provides an XML-based language for the specification of business processes and interaction protocols. BPMN specification contains a mapping to BPEL. Hence, BPMN and BPEL can be used together for the modeling and web services-based implementation of processes.

There are three basic types of BPMN process models. The *private processes* are internal to one organization and often describe accurate workflows. Private processes can be transformed to executable BPEL descriptions. *Public processes* describe the interactions and messages between the private processes and other participants, displaying the communication activities and the flow control mechanisms. *Collaboration processes* in BPMN describe interactions between two or more business entities (public processes). Collaboration process can be potentially mapped to collaboration protocols of electronic business such as ebXML or RosettaNet. BPMN can be extended using BPMN artifacts that provide extra information in modeling tools. Such artifacts can be used to support the requirements of a vertical domain such as healthcare. Although the graphical diagram is the most notable part of BPMN, non-graphical attributes play an important role when BPMN models are mapped to execution languages.

BPEL descriptions aim to support the automated execution of processes, defining the process from the standpoint of one participant only [9]. External partners of the process are defined as web services and contacted through interfaces described using WSDL (Web Services Definition Language). The external view of a BPEL description is also a web service. Public BPMN processes can also be transformed to *abstract BPEL descriptions* which describe interactions and hide other than communication-related aspects of the executable process. The BPEL specification does not include a graphical notation, but the notations have been mostly tool-specific. BPMN has been suggested as a generic graphical notation for BPEL [8].

HL7 Development Framework (HDF)

The *Health Level 7 (HL7) Development Framework Methodology Specification (HDF)* [10] is a framework for analyzing, designing, and documenting the processes, tools, actors, rules, and artifacts of the HL7 version 3 standards development. The approach is based on the use of a model-driven methodology and the derivation of specifications and other work products from a common set of reference models. It supersedes the earlier Message Development Framework [11].

An essential element in the HL7 Version 3.0 and the HDF is the *HL7 Reference Information Model* (RIM). All information models in the HL7 version 3 standards are based on the RIM and follow structural vocabulary specifications. The RIM is described using UML class diagram notation [11]. In addition, other modeling practices of the HDF apply some UML models directly. The HDF UML profile also defines extensions to the meta-model of the UML.

The *development process* of the HDF has seven phases [10]: project initiation, requirements analysis and documentation, specification modeling, specification documentation, specification publication, specification approval, and standard profiling. The requirements analysis and documentation, specification modeling and standard profiling phases are within the scope of this work. Several diagram types, models and design and documentation tools are utilized in these phases. The *requirements analysis and documentation* phase produces artifacts which describe the healthcare business in a given domain using its terminology. In this phase, the business processes are described using *storyboards*, *use cases* and *activity diagrams*. In addition, spreadsheets and class models are used for *message contents* and *domain analysis models (DAMs)*. The *business rules* are described by defining *relationships, triggers and constraints* to the exchange of data, and *state diagrams*. The domain experts usually develop these specifications and a *glossary* which are then used by technology experts for developing HL7 v3 messaging specifications. In the *specification modeling* phase, the reference models are constrained into design models, based on the artifacts from the previous phase. Some of the artifacts produced in this phase can be balloted as standards. Central information design models include *domain information model (DIM), constrained information model (CIM), and serialized constrained information model*. In addition, sequence and collaboration diagrams are used to describe the needed interactions and application roles and to link them to specific messages. The reuse of design model components and the harmonization of design models with the RIM are also considered. In the *standard profiling* phase, the specified standard can be profiled: its elements are annotated, constrained, extended, or left unchanged. This leads to a set of specification profiles and conformance statements. The specifics of the information exchanged, the dynamic model and the acknowledgement responsibilities are defined. In addition, this phase considers user documentation and Implementation Technology Specifications (ITS) such as XML schemas for the developers.

Results

The analysis of the three model-centric approaches is presented in Figure 1 and Table 1. Figure 1 relates the suggested modeling artifacts of each approach to the ISD value chain. Table 1 organizes the other considerations of our framework. In addition to these considerations, some specific features of each approach were observed during the analysis.

The MDA is an overarching approach. It aims to support all aspects of application development in a way which isolates technology changes from the logical solutions. The other approaches in this comparison could be included in an MDA-based approach. The MDA has been used in very different ways, and the meta-level foundations for the MDA are not required in various tools or projects. Hence, there is no "one MDA" but its application depends on situation-specific needs.

Process modeling with BPMN excludes many aspects of the holistic systems development, such as organizational structures and resources, functional breakdowns, data and information models, strategies and business rules [8]. Users are not especially considered beyond "user tasks" in BPMN. On the other hand, BPEL engines provide valu-

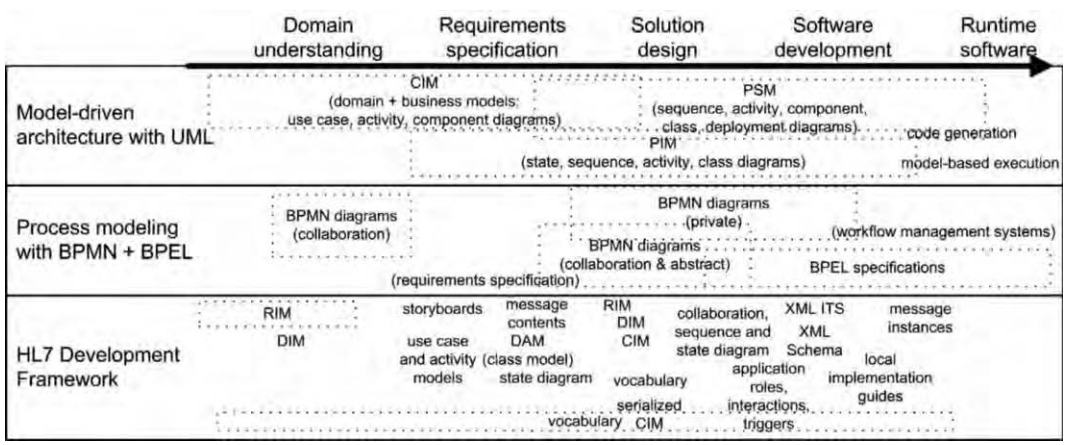


Figure 1 - The modeling artifacts of the three approaches in relation to the information systems development value chain.

Table 1 - Summary of the comparison of the three model-centric approaches.

Consideration	Model-Driven Architecture	Business Process Modeling with BPMN and BPEL	HL7 Development Framework
Main purpose	model-based software development, insulation of business from the changing technology	process modeling notation and process execution	model-based production and management of interoperability standards
Structure & information	class and other UML diagrams on CIM, PIM and PSM levels	BPMN: little description, BPEL: through XML/WSDL interfaces	RIM, domain analysis and information models, constrained models, XML schemas for messages
Functionality	use case, class and other UML diagrams	on process level only, internal functionality in private BPMN	interaction names, receiver responsibilities for messages
Interactions	sequence, activity and other UML diagrams	abstract and collaboration BPMN: generic, BPEL: specific	triggers, application roles, interactions, UML diagrams
Main users of the method	CIM: domain practitioners, PIM, PSM: IT experts	business analysts, business process owners and designers	accurate user roles specified for different phases
Semantic specification details	CIM focuses on information viewpoint, including vocabulary, more specific in PIM and PSM	little emphasis on information or semantics other than parameters of interfaces	glossaries, data types and vocabularies specified for more rigorous information support
Illustration to the end users	CIM: vocabulary, environment, functionality	BPMN: visualized processes can be examined	storyboards, use cases and activity diagrams
Requirements quality	documentation identified but not specified for the requirements	many requirements are implicit in the process descriptions	on generalized level (for all standards), but clearly traceable
Definitiveness	very loosely defined, accurate methods and tools needed	not in notations and languages but in tools and methods	accurate phases and outcomes specified, messaging presumed
Abstraction levels	CIM, PIM and PSM, lots of freedom within levels	BPMN: private, public and abstract; BPEL: execution level	models and tools defined for all specification phases and steps
Visibility of users / health professionals	evident in use case models, CIM and PIM emphasize communication with users	BPMN: user interactions, lanes, pools, activities; BPEL: no distinction between user and system steps	storyboards, use cases, activity diagrams, indirectly through application roles and triggers
Automated implementa-tion	transformations of models to executable code emphasized, various different possibilities	BPEL can be generated from private BPMN; BPEL is executed in workflow engines	implementation based on the specifications, little automation due to local variability
Dissemina-tion	promoted widely, solutions mostly tool-specific	promoted for business process management, increasing support	healthcare-specific, mostly used for standardization only
Other relevant aspects	requires more detailed methods, has been interpreted varyingly in different tools and methods	focuses on process modeling and management, relationships to other aspects remain undefined	currently under revision, new dynamic and static model approaches being refined

able information about the activities related to the processes. As the user interactions are typically required as a part of the processes, the approach can be complemented with models for user interface design.

The HDF focuses on the information aspects, and the functionality is mostly implicit in the application roles and their responsibilities. The dependencies of the related interactions and triggers are illustrated in some domains, but they are not always specified. The HDF process has clear participation roles, but the requirements dilemma of standards is evident: the requirements are generalized, and it is not always easy to find the origin of the solutions. HDF does not provide many guidelines or automated tools to support the implementation.

Discussion

The three modeling approaches clearly support the documentation and communication purposes and shared understanding in the development of HIS. However, none

of the studied approaches covers all the necessary aspects, or they do not provide a detailed support for different phases of the ISD chain. The description of processes, relationships and activities in healthcare is supported by clear notations and abstraction levels in the modeling approaches, but models do not generally suffice to provide atomic or unambiguous requirements. Furthermore, the reference from the solution models to the actual needs of the healthcare domain easily remains unclear.

In particular, improvements or local solutions are needed in the illustration of the solutions to the end users and stakeholders, the identification of aspects which are not covered by the approach and in the selection of graphical or textual notations for these aspects. The graphical models can always be introduced to the users, but this does not illustrate all the necessary aspects to them. In addition, the accurate specifications of the semantic aspects and the careful referencing to the needs and requirements are mostly left to the users of the approach.

The most notable differences in the studied approaches are related to the level of detail in the information models, functionality definitions, support for automated implementations and the definitiveness of the approach. These differences mainly stem from the different purposes of the approaches. All the approaches, however, have identified the distinction between the domain and IT expertise in the modeling efforts.

The many applications of the MDA to various specific needs make it an attractive approach as a reference framework, but also require detailed refinements which are not commensurable. The process modeling with BPMN and BPEL provides a lightweight and clear approach for end-to-end specification of processes, but leaves many aspects of the solutions unspecified. The HL7 Development Framework naturally provides the most advanced support for healthcare-specific requirements. It also emphasizes semantic aspects and provides most support for different phases of the ISD process. However, it focuses on standardization and messaging, and does not provide advanced automation or clear functional specifications.

Conclusions

The compared approaches are all viable modeling candidates for HIS. Their distinct scopes and strengths guide the selection according to different requirements. The approaches are extensible and non-exclusive. Besides the main scope, the selection depends on how definitive a guidance for the specification process and what kind of automation is desired.

All the approaches, however, require additional semantic definition and the illustration of the solutions to the end users. The approaches should be locally combined or complemented with the documentation of these aspects. In the future, accurate ontology-based approaches, user interface illustrations or domain-specific modeling extensions should be included in them. Many additional approaches such as two-level information modeling and various domain-specific modeling approaches have also been suggested to support accurate semantics and domain-specificity. They could be included in a comparison using the specified analysis framework.

Acknowledgments

This work is part of the SerAPI project, funded by the Finnish Funding Agency for Technology and Innovation TEKES grants no. 40251/06, 40353/05 and 40437/04, together with a consortium of companies and healthcare organizations.

References

[1] Staccini P, Joubert M, Quaranta J-F, Fieschi D, and Fieschi M. Modeling health care processes for eliciting user requirements: a way to link a quality paradigm and clinical information system design. Int J Med Inf 2001: 64 (2-3): 129-142.

[2] Szirbik NB, Pelletier C, and Chaussalet T. Six methodological steps to build medical data warehouses for research. Int J Med Inf 2006: 75(9): 683–691.

[3] Rubin KS, Beale T, and Blobel B. Modeling for Health Care. In: Demetriades JE, Kolodner RM, and Christopherson GA, eds. Person-Centered Health Records - Towards HealthePeople. Springer, 2005; pp. 125-146.

[4] Mykkänen J, Minkkinen I, Pöyhölä A, and Riekkinen A. Improving Traceability of Functional Requirements to Information Needs and Applications in Healthcare. Abstract. In: Doupi P, ed. Proceedings of the 6th Nordic Conference on eHealth and Telemedicine - NCeHT2006. Helsinki: Valopaino Oy, 2006, pp. 183-186.

[5] Iivari J. Object-oriented information systems analysis: A framework for object identification. In: Shriver BD, ed. Proceedings of the 24th Annual Hawaii International Conference on Systems Sciences, Vol. II. IEEE Computer Society Press, 1994; pp. 205-218.

[6] Miller J, Mukerji J, eds. MDA Guide version 1.0.1. OMG document omg/2003-06-01. OMG, 2003.

[7] Fado DH. A review of the MDA toolkit. MDA Implementers' workshop, May 17-20, Orlando. OMG, 2004.

[8] Business Process Modeling Notation Specification. OMG document dtc/06-02-01. OMG, 2006.

[9] Thatte S, ed. Business Process Execution Language for Web Services version 1.1. Accessed 20 Nov 2006. http://dev2dev.bea.com/webservices/BPEL4WS.html.

[10] HL7 version 3, September 2006 ballot package, HL7 Development Framework. Health Level Seven, Inc, 2006.

[11] Beeler GW. HL7 Version 3 - An object-oriented methodology for collaborative standards development. Int J Med Inf 1998: 48 (1-3): 151–161.

Address for correspondence

Mika Tuomainen
University of Kuopio, IT Services Centre, HIS R & D Unit
P.O.B. 1627, Fin-70211 Kuopio, Finland
mika.tuomainen@uku.fi

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Software Engineering Principles Applied to Large Healthcare Information Systems A Case Report

Fabiane Bizinella Nardon[a], Lincoln de A Moura Jr.[a]

[a] *ZILICS – Health Information Systems, São Paulo, Brazil*

Abstract

São Paulo is the largest city in Brazil and one of the largest cities in the world. In 2004, São Paulo City Department of Health decided to implement a Healthcare Information System to support managing healthcare services and provide an ambulatory health record. The resulting information system is one of the largest public healthcare information systems ever built, with more than 2 million lines of code. Although statistics shows that most software projects fail, and the risks for the São Paulo initiative were enormous, the information system was completed on-time and on-budget. In this paper, we discuss the software engineering principles adopted that allowed to accomplish that project's goals, hoping that sharing the experience of this project will help other healthcare information systems initiatives to succeed.

Keywords:

Healthcare information systems, software engineering, electronic health record.

Introduction

Several studies show that most software projects fail. An important survey in the past pointed out that only 16.2% of the software projects are completed on-time and on-budget [1]. If a large organization is involved, as it is the case of most public healthcare information systems, statistics shows that only 9% of their projects will finish on-time and on-budget. Due to its inherent complexity, healthcare projects face even higher risks and there are many reports of failed healthcare information systems. Recently, a report on the NHS software initiative, one of the most ambitious public healthcare information systems, described that the project has been failing on delivering what it was promised, with some of its contractors, formed by large and reputable companies, giving up and leaving the project [2]. In contrast, large amounts of money are spent on such projects every year, with a low success rate.

As suggested by Robert Grady in his book [3], as in medicine, great progress occurred in a relative short time when the rules against dissection were eased, we must experience a similar renaissance period in software development, making software autopsies to gather information that will help us to decrease the rate of software failure in healthcare. In this spirit, this paper describes how the public healthcare information system of São Paulo city was built from a software engineering point of view, aiming to show the techniques adopted that allowed to deliver one of the largest healthcare information systems ever built on time and on-budget.

The City of São Paulo healthcare information system

São Paulo is one of the largest cities in the world, with 10.3 million people in the city and some 18 million in the Metropolitan Area.

In June 2003, São Paulo Public Health came to operate as a "full managed-care" city, which means that resources from the National Health Fund are transferred directly to São Paulo City Department of Health (SPCDH) on a capitation basis. In exchange, the city agrees to meet certain production and quality goals, and has to send monthly reports that allow the Ministry of Health to assess if those criteria are being met.

By becoming a "full managed-care" city, São Paulo Dept of Health came to manage the following yearly figures:

- 10 million primary care consultations
- 8.5 million specialized consultations
- 550 thousand hospital admissions

São Paulo decided to invest in an Information System that would support Patient Flow Control and provide an Ambulatory Electronic Health Record System [4]. The core conditions that underlined the project can be summarized as:

- The information system would not only be fully compliant with all National Standards; but also help to promote them;
- Open standards and open source-code should be used at all levels, whenever feasible;
- Whenever possible the project should use the results of previous projects, existing technologies and concepts;
- São Paulo City would receive source-code and consulting support from the Health Ministry and, in return, would send back to the Ministry all deliverables from the project;

- Finally, and most importantly, the system should be fully integrated and provide a framework for continuously embodying new functions in an easy and natural way.

The requirements and concepts summarized above were described at length in a Term of Reference prepared by São Paulo City Department of Health. The Term of Reference defined 4 major sub-projects that should result in one and only one Information System:

Municipal health register, whose objective is to handle and process the identification data for health care users, workers and organizations, as well as the relationships among them. The Register is the prime data source for all other modules, as no operation can be carried out unless its actors are registered.

This subsystem also stores and processes all standard vocabularies in use within SUS. All data within the Municipal Health Register are fully compliant with SUS.

Patient flow control handles all requests for health care services (consultations, procedures, inpatients admission and emergency) and finds the best possible match, based on criteria such as budget, distance, availability and waiting time. This module also processes authorization requests for high-cost high-complexity procedures. This is standard-procedure under SUS. It also handles exceptions, i.e. whenever resources use exceed predefined limits or are unavailable, an accredited doctor handles the exception, either by extending the budget, finding available resources or holding the request for some time, if that is the case.

This module also processes patient flow control within the health care unit and copes with waiting lists, either local or in other reference layer.

Ambulatory electronic health record collects an essential dataset from the encounter and triggers related actions, such as notifying diseases or work-related diseases, when such conditions are met.

Role-based access control system is a single sign-on system that identifies the user and its user profile, thus enabling or disabling access to system's functions. Of course, all users have to be recorded in the Register, before being authorized to access the system. Initially, some 40 profiles were defined. Through system usage the Dept of Health decided to delete some and created other profiles, totaling more than 60 profiles currently.

To develop the project, SPCDH hired several companies that, under its management, focused on specific subprojects or on threads such as hosting, communications, equipment, software development, training and support.

The resulting information system came to be known as "SIGA Saúde", which means "following health" and is an acronym for "Integrated System for Health Care Management".

As of this writing, SIGA Saúde is in use in 372 primary care units for a) registering patients, workers, health units and their services; b) scheduling local appointments.

The actual numbers of SIGA deployment as of November 6 to 12, 2006 as informed by São Paulo City Health Department are described in Table 1, below.

From a software engineering point of view, the project had almost all the risks that usually take projects to failure: the requirements were not well defined, the scope was large, the timeframe to deliver the first version of the project was very short (about 9 months), the project team was huge (about 80 people) and informaticians had different skill levels in the technology involved, ranging from beginners to experts. Most of the team members had never worked together before and the team was not all in the same geographic location.

Table 1 - SIGA Saúde production, October 2006

Item	Amount
Primary Care Units Using SIGA Saúde	372
Primary Care consultations scheduled	1,657,023
Specialized Consultations scheduled	35,250
High-Cost Procedures Requests Processed	223,225
Patients Registered	8,357,863

Despite all odds, the team succeeded in delivering the system on time. As shown in the timeline of figure 1, the contract was signed in January of 2004, it took about two months to assemble the team and define the use cases that would be implemented. In March 2004 the first use cases started to be specified. The implementation of the software started in May 2004, and in September of that year the information system was deployed and in production with all the features planned completed. In about 4 months, more than 2 million lines of code were produced and about 300 use cases were implemented. Since then, new features and improvements have been added to the system.

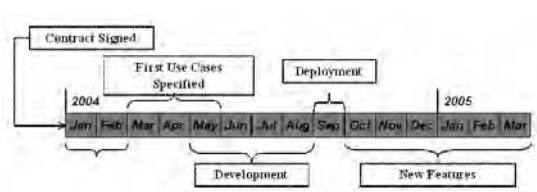


Figure 1 - The project timeline

The following sections will describe the methods and techniques used for achieving these results, in the hope to give an insight on how software engineering principles applied to large health care information systems can be a key success factor for any large project.

Methods

In a large project such as São Paulo City's Health Information System, it is important to have the methods and the architecture of the system well defined before the first requirement is captured and the first line of code is written. Our strategy for that project was based on six pillars:

- Well-defined software development process;
- Component-driven software development;
- Well-defined project guidelines and coding standards;
- Continuous integration;
- Project management;
- Knowledge sharing.

Software development process

Our development process was based on the principles of the Unified Process [5]. The software was produced in an incremental process, in such a way that new parts of the information system were delivered at each iteration. The system specification was made using the use case methodology proposed by the Unified Process. However, our use case document templates were modified to allow non-technical stakeholders to better understand what was specified and how that would look like in the real system. Thus, we added to the template a draft of the screen prototypes with a clear description of the behavior expected for each screen visual element. These use case documents were discussed with the stakeholders and after they were formally approved by all, the same document was used by the implementation team to implement the use case.

The iterative process worked very well and allowed to diminish the risks of the project, since stakeholders were able to see the results early in the project and thus they understood the development process better and communication became easier as the project evolved. The first module of the system went into production by the end of May 2004 and by then the team and the stakeholders had a clear understanding of how the development process worked.

Each iteration took about 6 weeks, with typically 4 weeks of coding, one week of testing and one week for the stakeholders testing and approval. At each iteration, while the development team was coding, the system analysts were specifying the use cases of the next iteration.

Component-driven software development

One of the principles we set out for this project was that no piece of functionality should be coded twice. To accomplish that, we adopted a component-driven software development. As each use case was ready to be implemented, a group responsible for the architecture of the project would identify which existing components could be reused, which ones should be modified and which new components should be created.

There were well defined and well documented rules about how a component should be constructed, so we would guarantee that it could be reused by other parts of the system. When a new component was developed, all the team was automatically informed of it, avoiding duplication by lack of knowledge on the components available in the component library.

In the few situations where we identified a feature implemented twice, a "refactoring" process was made on the system to eliminate the duplicated component.

The component-driven strategy increased the productivity of the team, not only because it improved reuse of code, but also because it decreased the number of errors. Once an error was identified, the corresponding component was fixed and automatically all the modules of the system that used it were also fixed.

Project guidelines and coding standards

Since different teams, with different background and skill levels were working on the project, it was important to enforce that all the teams followed defined guidelines and coding standards. The main goal here was to have a homogeneous source code, making it easier to modify and reuse the software.

These were the main guidelines defined for the project:

- A standard for directory structures, defining where each artifact, from specification to code should be stored;
- A standard for compiling and packaging the system, defining standard scripts to build all the components using the same procedure;
- A standard for code style, that was enforced automatically by tools that verified if the code produced was compliant to the standard;
- A standard for documentation, creating templates either for software specification and for code documentation;
- Guidelines for implementing new use cases, defining the technologies that would be used and how they should be used.

The project was based on open-standards and on open-source paradigms. Java Technology was chosen from the first moment for its ability to generate systems that run on any platform; to make extensive use of object-oriented analysis, and to create reusable software components. Although the system should run on a variety of equipment, a basic platform was defined with Linux as the standard operating system, JBoss as the application server and Oracle as the database management system. The only proprietary piece of software chosen was Oracle, as for the foreseen volumes no free-software database management system was considered suitable for the task. The system was developed in three-layers as depicted in Figure 2.

The application is web based, which lowered costs, since new versions are installed at only one place – the server. Also, users use simple diskless thin-clients to access the system, which also decreased maintenance costs.

The persistence and business layers of the system were implemented using Enterprise Java Beans 2.0 technology. The web layer was implemented using Struts and Tiles frameworks.

As happens with any complex technology, this architecture allows to use many different paths to implement the same functionality. If the developer is free to choose its own solution, the result is a heterogeneous code, with hard maintenance, and a potential increase in the number of errors. Also, when not well guided, beginners in the technology tend to make mistakes that well experienced developers would not.

To avoid this problem our strategy was to capture the knowledge of the most experienced developers and code this knowledge as metadata. Then, the developers would annotate the code with metadata stating the desired behavior of the use case and our framework would process these annotations and generate "expert code". The code generation strategy eliminated boilerplate code, enforced the guidelines and standards, increased productivity and made beginners to produce code as an expert would.

Differently from other coding generation tools available, however, our code generation strategy was conceived to allow extensions, making the system flexible.

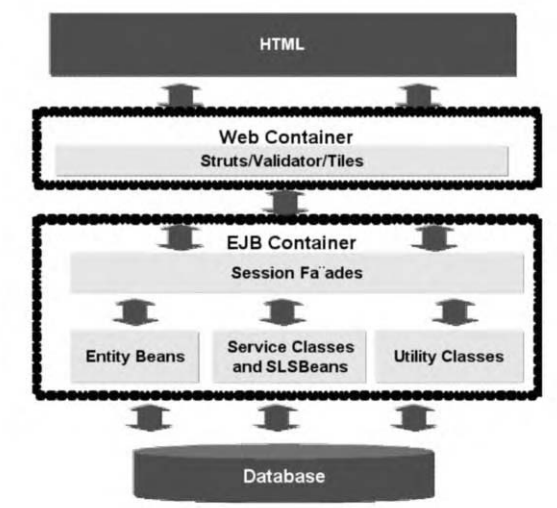


Figure 2 - System architecture

After analyzing three years of statistics from the project, it is interesting to notice that the most modified files of the system were files that implemented the user interface. Not coincidently, the user interface implementation was the only part of the architecture that didn't have a well-defined guideline or code standard. Although we had defined that the user interface should be coded using the Struts Framework and the Tiles technology, this was not enough to avoid beginners' mistakes. This clearly shows that besides defining the frameworks that should be used, it is important to create guidelines on how to use them.

Continuous integration

Continuous Integration [6] is a software development practice through which members of a team integrate their work frequently. The goal is to identify integration errors as soon as possible, reducing integration problems.

This practice is very important when multiple teams are involved and when component-driven development is used.

Our strategy was to integrate the software daily, using an automated tool. All the projects artifacts were stored in a central versioned source code repository. Each day, our continuous integration tool retrieved all the code produced and integrated it. If errors occurred, the developer responsible for it would receive an email alerting that his/her work had generated an integration error and this should be fixed.

As a result, in the last day of the iteration, we were certain that when joining the pieces together the software would not have integration problems. This contributed to decreasing the time spent solving integration problems and accelerated the development time.

Project management

Managing a large project as this one is always a challenge. Managing the human resources involved is certainly the most difficult task for project managers. The most subjective aspects of human resources management are out of the scope of this paper, but it is important to share some techniques that played an important role in the success of this project.

The project had a well-defined issue tracking process. This process stated clearly how each issue (an error, an improvement, a new task) should be addressed on the project. It started with someone registering the issue in a web based project management software. This generated an email to the project manager. The project manager would analyze the issue and either reject it or schedule the issue to resolution, assigning it to a member of the team and registering a time estimate to the issue resolution. Once solved, the owner of the issue would register what had been done and how much time was spent. Since all the process was controlled by a web-based software tool, it was possible to verify in real time, how the project was evolving, who was late on the tasks, who had time to work on other tasks, what was scheduled to be done on each iteration, and so on. Since this project management tool was integrated with the version control repository, each time a issue was solved, it was possible to track back exactly what changes had been made on the project artifacts because of that issue, what helped to identify why things changed.

This technique gave us important metrics that helped to keep the project on schedule and improved the communication among the team members.

Knowledge sharing

Knowledge sharing is very important on a large IT project. The best way to keep information about a project available for all is to enforce documentation.

In the São Paulo City Health Information System project, all project artifacts were stored in a central versioned repository. This repository was available online and could be accessed by all the members of the team.

When a new requirement was identified and the software had to change, the documentation was changed and only then the software was changed.

This approach reduced the development time, by eliminating the need for fixing poorly documented requirements. It also improved system maintainability, since even when a member of the team left the project, the knowledge was well documented and the other members could continue the work on it.

Results and conclusion

Impact data related to using SIGA Saúde is still under collection. However, current assessment reveals that outpatient services productivity has increased about 35%.

Patients' perception has changed for better, as patients have stated they can schedule appointments within the same month, against a three-month wait that was typical before SIGA Saúde. The Brazilian Press, usually very critical of public IT services published an opinion article stating their view that SIGA Saúde was a major contribution to the Bazilian Health System [7]. The project has also won a Duke's Choice Awards, granted by JavaOne 2005 Conference Meeting [8].

SIGA Saúde architecure is very innovative. By being fully web-compliant and using public Internet for providing safe role-based access to the system, SIGA Saúde can easily be deployed in other cities or regions. Also, as SIGA Saúde is fully compliant with all Brazilian standards and SUS practices, it is a tool to be used to help SUS itself be implemented throughout Brazil. Finally, São Paulo City's dimensions and complexity are so evident that taking SIGA Saúde to other cities and states requires simplification rather than new functionalities.

The techniques described on this paper allowed to deliver a large healthcare information system on time and had a significant impact on decreasing the number of errors.

The results of this project show that using software engineering principles is a key factor to avoid project failure.

There are many other aspects that could lead a project to failure, such as political forces, management changes, and others, but it is known that most IT projects failed because they can not deliver what was promised on time.

It is very important to choose the methods and the architecture of the information system before the first line of code is written. It is also important to communicate to all members of team clearly what the guidelines and methods are.

Although the techniques described on this paper were very important to the success of the project, it is also clear that they require discipline from the team and strong management, what is not easily accomplished. In this sense, choosing the right team is also an important aspect of the project.

The experience of the City of São Paulo Healthcare System should be used as an example to other public agencies that are striving to build a reliable information system. Our experience shows that applying the right software engineering principles help to achieve the goal of having an information system that will help to provide a better quality of care.

References

[1] Standish Group. The Chaos Report. Ed. The Standish Group. 1995.

[2] McCartney L. U.K. Dept of Health: Prescription for Disaster. In: Baseline Magazine. November, 2006. Available at: http://www.baselinemag.com/article2/0,1540,2058194,00.asp.

[3] Grady RB, Practical Metrics for Project Management And Process Improvement. ed. Prentice-Hall, 1992.

[4] Moura LA, Vecina-Neto G, Lira ACO, Leão BF, Hirano EM, Troccoli FT, Pacheco R, Castanha LHA. A Construção do Sistema de Informação da Secretaria Municipal de Saúde de São Paulo. Proceedings of the IX Brazilian Conference on Health Informatics, Ribeirão Preto, SP, Brazil, November 2004.

[5] Jacobson I, Booch G, Rumbaugh J. The Unified Software Development Process. Ed. Reading, MA: Addison-Wesley, 1999.

[6] Fowler M. Continuous Integration. May, 2006. Available at http://www.martinfowler.com/articles/continuousIntegration.html

[7] Opinion, A Saúde na Era da TI (Health Care in the Age of IT), O Estado de São Paulo, 23 August 2006, São Paulo, SP, Brazil.

[8] Duke's Choices Award, www.sun.com/javaone/

Address for correspondence

Fabiane Bizinella Nardon, Av. Luis Carlos Berrini, 962 – Cj. 61, São Paulo – SP – Brazil. E-Mail: fabiane@tridedalo.com.b

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Towards Modeling and Simulation of Integrated Social and Health Care Services for Elderly

Alexander Horsch[a], Daryoush Khoshsima[b]

[a] *Technische Universitt Mnchen, Germany, University of Troms, Norway*
[b] *University of Applied Sciences Heilbronn, Germany*

Abstract

In order to estimate the impact of an innovation on a segment of the health care system under certain assumptions such as different possible regulatory or financing schemes (scenarios) prior to its diffusion, one must understand the dynamic behavior of the entire system with its essential control loops. Aim of this feasibility study was to explore the potential of System Dynamics (SD) modeling for this purpose. First, a UML-based modeling of an Innovative Care for Elderly (ICE) system for provision of integrated social and health care services to elderly living at home was done. Then monetary and quality of life aspects of the social and health care system were described by two coarse SD models. On these models the impact of the introduction of the ICE system under certain assumption (scenarios) was studied, based on data from the German Health Expenditure and German Federal Statistics Office. The simulations show plausible behavior, however, are not yet detailed enough for a final conclusion. A major problem is missing data for setting model parameters: estimates had to be made. In conclusion, SD modeling might be a useful method for studying impacts of the diffusion of an innovation in the health for elderly sector, but more research is needed.

Keywords:

Health services; eHealth; telemedicine; elderly; system modeling; dynamic modeling; computer simulation.

Introduction

The demographic trends in most countries show an overproportional increase of elderly in relation to the whole population, coming along with an increase of chronically ill or multi-morbid persons, and persons that cannot live without support from social and health care. The social and health care systems have serious problems with the increasing costs of services. The diversity of care services is an additional threat for geriatric clients. The current trends will lead to a situation where it will be mandatory to utilize information and communication technology to support an independent living of elderly at their own homes as long as possible, including not only health care services, but also social services with the overall goal to intensify the personal involvement of elderly in the social life of a society as well as to deliver high-quality medical care.

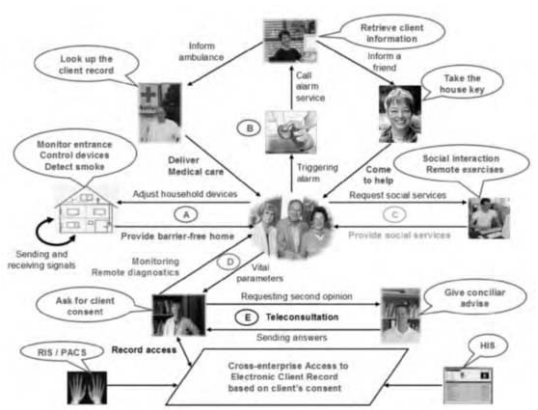


Figure 1 - The Innovative Care for Elderly (ICE) System

At the Troms Telemedicine Conference 2004 we presented the system analysis for a comprehensive telematics support of social and health service chains for elderly [7] (Figure 1) including the analysis of requirements, business processes, business use cases and system use cases of a system which in its full range covers the following components: automated house; emergency call; virtual home for elderly; telecare service. There are pilot projects and accompanying studies such as [1][2] dealing with the integration of telecare services into service delivery. And there is a variety of efforts addressing care innovations such as assistive technology or telecare in different settings. But, to our knowledge, there is no deep understanding of the entire systems dynamic behavior and the impact of innovative services on the clients quality of life under the technological and organizational intervention of introducing health care at home. This might be due to the fact that sufficiently detailed models are fairly complex and need a lot of data to be parameterized in the right way (see e.g.[1][2]). Often, the necessary information is not available and has to be estimated.

Being aware of these difficulties, we have tried to build two simulation models using the methodology of System Dynamics (SD) [10][11][9], with the goal to achieve, on a moderate level of granularity, the basis for understanding

the systems behavior and a starting point for further refinement and adjustment to relevant eHealth scenarios [8][5]. The first model is designed to investigate the economic issues of the system trying to answer the study question:

How should the costs be distributed among the different stakeholders, so that a win-win situation can be achieved? Is such a benefit-for-all situation achievable during the lifetime of the system?

The general system behavior has been explored to observe the behavior it can produce and to observe the common mode of the system behavior. The second model was created to investigate the quality aspect of the system in order to answer the study question:

How may the system affect the clients quality of life?

Main goal here was to identify the key factors influencing quality of life and to model their interrelationship in order to study the change of these entities over time, in dependence of the model parameters which describe the type of service implementation chosen in a region or country.

Material and methods

The Innovative Care for Elderly (ICE) system

As innovation to be studied in terms of diffusion into the health and social care sector for elderly, the Innovative Care for Elderly (ICE) System is used [7][8]. The system is fictitious, since this kind of overall systems does not exist, yet. It is a comprehensive telematics support of social and health service chains for elderly, which is customized to the needs of the individual client. A health manager or the ICE service provider together with the client chose the appropriate subset services, mandatory and optional ones. The Unified Modeling Language (UML) and the modeling software Together 6 (Borland) have been used for the analysis and description of requirements; business processes; business use cases and system use cases.

The main idea is to have a system that in its full range comprises the following four components (see Figure 1):

1. *The Automated House.* Such a house can offer, through simple to use interfaces adapted to the degree of disability of the occupant, e.g. various control functions for operating parts of the house (doors, jalousies, etc) and devices in the household, as well as for device communication. It includes the connection to external data networks.
2. *The Emergency Call.* This component mainly addresses sick and/or disabled persons. In case of an emergency the occupant triggers an alarm and a connection to a central help desk is established for further action. This trigger event can for example be the push on a button or a signal from a fall detection device.
3. *The Virtual Home for Elderly.* This component provides a portfolio of social and health care services. Social services comprise, for example: videophone communication with the care personnel; social interaction; consulting in critical situations; mediation of contacts to e.g. self-aid groups; support of relatives involved in the care process.
4. *The Telecare Service.* Through this service mainly the medical needs are covered. It comprises telemonitoring of the health status through e.g. vital parameters such as blood pressure, heart frequency or temperature, and all actions and services needed to handle medical or nursing tasks identified by the responsible professionals at the service center.

In addition, a teleconsulation service is included for supporting a smooth professional to professional communication.

Systems thinking

Considering a system "as a set of elements or components that work together in relationships for the overall good and objective (or vision) of the whole" [4] there are different viewpoints on SD (Figure 2). Descartes analytic thinking, based on the postulate that in separating the system in manageable pieces and understanding these pieces will lead to an understanding of the whole, when put together, is reflected by disjointed viewpoints on subsystems. This straightforward reductionist thinking usually fails to address complex systems because it is not able to capture the complexity adequately and deal with interrelationships and interactions between subsystems.

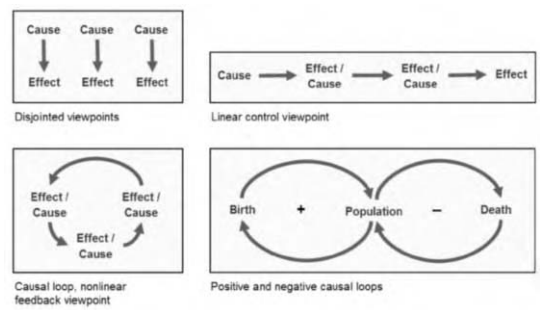


Figure 2 - Viewpoints on System Dynamics (SD)[5]

Another approach is the linear thinking, which tends to see simple sequences of cause and effect (Figure 2, linear control viewpoint). But in real systems, causes and effects are circular in their nature, not linear, forming positive or negative feedback loops (Figure 2, causal loops). Therefore, also linear thinking is inappropriate for modeling complex systems. This brings us to the point where interrelationships and control mechanisms come into the consideration, which the formal description of complex systems composed by subsystems which influence each other in a specific dynamic way.

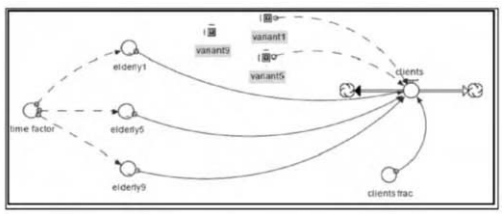


Figure 3 - Sector Frame Demographics

For our feasibility study, we used a method known as Systems Thinking [3]. This method provides:

1. *Stocks.* Represent a certain amount of units-of-measure. Four types of stocks are used: Reservoir, Conveyor, Queue, and Oven.
2. *Flows.* Represent transmission of units-of-measure from one stock to another, and can be unidirectional, bidirectional, or unit converters
3. *Converters.* Change the unit-of-measure of a flow.
4. *Connectors.* Represent information or inputs that regulate flows.

For our feasibility study we only used reservoirs, uni- and bidirectional flows, and connectors. As software tool, iThink Version 8.0 (High Performance Systems, Inc., Lebanon, NH, USA) was used. For the simulations, this software solves the finite difference equations determined by the model using iterative numerical methods (Euler, Runge-Kutta).

Results

The economic model

The unit-of-measure for the economic model is a monetary unit. Three so called *Sector Frames* these are model components which can be simulated separately have been modeled:

1. Demographics (here: Germany, see Figure 3)
2. Expenditure on system components (services)
3. Financial sources of system components (services)

For Sector Frame 1, data from the 10th coordinated population prognosis for Germany (published 2003) have been used. Figure 3 shows the Flow clients controlled by the demographic variant chosen (elderly1, elderly5, or elderly9) and the fraction of elderly using the ICE system (clients frac). To give an example of the modeling of ICE components, Figures 4 and 5 show the Social Alarm (SA) system expenditures and the financial sources modeling, respectively, with the SA cost stock being the connecting model entity. The flow into the SA cost stock is controlled by initial costs, fixed costs, extra costs, etc., the flow out of the stock is controlled by the different financing sources such social security funds (SS) and general government excluding security funds (GG) on the public funds side (PF FS), and private social insurance (PSI), private insurance enterprise (PIE), social security funds (SS), private out-of-pocket payment (PP), non-profit institutions (NPI) on the private sector side (PS FS). The complete model with the complete Sector Frames is controlled by a total number of 85 hard variables. For details and a description of the complete model in iThink see .

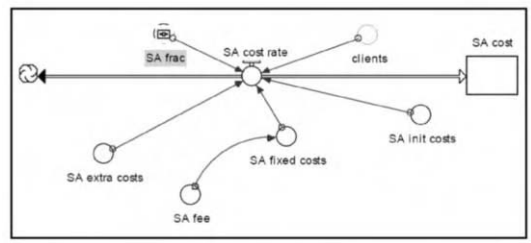


Figure 4 - Social Alarm (SA) System Expenditures

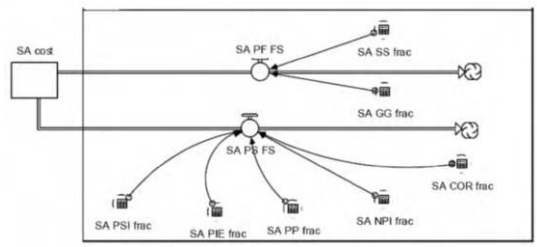


Figure 5 - Financial Sources of SA System

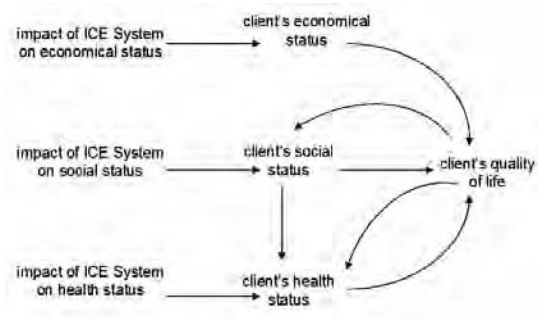


Figure 6 - Causal loops of the Quality of Life model

The Quality of Life model

The Quality of Life (QoL) model is constructed without sector frames and comprises solely soft variables. The clients QoL may be affected by many different factors, such as recognition, health, social relationships, richness of daily life, satisfaction with work. Influence on an individual changes over time. Influence on a population may be considered in a statistical way and described in terms of averages. In this feasibility study, we are interested on the large-scale influence on the QoL of a population. Our model is based on the three key factors driving the impact of the ICE System on the QoL: economical status, social status, and health status of the client. Figure 6 shows these factors which represent Stocks in the iThink and the essential causal loops QoL social status, and QoL health status. The model is controlled by a total number of 23 soft variables. For details and a description of the complete model in iThink see .

Table 1 - Parameter settings for first economic scenario

Parameter	Value	Parameter	Value
SA frac	0.1	VH NPI frac	0.25
VH frac	0.1	AH PSI frac	0.25
AH frac	0.1	AH PIE frac	0.25
TeC frac	0.1	AH PP frac	0.25
SA SS frac	0.95	AH NPI frac	0.25
SA PP frac	0.05	TeC PSI frac	0.25
VH PSI frac	0.25	TeC PIE frac	0.25
VH PIE frac	0.25	TeC PP frac	0.25
VH PP frac	0.25	TeC NPI frac	0.25
All other parameters are set to 0.			

The simulation runs

Based on statistical data from the German Health Expenditure and German Federal Statistics Office, simulations have been run under varying circumstances and assumptions over different periods of time for different scenarios in order to demonstrate how a change of parameters (population progression, financing sources, time factors, fraction of clients assigned to the different services) changes the behavior of the whole system.

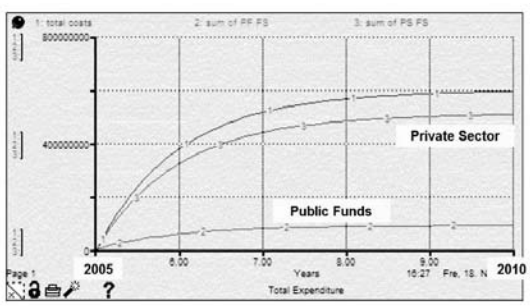


Figure 7 - Simulation Run of First Economic Scenario

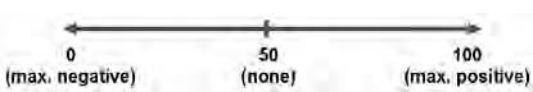


Figure 8 - Scale for Impact on Quality

For the first economic scenario, it was assumed: Germany with the middle variant of population progression, leading to middle number of elderly; 10% assigned to each service; the SA system 95% financed publicly, 5% privately; other services 100% financed privately, with a share of 25% for NPOs, private social insurance, private insurance enterprises, and private out-of-pocket payment. Table 1 shows the setting of parameters for Social Alarm (SA), Virtual Home (VH), Automated House (AH), and TeleCare (TC) service, with *frac* indicating the fraction of the population receiving a service (e.g. SA frac = 0.1 indicating 10% receive SA service) or the fraction of financing source (e.g. AH PSI frac = AH PIE frac = AH PP frac = AH NPI frac = 0.25 indicating the automated house being financed 25% by the four sources PSI, PIE, PP, and NPI), with SS indicating social security funds, PP indicating private out-of-pocket payment, NPI indicating non-profit institutions, PSI indicating private social insurance, PIE indicating private insurance enterprise as source of financing. Initial costs and running costs, as well as further parameters are included as well, see [8] for details. For the second economic scenario (not described in more detail, here), public funds reimburse the highest share of expenditure on the ICE system.

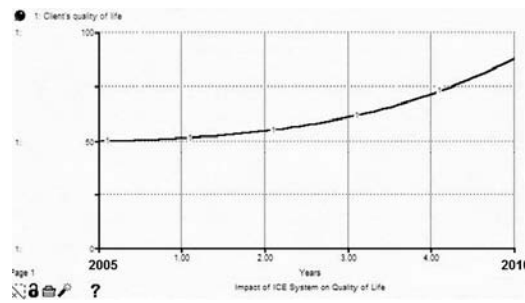


Figure 9 - Increasing QoL

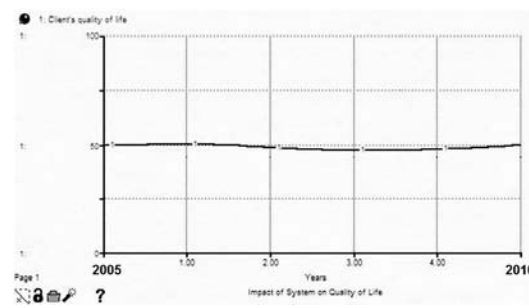


Figure 10 - Oscillating QoL

Two QoL scenarios have been considered (represented according to the scale shown in Figure 8):

1. First scenario: Dominant impact of *social aspects* on QoL, with only 10% maximal change of QoL by economic status, but 30% by social status, and 50% by the health status. Assuming the first economic scenario (private financing dominant, see above), which has a negative impact on the economic status of the clients, parameters were set to: impact of ICE system on economic status = 20 (big negative impact); on social status = 55 (small positive impact); and on health status = 55. Simulation over 5 years shows increasing positive impact of the ICE system on the QoL (Figure 9).

Assuming the second economic scenario (public funding dominant), the impact increases faster.

2. Second scenario: Dominant impact of *economical aspects* on QoL, with only 10% maximal change of QoL by social status, but 30% by economic status, and again 50% by the health status. Assuming the first economic scenario and the same impact of the ICE system on the three components of the clients status (economic, social, health), simulation over 5 years shows slight oscillation around the no-impact baseline (Figure 10). Assuming the second economic scenario (ICE system has no impact on economic status), QoL increases comparably to the first QoL scenario.

Discussion

Total costs for the ICE system to the purchaser in both economic scenarios start by zero and grow very rapidly during the first year, and increase further during the next three years. After four years there is no significant increase. This dynamic behavior of the model is reasonable since costs and expenditures on the ICE system have been taken as constant, which is not realistic: production costs for devices will probably drop, while costs for human resource may rise. It is also desirable to extend the model by a sector frame for the current care delivery system for elderly, so that it would be possible to switch between both service delivery systems.

The QoL simulations have shown exponential increase in three of four cases: if dominance of social aspects on QoL is assumed, under both financial schemes (primarily private vs. primarily public) QoL increases; if dominance of economic aspects on QoL is assumed, then QoL increases only in case of primarily public funding, but essentially stays constant in the other case.

The problem of missing data for the economic model could be tackled by involving e.g. health insurances and health care service providers of a test region in further studies with refined models. Concrete innovations introduced in a test region could serve as framework for validation studies. An empirical study based on questionnaires could help in getting realistic parameters for the QoL model.

Conclusion

The current models presented in this paper reflect the systems core components and possible dynamical characteristics. The models need substantial enhancement to make them realistic enough to give answers to our research questions. Details on costs and expenditure processes have to be included, and more variables (e.g. market saturation, clients acceptance, quality of service) have to be included. The quality of life model requires more research. It should also be refined, for example by adding direct influence of the clients economic status on the social status.

Future efforts must be put on the continuing development and validation of the models by empirical studies.

Acknowledgments

We thank the colleagues from Technische Universitt Mnchen, Ludwig-Maximilians-Universitt Mnchen, University of Troms, and Norwegian Center for Telemedicine, for the inspiring discussions. Furthermore, we thank the reviewers for their valuable comments on the first version of the paper.

References

[1] Barlow J; Bayer S; Curry R: The design of pilot telecare projects and their integration into mainstream service delivery. J Telemed Telecare 9 (Suppl 1) S1: 1-3, 2003.

[2] Bayer S, Barlow J, Curry R: Assessing the impact of a care innovation: telecare, Imperial College, London, unpublished, 2004.

[3] Carpa F: Ecology, Systems Thinking and Project-Based Learning. Procs Ann Conf Project-Based Learning, 1998.

[4] Haines S: Systems Thinking and Learning From chaos and complexity to elegant simplicity, Center for Strategic Management, 1999.

[5] Hitchins DK, http://sysdyn.clexchange.org

[6] Horsch A, Khoshsima D: Towards a Dynamic Modeling of Integrated Social and Health Services for Elderly at Home. Abstract, TTeC 2006.

[7] Khoshsima D, Horsch A, et al.: System Analysis for a Comprehensive Telematics Support of Social and Health Service Chains for Elderly. Abstract, TTeC 2004.

[8] Khoshsima D: Assessing the Impact of Telematics and Care Innovations for Elderly at Home using System Dynamics Simulation. Diploma thesis, FH Heilbronn, 2005.

[9] Milstein B, Homer J: Background on System Dynamics Simulation Modeling. With a Summary of Major Public Health Studies. Rev. April 8, 2005. Syndemics Prevention Network www.cdc.gov/syndemics.

[10] Royston G, Dost A, et al.: Using system dynamics to help develop and implement policies and programmes in health care in England. Syst Dyn Rev 15: 293-313, 1999.

[11] Wolstenholme EF: Towards the definition and use of a core set of archetypal structures in system dynamics. Syst Dyn Rev 19: 7-26, 2003.

Address for correspondence

Alexander HORSCH, Prof. II, Dr.rer.nat., Dr.med.habil.
Email: alexander.horsch@tum.de, or horsch@cs.uit.no
Phone: Tel +49-89-4140-4330, Fax +49-89-4140-4974

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Healthcare Delivery Systems: Designing Quality into Health Information Systems

Phil Joyce [a], Rosamund Green [b], Graham Winch[c]

[a] *Faculty of Information and Communication Technologies, Swinburne University of Technology, Melbourne, Australia*
[b] *Barwon Health, Geelong, Australia*
[c] *University of Plymouth Business School, Plymouth, England*

Abstract

To ensure that quality is 'engineered in' a holistic, integrated and quality approach is required, and Total Quality Management (TQM) principles are the obvious foundations for this. This paper describes a novel approach to viewing the operations of a healthcare provider where electronic means could be used to distribute information (including electronic fund settlements), building around the Full Service Provider core. Specifically, an approach called the "triple pair flow" model is used to provide a view of healthcare delivery that is integrated, yet detailed, and that combines the strategic enterprise view with a business process view.

Keywords:

Health Information Systems, TQM, *triple pair flow* construct

Introduction

Health care providers in western society now command and require increasingly large budgets to provide health care for their patients (or clients) and their communities. Indeed, government funding to health care has been steadily growing during the last ten years. This has required health care providers to clearly define their resource usage and provide information (e.g., patient level data) on their organisational and operational processes in the provision of these services. This has seen the development of funding models that are based on the development of "best practices in health care."

Organisations responsible for healthcare delivery are presently facing competing challenges. Health managers are seeking to satisfy the ever increasing number and size of stakeholder groups with quality healthcare delivery that meets the patient's needs, whilst delivering quality healthcare data to the hospial [1]. Healthcare delivery quality, as defined in [2] "as that kind of care which is expected to maximise an inclusive measure of patient welfare, after one has taken into account the balance of expected gains and losses (variability) that attend the process of healthcare in all its parts." Clearly, the complexity of this task cannot be under estimated. It requires the development of effective management and operational processes that are capable of capturing information to support the management and control of healthcare delivery. The captured information allows healthcare providers to evaluate the effectiveness of actual patient care, the efficiency of the hospital operations, the appropriate usage of resources and the expectations of patients, physicians, other hospital staff, etc, to develop measures of quality management in healthcare provision. Information and Communication Technology (ICT) support and help management of complex processes and operations with Enterprise Systems (ES) and Enterprise Resource Planning System (ERP) providing organisational wide systems that capture the processes of the organisation (horizontal flows) [3]. Moreover, eBusiness (and eCommerce) systems support the provision of services and/or goods electronically, e-fulfilment and these systems provide management with not only an effective tool to deliver product, information and funds but provide management with a method to model the strategic possibilities and implications to support healthcare delivery.

The need for strategic vision in the executive group to be communicable to and useable by the other stakeholders in their developments is essential. This allows all stakeholders to gain a mental picture or model of what is being achieved and focuses management on: envisioning, specification, design and implementation. This can be surfaced and articulated with other stakeholders in such a way that they can see within their area, business processes and technological infrastructures that are the basis of their thinking. This will ensure that quality is 'engineered in'. A holistic, integrated and quality approach is required, and Total Quality Management (TQM) principles are the obvious foundations for this. TQM is a business philosophy that encourages an over-arching responsibility - both individual and collective - to achieve quality and customer satisfaction. Commitment must be at every stage of the delivery. Importantly, the successful implementation of TQM will require the alignment of the organisation's information systems and other management systems with the new TQM environment.

Integration of process and information technology integration in quality healthcare delivery

Many large integrated information systems, particularly in public service projects, have notoriously underperformed and disappointed [4]. Often the majority of healthcare organisations do not understand the impact and effect of operating an integrated information system that captures the core processes of the organisation. Moreover, systems of this nature dramatically and fundamentally change the

way the organisation operates and interacts with its primary organisational objective and outcome: patient care. Similarly, healthcare managers may often have different objectives in their roles to support the organisational objective [5]. In this sense: clinical practitioners will primarily focus on the processes concerning patient outcomes; senior management on trend analysis and successful long term strategies and strategic planning; middle level management and reviewers with monitoring, review, productivity and resources utilisation; and business services management with billing, budgeting and accounting. Similarly, each group will have specific key performance indicators of their area's performance in meeting the organisational objectives and outcomes.

Healthcare managers are responsible for the envisioning, specification, design and implementation of new electronic processes and must therefore take a holistic, integrated and quality approach. TQM principles are the obvious foundations for this. Total Quality Management (TQM) is essentially a business philosophy. For TQM to be successful, management decisions must be aligned and integrated into a system of continuous quality improvement to meet the expectation of the customer.

There is a large amount of literature on the topic, and there are at least two major models that are used for business excellence assessment based on TQM principles – the EFQM Excellence Model [6] and the Baldridge Award [7]. These two world benchmarks use very similar criteria and dimensions for assessment, and both include Leadership, Processes, Information, and People Involvement as critical elements. However, much of this is essentially diagnostic or aspirational – 'do we think we are a quality organisation?' or 'what should a quality organisation look like?'. The appropriateness of a direct TQM approach in healthcare is long established (for example, [1, 2, 5]) though little is instrumental in the sense that it provides tools that can directly support the development of effective, coherent, and purposeful systems within an organisational quality framework. This is particularly the case when it comes to the design of integrated fulfilment, information and fund transaction systems in complex information systems in a healthcare setting.

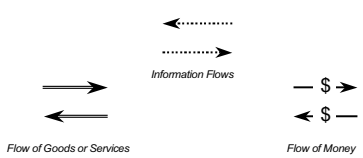


Figure 1 - The 'triple-pair' process flow model of supply chain fulfilment

Repenning and Sterman [8] correctly reported that there is a growing emphasis on business processes rather than business functions as the keystone to organisational improvement and quality enhancement. Process oriented improvement techniques such as Total Quality Management and Business Process Reengineering have proven to be powerful tools for improving the effectiveness of many organizations. However, despite the fact that it is easier than ever to learn about these performance improvement techniques and where they are being used (presumably beneficially), there appears to have been relatively little improvement in the ability of organisations to incorporate these innovations in their everyday activities. Similarly, the authors observe that the ability to identify and learn about new improvement methods no longer presents a barrier to most managers, but rather the biggest challenge is successful implementation. Repenning and Sterman [8] also assert: "You can't buy a turnkey six-sigma quality program. It must be developed from within."

A modeling approach to integrate the enterprise view with a business process view

Previous work by the authors has integrated thinking from strategic management, business process engineering, and resource-based view (RBV) of the firm and balanced scorecard (BSC) analysis to produce an integrating framework for e-business design. This framework reflects both a top-down, entrepreneurial/customer-facing view with a bottom-up, instrumental, capability-based view of what can be done [9]. The original framework, which itself essentially reflects aspirations, has then been integrated with the e-business models of Weill and Vitale [10] and the process flow modelling from system dynamics used to present a method for visualizing, communicating and then developing a shared view or consensus on the critical flow processes that can operationalise a business vision [9]. The visualisations support the critical examination of base ideas by different stakeholder groups and different domain experts; the prototype systems can then be expanded and refined of to best serve the precise needs of the organisation and its stakeholders. This, it is argued, supports powerfully the process of internal development of systems that are part of and support a quality environment within the organisation.

The 'Triple- Pair Flow' Construct for Envisioning Fulfilment Systems

All business transactions are in one way or another a supply chain fulfilment system, and healthcare delivery systems are no exception. A goods or services need is fulfilled (i.e. satisfied) by the good or service being delivered or provided with payment being received by the supplying organisation in exchange. Effectively, three flow processes comprising all such systems:

- information flows;
- money flows, payment for goods or service; and
- delivery of goods or services.

These are just the primary flows, in a healthcare system delivery of patient-care comprises a variety of service elements including medications and other consumables, and similarly money flows may be direct from patients, from insurance companies and government funding agencies. These may be seen as refinements, alternatives or extensions of the primary flows above. However, an important consideration is that each of these flows can be two-way:

Reverse information flows might include order acknowledgements, delivery notices, invoices, out-of-stock notifications, etc. It might also include information not directly related to individual order fulfilments, for example, stock position advisories and so on;

Reverse money flows might be refunds, cash-back, commissions, etc; and

Reverse goods flows might be returns, trade-ins, etc.

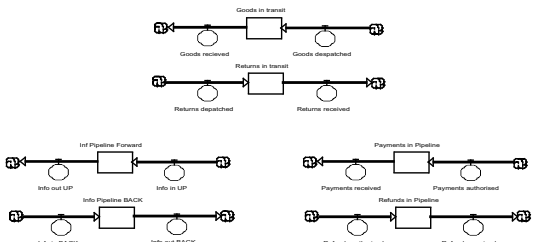


Figure 2 - The triple pair model with detail for each flow process

Joyce and Winch (2004) have described a novel construct – the "triple-pair flow" construct for envisioning such systems based on six main flows: two-way information, two-way goods and services, two-way money flows, as shown in Figure 1. If all the relevant flows relating to a particular healthcare provider could be represented within this triple pair model, then the configuration of the six flows can be mapped onto any business model that the healthcare company has in place, or wishes to adopt. Using the stock-flow diagramming convention of System Dynamics extends their construct. We will investigate how business process models can be mapped onto business models. System Dynamics is well suited to this application. There are many excellent texts that describe the principles and processes in system dynamics modelling and [11] is a leading example text. From system dynamics earliest day of development [12], has always explicitly reflected industrial and business structures as a complex inter-related set of flows of money, materials and information, and has always been concerned with the structural relationships that make up business processes as well as the softer processes. This dynamic process model perspective leads to a representation of the enterprise as a set of six sub-systems comprising stocks (or levels or inventories) and flows, as in Figure 2. This basic enhancement simply establishes that extension with detail is possible, and that the sub-structures will comprise a set of stock-flow chains. The process of expanding this to a full representation of the supply chain(s), funds transaction chain(s) and information flows can then proceed as an iterative process involving all key stakeholders. The graphical processes involved in system dynamics are proven to be of significant effectiveness in supporting team building, alignment, communicating views and thoughts with different specialist groups, and consensus building (see for example, [13-17]). It is also the basis for what [14] described as "operational thinking" within what he considered the amalgam of critical thinking skills.

An integrative design process for quality healthcare delivery and stakeholder need fulfilment

One Weill and Vitale [10] model is the Full Service Provider (FSP), which they define as "offering a full range of services in one domain … directly as well as via 'complementors' ". Critically, they include health care provision in this group. They discuss how this sector can be served by enterprises with stakeholder groups and interconnecting goods/services, information and money flows as well as characterise the FSP model by a simple schema linking these and characterise the actor and flow structure as in Figure 3. This business model reflects the situation where access to a range products or services is provided through a primary provider who might not only supply its own products or services, but also sources related products and services from partner organisations. The primary relationship in this system is between the provider and the patient, but there are additional relationships involving flows of money, product/services, and/or information between the provider and its second-level supply network partners – which Weill and Vitale consider could be suppliers or complementors (resellers and other suppliers of complementary products and services) – and between the second-level suppliers and customers.

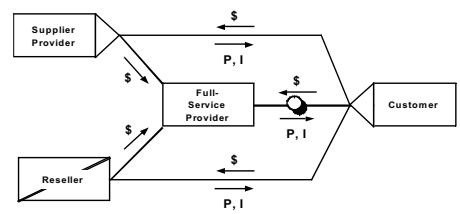


Figure 3 - Weill and Vitale's Characterisation of their Full Service Provider Model

From an overview of the fulfilment processes in a health service - patient care delivery, fund transaction and information flows, including suppliers of medications and other consumables, private and public funding bodies, and regulatory bodies, it is easy to see how these map onto the FSP model in its simple form. This then is the starting point for using the triple-pair construct in the envisioning, specification and design of systems that will engineer in the customer focus, stakeholder alignment and system coherence necessary to achieve quality healthcare delivery.

Dimensions of healthcare delivery within a full service provider perspective

Through TQM management system managers are able to integrate the vision of quality into the structure of the organisation. This must been seen at the clinical level, in the direct support of suitable patient outcomes, and from a management level, in the direct support of financial and business services and to ultimately create an accountable strategic plan for long term quality achievements. If we consider the role of a hospital as a domain in its attempt to provide quality healthcare we are able to examine the implications of the Full Service Provider Model providing a clear picture of the process of healthcare provision and the implications of TQM. The development of information system architecture with an emphasis on TQM is possible by the development of new systems or the redevelopment of older bureaucratic systems. There is a tendency when reviewing information systems in the healthcare area to create separate decision systems and management systems. Information systems within healthcare domain often fall into three categories [2]. These are:

1. Clinical or medical information systems – designed to support the activities of patient support (i.e., hospital admissions, medical records, etc)
2. Operational administrative systems – designed to provide necessary non patient care activities for the organisation (i.e., financial, personnel, payroll, etc)
3. Decision support systems – designed to provide management with information for decision making (i.e., strategic planning, analysis and evaluation, etc)

Examples of key healthcare sub-systems

If a complete healthcare organisation can be seen as a (FSP), within the Weill and Vitale classifications, then potentially new or reorganised structures can essentially be viewed. At the core will be a central service delivery subsystem, which, using the 'triple pair' flow interpretation. This reflects the technical details of the individual healthcare organisation, with the provider and the patient as the key players. In addition there would be other subsystems presenting the other inter-relationships between the FSP provider – the healthcare organisation itself – the patients and the other stakeholders.

Patient management system and government reporting

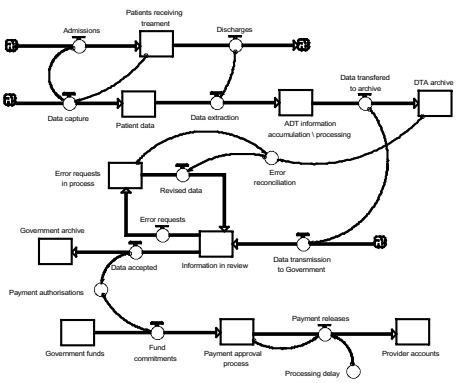


Figure 4 – Integrated process view of the ADT sub-system

The patient management system maintains the primary source of patient demographic information for the organization. It is a detailed, date related database that is an integral corner stone for health care providers as they move towards an electronic health record. Enhanced systems also maintain data encompassing patient admissions, ward transfers, discharges, appointments, operations and other clinical interactions. In one specific public health funding model the total procedures and diagnosis attributed to a patient during their inpatient stay are assigned a single code. The government health body responsible for funding and health quality uses this code in conjunction with the length of the patient's stay (LOS), time in intensive care and other critical clinical issues. The LOS is compared to the statistical mean of other patients accorded the same code from other health services and the level of funding determined. Funding is assigned to the health service retrospectively to the patients discharged from hospital. In some clinical areas such as the emergency department and surgical waiting list bonus funding can be achieved by meeting preordained patient related performance targets.

A Triple Pair Flow View of the Processes in the Patient Management System and Government Reporting

Using the construct described above, a first representation of the flow processes involved might appear as Figure 4. The sub-system is driven by the patient treatment process – this is shown at its most basic level of detail at the top as a simple admission / in-treatment / discharge process. (In a full model, this would involve a much greater level of detail as part of the core patient care delivery system.). Patient data is captured on admission, during treatment and on discharge, in the Admission Discharge Transfer (ADT) module of the clinical information system. After the data is accumulated, analysed, corrected and coded it is archived but also transmitted to the appropriate government body. This data may be checked against previous files and errors or anomalies returned to the hospital for review, after which it will be returned to the government body – this process is shown in an information recycling route. When the information is finally accepted it will be absorbed into the government body's database and, where appropriate, will trigger related payments to the hospital.

This is a first simple diagram and is based on certain assumptions about how health care activities are undertaken and what triggers particular activities. For example, in this representation, data is accumulated and analysed in the ADT system at the hospital and transmitted subsequently, possibly on a weekly or monthly basis, to the government; it is conceivable that at the present time or in the future this data could be transmitted in real time, in which case the model representation would be amended slightly. The model reflects an integrating view of how this subsystem presently operates or could operate in the future. It links the three critical process flows – patients, information, and funding, and can form the basis for debate and discussion on the effectiveness of the system and how the various stakeholders' needs and responsibilities inter-relate. While embedded within established business process engineering principles, the diagram is easily envisioned in terms of operations on the ground, enabling the focus of attention to be on the design of systems that will effectively deliver to all stakeholders. It can incorporate all-important data flow and identify trigger points which will impact on other stakeholders – for example when payments are to be triggered.

Diagnostic testing and payments sub-system

Diagnostic test order entry has ethical and security implications but in the appropriate system is a quality initiative for data entry into a diagnostic management system. It also has the advantages of reducing data entry clerk errors, save clinical time eliminating paper based sitting in a box waiting for collection and provides the clinicians ordering diagnostic tests immediate feedback on the cost of the tests being ordered and previous diagnostic interventions. Diagnostic test requests are delivered electronically to other systems both internal and external to the organization where diagnostic tests have been outsourced. Once the tests have been completed for the patient, information is returned to the diagnostic management system in the form of results. These results if in particular are of a pathology

nature can be in a cumulative format to provide instant access to trend analysis. Cost of the tests performed and billing requisition is also returned in a combination of electronic and hard copy means to the health care provider, government body, insurer or patient him or herself.

An integrated view of the processes involved in diagnostic testing

The second sub-system discussed above concerns diagnostic testing, and again a simple representation of the integrated set of process flows is shown in Figure 5.

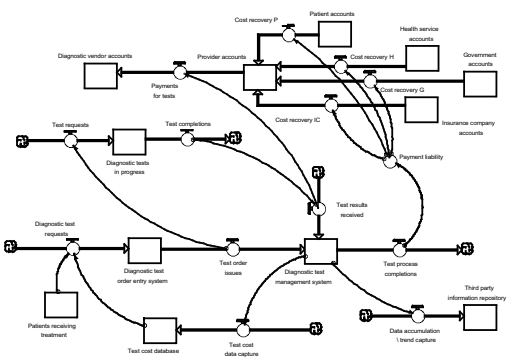


Figure 5 - Service, information and money flows involved in diagnostic testing

The central spine of the diagram reflects the process of requests for diagnostic tests being raised and then completed. Depending on hospital procedures, tests orders might be raised at the point of treatment or through a central function. When test orders are issued they trigger the request for the test at the test service providers – an internal department or external vendor – and are then logged into the diagnostic test management system. In terms of understanding the inter-related processes from the hospital's perspective, the depiction here is possibly adequate, though the vendor itself will have its own interacting system of test backlogging, information management and so on.

When the test results are received back by the management system, payments to the vendor can also be triggered. The management system will also analyse all tests completed and bill out as appropriate to the responsible institution, the patient him or herself or their insurer. As part of the system, represented as a small structure at the bottom left, a database can be maintained to provide feedback to clinicians and possibly to inform them for their future clinical decisions. Data may be accumulated, analysed for trends and so on as well as provide external agencies involved in disease containment and healthcare policy. As in the previous example, this is only a very simple representation capturing processes only at the coarsest of levels in terms of aggregation, detail, and alternate or parallel processes. Such an initial model as this will likely be the starting point for an integrated design processes for emerging systems.

Conclusions

The approach described here reflects that the highest quality levels in electronically based fulfillment systems in healthcare can indeed be best achieved where an integrative approach to system design is applied throughout the system visioning, processes specification, and detailed design process. There have been many documented failures of large ICT and electronic fulfillment systems especially in the public and health sectors where the perception was that the design processes were unable to cope with the system complexity, the diverse and sometimes competing objectives held by different stakeholders, and multiple political agendas surrounding the service.

References

[1] A. K. Aggarwal and M. Zairi, "The role of total quality management in enabling a primary health-care orientation," *Quality Management*, vol. 8, pp. 347 - 360, 1997.

[2] P. A. Rivers and S. Bae, "Aligning information systems for effective total quality management implementation in health care organisations," *Total Quality Management*, vol. 10, pp. 281-289, 1999.

[3] C. N. Madu and A. A. Madu, "E-Quality in an integrated enterprise," *The TQM Magazine*, vol. 15, pp. 127-136, 2003.

[4] R. H. Lowson and N. F. Burgess, "The building of operations strategy for e-business," *The TQM Magazine*, vol. 15, pp. 152-163, 2003.

[5] P. Adinolfi, "Total quality management in public health care a study of Italian and Irish hospitals," *Total Quality Management*, vol. 14, pp. 141-150, 2003.

[6] EFQM, "EFQM," 2005.

[7] NIST, "Baldridge National Quality Program - National Institute of Standards and Technology," 2001.

[8] N. Repenning and J. Sterman, *Getting Quality the Old Fashion: Self-Confirming Attributions in the Dynamics of Process Improvement*. Thousand Oaks, CA: Sage, 2000.

[9] P. Joyce and G. W. Winch, "A Framework for Codifying Business Models and process Models in e-Business design," in *Value Creation from e-Business Models*, W. Currie, Ed. Jordan Hill, Oxford: Elsevier, 2004.

[10] P. Weill and M. Vitale, *Place to Space*. Boston: Harvard Business Press, 2001.

[11] J. D. Sterman, *Business Dynamics: Systems Thinking and Modelling for a Complex World*: McGraw-Hill, 2000.

[12] J. Forrester, *Industrial Dynamics*. Cambridge: MIT Press (Republished by Pegasus Communications, Waltham Mass.), 1961.

[13] C. Chandra, S. Kumar, and A. Smirnov, "E-management of supply chain: general models taxonomy," *Human Systems Management*, vol. 22, pp. 95-113, 2002.

[14] B. Richmond, "Systems thinking: Critical thinking skills for the 1990s and beyond," *System Dynamics Review*, vol. 9, pp. 113-134, 1993.

[15] J. Vennix, *Group Model Building: Facilitating team learning using system dynamics*. Chichester: John Wiley & Sons, 1996.

[16] G. W. Winch, "Consensus Building in the Planning Process: Benefits from a 'Hard' Modeling Approach," *System Dynamics Review*, vol. 9, 1993.

[17] G. W. Winch, "Developing Consensus: Reflections on a Model Supported Decision Process," *Management Decision*, vol. 33, 1995.

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

A Mobile Data Collection Tool for Workflow Analysis

Jacqueline Moss[a], Eta S. Berner [b], Kathy Savell[a]

[a] *School of Nursing, University of Alabama at Birmingham, Birmingham, USA*
[b] *School of Health Professions, University of Alabama at Birmingham, Birmingham, USA*

Abstract

Faulty exchange and impaired access to clinical information is a major contributing factor to the incidence of medical error and occurrence of adverse events. Traditional methods utilized for systems analysis and information technology design fail to capture the nature of information use in highly dynamic healthcare environments. This paper describes a study designed to identify information task components in a cardiovascular intensive care unit and the development of an observational data collection tool to characterize the use of information in this environment. Direct observation can be a time-consuming process and without easy to use, reliable and valid methods of documentation, may not be reproducible across observers or settings. The following attributes were found to be necessary components for the characterization of information tasks in this setting: purpose, action, role, target, mode, and duration. The identified information task components were incorporated into the design of an electronic data collection tool to allow coding of information tasks. The reliability and validity of this tool in practice is discussed and an illustration of observational data output is provided.

Keywords:

information, systems analysis, systems design, observation.

Introduction

The communication of clinical information is the basis for coordination of patient care (1). Healthcare professionals work in teams that require circulating information to maintain team reliability. Team members use each other and Information Communication Technology (ICT) as repositories and depositories of their collective group knowledge, maintaining their own situational awareness for team coordination. Not surprisingly, faulty clinical communication and information access is a major source of error that can result in adverse events and poor patient outcomes (2). In a study designed to analyze the causes of medical errors, Leape et. al.(3) found that all seven of the most frequent medical errors identified were due to impaired access to information. These errors were primarily the result of system design faults and accounted for 78% of the total errors uncovered. Progress toward decreasing the error associated with medical care will depend heavily on better systems analysis and design.

Traditional systems analysis techniques to determine information and communication technology requirements for healthcare were originally developed for the design of business information systems. Generally, business procedures are analyzed through the use of interviews, questionnaires, focus groups, and the examination of organizational documents such as job descriptions, organizational protocols, and organizational policies (4).

There are several problems with using this approach to designing information and communication systems for healthcare. One problem is the extremely dynamic and mobile nature of healthcare work. The variability of healthcare work allows for only component segments of work to be standardized, and even these are subject to change from one instance to the next (5). Users jump from one activity to another, perform activities differently under stress, and with different levels of experience (6). In addition, healthcare practitioners are highly mobile and favor synchronous communication (7). However, healthcare team members are geographically distributed and not always in close enough proximity for face-to-face verbal communication. This is in contrast to the environment that most information systems were originally designed for; stationary workers at a desk using a desktop computer (8).

Another problem is that the use of surveying, interviewing, and focus groups alone implies that the users can articulate their work processes in an unambiguous manner. Unfortunately, discrepancies arise between reported and actual work practices, making direct observation the most reliable method for data collection (9). Direct observation can be a time-consuming process and without easy to use, reliable and valid methods of documentation, may not be reproducible across observers or settings. Clearly there is a need for tools that can make the observation processes more reliable and standardized. The purpose of this paper is to describe the development and evaluation of a mobile electronic data collection tool designed to capture information tasks of healthcare workers. It was originally developed by direct observation in a cardiovascular intensive care unit (CICU), but has since been adapted for other settings.

Methods

Development of mobile data collection tool

The electronic observational data collection tool designed for use in this study builds on previous work to develop an observational data collection tool designed and pilot-tested by Moss, Xiao, and Zubaidah (10) to determine information needs through the observation of charge nurses in a six-room trauma center operating room (OR) suite. The tool was pen and paper based and allowed coding of communication between healthcare workers. In the present study the tool was not only automated, but was refined to allow the documentation of all clinical information use, not just verbal communication.

Tool development and refinement

The original categories were expanded and new categories were added in order to characterize the totality of care in the CICU. Initially, observations were recorded on pencil and paper documenting the communication activities and information use of individual healthcare team members in their usual activities. These observations served to confirm the preliminary logical data collection codes and to identify additional codes. Observations were made during different periods of the day in 3-4 hour increments of time in order to capture various activities of each team member. During this initial stage, statements made by participants were written down as comments along with the corresponding activity.

In addition to the observations, team members were interviewed to obtain a description of unit activities (i.e., how patient transfer and family visitation is accomplished) and role responsibilities. The unit was toured to ensure the all modes of information exchange utilized were included as a tool category choice. Observations in the clinical setting were repeated until stable categories were established for each participant's information use. Approximately 50 hours were spent in this initial paper and pen phase. The observations in this phase were important in identifying who regularly participates in the care of the critically ill patient, modes of information and communication utilized, and the nature of information tasks in this environment.

The accumulated data were analyzed through the use of inductive content analysis. Content analysis involves the use of two interrelated processes: identifying specific characteristics of concepts to be measured, and employing explicit rules for identification, coding, and recording of concept characteristics (11). Categories can be identified inductively through an analysis of the content, or deductively wherein theory-based categorical schemes are defined prior to data analysis (11). First, two experienced critical care nurses reviewed and sorted the data into information categories and associated category selections individually. Then, each individual categorization was compared and a final set of categories with associated definitions was developed. Through an iterative process, selections under each category were further developed and validated in the clinical setting, until the categorical set was stable.

Tool automation

The paper and pen tool was transferred into an electronic format utilizing Microsoft Access to build a forms-based tool on a tablet computer. The observer was able to follow healthcare team members during the course of their work entering data directly into the study database through the use of a stylus pen. The data collection tool allowed rapid coding of information tasks in a uniform and consistent manner. Collecting data in this manner insured that the data were in a format amenable to analysis of information tasks.

Reliability and validity

Healthcare team members participated in every phase of the tool development. Categorizations of information tasks as well as the selections under these categories were reviewed by CICU healthcare team members for accuracy and comprehensiveness. The completed electronic data collection tool was also reviewed by CICU team members for face validity.

To insure reliable coding of information tasks in the setting, extensive training of data collectors was completed. Data collectors practiced data collection in the CICU and met with the principal investigator (JM) after each session to discuss the proper coding of each information task situation. In addition, the forms-based design of the data collection tool guaranteed that data would be collected completely and in a standardized manner.

In a study of operating room communication, an earlier version of this tool demonstrated that the tool's categorical set was able to show differences between characteristics of communication in two different settings and similarities in characteristics of communication in similar settings (12). The sensitivity of the tool's categorical set was demonstrated by comparing the original data with data obtained from a general OR suite in a comparison hospital (12). These data demonstrated that the identified categories were able to discriminate between information needs in different types of operating room suites. (i.e., demonstrated construct validity). The same technique was used in this study to determine if data would vary in an expected and explainable manner demonstrating construct validity. Observations were conducted on a medical nursing unit and compared to observations conducted in the CICU. The patient population, physical work environment, and information communication technology were all very different from the CICU.

Results

Information task attributes

The following attributes were found to be necessary components for the characterization of information tasks in this setting: purpose, action, role, target, mode, and duration. These attributes are described in detail below.

Purpose

Information task purpose is the reason for initiating the information task and was determined to fall into one of

five categories: team status, patient status, task status, instruction, and equipment management

- Team status- information task that is essential for the coordination of team members (i.e. reporting patient transfers, admissions, staff assignments; overhead page of staff member; writing patient's on-call time from OR on whiteboard, on-call assignment; patient assignment to nurse/team)
- Patient status- information task related to a particular patient's physiologic status (i.e. reporting vital signs, blood gas results; viewing monitor displays; viewing x-rays; reading test results; documenting patient assessment)
- Task status- information task that results in the performance of work or an action (i.e. verbal request for a medication, treatment, lab, or x-ray order; scheduling/ordering a medication, test, surgery, or consult in the computer; checking on/delegating a task)
- Instruction- information task related to the exchange of knowledge (i.e. directing an orientee; drug/diagnosis reference; studying/teaching/testing)
- Equipment management- information task related to the preparation, function, placement, or location of equipment (i.e. location of equipment in the unit, reporting of malfunctions; directing use of equipment; cleaning/sterilizing equipment)

Action

The information task component 'action' is a verb used to modify the information task purpose. Four attributes were found to be necessary for the description of information task action in this environment: directing, receiving, reporting, and seeking.

- Directing- information tasks related to delegating or coordinating the work of others
- Receiving- information tasks related to the reception of information
- Reporting- information tasks related to the presentation or output of data
- Seeking- information tasks related to the active search for information

Role

Role describes the functional role of the central participant being observed. These functional roles included: nurses, pharmacists, patient care technicians, unit clerks, and laboratory technicians.

Target

The target of the information task is the source or conduit of communication or information exchange. For example, information might be sought from a physician, nurse, or system database. Twenty two information task targets were identified in this study. These targets included both individuals such as physicians, nurses, and laboratory technicians as well as electronic sources such as information system databases and patient monitoring devices.

Mode

The mode of the information task is the method used by participants for the transmission or procurement of information. Seventeen separate modes of communication were identified in this context including interpersonal (e.g. .face-to-face, wireless phone), information communication systems (e.g. computerized information system), public display artifacts (e.g. whiteboard), printed material (e.g. printed reference), and personal notes.

Duration

The duration of each information task was recorded on the tool by starting the tool timer at the beginning of the task and stopping the timer at the completion or abandonment of the information task.

Reliability and validity

Tests of inter-rater reliability of the observational coding in the CICU were conducted between the study data collector and one of the investigators (JM). Both study team members observed the same nurses at the same time, entering data into separate tablet computers. This process was completed on approximately 7% of the information task observations. Analysis of these data indicated 91% percent agreement between the observers in the CICU (13).

Data analysis revealed that the tool was able to show variation and similarities in information tasks between the two types of units in an expected manner. For example, CICU nurses' information tasks were more likely to be related to managing patient data (50%) than Medical unit nurses (38%) and Medial unit nurses were more likely to be involved in information tasks related to directing the work of others (14.6%) than CICU nurses (8%). Table 1 compares nursing information tasks by purpose. Table 2 provides a comparison of nursing information tasks by action.

Table 1 - Comparison of nursing information tasks by purpose

Purpose mgt	patient	team	task	instruction	equip
CICU	50%	21.7%	11%	10.9%	6.3%
Med Unit	38.3%	23.7%	24%	10.5%	4.9%

Table 2 - Comparison of nursing information tasks by action

Action	Directing	Receiving	Reporting	Seeking
CICU	8%	21.3%	36.2%	10.9%
Med Unit	14.6%	21.9%	37.9%	25.4%

Characterization of nursing information tasks

Although data were collected on the information tasks of several types of healthcare team members, nursing information tasks are characterized in this paper to provide an example of potential data collection tool output. Twelve CICU nurses were observed for approximately 120 hours. From these observations data were recorded regarding a total of 1,631 information tasks. Duration of information tasks ranged from 1 second to 14.11 minutes with a mean duration of 33.41 seconds. Due to space limitations only a sample of these data are presented.

Patient status

Reporting information was the most frequently performed information task (43%) related to patient status, followed by seeking information (39%), receiving information (17%), and directing (<1%). More than 47% of the time nurses reported patient status information via a computerized information system. Also, nurses were more likely to seek patient status information from the computerized information system and patient monitors than any other source (81%). Unsolicited information regarding patient status was most frequently received via patient equipment alarms and alerts (20%). Nurses' most frequent reporting of patient status information verbally was to patient family and significant others (13%) and other unit nurses (10%). Surprisingly little patient status information was reported directly to residents (3%), or attending physicians (1%) in any manner.

Team status

A significant percentage of observed information tasks (21.7%) were related to directing, receiving, reporting, or seeking information for healthcare team coordination. The vast majority of these information tasks occurred face-to-face (76%) between healthcare team members. Other modes for the communication of team status information were public display artifacts such as whiteboards (5%) and telephones (4%). Almost no team status information was communicated via the computerized information system (<1%).

Discussion

Categorical structure

The categorical structure of the observational data collection tool developed by Moss, Xiao, and Zubaidah to characterize communication in Ors (10) bears some similarities to both the conceptual schema for communication space portrayed by Stetson et al. (1) and a categorization used by Coiera et al. (14) to study healthcare team communication in an emergency room. In the study of OR charge nurse communication (10) the researchers were able to determine information exchange through verbal communication and potential sources of communication error. Missing, however, was an important source of information exchange in the clinical setting: that between workers and other sources of information such as information systems, monitors, and public display artifacts. To better understand the use of clinical information in practice, we have expanded our conceptualization of communication to include any exchange of information whether between healthcare team members or team members and inanimate information sources.

To increase the granularity and usefulness of data collected we have added the category 'action' to modify and further describe the purpose of information tasks. Addition of the action modifier has allowed us to analyze the data and describe the healthcare information tasks at a level that can better inform the design and evaluation of ICT in practice illustrated by the characterization of information tasks for patient status and task status in the CICU data example.

Information task comparison

A comparison of data collected in the CICU and general medical unit illustrates the data collection tool's ability to discriminate between information tasks in different settings. For example, nurses on the medical unit were less frequently involved in information tasks related to patient status and more likely to be involved in information tasks related to task status. These variations are to be expected in this environment. Nurses in the CICU are constantly collecting and documenting patient information from patient monitors, equipment, laboratory, and physical assessments. The amount of patient data a less acutely ill medical patient generates is considerably less than a patient in an intensive care unit.

Nurses in a general medical unit are more likely to be involved in information tasks related to task status because they provide patient care in cooperation with unlicensed medical personal, delegating and coordinating the completion of patient care tasks. Another example of the increased need for general medical unit nurses to delegate and organize patient care is that only 4% of CICU nurses' actions related to task status involve directing tasks while almost 23% of these same information tasks require direction by the nurses on the general medical unit. Along the same vein, are the differences found in the frequency of information tasks associated with team status. While overall nurses in both units were involved in information tasks needed to coordinate the healthcare team at approximately the same rate, nurses working the general medical unit were much more likely to be directing team work than CICU nurses.

Implications for design

Information and communication system design in healthcare requires an understanding of how healthcare information exchange and communication occurs to enable the delivery of the correct information in the correct format. Little consensus exists in the literature as to the information needs of healthcare providers. This is not surprising, given that information needs are context specific (15) and will change with the clinical environment studied (12, 16).

Ammenwerth et al. identify the need to analyze healthcare institutions from five views: roles and responsibilities, information processing and information processing tools, communication between healthcare professionals, busi-

ness processes, and team structure and cooperation within teams (17). The tool developed in this study can serve as an aid to achieving these views of the healthcare environment. The categories of purpose, action, role, target, mode, and duration allow us to capture the use of information tools and communication between healthcare professionals during the course of their work to identify areas for system reengineering. The capacity to examine information system requirements through the use of an observational data collection tool that can be easily adapted for different settings will strengthen our ability to design systems that meet the information needs of healthcare workers.

Conclusion

In this paper we described the design and testing of an electronic data collection tool used to code observational data in a cardiovascular intensive care unit. The use of the electronic data collection tool allows a researcher to quantify and categorize the use of information during the course of clinical work, adding another dimension to traditional systems analysis. The forms-based tool design allows the observer to enter data directly into the study database using a lightweight tablet computer. Set responses and structured screen design insure that data are completely and uniformly recorded.

The tool can be used not only to inform information system design, but also to evaluate changes in interactions after the implementation of new information systems. The tool has been adapted for characterization of information needs in cardiovascular intensive care, a general medical unit, and pediatric care teams. Currently, the tool is being adapted to characterize the process of intravenous medication administration in intensive care units.

Acknowledgments

We would like to acknowledge the cooperation of all the CICU healthcare team members for their support and willingness to be a part of this study. We would also like to acknowledge the consultation of Dr. Yan Xiao on the development of this data collection tool. This study was supported by a grant from National Library of Medicine (NIH-LM008240-0).

References

[1] Stetson P, McKnight, K., Bakken, S., Curran, C., Kubose, T., Cimino, J. Development of an ontology to model medical errors, information needs, and the clinical communication space. In: AMIA Annual Symposium; 2001; Washington, DC: AMIA; 2001. p. 672-676.

[2] Parker J, Coiera, E. Improving clinical communication: a view from psychology. Journal of the American Medical Informatics Association 2000;7(5):453-461.

[3] Leape L, Bates, D., Cullen, D., Cooper, J., Demonaco, H., Gallivan, T., Hallisey, R., Ives, J., Laird, N., Laffel, G., Nemeskal, R., Petersen, L., Porter, K., Servi, D., Shea, B., Small, S., Sweitzer, B., Thompson, B., Taylor, B., Vander Vliet, M. Systems analysis of adverse drug events. JAMA 1995;274(1):35-43.

[4] Hoffer J, George, J., Valacich, J. Modern Systems Analysis & Design. 2 ed. Reading, MA: Addison-Wesley; 1998.

[5] Strauss A, Fagerhaugh, S., Suczek, B., Wiener, C. Social Organization of Medical Work. New Brunswick: Transaction; 1997.

[6] Brender J. Methodology for constructive assessment of IT-based systems in an organisational context. International Journal of Medical Informatics 1999;56:67-86.

[7] Coiera E, Tombs, V. Communication behaviours in a hospital setting: an observational study. British Medical Journal 1998;316(7132):673-676.

[8] Bardram J, Christensen, H, Olsen, AK. Activity-driven computing infrastructure-Pervasive computing in healthcare. 2002: http://www.pervasive.dk/publications.

[9] Beuscart_Zephir MC, Brender, J., Beuscart, R., Menager-Depriester. Cognitive evaluation: how to assess the usability of information technology in healthcare. Computer Methods and Programs in Biomedicine 1997;54(1-2):19-28.

[10] Moss J, Xiao, Y, Zubaidah, S. The operating room charge nurse: Coordinator and communicator. Journal of the American Medical Informatics Association 2002;9(6): S70-S74.

[11] Waltz C, Strickland, O, Lenz, E. Measurement in Nursing Research. Philadelphia: F. A. Davis; 1991.

[12] Moss J, Xiao, Y. A comparison of communication needs in two operating room suites. In: American Medical Informatics Association Annual Symposium; 2002; San Antonio; 2002. p. 543-547.

[13] Moss J, Steadman, A. Using electronic data collection tools in the clinical setting. Computers, Informatics, and Nursing Plus 2005;23(5):243-245.

[14] Coiera E, Jayasuriya, R., Hardy, J., Bannan, A., & Thorpe, M. Communication loads on clinical staff in the emergency department. The Medical Journal of Australia 2002;176(9):415-418.

[15] Forsythe D. Using ethnography to investigate life scientists' information needs. Bulletin of the Medical Library Association 1998;86:402-409.

[16] Reddy M, Pratt, W., Dourish, P., & Shabot, M. Asking questions: Information needs in a surgical intensive care unit. In: American Medical Informatics Association, Annual Symposium; 2002; San Antonio: 2002. p. 647-651.

[17] Ammenwerth E, Eichstadter, R., Haux, R., Pohl, U., Resch, F. Systems Analysis in Health Care: Framework and Example. Methods of Information in Medicine 2001;41:134-140.

[18] Xiao Y, Lasome, C., Moss, J., Mackenzie, C., & Faraq, S. Cognitive properties of a whiteboard: A case study in a trauma center. In: W. Prinz MJ, Y., Rogers, K., Schmidt, & Wulf V., editor. Proceedings of the Seventh European Conference on Computer-Supported Cooperative Work; 2001; Bonn, Germany; 2001. p. 259-278.

Address for correspondence

Jacqueline Moss PhD, RN
NB GMO26
1530 3rd Ave. South
Birmingham, AL
USA, 35294
Email: mossja@uab.edu

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Towards Sustainable e-Health Networks: Does Modeling Support Efficient Management and Operation?

Martin Staemmler[a]

[a] *University of Applied Science, Stralsund, Germany*

Abstract

e-Health Networks require cost-effective approaches for routine operation to achieve long-lasting sustainability. By using a model to represent (i) the network's enterprise functions, (ii) the applications used and (iii) the physical implementations, the tasks of management, adapting to changes and providing continued maintenance can be effectively supported. The paper discusses approaches for modeling, assesses their usefulness for the above tasks and decides on the use of the 3LGM meta model. Based on this concept, three ways for modeling the specific properties of an e-Health network are presented, leading to the decision to represent the hospitals involved in only one layer. As a result the model derived is presented, assessed and proved to support strategic management, day-to-day maintenance and documentation.

Keywords:

e-health network, health telematics, health information system, meta-model.

Introduction

Establishing e-Health networks tends to follow a typical paradigm: (i) planning and development, (ii) implementation and (iii) routine operation [1] with periods (i) and (ii) typically receiving partial support from regional or national grants, funding agencies or commercial organizations (e.g., health insurance company, hospital trust, etc.). As a result these two periods lead to a successful implementation, whereas period (iii) – the routine operation – reveals the sustainability of the approach taken. Maintaining routine operation proves to be the cornerstone of e-Health networks and, sadly to say, has lead to numerous failures. Routine operation of an e-Health network needs to cope with:

1. regulatory policies for health care provision
2. partners joining and leaving, changes in the way of cooperation between partners, e.g., requests for specific services for a subgroup of partners, new services
3. limited financial resources compared to the initial periods and
4. constant changes both in the technology used for the e-Health network and at each partner's site.

Item 1 to 4 have motivated an attempt to representing the e-Health network by a model in order to assist in the management decisions, to implement an effective change-management and to support the day-to-day operation including error conditions.

Modeling has been used in various ways: (i) to automatically derive a network infrastructure using ICMP and SNMP, (e.g., Scotty [2]) as public domain tool or commercial tools [3], (ii) to describe information system architectures[4, 5], (iii) to provide health specific reference information system models e.g., HISA[6], HL7-RIM[7], CORBAmed[8] and (iv) to document and analyze business processes e.g., with ARIS[9] or Bonapart[10]. However, most of the model approaches address only a part of the requirements to allow strategic management and day-to-day operation of an e-Health network. With the 3LGM the Institute for Medical Informatics, Statistics and Epidemiology, Leipzig, Germany [11,12] have developed an extensive meta model together with a tool for modeling hospital information systems. It uses three layers to represent enterprise functions on a domain layer, their mapping to application components on a logical tool layer, and data processing components and communication on a physical layer.

The objective of this paper is to use the 3LGM methodology for modeling an e-Health network and to assess its usefulness for the strategic management as well as for the day-to-day support and maintenance.

Materials and methods

The e-Health network

The "Health Telematics Network for the Support of Tumor Diagnosis and Treatment in the Euroregio Pomerania" extends to more than sixteen hospitals in the north-eastern part of Germany and in the western part of Poland [13, 14]. In general, the hospitals involved utilize a subset of the whole set of enterprise functions available:

- teleradiological services (emergency consultation, second opinion),
- telepathological services (perioperative instantaneous section, second opinion),
- telecardiological services (ECG analysis and reporting, second opinion prior to cardiac intervention),

- teleconferencing (with two and more partners).

These interdisciplinary enterprise functions are mapped to dedicated applications, thereby introducing mutual dependencies and requirements for the network topology (point-to-point versus hub-and-spoke). Additionally, the communication is based on a variety of access technologies, like dial-up lines using up to eight ISDN channels, leased lines, Internet based communication and radio links.

The revised three-layer graph-based meta model (3LGM)

The 3LGM has been primarily developed to model enterprise functions and associated applications and systems for information managers in hospitals [11, 12]. It holds three layers and defines relations to link adjacent layers.

- The domain layer models enterprise functions and entity types. Enterprise functions may use or create information about an entity of a given entity type, e.g., ADT functions use/create information of the entity patient.
- The logical tool layer holds application components, which represent dedicated tasks within an application. Communication is achieved by component to component interfaces or between a component and the user by means of an UI. As an example one could consider the ADT component sending messages to the laboratory system or the laboratory using a web-based GUI to present lab results.
- The physical tool layer contains components to support the two upper layers. This may be hardware systems (e.g., computer, network components, etc.) or even persons e.g. for handling a paper document based archive.
- Links between functions of the domain layer with application components of the logical tool layer reflect the relations “can support”, “based on” or “triggered by”, e.g., the ADT function is supported by an ADT component customized for this usage with a configuration. Similarly a link from the logical tool layer to a processing component in the physical layer represents the mapping of an application component via a configuration to a processing component.

The 3LGM tool does not only allow for the visualization of the above-mentioned items but provides analysis means to follow the execution of an enterprise function, as it is worked on by one or more - potentially communicating - application components and implemented using hardware and network resources. In addition, these analysis facilities motivated the use of 3LGM for an e-health network in order support questions like “What is required for providing a second opinion service for a CT-scan between hospital A and B?”.

Deriving the e-Health network model

Comparing the 3LGM target domain “hospital” with an e-Health network it refers not only to one hospital but to several hospitals. Three different approaches to represent this fact in the model have been investigated.

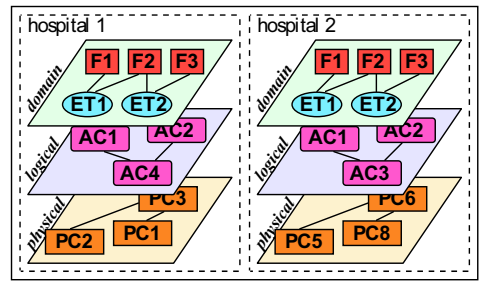


Figure 1 - Approach 1 with separate models for each hospital

The first approach relies on using separate models for each hospital as indicated in Figure 1. As a result the models shown represent two hospitals with identical enterprise functions (F1, F2, F3) and entity types (ET1, ET2) in the domain layer, some minimal differences between the application components (AC1 … AC4) in the logical layer and a mapping on separate physical components in the physical layer (PC1 … PC8). The drawbacks of this approach are (i) a repetitive modeling for at least the domain layer and (ii) more importantly the inability to perform an analysis across separate models e.g., for answering the question above.

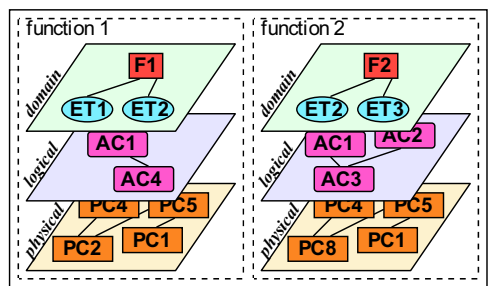


Figure 2 - Approach 2 with separate models for each enterprise function

The second approach groups into separate models according to the enterprise function F1 and F2 and is shown in Figure 2. Consequently these functions may use or create information for entities types, which are quite similar. Looking at the logical and the physical layer one finds different applications and physical components reflecting the resources of each hospital provided for a particular function. This approach causes duplicated items (e.g., AC1, PC4, PC5 and PC1) and thereby conceals a clear view.

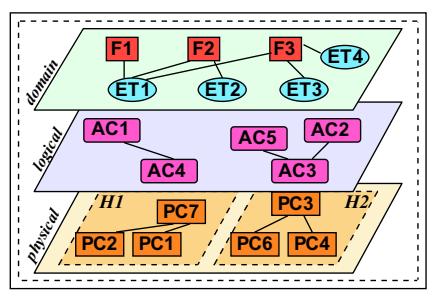


Figure 3 - Approach 3 with a model including all hospitals

The third approach focuses on avoiding the disadvantages from the ones above, however with the implication, that the definition of a physical component had to be extended to include a hospital as a grouping element (identified by the rectangles H1 and H2 in Figure 3. From this approach it becomes obvious, which enterprise functions are supported within the e-Health network (as depicted in the domain layer) and their relationship to application components.

Having modeled the inter-layer relations and dependencies an answer to a particular question could result in Figure 4, which maps the function F2 to the application component AC4 and the physical component PC6 as well as the function F3 to the application component AC5 and the physical component PC1. It is worth noting, that this approach would result in two paths if an identical enterprise function is performed between two hospitals in a cooperative session.

Since this approach provides a clear view on the e-Health network functionality it has been chosen for modeling.

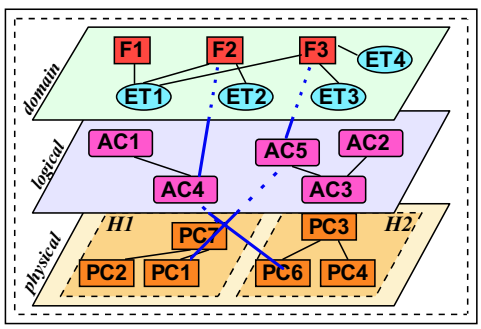


Figure 4 - Approach 3 together with inter-layer relations

Results

Figure 5 shows the model derived as a layered view. On the top layer the enterprise functions (teleradiology, telepathology, teleconference and communication/data security) have been detailed e.g. in order entry, image generation, reporting, second opinion, pseudonymization (patient identity replaced by a pseudonym) and secure transmission. Entity types identified are, for example, order, JPEG images, DICOM images, patient and tissue sample. The two types of images reflect two different usages: (i) JPEG for viewing only and (ii) DICOM for viewing and reporting at full image quality.

Application components are, for example, grouped by "Web-Viewer" and "DICOM Viewer" and hold sub-components like "Query", "DICOM Download", "Viewer". Finally the physical layer depicts a view on five hospitals with a detailed representation of their individual network infrastructure. The arrows between these hospitals and communication providers exhibit different routes between partners and clearly show potential and build-in fallback communications paths in case of a fault. Additionally, Figure 5 highlights a subset of the inter-layer relations modeled according to the selected function "emergency consultation".

Discussion

e-Health networks as well as health information systems exhibit quite some diversity reflecting heterogeneous institutional policies and communication links. Using the 3LGM tool this variety could be represented in the resulting model**.** Compared to typical network management tools [2,3], which fail to refer to applications, the 3LGM model widens the scope from enterprise functions and their relation up to network components. UML [15] diagrams allow to represent enterprise functions by means of use cases, to map processes to activity or sequence diagrams and to depict physical resources by deployment diagrams. However, UML diagrams tend to focus on specific views and provide a less integrated view on an information system with its internal dependencies and layers. Comparing 3LGM to business process oriented tools like ARIS[9] or Bonapart[10] which support simulation in time, the 3LGM model represents only static properties of the items of each layer. However, since the dependencies between the model items are known the 3LGM tool might be enhanced to support simulation for enterprise functions as well.

One of the main purposes was to perform an assessment towards the applicability of the 3LGM to an e-Health network Investigating three different approaches resulted in an approach with a coverage of all institutions in one model, but interpreting each hospital as a physical grouping component on the lowest layer. This is somewhat contradictory to the definition given with the 3LGM but simplifies the assignment of real physical components to each hospital. An alternative approach would have been to avoid the grouping and to assign real physical components by using a naming convention to a grouping for each hospital, which however would make the graphical representation less meaningful.

With the model being in place it assists in the management by various aspects: (i) effective documentation due to references included as attributes for each model item (ii) analysis functions to search for dependencies between model items, and (iii) support for error analysis and maintenance, in particular to derive fallback solutions, e.g., alternative communication paths. For enhancing the applicability towards (i) and (ii) the 3LGM would benefit from a version control mechanism.

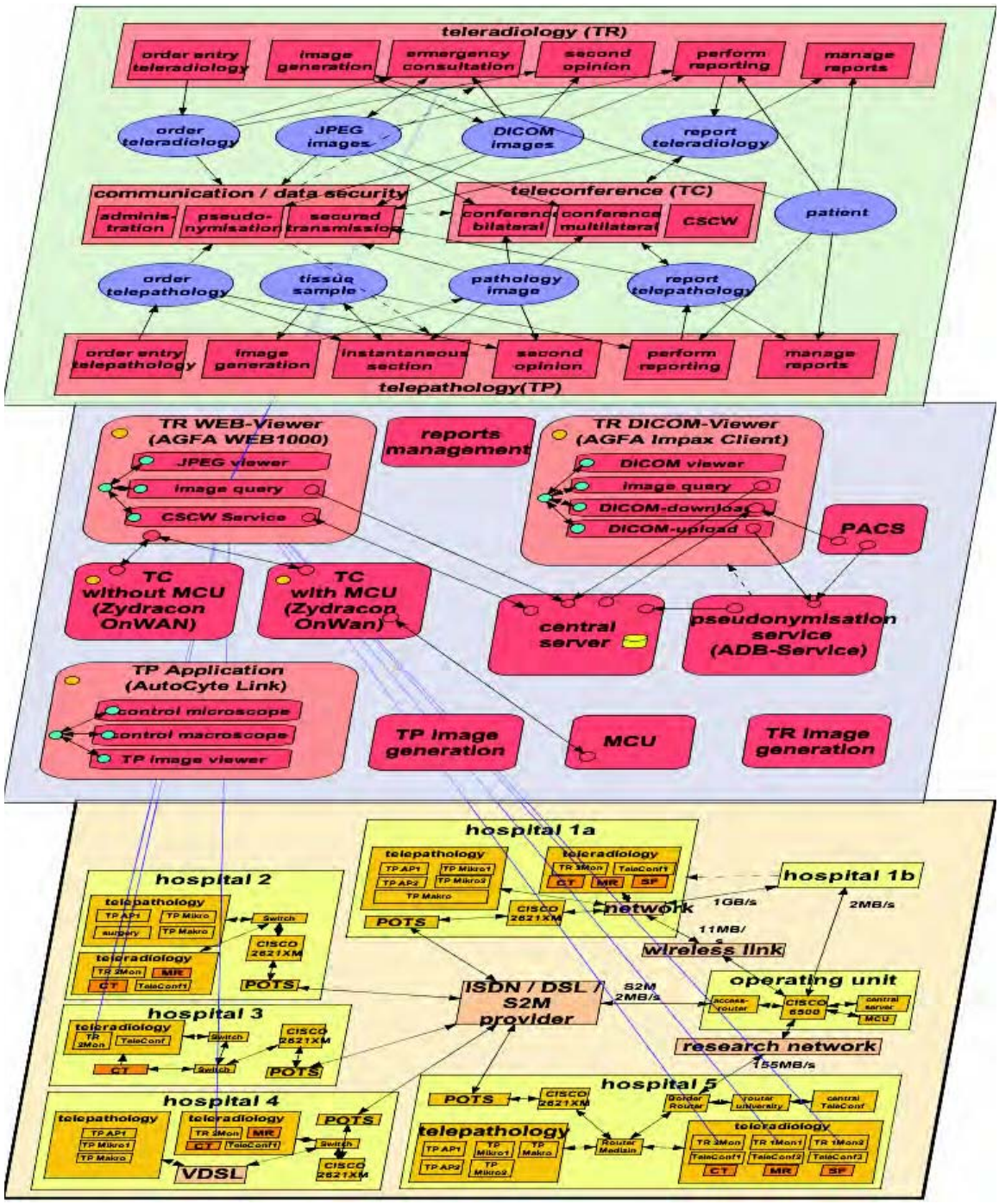


Figure 5 - The three layers as a result of the modeling

The initial effort to derive the model required a substantial amount of work, however in return, the model served both management and day-to-day maintenance. Current extensions of the network – like a dedicated sub-network between two hospitals – have been planned on the basis of this model resulting in an immediate documentation of the new enterprise function and its mapping to further and existing applications as well as to the physical layer.

Some minor handling problems, e.g., a limited workspace, missing grid lines, limited drawing functions and/or an adjustable level of detail did not have much impact on the effective work with the 3LGM.

Conclusion

The 3LGM successfully proved its applicability to model an e-health network. The model obtained serves for at least three aspects: strategic management of such an e-Health network, day-to-day maintenance and in depth documentation of the typical complex network setup. As such it contributes to the sustainability in the "routine operation" period and adds to a cost-effective support and operation.

Acknowledgments

The modeling would not have been possible without access to the project "Health Telematics Network supporting Tumor Diagnosis and Therapy in the Euroregio Pomerania" which has been granted by the European Commission for regional development (INTERREG A) together with the support of the federal state Mecklenburg-Vorpommern. Furthermore, the author wants to thank Lars Kubowicz for the fruitful discussions and the support in implementing the model.

References

[1] Stroetmann KA, Jones T, Dobrev A. Evaluating the Economic Impact of eHealth Applications – Approach and Method, http://www.ehealth-impact.org → download → documents → eHI_papers_malaga.pdf, (last accessed Dec. 4, 2006)

[2] http://www.ibr.cs.tu-bs.de/projects/scotty/ (last accessed Dec. 4, 2006).

[3] http://www.concord.com/products/network.shtml (last accessed Dec. 4, 2006)

[4] Bernus P, Mertins K, Schmidt G. Handbook on Architectures of Information Systems, Berlin, Springer, 1998

[5] Van de Velde R, Degoulet P. Clinical Information Systems, Health Informatics Series, New York, Springer 2003

[6] Ferrara FM. Healthcare Information Systems Architecture. Dudeck J, Blobel B, Lordieck W, Bürkle T. editors. New Technologies in Hospital Information Systems. Amsterdam, IOS Press 1997, 1-10

[7] HL7-RIM (Reference Information Model), http://www.hl7.org -> DataModels -> RIM (last accessed Dec. 4, 2006).

[8] http://www.omg.org/technology/documents/spec_catalog.htm (last accessed Dec. 4, 2006)

[9] Seidlmeier H. Process Modelling with ARIS, Wiesbaden, Vieweg, 2004

[10] Bonapart, Emprise Process Management GmbH, www.emprise.de (last accessed Dec. 4, 2006)

[11] Wendt T, Brigl B, Winter A. Modeling Hospital Information Systems (Part 1): The Revised Three-layer Graph-based Meta Model 3LGM. Methods Inf Med 2003, 42(5):544-51

[12] Wendt T, Häber A, Brigl B, Winter A. Modeling Hospital Information Systems (Part2): Using the 3LGM Tool for Modeling Patient Record Management. Methods Inf Med 2004, 43(3):256-67

[13] Staemmler M, Ehricke H, Dräger J, Hosten N. Interdisziplinäres Netzwerk zur Unterstützung der Tumorversorgung in der Euroregion Pomerania. Forum der Medizin_Dokumentation und Medizin-Informatik 2004; 1(6):14-18

[14] http://telemedizin-mv.fh-stralsund.de (last accessed Mar. 30, 2007)

[15] http://www.uml.org/ (last accessed Mar. 30, 2007)

Address for correspondence

Martin Staemmler, University of Applied Sciences, Stralsund, Medical Informatics, Zur Schwedenschanze 15, D-18435 Stralsund, Germany, martin.staemmler@fh-stralsund.de

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

The *open*EHR Java Reference Implementation Project

Rong Chen[a], Gunnar Klein[b]

[a] *Department of Biomedical Engineering, Linköping University, Sweden*
[b] *Department of Medicine, Karolinska Institutet, Sweden*

Abstract

The openEHR foundation has developed an innovative design for interoperable and future-proof Electronic Health Record (EHR) systems based on a dual model approach with a stable reference information model complemented by archetypes for specific clinical purposes.

A team from Sweden has implemented all the stable specifications in the Java programming language and donated the source code to the openEHR foundation. It was adopted as the openEHR Java Reference Implementation in March 2005 and released under open source licenses. This encourages early EHR implementation projects around the world and a number of groups have already started to use this code.

The early Java implementation experience has also led to the publication of the openEHR Java Implementation Technology Specification. A number of design changes to the specifications and important minor corrections have been directly initiated by the implementation project over the last two years. The Java Implementation has been important for the validation and improvement of the openEHR design specifications and provides building blocks for future EHR systems.

Keywords:

Electronic Health Records, Health Information Systems, *open*EHR, Java, open source.

Introduction

The *open*EHR foundation, a not-for-profit organization has since 2004 published a series of design specifications [1] for semantically interoperable and future-proof Electronic Health Records (EHR) systems based on more than 15 years of research and development projects in the EU and Australia. The most distinctive feature of the *open*EHR design is that it separates the clinical concerns and the technical design of EHR systems by an innovative design approach called two-level modelling [2]. The first level model takes care of the technical concerns and deals with the information structure and data types with a small set of information model classes, upon which the core of EHR systems can be built. The second level of model handles the concerns of the clinical domains, which are about how to represent and communicate the semantics of the clinical content. Because of this separation of concerns, stable EHR systems can be built without knowledge about specific clinical content necessary in different fields. The specification of clinical content can be authored and amended later, shared between health organizations since it can be loaded into EHR systems at runtime. The result is highly adaptive EHR systems and vastly improved semantic interoperability [3]. Such systems are sustainable since they can evolve as the clinical requirements changes.

A group of developers from Sweden specialized in EHR systems became inspired by the design ideas behind *open*EHR and started initial exploration in 2004. This led to the initial Java implementation of the core *open*EHR specifications. The group was invited to release the software under open source license so the work could be reused and possible enhanced by other developers. The group favoured the idea and announced the availability of the Java code initially released under GPL license. Later the same year, the software was donated to the *open*EHR Foundation and the license changed to the less restrictive MPL/GPL/LPGL triple-license to be compatible with the *open*EHR Foundation's licensing model. The action was welcomed by the *open*EHR community and the software was formally adopted as the Reference Implementation in the Java programming language in 2005.

Materials and methods

The Java implementation is implemented with the Java programming language with the Java 5.0 platform to take advantage of the "generics" and other new language features. Several well-established open source libraries have been used, e.g. log4j and commons-lang from the Apache Software Foundation. The implementation does not make any assumption of where and how the components would be used, e.g. client or server side application. A principle of the Java implementation is to be as faithful as possible with the specification while keeping the Java look-and-feel. Being faithful with the original design specifications means that the complete semantics of the design should always be correctly implemented in the Java language to ensure interoperability with *open*EHR systems implemented in another programming language. It involves mapping between *open*EHR assumed data types and native Java types.

One of the difficulties was how to properly handle invariants checking in the class specifications. It is well-known

that Design by Contract is not natively supported by the Java platform although there are 3rd party libraries aiming to support it in Java. The same effect is achieved through the use of immutable Data Value object pattern with incoming parameters checking by the object constructors. Keeping the Java look-and-feel makes the software familiar to Java developers and involves use of the standard naming convention and common Java idioms. The details of the work have been documented and explained in the *open*EHR Java Implementation Technology Specification (ITS) [4] and the design document of the Java Implementation.

The software from the Java Reference Implementation project is released under open source licenses, namely Mozilla Public License (MPL), Gnu Public License (GPL) and Less Gnu Public License (LGPL). The users can chose any one of these three that suits them best. It is worth to mention that among these three, MPL is the least restrictive and GPL is the most prohibitive for commercial use. This means that the even commercial suppliers can use this software as part of their offerings without any commitment to *open*EHR. The software is freely accessible through Internet and all changes are public and accessible through the source repository. Software bug fixes and patches contributed by developers are verified and committed to the source repository.

A set of rules have been applied in the early stage of the project to guarantee the quality of the software. Unit testing is used extensively to make sure the implementation meets the intention of the design. It provides the extra value as further documentation beside the specification since it exercises the code in real use scenarios. Unit testing is usually integrated into the build script so that the test cases can be executed by anyone at any time. Since it can be automated as part of the build cycle and run often, unexpected errors can be caught when they occur. The development guidelines require that each bug fix in the production code should be accompanied with a minimum test case that demonstrates the bug before the fix and gets passed when the bug is fixed. This so called regression test provides a safety net to the code base to make sure the old bugs will not re-appear in the subsequent versions.

The software build is managed by Maven from the Apache Software Foundation. It provides comprehensive support of software projects and handles software library dependency particularly gracefully. Unit testing of the software is seamlessly integrated with build. A Continuous Integration tool called Continuum also from Apache is used to execute the Maven build script directly for monitoring the source code changes in the source repository and make automated software builds.

The Java implementation projection software is currently change controlled by a Subversion repository hosted on the *open*EHR site. Like CVS, Subversion is an open source product for managing software changes and favoured by many open source software projects around the world. The URL to reach the root directory of the Java project in Subversion is http://svn.openehr.og/ref_impl_java.

Results

Table 1 - Software artifacts provided by the Java Reference Implementation Project

Component Name	Implemented Specifications	Brief Description
openehr-rm	*open*EHR Reference Information Model – Common, Support, Data Types, Data Structures, Demographics, EHR	Base component that provides the Java implementation of all *open*EHR Reference Information Models
openehr-aom	*open*EHR Archetype Object Model	Object validation and construction
openehr-ap	*open*EHR Archetype Profile	Implementation of the domain data types
adl-parser	*open*EHR Archetype Definition Language	Translates ADL to AOM
adl-serializer	*open*EHR Archetype Definition Language	Serializes AOM to ADL
rm-builder	*open*EHR semantics	Reference Model objects builder

The software from the Java implementation is divided into several software components each with its own clearly defined responsibility. The openehr-rm, which stands for *open*EHR Reference Model (RM), is the base component that implements all Reference Information Model specifications, namely Data Types, Data Structures, Common, Support, EHR Extract, and Demographics. It maps *open*EHR assumed data types into Java native data types and implements other higher level data types. It also implements path based queries for finding leaf nodes in a large object tree with single crafted paths. All constructors of the RM classes have annotations for the parameters so that the invocation of object construction can be done automatically. This feature is used by the rm-builder for RM objects construction with Archetypes. This component is the fundamental building block for all subsequent components. According to the *open*EHR two-level modelling approach, the core of EHR systems can be built entirely based on the RM components, e.g. EHR data query and storage mechanism can be built on top of the RM component.

We also implemented the openehr-aom and openehr-ap, which stands for *open*EHR Archetype Object Model (AOM) and *open*EHR Archetype Profile, respectively. These components provides in-memory representation of Archetypes, which are usually authored and transmitted in Archetype Definition Language (ADL) format but parsed into AOM in-memory format to be used in runtime. They also provide support for object validation and creation on single archetype constraint level.

The rm-builder component is based on the openehr-rm, openehr-aom and openehr-ap components and provides Archetype based object validation and creation support. The rm-builder guarantees that the semantics in the archetypes are faithfully reflected in the process of RM objects validation and construction. The implementation of RM, AOM and the support for Archetypes based validation and construction is often referred to collectively as the *open*EHR Kernel.

The adl-parser component implements Archetype Definition Language and transforms Archetypes in ADL textual format into in-memory AOM form. It is built on a JavaCC, which generates the parser based on a grammar file. The grammar file is directly translated from the ADL specification with necessary changes. The adl-parser is the entry point for Archetypes into any EHR system so it is of vital importance for the correctness of the behaviour of an Archetype based system. It can also be used, e.g. in an Archetype editor, an authoring environment for Archetypes to check the correctness of the Archetypes in ADL format.

The adl-serializer provides conversion of Archetypes from in-memory AOM form to textual representation in ADL format, the reverse of what the adl-parser does. It is often used before an authored archetype can be stored and transmitted between systems.

Java implementation technology specification

The Java implementation experience has been summarized and written up as the Java Implementation Specification (ITS) of *open*EHR design, [4] which is now included as part of the official release of the *open*EHR specifications. The document is still in beta version and being updated to reflect the feedback from the implementation activities. The purpose of this ITS is to provide guidance on perspective implementation of the *open*EHR design on the Java platform and make sure the semantics of the design is kept faithful in the Java programming environment. More importantly, it should also ensure that *open*EHR systems from different computing platforms are interoperable.

During the last two years, a number of design improvements have been directly initiated from the Java implementation project and indirectly from other development projects that are based on the Java reference implementation components. A number of additional minor issues, e.g. inconsistency in the documentation have also been raised, particularly during the first stage of the implementation when the specifications were just released before version 1.0.

Table 2 - Selection of specification changes accepted by openEHR due to the Java implementation

Related Specification	Description
Data Structures	EVENT should inherit from LOCATABLE.
Data Types	Date/Time classes improvements
Archetype Object Model	Missing adl_version in ARCHETYPE
Archetype Object Model	Table for missed class ASSERTION_VARIABLE added. Assumed_value assertions corrected; standard_representation function corrected
*open*EHR Archetype Profile	Fix invariants in C_QUANTITYclasses. Correct C_QUANTITY.property to CODE_PHRASE. Correct invariants for C_CODED_TEXT; correct inheritance for C_DV_ORDERED. Corrected C_QUANTITY_ITEM class. Corrected errors in DV_STATE model by adding 2 new classes
Most of the models	Use constants instead of literals to refer to terminology in RM, mainly used in the invariants part of the class descriptions
Common, EHR	Make DIRECTORY reusable, add new directory package
Demographic	Remove details /= Void invariant from PARTY
Archetype Definition Language	Numerous clarifications of the ADL syntax and semantics

Some design issues were not obvious initially but became apparent when more implementation experiences were gained. For instance, the modelling of EHR package initially included support for a directory based way of organizing records. When the requirement was identified to support a good way of organizing demographic objects, it seemed natural to extract the generic directory support into a more generic package so it can be re-used for organizing objection other than EHR compositions.

Sometimes, implementation in the programming language other than the one used or envisioned by the original designer can lead to improvement of the design itself. One example of this is within invariants checking, present in many places of the specifications, built-in terminology names are referenced in plain text format. This is recognized as an anti-pattern among Java developers since such way of handling textual constants would lead to more brit-

tle software since any deviation, e.g. a typo of the terminology name will not be caught by the compiler during software compilation phase but only show up during system runtime. A better way of doing that is to introduce a type safe enumeration pattern. This design improvement has been well taken and incorporated into Release 1.0.1.

Building blocks for EHR applications

The original Swedish team has built a pilot Child Leukaemia Treatment management application using the Java components for Karolinska University hospital in Stockholm. It is a web-based system designed to facilitate chemotherapy management for child Lymphoblastic Leukaemia patients in maintenance stage. Unlike conventional EHR applications, the clinical content is defined by archetypes and the screens are generated dynamically.

The Medical Informatics group from Linköping University has built a number of applications based on the Reference Java Implementation and made substantial contributions to the Java project. The Java Archetype Editor [5] has particularly received good comments and is now released under open source license. A web-based patient portal application has been built as part of a master student's thesis work. An ongoing EHR visualization project [6] uses the Java implementation as a way of representing EHR data.

A Biomedical Engineering Group from Uruguay funded by the Ministry of Education is building an Intensive Care Unit application, which is based on the existing Java components. The system covers admission to ICU, clinical variables monitoring, laboratory requests and tracking, instructions on medication, nutrition, fluids and tracking of instructions status, clinical logbook for each patient. The system will be installed for a one year field test in three ICUs in Uruguay during 2007.

In the Netherlands, a team of developers is building a commercial solution to support elderly care based on the openEHR Java implementation. The system supports a wide range of users from General Practitioners (GP), nurses, to elderly / disabled people and other involved heath professionals. The solution is particularly favoured by the customers since it is based on public specifications and open source software.

A research team from the Health Informatics Group, Central Queensland University, Australia is building a new version of the Archetype Finder that uses the Java ADL parser.

Since the software was released as open source in 2004, it has been constantly updated and maintained by the Java Reference Implementation team. Developers around the world have been giving feedback, bug reports and software patches resulting in overall increasing quality of the code base over time. The initial release was based on the specification release 0.95. During the last six months, the team has spent considerable effort to update the implementation towards the latest specification 1.0.1. At the same time, the code base has been re-organized to facilitate re-use and improve testability even further.

Discussion

Nowadays health information standards are becoming increasingly complicated. This makes it very difficult to rely on a textual document only for the description of the semantics of the design. If the intention of the standard is misinterpreted, the risk of creating non-interoperable software is very high. Unfortunately the formal standardization bodies CEN and ISO have not yet allowed publication of standards artifacts in forms other than human readable documents. Data structures intended for use by software development tools can sometimes be distributed as informative associated material but not as the core of the standard. Open source reference implementations can be very useful to ensure a common and complete understanding of a written specification facilitated by the fact that such software like ours is usually developed in close collaboration between the specification designers and the implementers. The software can be more detailed and precise than the textual specification and be easily understood and put to production by target developers.

The fact that a particular implementation is done on a specific platform means not only that it proves the specification is possible to implement but also that the implementation details have been worked out for that specific platform. There are a large number of similar examples in the IT industry. The Java specifications are released together with an open source reference implementation sometimes from Sun itself but also from open source organizations like Apache.

Innovative ideas need to be experimented with and verified before they become part of a formal specification. The IT industry has seen too many “committee-made” design specifications, which can be very hard to implement and truly annoying to work with if they unfortunately become formal standards and last for years. So it is really important to get some real implementation experience and let the feedback from that guide the refinement of the design specifications. Because open source implementations are readily accessible by large user base and perspective implementers, they are particularly suitable for validating new design ideas. The *open*EHR community, which consists of both clinical professionals from many specialities, technical developers and also medical informatics researchers, is a particularly good example of this. With experience from international standardization activities and open source software projects, the authors argue that an open source reference implementation is an ideal way of ensuring the quality of international standards for communication.

The open source licenses used by the Java Implementation Project are very relaxed, which means that they allow commercial applications to be built on top of these components. These licenses effectively remove entry barriers for small companies and academic researchers that may not have the resources to implement the design from scratch. If more companies choose to use this solution, it will undoubtedly increase the interoperability between EHR systems. Because of the availability of the open source

implementation, more medical informatics research activities could be performed in connection with EHR data e.g. Decision Support and Epidemiological statistics. The ground breaking design proposed by *open*EHR opens a lot of opportunities, which were not possible previously due to poor semantic interoperability between systems and brittle system design. One well-known problem in Decision Support Systems (DSS) is that the data query mechanism is dependent on the information model of the local EHR system and has to be adapted each time the DSS is integrated with a new EHR system. This issue is effectively resolved if *open*EHR two-level models are used since DSS can query an EHR system with a standardized reference information model with the aid of clinical content definitions explicitly expressed in Archetypes. These research areas are of high interest at the moment, and collaboration on these using the open source implementations are ongoing.

The open source implementation creates synergies between EHR system developers for collaboration on common EHR software components. More access to end user base and beta testers, and focused development efforts on common parts of EHR systems will most likely result in more reliable and re-usable software components that are building blocks of future EHR systems. With enough core building blocks, a common health information platform can be established, which provides essential features for any full-blown EHR system, e.g. standards compliant interfaces and semantic interoperability.

Even the application level presentation layer can benefit from generic screen form generation components [7]. EHR applications could then be built rapidly on top of the common platform with clinically defined archetypes and certain application level customizations in ways previously unimagined. Also the EHR application development can focus on solving application related problems rather than generic EHR problems. We believe this would result in a paradigm shift in how to develop EHR applications in the near future, which would dramatically increase the productivity of EHR development and therefore lead to more sustainable solutions [8] than the current way of building EHR systems.

Conclusion

The *open*EHR Java Reference Implementation project has been very useful for validation and improvement of the *open*EHR design specifications. The Open Source Software development model used by the project fits very well with the purpose of publicly available specifications. This is a sustainable way of providing major healthcare standards for interoperability that we argue should be used more.

The project creates synergies between EHR developers and encourages software re-use aiming to provide building blocks for EHR applications. Not only would this reduce costs of the software and lead to more reliable health information systems, but also improved interoperability between systems. We believe this can contribute to the overall sustainability of health information systems.

Acknowledgements

We want to firstly thank the *open*EHR Foundation for their pioneering work and Thomas Beale in particular for his invaluable guidance on the implementation work and review of the manuscript. We also want to express our appreciation of the UCL team for good collaboration and Yin Su Lim in particular for her meticulous efforts, the Linköping Medical Informatics group. and last but not least, all the early *open*EHR implementers who have tested the Java components and came back with feedback and patches. We are also grateful for the financial support for some of this work from the Semantic Mining EU- project and from Cambio Healthcare Systems.

References

[1] Beale T, Heard S, Kalra D, Lloyd D. Architecture Overview. http://svn.openehr.org/specification/BRANCHES/Release-1.1-candidate/publishing/architecture/overview.pdf [cited 2006 December 03]

[2] Beale T. Archetypes: Constraint-based domain models for future-proof information systems. In: Eleventh OOPSLA Workshop on Behavioral Semantics

[3] Garde S, Knaup P, Hovenga E, Heard S. Towards Semantic Interoperability for Electronic Health Records: Domain Knowledge Governance for openEHR Archetypes. Accepted for Methods of Information in Medicine 2007

[4] Chen R. openEHR Reference Model Java ITS. http://svn.openehr.org/specification/BRANCHES/Release-1.1-candidate/publishing/its/Java/openEHR-JavaITS.pdf [cited 2006 December 03]

[5] Sundvall E, Qamar R, Nyström M, Forss M, Petersson H, _hlfeldt H, A Rector. Integration of Tools for Binding Archetypes to SNOMED CT. Proceedings of SMCS2006; 1-3 Oct, Copenhagen, Denmark, p 64-68

[6] Sundvall E, Nyström M, Forss M, Chen R, Petersson H, _hlfeldt H. Graphical Overview and Navigation of Electronic Health Records in a prototyping environment using Google Earth and openEHR/Archetypes. Accepted by Medinfo 2007

[7] H van der Linden, Schuler T, Chen R, Talmon J. Generic screen representations for future proof systems, is it possible? Accepted by Medinfo 2007

[8] Garde S, Hullin CM, Chen R, Leslie H, Heard S, Schuler H, Gränz J, Knaup P, Hovenga E. Towards Sustainability of Health Information Systems: How Can We Define, Measure and Achieve It. Accepted by Medinfo 2007

Address for Correspondence

Rong Chen,
Dept. of Biomedical Engineering,
Linköping University,
SE-581 85 Linköping,
Sweden.
E-mail: rong.chen@imt.liu.se

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

A Conceptual Model of Computerised Hospital Information System (CHIS) use in South Africa

Lyn A Hanmer[a], Sedick Isaacs[b], J Dewald Roode[c]

[a] *Medical Research Council, South Africa*
[b] *Consultant Statistician, South Africa*
[c] *Department of Information Systems, University of Cape Town, South Africa*

Abstract

The aim of this project is to contribute to limiting the risk of CHIS (computerised hospital information system) failure by identifying factors which are associated with the successful implementation of CHISs in district and regional hospitals in South Africa (SA). Case studies were conducted in four regional hospitals in the Western Cape province of SA to obtain data about user perceptions of the success or lack of success of the CHISs in use. A conceptual model of CHIS use has been developed based on the results of the case studies, in order to assist in the interpretation of the differing experiences across the hospitals. Key factors in the conceptual model are perception of usefulness of the CHIS and management commitment to ensuring CHIS success, which in turn are related to effective use of CHIS and/or CHIS outputs, and allocation of resources for CHIS further development. Further development of the model will be influenced by the next phase of this project: a survey of district and regional hospitals in two SA provinces.

Keywords

Hospital information system, evaluation, model

Introduction

Increasing numbers of computerised hospital information systems (CHISs) are being implemented in public hospitals in South Africa (SA), as part of a concerted effort to improve the data available to decisionmakers at all levels of the SA public healthcare system [6]. A proportion of these hospitals is situated in environments with limited access to the resources required to run CHISs, including skilled support personnel for hardware and software, and internal resources in the form of trained personnel in various roles to enable hospitals to obtain maximum benefit from the CHIS. Vulnerability to disruption of services can result from lack of direct access to people with the required skills to support the CHIS software and hardware, since support staff may be based in a central location and have to travel significant distances to reach hospitals which may require their services. Where the required physical infrastructure is available, services can be subject to disruption, sometimes due to lack of resources (including human resources) for maintenance and repair. The successful implementation of CHISs under these conditions is a particular challenge. District and regional hospitals in environments of limited or vulnerable resources, as described, are the focus of this project.

To date, there have been few published evaluations of the CHISs which have been implemented in SA. However, the results of some of the evaluations which have been conducted [15, 17], as well as anecdotal evidence, indicate that problems have been and are being experienced with some of the CHISs which have been implemented. In the Limpopo Province (previously the Northern Province) two of the major reported problems were system performance, and system support, especially in locations remote from the main centres. In some provinces, system implementation and rollout have taken much longer than anticipated, due to problems with customisation, system performance and change management among others. Since the monetary costs of CHISs are generally high in relation to the available budget, and CHIS implementations can require significant organisational change, the implementation of systems which do not meet expectations has major implications for hospitals [21].

Decisions about the acquisition of CHISs for public hospitals in SA are made at provincial level, following a formal tender process, and the same CHIS is typically implemented at multiple hospitals in a province. All new CHISs acquired for use in the SA public healthcare sector have to conform to a common set of requirements which have been agreed to at national level. While in practice at present there is more than one CHIS in use in some provinces, several provinces (including Limpopo and the Western Cape, for example) have already decided to move towards the implementation of a single CHIS in the province, with variations in the scope of the implementation depending on the needs of different hospitals [6,7].

The development of criteria for the selection of a CHIS is particularly challenging when the same system is being selected for use in specialist referral hospitals and in district and regional hospitals, since the district and regional hospitals typically have access to fewer resources for CHIS implementation and maintenance than do the specialist referral hospitals. Current criteria used to inform the selection of CHISs for implementation in public sector hospitals in SA do not explicitly address factors which

could affect (positively or negatively) the chance of a future successful implementation [10].

Related projects

There are few practical and/or theoretical guidelines available to support decisionmaking about CHISs. One notable exception to this lack of guidelines is the set developed by Heeks and colleagues, who, in a paper on why health information systems succeed or fail, provide an important analysis of factors which have the potential to affect CHIS success [9]. They use the concept of conception-reality gaps to describe the extent to which the aims of an information system can be matched with the reality of the environment in which it will be implemented.

The literature on IS success and HIS (health information system) evaluation has provided the major theoretical basis for this project. The IS success literature reflects an approach to theory development which is based largely on analysis of relevant empirical studies, with the work of DeLone and McLean [4,5] forming the basis for much further theory development.

In the HIS domain, much of the work has been based on qualitative case studies, in some cases supported and/or supplemented by lessons from the IS success literature. Some of the work on success of HIS is reflected in the literature on evaluation of HIS. The emphasis in much of the evaluation literature is on the organisational aspects of HIS implementation, including the essential role of people as system users and system managers, and is therefore very relevant for this study [1,8,16,18,21,24].

The model of HIS implementation developed by Korpela and colleagues for the INDEHELA-methods project [12] and the ITPOSMO model of Heeks and colleagues [9] are among the models reviewed to date which specifically take account of context, although neither of them explicitly addresses the issue of limited and vulnerable resources for IS and HIS implementation and use. The INDEHELA-context project is aimed at comparing experiences between developing and developed environments [12]. Braa and Hedberg [3] and Shaw [20] report on the HISP project, which specifically addresses approaches to IS development which aim to ensure sustainability of systems in developing environments, with limited infrastructure and other resources.

Problem identification

There is a need to provide support for decisionmakers to reduce the risk of lack of success in the acquisition and implementation of CHISs for district and regional hospitals in South Africa. A model of CHIS use which is applicable to environments of limited or vulnerable resources could provide decisionmakers with a framework for assessing the potential for success of CHISs in these hospitals.

There is a general lack of 'gold standards' against which to measure the performance of HISs in hospitals [23]. This is related to the more general problem of describing, explaining and measuring information system (IS) success [2,5,19]. In their revised IS success model, DeLone and McLean [5] identify the extent to which the IS is used (or there is an intention to use it) as being one of the factors related to IS success. LeRouge and colleagues [14] have extended the DeLone and McLean concept of system use by analysing use quality as a factor in IS success. Of particular interest for this project is the extent to which the CHIS outputs are used by hospital- and provincial-level managers to support their planning for and management of hospitals. The conceptual model developed in this project describes factors associated with CHIS use, since one essential component of CHIS success is that the system, or its outputs, should be used by the intended users.

Methods

The aim of this project is to contribute to limiting the risk of CHIS failure by identifying factors which are associated with the successful implementation of CHISs in SA public sector district and regional hospitals.

Case studies

Case studies were conducted at four public sector regional (level 2) hospitals in the Western Cape province.

The objectives of the case studies were

- to describe and analyse the effects on the hospital of CHIS implementation
- to identify those factors which are associated with perceptions of the success or lack of success of the implementation, and
- to identify those factors which could be associated with the success or lack of success of the implementation.

A combination of observation of the CHIS in use and semi-structure interviews with representatives of hospital management (clinical, nursing and/or administrative), any specialist information management personnel, case managers responsible for co-ordinating services for any private patients in the hospitals, and CHIS end users was used to obtain data for the study. A standard set of questions was used as a guide for all interviews. Between four and eight interviews were conducted at each study hospital. The case study protocol and the interview guide were made available to the management of the study hospitals in advance.

Study hospitals were selected on the basis of accessibility, and in order to obtain a set of results which is broadly representative of conditions in district and regional hospitals in SA in which a CHIS has been implemented. The study hospitals had approximately 400 beds each, with similar management structures. Two of the hospitals are situated in a major urban centre, the third is approximately 100 km from the urban centre, and the fourth is approximately 400 km away.

Since the aim of successive cases in a qualitative study like this one is to extend and provide richness to the information gleaned from earlier studies, the fourth case study hospital was selected in order to assess the effect on the operation of the CHIS in a hospital which is so far from the support centre for both system software and hardware that

the distance could affect the success of the CHIS implementation.

The scope of the fourth case study was further extended to examine the use of the CHIS to support the work of clinical managers in the hospital, i.e. clinicians responsible for the management of clinical services such as family medicine and emergency medicine. This issue was not addressed in the other case studies.

All study hospitals were using the same CHIS with the same scope, and the CHIS had been in use at each hospital without major changes for at least six months. The CHIS included ADT (admission/discharge/transfer) and billing functions. Van der Loo and colleagues [22], in developing a framework for the evaluation of healthcare information systems (HISs), formulated a useful classification of the health care process. In terms of this classification, a distinction is drawn between care processes – divided into the medical care process and the supporting process – and auxiliary processes, which do not contribute directly to the care process. In terms of this classification, the CHISs typically in use in SA district and regional hospitals support aspects of the supporting process (ADT component) and an auxiliary process (billing).

Results and discussion

Comparisons between the hospitals in order to identify similarities and differences between the environments and the experiences of implementing the CHIS provided a rich picture of the use of the CHIS in the hospitals, and the perceptions of users about the CHIS.

The experience at all hospitals was that the CHIS is generally stable and reliable. Since CHIS hardware and software support is only available during office hours (although all the study hospitals provide 24-hour emergency services), delays had been experienced when problems had arisen. Delays had also been experienced due to limited availability of software support personnel, even during office hours. At the remote study hospital (400 km from the service centre) the management expressed serious concerns at their vulnerability due to the distance from the service centre, although in practice major problems had fortunately seldom arisen.

At all hospitals, there was concern about the quality and completeness of the data being input to the CHIS.

All the hospital managers interviewed were clearly aware of the importance of information as a resource for decisionmaking, and practical efforts were being made to use data from the CHIS in combination with data from other information systems to support management decisionmaking. However, managers were also concerned that they had only a limited understanding of the potential capabilities of the CHIS, and few if any opportunities to gain a better understanding of the CHIS. Due to the limited collection of clinical data via the CHIS, the available clinical data was not used at all by clinical and nursing personnel.

Despite these concerns, there were plans at all the study hospitals to extend the usefulness of the CHIS by installing terminals in wards at three hospitals, and by acquiring an additional module of the system at one of the hospitals.

The most striking differences between the hospitals were in management approaches to the allocation of resources for ensuring data quality and completeness in the CHIS, and the extent of reliance on data from the CHIS for reporting within and beyond each hospital.

Conceptual model of CHIS use

A conceptual model of CHIS use has been developed, based on the results of the case studies and theoretical models reviewed to date, including those of DeLone and McLean [5], Ballantine et al [2] and Heeks, Mundy and Salazar [9]. The aim of this conceptual model is to clarify those issues identified during the case studies which have the potential for explaining differences between the experiences of CHIS implementation in the four hospitals.

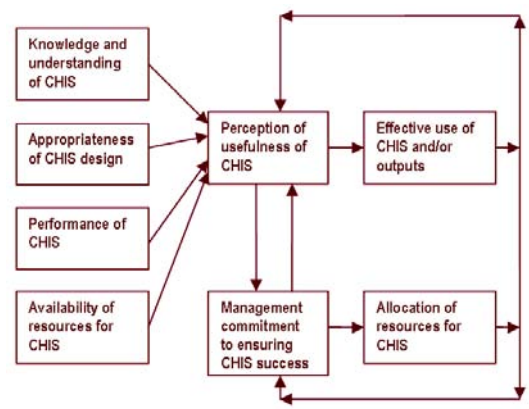


Figure 1 - Conceptual model of CHIS use

The model is based on the premise that user perception of the usefulness of a CHIS is a key determinant of whether or not the system will be used effectively (see figure 1). Factors which affect user perceptions of the usefulness of the CHIS have been identified as **knowledge and understanding of the CHIS, **appropriateness of CHIS design **CHIS performance and **availability of resources for CHIS.

At a feedback session with the interviewees at one of the case study hospitals, and during related interviews with key informants outside the study hospitals, all comments related to CHIS use at hospital level could be assigned to one of the factors identified in this conceptual model.

Perception of usefulness

The attitude of users is a key element in determining whether the potential of a computerised health information system (CHIS) implementation is realised. If users believe that a CHIS is useful for them, they will make an effort to ensure that the system works, and will use the outputs from the system. Conversely, if there is a perception that a CHIS is not useful, there will be little or no commitment by users to ensuring that the system is used correctly, and outputs from the system will not necessarily be used, especially when similar information can be obtained from other sources.

For example, the case managers at two of the hospitals, responsible for ensuring that fees were collected for private patients, were particularly enthusiastic about the potential of the system to provide the data required to enable them to ensure that accurate accounts were raised. They were frustrated by the limited understanding of some of their colleagues about the need for accurate and complete data input, and by limitations in their own understanding of the functionality of the CHIS, and were looking for ways of improving CHIS use, in order to improve the benefits gained from the system.

Effective use of CHIS and/or outputs

One measure of the effectiveness of a CHIS is whether it is used effectively. DeLone and McLean [5] use the concepts of 'use' and 'intention to use', leading to the achievement of 'net benefits' from the system to describe IS success. Heeks, Mundy and Salazar ([9] have identified information needs of the users as one of the dimensions along which a mismatch between the design of a health care information system and the reality of the environment in which it is implemented could occur. If the information needs of users are not met by the system, it is unlikely to be effectively used, and therefore could not be regarded as a success.

Management commitment to ensuring CHIS success

The context in which the CHIS has been implemented is reflected by including 'management commitment to ensuring CHIS success' as a component of the model. Allocation of resources for implementation, training and maintenance of the system is modelled as the main indicator of management commitment prior to and during system implementation. The allocation of further resources for system development would be a reflection of ongoing management commitment to the use and development of the system.

There were striking differences between the study hospitals in the allocation of personnel resources to ensuring the effective use of the CHIS: The medical superintendents of two of the hospitals had assigned specific responsibility for the preparation of management reports, based on data from the CHIS and from other systems, to full-time clerks responsible directly to them. At the third hospital, the responsibility for preparation of management reports was divided among members of the management team, and the staff member responsible for reception and fees office staff. At this hospital, the hospital manager, the most senior administrative staff member, was responsible for the final collation of hospital reports required by the provincial department of health. At the fourth hospital, an information manager had been appointed at senior management level to take overall responsibility for all reporting for the hospital. Comparing the study hospitals, the degree of user satisfaction with the CHIS seemed to correlate directly with the allocation of personnel time to ensure accuracy and completeness of the data on the CHIS. At the hospital at which there was no full-time person responsible for data management, there was the greatest degree of dissatisfaction with the CHIS, and the impression gained was that there was little management commitment to CHIS success.

Relationship between the user and management components of the model

'Management commitment' influences 'perception of usefulness' in this model. Since 'perception of usefulness' reflects user attitude to the CHIS, this link provides a connection between management and user attitudes to the CHIS.

'User perception of usefulness' could persuade management to commit resources to ensure the effective operation of the CHIS. This could also reflect the attitude of members of hospital management who use the CHIS (either directly or indirectly, via CHIS reports). This direction of influence is reflected in the model by having a two-way arrow between these two components of the model.

Conclusion

On a practical level, the aim of this project is to inform decisionmaking about the selection of CHIS for use in district and regional hospitals in SA, by identifying factors associated with the effective use of CHISs in these environments.

On a theoretical level, this project aims to contribute to the theoretical basis for evaluating CHISs and, hence, the evaluation of ISs in general. An attempt has been made to synthesise findings from multiple environments which are applicable to the complex environment of computerised information systems in hospitals, as has been recommended by Kukafka et al [13] and van der Meijden et al [23], among others. The recommendations consistently made by Kaplan [11], Lorenzi and Riley [16] and Southon et al [21], for example, that organisational and people issues must be taken into account in analyses of the use of information systems in health care, and in other environments, have been followed.

Further plans for this project include the refinement and extension of the conceptual model, based on the results of a planned survey of CHIS implementations in district and regional hospitals in SA. The aim of the survey is to test the conceptual model, and to determine the relative significance of the factors identified in the model.

Acknowledgments

This project is funded by the Medical Research Council. The support of colleagues at the study hospitals and in the Provincial Government Western Cape: Department of Health is gratefully acknowledged.

References

[1] Ammenwerth E and de Keizer N. An inventory of evaluation studies of information technology in health care: Trends in evaluation research 1982 – 2002. In: Feischi M, Coiera E, Li Y-C J, eds. MEDINFO 2004. Proceedings of the 11th World Congress on Medical Informatics. Amsterdam: IOS Press, 2004; pp 1289-1294.

[2] Ballantine J, Bonner M, Levy M, Martin A, Munro I and Powell PL. Developing a 3-D Model of Information

Systems Success. In: Garrity EJ, ed. Information Systems Success. Idea Group Publishing, 1998; pp. 46-59.
[3] Braa J and Hedberg C. Developing district-based health care information systems: the South African experience. In: Proceedings of the IFIP WG9.4 Conference on Information Flows, Local Improvisations and Work Practices, Cape Town, South Africa. Computer Society of South Africa, 2000.
[4] DeLone WH and McLean ER. Information systems success: The quest for the dependent variable. Information Systems Research 1992: 3(1): 60-95.
[5] DeLone WH, McLean ER. The DeLone and McLean Model of Information Systems Success: A Ten-Year Update. J Man Inf Systems 2003: 19(4): Spring 2003: 9-30.
[6] Department of Health. National Health Care Management Information System of South Africa NHC/MIS, 1995. http://www.doh.gov.za/nhis/ (accessed May 2006)
[7] Department of Health and Social Development. Tender for the supply, implementation and maintenance of the provincial health information system (PHIS). Tender number HEDP 694/05. Department of Health and Social Development, Limpopo Province, South Africa, 2005.
[8] Hanmer LA. Criteria for the evaluation of district health information systems. Int J Med Inf 1999: 56: 161-168.
[9] Heeks R, Mundy D, and Salazar A. Why Health Care Information Systems Succeed or Fail. University of Manchester Institute for Development Policy and Management, 1999: Working Paper Series no 9. http://www.sed.manchester.ac.uk/idpm/publications/wp/igov/igov_wp09.htm (accessed May 2006)
[10] Isaacs S. Choosing an information system for a health care facility. Presented at HELINA 1999. 3rd International Conference on Health Informatics in Africa. Harare, Zimbabwe; unpublished.
[11] Kaplan B. Evaluating informatics applications – some alternative approaches: theory, social interactionism, and call for methodological pluralism. Int J Med Inf 2001: 64: 39-56.
[12] Korpela M, Hanmer L, de la Harpe R, Macome E, Mursu A, Soriyan HA. How can African healthcare facilities get appropriate software? Sociotechnical research in the INDEHELA-context project. Presented at Second International Conference on IT in Health Care: Sociotechnical Approaches (ITHC 2004), Portland, Oregon, USA, 13-14 September 2004. http://www.uku.fi/tike/indehela/inde-ITHC2004.pdf (accessed May 2006).
[13] Kukafka R, Johnson SB, Linfante A, Allegrante JP. Grounding a new information technology implementation framework in behavioral science: a systematic analysis of the literature on IT use. J Biomed Inf 2003: 36: 218-227.
[14] LeRouge C, Hevner AR and Collins RW. It's more that just use: An exploration of telemedicine use quality. Decision Support Systems 2006: In press.
[15] Littlejohns P, Wyatt JC, Garvican L. Evaluating computerised health information systems: hard lessons still to be learnt. BMJ 2003: 326: 860-863.
[16] Lorenzi NM, Riley RT. Organizational issues = change. Int J Med Inf 2003: 69 (2-3): 197-203.
[17] Mbananga N, Madale R, Becker P. Evaluation of hospital information system in the Northern Province in South Africa. Health Systems Trust, May 2002. http://www.hst.org.za/publications/504 (accessed May 2006).
[18] Odhiamo-Otieno GW. Evaluation criteria for district health management information systems: Lesson from the Ministry of Health, Kenya. Int J Med Inf 2005: 74: 31-38.
[19] Seddon PB. A respecification and extension of the DeLone and McLean model of IS success. Information Systems Research 1997: 8(3): September 1997: 240-253.
[20] Shaw V. Health information system reform in South Africa: developing an essential data set. WHO Bull 2005: 83(8): 632-639.
[21] Southon G, Sauer C, Dampney C. Lessons from a failed information systems initiative: issues for complex organisations. Int J Med Inf 1999: 55: 33-46.
[22] 'Van der Loo RP, van Gennip EMSJ, Bakker A, Hasman A, Rutten FFH. Evaluation of automated information systems in health care: an approach to classifying evaluative studies. Comp Meth Prog Biomed 1995: 48: 45-52.
[23] 'van der Meijden MJ, Tange HJ, Troost J, Hasman A. Determinants of success of inpatient clinical information systems: A literature review. J Am Med Inform Assoc 2003: 10(3): 235-243.
[24] van Gennip EMSJ and Bakker A. Assessment of effects and costs of information systems. Int J Bio-Med Computing 1995: 39: 67-72.

Address for correspondence

Lyn Hanmer
Medical Research Council
PO Box 19070
Tygerberg
7505 South Africa

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Are Problem-Oriented Medial Records (POMR) Suitable for Use in GPs' Daily Practice?

Etienne De Clercq[a], Viviane Van Casteren[b], Pascale Jonckheer[c], Peter Burggraeve[d], Marie France Lafontaine[b], Karen Degroote[b], Francis Roger France[a]

[a]*School of Public Health, Health System Research, Université Catholique de Louvain, Brussels, Belgium*
[b] *Scientific Institute of Public Health (IPH), unit of Epidemiology, Brussels, Belgium*
[c] *Société Scentifique de Médecine Générale (SSMG), Brussels, Belgium*
[d] *Domus Medica (Flemish Institute of General Practitioners), Berchem, Belgium*

Abstract

Problem-oriented functions have been implemented in almost all Belgian GPs' software systems since 2003. We therefore investigated whether some of them – especially the explicit linking procedure between treatments or referrals and the relevant problems – can be used by GPs in their current daily practice.

In 2005, within the Belgian ResoPrim project, we organized data collection, mainly around the theme of "hypertension and cardiovascular risk factors", by 26 volunteer GPs' practices using three different software systems. Data were collected prospectively over six weeks in early 2005, and retrospectively for 2004. In this paper we report only on the part of the study that aimed to assess the linking procedure. For all patients and hypertensive patients alike, the key indicators used were the percentage of (problem-)linked drugs among the drugs extracted, the percentage of anti-hypertensive (problem-)linked drugs among anti-hypertensive drugs extracted, and the percentage of (problem-)linked referrals among the number of referrals extracted.

For all patients, the data collected relate to 10 914 contacts (7 831 patients) in 2005, and to 74 878 contacts (16 813 patients) in 2004. Large variations were observed per software system and GP, and also over time. The percentage of linked drugs rose from 2% (2004, two GPs) to 36% (2005, fourteen GPs). For linked referrals the percentage was 65% in 2004 vs. 75% in 2005.

Our study shows that some functions related to the problem-oriented patient record were spontaneously used by GPs in daily practice. This use increased during collaboration with the primary care research network. This increase was not restricted to the theme of data collection (i.e. not restricted to hypertensive patients, to anti-hypertensive drugs or to links with cardiovascular problems).

Keywords:

Computerized Patient Records, Primary Health Care, Problem-Oriented Medical Records

Introduction

The GPs' Electronic Patient Record (EPR) can be a rich source of information for various research purposes, such as epidemiological studies, health-care quality assessment or socio-economic studies. Although it has been questioned whether data from EPR can be used for such purposes [1-3], various advantages have also been described: the ability to record over the long term, the potential availability of numerous kinds of data, and the options for collecting data from many GPs' practices and on many patients [4-7].

The problem-oriented architecture of the patient record can provide additional advantages, not only for the care process but also for researchers, as data linkage (i.e. when treatments and actions are linked to the relevant problems), can help with data interpretation, self audit and communication between practitioners [8-13].

Since 2003, this problem-oriented architecture has been included nation-wide in Belgium's quality evaluation process. Its subsequent implementation in almost all GPs' software systems involved a major update [14].

Against this background, we undertook research within the framework of the ResoPrim project, a Belgian GPs' research network [15], to investigate whether particular functions related to the problem-oriented architecture could be used by GPs in their daily activities. Three research questions were defined:

- Are some problem-oriented functions spontaneously used by the GPs without specific IT skill or training?
- What is the impact of participation in the ResoPrim research project on the use of the problem-oriented functions?
- If present, is this impact restricted to the specific theme of the ResoPrim project, i.e. hypertensive patients and cardiovascular risk factors and management?

Methods

This paper is based on the consolidated results of the first phase of the ResoPrim project whose main theme was hypertension and cardiovascular risk factors.

Data were collected from 26 volunteer GPs' practices, which between them used three different software systems (13 using software 1, five using software 2, and eight using software 3). Data were collected anonymously through an automatic extraction procedure and through a manual procedure whereby GPs filled in an electronic questionnaire at the end of each contact. Data were collected retrospectively for 2004 and prospectively over six weeks in early 2005. Fuller details are provided elsewhere [15].

In order to study the usability of some problem-oriented functions, we explicitly extracted the single or multiple links between the referrals or drugs prescribed and the problems to which they applied. All referrals and problems (ICPC codes) were considered. However, only a small number of specific drugs (ATC codes) were extracted from the EPR, i.e. aspirin, statins and drugs related to a number of specific problems (hypertension, diabetes, epilepsy, migraine, glaucoma, herpes, gout, hypo/hyperthyroidism). Only coded drugs and coded problems entered in the patient record by the GPs were automatically extracted.

We used retrospective data (2004) to assess the spontaneous use of the problem-oriented functions, and 2004 and 2005 data to study the impact of the ResoPrim project. For hypertensive patients, we used 2005 data to compare certain automatically extracted data with the answers to questions addressed to the GPs by the electronic questionnaire (prospective study). These questions concerned either the specific drugs prescribed for hypertension (central working agents, alpha-blockers, diuretics, beta-blockers, calcium-antagonists, ACE-inhibitors and sartanes), or whether a referral had been made for hypertension in the current contact (2005) or in 2004.

The key indicators used were the percentage of linked drugs among the drugs extracted, the percentage of anti-hypertensive linked drugs among the anti-hypertensive drugs extracted, and the percentage of linked referrals among the number of referrals extracted. We applied these indicators to two populations: all patients and hypertensive patients identified by a direct question to the GPs (electronic questionnaire). Only drugs and referrals entered into the EPR during the periods 2004 and 2005 were taken into account. We used numbers of patients for the comparisons between extracted data and answers to the questions addressed to the GPs.

Quality control (four weeks) and quality assessment procedures (using a dummy patients technique) were conducted for the extraction modules developed by each software package.

Results

Prospectively (2005) we used data derived from all three software systems and from 26 practices, 7 831 patients and 10 914 contacts. For 2004 we obtained data from only two software systems for 18 practices, 16 813 patients and 74 878 contacts. For the comparisons between 2004 and 2005, only these two software systems were taken into account, with 5 090 patients and 7 265 contacts.

Prospective data collection (six-week period, 2005)

For 2005, table 1 presents the percentage of drugs, anti hypertensive drugs and referrals linked with a problem. Figures are shown for all patients (7 831 patients) and for hypertensive patients (1 554 patients) identified by a question addressed to the GPs (electronic questionnaire) for each patient attending the GPs' surgery during the six-week data collection period. Of the 17 GPs who had encoded referrals, 13 encoded linked referrals (percentage ranging from 40% to 100%); of the 21 GPs who had encoded drugs, 16 encoded linked drugs (percentage ranging from 4% to 98%).

Table 1: Percentage of linked referrals, drugs and anti-hypertensive drugs (HTdrugs) for all patients and hypertensive patients (2005)

Population	2005 All patients (7 831)			2005 Hypertensive patients (1 554)		
Software	referrals	linked referrals.	%	referrals	linked referrals	%
Soft. 1	114	92	81%	37	33	89%
Soft. 2	48	30	63%	22	14	64%
Soft. 3	182	43	24%	29	6	21%
Total	344	165	48%	88	53	60%
	drugs	linked drugs	%	drugs	linked drugs	%
Soft. 1	1 523	383	25%	1 009	271	27%
Soft. 2	564	367	65%	336	216	64%
Soft. 3	860	41	5%	533	17	3%
Total	2 947	791	27%	1 878	504	27%
	HTdrugs	linked HTdrugs	%	HTdrugs	linked HTdrugs	%
Soft. 1	870	219	25%	668	189	28%
Soft. 2	359	248	69%	230	147	64%
Soft. 3	485	27	6%	345	12	3%
Total	1 714	494	29%	1 243	348	28%

The many-to-many relationships (linkage of a drug or a referral to several problems) was seldom used and only within the first two software systems. In 2005, two GPs were using them for referrals, accounting for 1.66% of all the linked referrals (two software systems). Four GPs used them for drugs (0.67% of all the linked drugs in the first two software systems).

Among the 1 554 hypertensive patients attending GPs' surgeries in the six-week period in 2005 and identified by a direct question to the GPs, 1 491 were treated for hypertension (treatment identified by a direct question or by extracted drug codes). For these patients, asking the GPs if they were taking drugs had a sensitivity of 98%. Among these 1 491 patients, 891 (60%) had a coded anti-hypertensive drug and 196 (13%) had a coded anti-hypertensive drug linked with a hypertension code (K85, K86, K87). The percentages of the various types of coded drugs are preserved, as shown in figure 1.

Among the 494 anti-hypertensive linked drugs, 451 were linked to valid ICPC codes, 173 linked to non-hypertensive ICPC codes, and 71 linked to non-cardiovascular codes (not 'K' ICPC code).

For the 1 554 hypertensive patients mentioned above, we identified the referrals made during the data collection period either by asking a direct question to the GPs (electronic questionnaire) or by an extracted code. Among these patients, 180 had a referral for their hypertension in 2005 (six weeks period). Asking the question to the GPs had a sensitivity of 99%. Only 1.67% of the patients who had been referred in 2005 had a coded referral linked with hypertension in their EPR.

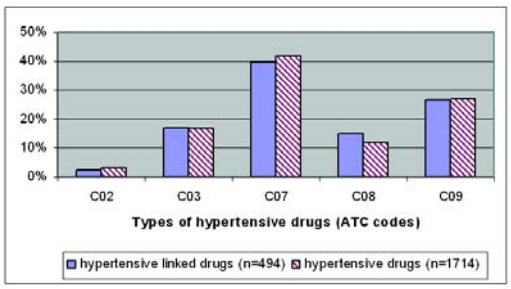


Figure 1- Percentage of the various types of anti-hypertensive drugs (N=1 714) and anti-hypertensive linked drugs(N=494) for all patients (2005): C02 (central working agents & alpha-blockers), C03 (diuretics & associations), C07 (beta-blockers & associations), C08 (calcium antagonists), C09 (ACE-inhibitors, sartanes & associations)

Comparison between retrospective and prospective data collections

For software 1 and 2 (18 GPs), table 2 presents the percentage of drugs and referrals linked with a problem in 2004 (retrospective) and 2005 (prospective). For 2004, 11 out of 13 GPs who had encoded referrals had encoded linked referrals (percentage ranging from 44% to 100%). Only two out of 15 GPs who had encoded drugs in 2004 (against 14 out of 15 in 2005) had encoded linked drugs (0.09% and 82%).

Table 2 - Percentage of linked referrals and linked drugs (2004 - 2005)

Population	2005 All patients (5 090)			2004 All patients (16 813)		
Software	referrals	linked referrals	%	referrals	linked referrals	%
Soft. 1	114	92	81%	839	501	60%
Soft. 2	48	30	63%	297	236	79%
Soft. 1+2	162	122	75%	1 136	737	65%
	drugs	linked drugs	%	drugs	linked drugs	%
Soft. 1	1 523	383	25%	8 693	1	0%
Soft. 2	564	367	65%	255	208	82%
Soft. 1+2	2 087	750	36%	8 948	209	2%

The reasons for which a patient was referred in 2004 (529 valid ICPC codes for software 1 and 2) included 76 referrals (71 patients) for a cardiovascular problem (14%). In 2005, the reasons for which a patient was referred (104 valid ICPC codes for software 1 and 2) included 15 referrals (14 patients) for a cardiovascular problem (14%). This is illustrated in figure 2 for the year 2004.

For software 1 and 2, among the 1 084 hypertensive patients attending GPs' surgeries in the six-week period of 2005 and identified by a direct question to the GPs, 335 patients had been referred for their hypertension in 2004 (referrals identified by asking a direct question to the GPs or by an extracted code). Asking the question to the GPs had a sensitivity of 99%. Only 2.69% of the patients who had had a referral in 2004 had a coded referral linked with hypertension in their EPR.

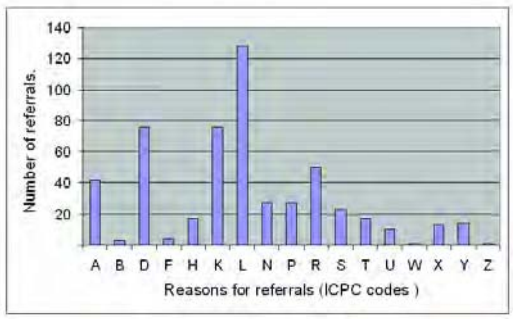


Figure 2 - Reasons for which a patient was referred (all patients, 2004, N=529, ICPC codes): A (general and unspecified), D (digestive), K (cardiovascular), L (musculoskeletal), R (respiratory)

Discussion

We found large variations between GPs, software systems and types of linked data (referrals vs. drugs). We also observed large variations over time (2004, not shown). Therefore, given our confidence level in the quality of the data (based on the quality assessment procedure), no statistical method can easily be used for a thorough examination of our findings. Yet, our (not representative) sample provides us with interesting findings regarding the use of functions related to the problem-oriented architecture of the EPR.

The linkage between referrals and problems was spontaneously used by 11 GPs before the ResoPrim project (cf. table 2, 2004 data). Yet, at that time, only one GP (table 2, software 2) extensively used the linkage between drugs and problems.

Participation in the ResoPrim project had a positive effect on the linkage between drugs and problems (36% linked drugs in 2005 vs. 2% in 2004; see table 2). Twelve GPs started using this linkage function during their participation.

For software 2 (table 2), there was a decrease in the percentages of linked drugs and linked referrals between 2004 and 2005. This was due to two GPs. The first of these never coded any drugs in 2004; while he started to code in 2005, he did so without using the linkage function. The second GP coded a few referrals in 2004 and quite a lot more in 2005, but never used the linkage function. If we withdraw these GPs from the comparison, the percentage for the linked drugs for software 2 is 96% in 2005 vs. 81% in 2004. For the referrals the figures were 88% for 2005 vs. 81% in 2004.

When investigating the potential impact of the theme of the data collection ("hypertensive patients and cardiovascular risk factors and management"), several issues were considered:

- Did the GPs link mainly drugs and referrals for patients with hypertension (as a result of their participation in the ResoPrim project)?
- Did the GPs link mainly anti-hypertensive drugs (as a result of their participation in the ResoPrim project)?
- Was a greater percentage of referrals linked with a cardiovascular problem in 2005 than in 2004?

In 2005 (see table 1), the percentage of linked referrals for hypertensive patients (60%) was higher than for all patients (48%). Yet a similar divergence can also be observed in 2004 between linked referrals for hypertensive patients identified by a hypertensive code (79%, result not shown in this paper) and all patients (65%; see table 2). It is also noteworthy that the growth in the proportion of linked referrals between 2004 and 2005 (15%) is similar for all patients (see table 2) and for hypertensive patients identified by a hypertensive code (not shown in this paper). At this stage, we cannot assume that there is a thematic effect (i.e. an impact of the theme of the data collection) on the GPs' behavior regarding linkage function between referrals and problems.

For linked drugs and linked-HTdrugs (anti-hypertensive linked drugs), table 1 shows no differences between hypertensive patients and all patients. Anti-hypertensive drugs were linked slightly more often than all the drugs (29% vs. 27 % in table 1). As shown in the results section, there was not a higher percentage of referrals linked with a cardiovascular problem in 2005 (14%) than in 2004 (14%). We can therefore assume that the effect of participation in the ResoPrim project on the GPs' use of the linkage functions is not limited to hypertensive patients, anti-hypertensive drugs or linkage to hypertensive codes.

A thematic effect of research participation on GPs' coding behavior had already been demonstrated in a previous Belgian study [16]. Our results suggest an additional extra-thematic effect related to participation in a research network such as ResoPrim. This effect might be related to GPs' interest in individual feedback, in comparisons with peers, in participating in software users' groups or in contributing to a research project. Preliminary additional findings within the ResoPrim project suggest a possible persistence of these effects.

At this stage it is still questionable whether there can be any secondary research usage of the links between drugs or referrals and problems (e.g. for quality of care or socio-economic studies). If most of the encoded referrals are linked (see tables 1 and 2), many referrals are not encoded (<12% of referral were encoded for hypertensive patients in 2005; figures not shown in this paper). If many drugs are encoded (60% of the hypertensive patients with an anti-hypertensive drug were identified), many encoded drugs were not linked to a problem (see table 1 and 2).

Some of these findings, however, are interesting because they suggest that the links have a promising secondary use. For example, the links make it possible to identify the reasons for which a drug has been prescribed. Thirty eight percent of the anti-hypertensive drugs related to a valid ICPC code (451 in 2005) were linked to a non-hypertensive problem. Sixteen percent were linked to a non-cardiovascular problem. Figure 3 shows the non-cardiovascular reasons for which an anti-hypertensive drugs was prescribed. Even for hypertensive patients, 25% of such drugs were directly and exclusively related to a non hypertensive problem (result not shown). This unexpected finding might be solved partly by a broader use of the many-to-many relationship (currently less than 1% of all the links). An anti-hypertensive drug can simultaneously be prescribed for hypertension and for another problem.

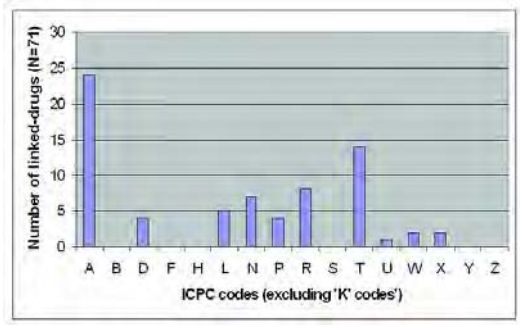


Figure 3 - Number of anti-hypertensive drugs related to ICPC coded problems, excluding cardiovascular problems (all patients, 2005; N=71): A (general and unspecified), N (neurological), R (respiratory), T (endocrine/metabolic and nutritional)

With regard to referrals, very little coded information is currently available in the patient records. Even hospital requests can not be disentangled from referrals to specialists. In this context, the link between a referral and a problem can be a rich source of information. This is illustrated in figure 2. In future, such links might also provide the profile of referrals for various groups of patients (such as hypertensive patients or patients with low income), thereby creating interesting perspectives for socio-economic research.

In order to be more useful for large-scale quality of care or socio-economic studies, the linked drugs or linked referrals should be representative samples of all the drugs prescribed or all the referrals made. At this stage, we have already shown (see figure 1) that anti-hypertensive linked drugs have the same prescription profile as all the hypertensive coded drugs, which is encouraging for further research based on this method.

Conclusion

Our study shows that, once implemented in previously non-problem-oriented software systems, some functions related to the problem-oriented architecture of the patient records can be used spontaneously by GPs in current daily practice without any specific IT skill or training.

These findings generate the hypothesis that participating in a primary care research network such as ResoPrim can stimulate the use of such functions. Within a thematic

research project such as that involving hypertension and cardiovascular risk factors, this behavioral change does not seem to be restricted to the theme (i.e. hypertensive patients, anti-hypertensive drugs, and links with cardiovascular problems). The persistence of this effect should be further explored.

Further research is still needed before the problem-linked data currently available can be used for secondary research purposes such as quality of care or socio-economic studies. Therefore it must still be determined whether these data can be treated as representative samples of the drugs prescribed or referrals made. In coming years, the problem may be solved partly by a higher number of electronic prescriptions or electronic referrals in the patient record.

Acknowledgment

We thank all the GPs and the industrial partners involved in the data collection. The ResoPrim project is supported by grants from the Belgian Federal Public Planning Service Science Policy (Nr I2/AE/207 and Nr I2/2F/207).

References

[1] Prins, H., Kruisinga, F. H., Buller, H. A., and Zwetsloot-Schonk, J. H. Availability and Usability of Data for Medical Practice Assessment. Int.J.Qual.Health Care 2002; 14(2): 127-37.

[2] Wilson, A. E., Pollock, C., Weekes, T., and Dowell, A. Can General Practice Provide Useful Information?--Evaluation of a Primary Health Care Information Project in Northern England. J.Epidemiol.Community Health 1995; 49(3): 227-30.

[3] Hiddema-van de Wal, A., Smith, R. J., van der Werf, G. T., and Meyboom-de Jong, B. Towards Improvement of the Accuracy and Completeness of Medication Registration With the Use of an Electronic Medical Record (EMR). Fam.Pract. 2001; 18(3): 288-91.

[4] Dresser MV, Feingold L, Rozenkranz SL, Coltin KL. Clinical Quality Measurement. Comparing Chart Review and Automated Methodologies. Medical Care 1997; 35(6): 539-52

[5] Okkes IM, Groen A, Oskam SK, Lamberts H. Advantages of Long Observation in Episode-oriented Electronic Patient Records in Family Practice. Method Inform Med 2001; 40: 229-35

[6] Pearson N et al. Collecting morbidity data in general practice: the Somerset morbidity project. BMJ 1996; 312: 1517-20

[7] Vlug AE, van der LJ, Mosseveld BM, van Wijk MA, van der Linden PD, Sturkenboom MC et al. Postmarketing surveillance based on electronic patient records: the IPCI project. Methods Inf Med 1999; 38(4-5): 339-344.

[8] Salmon, P, Rappaport, A, Bainbridge, M, Hayes, G, & Williams, J. Taking the problem oriented medical record forward. Proceedings AMIA Annual Fall Symposium, 1996; 463-467.

[9] Tange HJ, Schouten HC, Kester ADM. & Hasman A. The granularity of medical narratives and its effect on the speed and completeness of information retrieval. Journal of the American Medical Informatics Association, 1998; 5: 571-582.

[10] Rector AL, Nowlan WA & Kay S. Foundations for an electronic medical record. Methods of Information in Medicine, 1991; 30(3): 179-186.

[11] Carey IM, Cook DG, De Wilde S et al. Implications of the problem orientated medical record (POMR) for research using electronic GP databases: a comparison of the Doctors Independent Network Database (DIN) and the Genreral Practice Research Database (GPRD). BMC Fam Pract 003; 4: 14

[12] Bayegan E, Nytro O, Grimsmo A. Ranking of information in the computerized problem-oriented patient record. Medinfo 2001; 10: 594-598.

[13] De Lusignan S, van Weel Chris. The use of routinely collected computer data for research in primary care: opportunities and challenges. Family Practice 2006; 23(2): 253-263.

[14] De Clercq E, Piette P, Strobbe J, Roland M, Vandenberghe A, Steenackers J, Pas L. Setting up a common architecture for EPR in primary care: the Belgian experience. Stud-Health-Technol-Inform 2002; 90: 215-219.

[15] De Clercq E, Van Casteren V, Jonckheer P, Burggraeve P, Lafontaine M-F, Vandenberghe H, Lorant V, Artoisenet C, Degroote K. Research networks: can we use data from GPs' Electronic health Records. Studies in Health Technology and Informatics, 2006; 124: 181-186.

[16] Deliège D. A classification System of Social Problems: Concepts and Influence on GPs' Registration of Problems. Social Work in Health Care 2001; 34 (1): 195-238.

Address for correspondence

Dr E De Clercq,
Université Catholique de Louvain ESP-HSR-SESA,
30.41 Clos Chapelle aux Champs,
1200 Bruxelles, Belgium
Email: declercq@sesa.ucl.ac.be

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Can an EPR Support the Concept of Family-Centred, Individualized Developmental Care of Premature Infants and Newborns?

Christian D. Kohl[a], Claus Schott[b], Doris Verveur[b], Otwin Linderkamp[b], Petra Knaup-Gregori[a]

[a] *Institute for Medical Biometry and Informatics, Department of Medical Informatics, University of Heidelberg, Germany*
[b] *Children's Hospital Heidelberg, University of Heidelberg, Department of Neonatology, Germany*

Abstract

At the University Children's Hospital Heidelberg the concept of 'Developmental, Family-Centred, Individual Care of Premature Infants and Newborns' was introduced to support optimal growth of premature infants. This interdisciplinary concept requires cooperation of different specialists. A well operating communication is a precondition for such cooperation.

As a patient's record is not only used for storing information but also for exchanging information, the question was if a complete electronic patient record (EPR), in contrast to the existing patient's record, could sensibly support this new concept of care. To answer this question the whole communication of the staff in the infants ward was analysed using different observation methods. These observations delivered several issues which showed that an EPR could improve communication and workflow. Therefore an EPR for the neonatology at the University Children's Hospital Heidelberg can now be designed on the basis of our communication concept.

Keywords:

Computerized medical record system, NIDCAP, communication, premature infant, newborn infant.

Introduction

The concept of Developmental, Family-Centred, Individual Care of Premature Infants and Newborns (German term: **E**ntwicklungsfördernde, **f**amilienzentrierte, **i**ndividuelle **B**etreuung Früh- und Neugeborener – EfiB) was designed at the University Children's Hospital Heidelberg, Germany. The concept is based on a 'developmental care' [1] for premature infants and on the 'Newborn Individualized Developmental Care and Assessment Programme' (NIDCAP) according to Heidelise Als [2-3]. Treatment, care and environment are adapted to the needs of preterm infants. EfiB is a modular concept: Every module continues the approaches of the two core components. For instance the modules 'light reduction' and 'noise reduction' contain arrangements to provide a suitable environment for the premature infants. Another module contains a guideline which targets at a reduction of stress during painful examinations of the small patients. EfiB aims to optimize the growth of the preterm infants. With its different modules EfiB is a comprehensive concept which involves a lot of different professions (physicians, nurses, psychologists, physiotherapists…). The introduction of EfiB required fundamental changes affecting both organisation und communication: For example before EfiB, physicians and nurses worked at the premature infants at pre-defined points of time or when it was suitable for the staff. Now they perform - for instance a treatment that is not vitally important - only when the child is awake. The introduction of EfiB started in October 2005. The modules were established step by step.

A well operating communication – that means an exchange of information – is essential for health care [4-5]. A well operating communication is an important precondition that an interdisciplinary concept like EfiB can work properly. When introducing EfiB the communication processes were regarded as not optimal. Thus, accompanying to the introduction of the modules of EfiB in the premature infant ward at the University Children's Hospital Heidelberg a communication concept was created which aimed to establish an appropriate information logistics for EfiB. The objective of our paper is to answer the question whether an enclosing EPR could sensibly support EfiB. This question was answered by the results of the analysis of the present state.

Material and methods

The communication concept was established in three steps:

1. Analysis of the present state
2. To-be concept
3. Implementation strategies.

Pilot observation

In the pilot phase of our investigation all kind of professionals involved in newborn care were accompanied in daily routine for several days. These observations showed that communication can be divided into two classes:

- 'communication during individual workflows' and
- 'communication during team-conferences, ward rounds and handing-overs'.

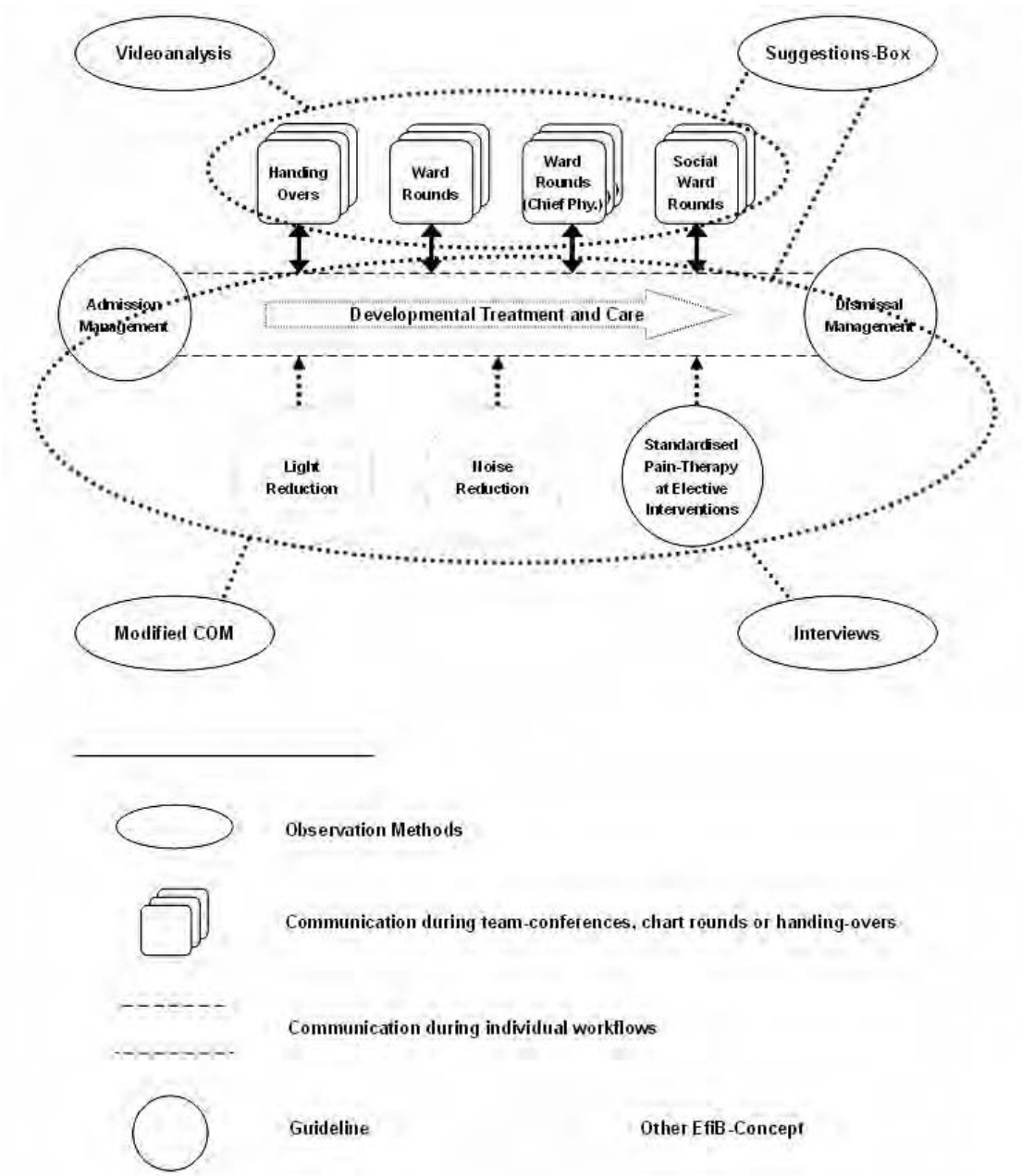


Figure 1 – Range of application of the used observation methods for analyzing the present state of communication

In contrast to communication during individual workflows during team-conferences communication takes place all the time and typically between one sender and many receivers. For these different classes of communication different observation methods were necessary. In addition particular methods were selected to reveal hidden information. Figure 1 shows the range of application of each method. All methods are introduced in the following.

Modified COM

To analyze the communication during individual workflows a modified version of the 'communication observation method' (COM) as described in [6] was used. The COM serves to measure communication load and patterns in the clinical environment. Therefore, the communication the observed person is involved in is recorded and transcribed. In the transcription communication events are identified, coded and stored in a database.

Queries on this database supply information to communication load and patterns in the observed area. Because of limited resources in our investigation we performed an abbreviated version of the COM: We refrained from recording and a transcription. The observer had to identify, code, and store communication events of the observed person during the observation (on-the-fly). That way we observed ward physicians, head nurses, staff nurses, supplying assistants and secretariat staff – three representatives of each group, each for a complete shift. In the end we had collected over 1,000 communication events in our database.

Video analysis

Analyzing communication during team-conferences with the modified COM was not feasible, as there was a too high density of communication events coding on-the-fly was not possible. So, related to the original COM, team-conferences were buffered on video. For analysis, we used besides the COM additional ideas from an 'assessment instrument for evaluating the quality of communication processes in health care' [7] and from an 'instrument to evaluate the communication process during multidisciplinary team conferences in rheumatology' [8]. That way we observed three handing-overs of the nurses, two chart rounds (which took place in a conference room and not at bedside to avoid disturbing the preterm infants), two chart rounds with the head physician of the department (also in the conference room) and three social ward rounds (participants are, beside others, a ward physician, the head nurse, the aftercare nurse and the psycho-social team consisting of a psychologist, a member of the social service and a family therapist).

Interviews

The modified COM could not be applied to all groups of professionals because some were only awhile per day in the ward (such as the aftercare nurse) or because of a too high interference of the observation (this problem for instance arose with the psychologist). To get information about the communication of these professionals, they were interviewed. We interviewed the chief physician, a ward physician, the member of the social service, a breast feeding advice assistant, three parents and the aftercare nurse.

All methods, the modified COM, the video analysis and the interviews, were accompanied by permanent, supplementary observations.

Suggestions-box

In addition to the other methods, we installed a suggestions-box to get also hidden information – information which is not apparent in observations. The suggestions-box was an offer to all staff: They could write problems, suggestions and ideas referenced to EfiB or their work on a paper and put it into the box anonymously.

Sample selection

The observed persons were randomly selected. We tried to observe as many different persons as possible of each professional group. Nevertheless, a few groups were only represented by one person in the ward, for instance the psychologist.

Results

Generally spoken, the combination of our applied methods showed a very heterogeneous communication structure. Nevertheless, the communication achieved almost ever the intended objectives: Information was mainly correct and mostly available in the right style at the right place for the right person. For instance the analysis of the videos showed, in contrast to the general observations of Verhoef et al. [8], that the participants of the analysed multidisciplinary team-conferences used a common language and haunted the same objectives. This helps to conduct a good interdisciplinary care. Furthermore, the video analysis showed that over all types of sessions (different team-conferences, ward rounds and handing-overs) an average of 3.22% of the duration was unused and that an average of 2.07% of the remaining duration was used for non-team-oriented communication (that means communication which concerns only a few persons of the group). That means, also in contrast to the general observations of Verhoef et al. [8], that there is not a huge exchange of discipline-specific information which is not relevant for the other participants.

Problems with complete new modules

We experienced, especially in our interviews and accompanying observations, that the reorganisation of existing structures and workflows including communication flows (for example the reorganisation from the conventional to the developmental care) was easier for the staff than the integration of complete new modules such as the psycho-social modules for instance. Since the psychologist, the member of the social service and the family therapist have overlapping duties and responsibilities on the one hand and different professional backgrounds and duties on the other hand, the staff nurses had to learn who is responsible for which problem and to whom they had to communicate which information.

Heterogeneous patient record

Not only the communication processes and structures were heterogeneous but also the patient record: When the introduction of EfiB in Heidelberg begun, mainly three different patient's records were used, two electronic and one paper based system. IS-H*med, the one electronic system, is a part of the hospital information system which is used in the whole University Hospital Heidelberg (beside patient's data, findings and reimbursement data is transmitted with this system). NeoDat, the other electronic system, is used only in neonatology to store a lot of specific data about newborns and to fulfil the legal requirement to deliver them to the German perinatal register. Furthermore, there is a paper based patient's record placed at the bedside. Physicians write their orders in this record and the nurses note their observations in it. If the paper based record at the bedside gets to large, older documents are swapped to a folder in the nurse's room (a separate folder exists for every patient in the ward). Lot's

of findings are transferred electronically to IS-H*med. For legal reasons the findings are printed, signed by a physician and archived in the patient's folder. When a patient leaves the hospital, all paper-based documents are archived in an additional archive-folder, except a few documents which are used when the patient comes back to the hospital for outpatient care. These documents are archived in a special outpatients-folder.

Insufficient Interfaces

In conversations with physicians we found that only demographic data can be transferred from IS-H*med to NeoDat. Electronic findings have to be copied from IS-H*med to NeoDat via copy-and-paste; some findings even have to be recorded in NeoDat separately.

Different media

Printing an electronic finding to archive it paper-based means using a different media. Vice versa, there is also a change of media when a physician enters data from the paper based patient record (for instance the weight of a child) into the NeoDat-application.

Discussion

Identified problems

The heterogeneous structure of the communication is not a weakness in general but may be caused by the cooperation of different professions. The heterogeneous patient record is a weakness which is consolidated by insufficient interfaces and different media. Copy-and-pasting findings or typing information into an electronic system costs working time and may cause mistakes. Apart from that, the paper-based record at the bedside has the typical disadvantages of a paper-based documentation (exclusive access, missing remote-access, endangerment of loss, problems caused by cloudy handwriting or dirt on the paper, effort at archiving and demand for adequate protected rooms). Moreover it was possible that parents read the patient records of their and other's children which is forbidden by law. In addition, the psycho-social team which is responsible for problems of the parents could not write their observations in the paper based record as the parents might read disagreeable information about themselves.

In the following we want to look at the strengths and advantages, as well as the weaknesses and limitations, of our applied methods.

Modified COM

Coding and storing the observed communication events on-the-fly according to our modified COM had the weakness that the acquired durations of the single events are not very precise. Therefore a quantitative analysis of the data was not reasonable. However the method was very useful to perform a structured observation and to recognize communication flows and structures. Furthermore, in comparison to the original COM, there might have been a higher attendance among the staff to participate in our study and a lower influence on their acting as we didn't perform an audio recording. Additionally, we saved resources with our modification.

Video analysis

The video analysis was highly required to facilitate the analysis of team-conferences, ward rounds and handing overs because of the high communication load. In addition to the buffering, the used video camera had the advantage that a time-index was recorded parallel. So we had the possibility to perform a quantitative analysis of these sessions. Although all participating staff were informed about the observation and gave their approval, the camera lying on the shelf didn't seem to influence their acting. It must be said that our sample size was rather low (two up to three observations per session type). Therefore, the quantitative analysis could only display a trend. Nevertheless, these analyses were also much helpful to discover communication flows and structures.

Interviews

Interviews were necessary to integrate persons and professional groups who could not be observed using the modified COM. As the interviewed persons might have been not aware of existing (communication) problems, accompanying supplementary observations were very important.

Suggestions-box

The staff used the Suggestions-Box to ask questions, to address problems and to make suggestions (e.g., to introduce a monthly group session for the parents with a senior physician and a head nurse). However, the utilisation was batch-wise. The staff had to be reminded about the offering from time to time.

Perspective

Introducing a comprehensive electronic patient record could abolish most of the weaknesses of the existing system. Additionally, an enclosing EPR could give more advantages:

- Other systems (e.g., monitoring systems or incubators) could be integrated by appropriate interfaces, so that corresponding data like heart rate or weight could be recorded automatically and permanently in the patient record.
- Data could be easily presented in a user oriented style or visualized for everybody at team-conferences.
- An electronic system could not only support communication but also workflows – for example, by calculating food quantities and compositions in addition to the weight and age of the preterm infants.
- Last but not least a complete electronic system could support clinical trials.

The Children's Hospital Heidelberg is going to move into a new building in 2008. That could be a chance to introduce such a comprehensive electronic patient record, when established workflows are going to change anyway.

Introduction of an EPR

Emanating from our communication analysis several things already changed. Small changes can be realized very quickly. For example, every staff member now wears a uniform name tag to ease communication. Bigger changes, like the introduction of a new comprehensive EPR-system, have to be designed carefully regarding the results of the presented communication analysis. An EPR can only work properly, solve the problems mentioned above and facilitate the described advantages, if the potential end-users are integrated in the design process. Additionally, it is crucial to develop a comprehensive system with communication interfaces to avoid double data entry. Our communication analysis helps to identify the communication needs of multidisciplinary health care professionals and the data to be transmitted between application systems by communication interfaces.

Conclusion

The analysis of the present state showed a lot of disadvantages of the current heterogeneous patient record. At the same time introducing a comprehensive electronic patient record seems to be, in general, very recommendable because of several reasons: On the one hand the ERP could lead to an efficient and (from the patient's point of view) good quality of care. On the other hand an comprehensive ERP optimally supports EfiB, that should finally lead to a high quality of patient care, too. Therefore the development of such an EPR should be initiated.

Acknowledgments

We would like to gratefully acknowledge all the staff of the Children's Hospital Heidelberg who supported the investigation in different ways, especially the staff and parents who took part in observations and interviews and the complete team of the premature infant ward H9. Special thanks also to Christina Weber and Victoria Ziesenitz.

References

[1] Schott C, Verveur D, Ziesenitz V, Linderkamp O. Die Gehirnentwicklung frühgeborener Kinder. Die Pflege 2006 (in print).

[2] Young J. Frühgeborene, fördern und pflegen. Wiesbaden: Ullstein Mosby, 1997.

[3] Als H. Program Guide, Newborn Individualized Developmental Care and Assessment Program (NIDCAP). Boston: NIDCAP Federation International, 2001.

[4] Toussaint PJ, Coiera E. Supporting communication in health care. International Journal of Medical Informatics 2005: 74: 779-81.

[5] Coiera E. When conversation is better than computation. Journal of the American Medical Informatics Association 2000: 7(3): 277-86.

[6] Spencer R, Logan P, Coiera E. Communication Observation Method, Manual. Sydney: University of New South Wales, 2002.

[7] Toussaint PJ, Verhoef J, Vliet Vlieland TPM, Zwetsloot-Schonk JHM. Improving the quality of communication in healthcare. Studies in health technology and informatics 2003: 95: 857-62.

[8] Verhoef J, Toussaint PJ, Putter H, Zwetsloot-Schonk JHM, Vliet Vlieland TPM. Pilot study of the development of a theory-based instrument to evaluate the communication process during multidisciplinary team conferences in rheumatology. International Journal of Medical Informatics 2005: 74: 783-90.

Address for correspondence

Christian D. Kohl
Institute for Medical Biometry and Informatics
Department of Medical Informatics
Im Neuenheimer Feld 400
D-69120 Heidelberg
Phone: +49 (0) 6221 / 56 - 28 92
Fax: +49 (0) 6221 / 56 - 49 97
Email: Christian.Kohl@med.uni-heidelberg.de

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Multiple Detection Modalities and Disease Natural History of Breast Cancer

Tony Hsiu-Hsi Chen[a,b], Amy Ming-Fang Yen[b], Grace Hui-Min Wu[b], Li-Sheng Chen[b], Yueh-Hsia Chiu[a]

[a]*Division of Biostatistics, National Taiwan University, Taiwan*
[b]*Institute of Preventive Medicine, National Taiwan University, Taiwan*

Abstract and Objective

Multiple detection modalities have increasingly gained attention in population-based screening. However, the disease natural history and its efficacy have been barely addressed. We reviewed a series of articles addressing multiple detection modalities including mammography, ultrasound and magnetic resonance image between 1995 and 2005. A stochastic model was developed to estimate transition parameters pertaining to the disease natural history defined by multiple detection modalities. The effectiveness of the combination of ultrasound or magnetic resonance image (MRI) with mammography was projected using a series of computer simulation models.

The results indicated that multiple detection modalities may lead to reduced mortality. However, the benefit and the selection of detection modalities are affected by biological factors including age, breast tissue type and histological type. In addition, other social factors may also affect the utilization of multiple detection modalities.

Keywords:

stochastic model, screening, breast cancer, modality

Introduction

In Asian countries, mammography has been recommended for the early detection of breast cancers because the incidence of breast cancer has been increasing rapidly. As the occurrence of breast cancer in most Asian countries peaks at around 40-49 years, population-based breast cancer screening using mammography became controversial. This was due to the fact that mammography screening in women aged 40-49 years is less beneficial than expected in older women, and some of the associated harm is more frequent in this age group.

Recently, a series of new technologies have been used as adjunct diagnostic methods to mammography or have been proposed as possible screening tests for younger women. These include ultrasound, digital mammography and magnetic resonance image (MRI). However, as concluded in two review papers, new screening modalities are unlikely to replace mammography for screening the general population given current evidence. Therefore, they suggest that large randomized controlled trials should be conducted by comparing two or more screening tests. However, the empirical results from a larger randomized trial require long-term follow-up and enormous costs. It is timely to have a better understanding of the appropriateness and role of these new technologies in screening for the general population. The current study aimed to develop a stochastic model to describe the disease natural history for breast cancer by using different time points and screening tools based on the sensitivity of multiple modalities. Based on the estimated results, the effectiveness of the combination of ultrasound or MRI with mammography is projected using a series of computer simulation models.

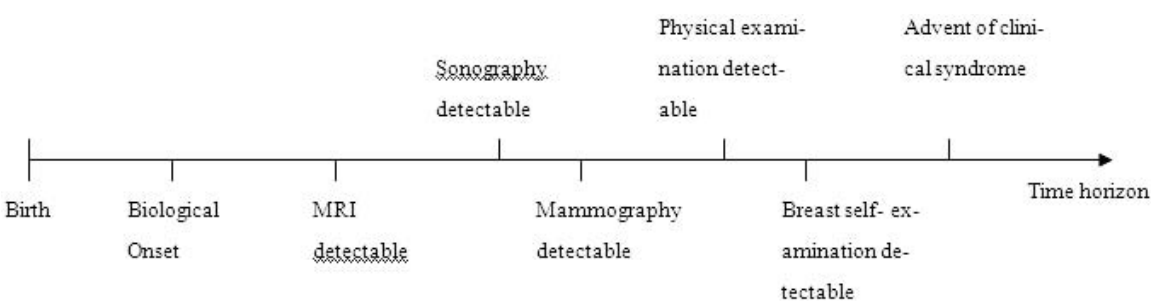


Figure 1 - Natural history of breast cancer defined by detection modalities

Methods

We reviewed a series of articles addressing multiple detection modalities for breast cancer including mammography, ultrasound and magnetic resonance image between 1995 and 2005 to obtain the performance of screening tools in terms of sensitivity and specificity. Different screening tools can detect breast cancer at different time points according to their ability to detect small or early stage breast cancer. The stochastic model to describe the progression from biological onset to the earliest detectable time point of the best screening tool, and the progress between detectable time points and between different screening tools is described below.

Estimation of transition parameters relating to disease progression with multiple detection modalities

A special case with known order of the performance of screening tools

The three available screening tools and the order of the time points with earliest detectable capability are MRI, ultrasound (US), and mammography (MA) as suggested in Figure 1. Let us treat MRI as gold standard; that is, the sensitivity of MRI is assumed to be unity. In the natural history of a special type of breast cancer, US has a better detection ability than mammography.

Patients with disease progressing beyond the earliest screening detectable point are treated as having the true disease. Since US is more capable than mammography of detecting breast cancer, the probability of detection by US or MA minus that of detection by MA only is equal to the difference of the sensitivities of the two modalities. This can be expressed as

$$P(\text{MA or US}) - P(\text{MA}) = Sen_{US} - Sen_{MA} \quad (1)$$

where P(M) denotes the probability of breast cancer being able to be detected with modality M.

The left-hand part can be written as follows:

$$\begin{aligned} &P(\text{MA}) + P(US) - P(\text{MA} \cap \text{US}) - P(\text{MA}) \\ &= P(US) - P(\text{MA} \cap \text{US}) \\ &= P(US) - P(\text{MA} \mid \text{US})P(US) \\ &= P(US)\{1 - P(\text{MA} \mid \text{US})\} \end{aligned} \quad (2)$$

Assuming constant hazard rate, the above equation is rewritten as follows:

$$\left(1 - e^{-\lambda_1 t}\right) \times e^{-\lambda_2 t} \quad (3)$$

Similarly, the relationship between MRI and MA can be expressed as follows:

$$\begin{aligned} &P(\text{MRI or MA}) \\ &= P(\text{MRI})\{1 - P(\text{MA} \mid \text{MRI})\} \\ &= 1 \times \left(P_{11}(t) + P_{12}(t)\right) = 1 - Sen_{MA} \end{aligned} \quad (4)$$

From expressions (3) and (4) together with knowledge of sensitivities of different screening tools, one can estimate λ_1 and λ_2.

Illustration

According to the Berg et al. (2004) study, the sensitivities of US and MA, and of MA only are 93.2% and 77.4%. Taking a 3-year screening interval, λ_1 and λ_2 estimated as 2.03 and 0.61. Given a constant hazard assumption, the sojourn time of staying in MRI detectable only and MRI+US detectable only are 4.6 months and 1.6 years.

Results

Types of new technologies

In addition to mammography, new technologies include ultrasound (US), magnetic resonance image (MRI), and full-field digital mammography (FFDM) with or without computer-aided detection (CAD). According to the review from Irwig et al. (2004), US together with mammography and clinical breast examination was applied to women with dense breast or normal, 'high-risk' female relatives of breast cancer patients, or high-risk women with BRCA mutation, or severe family history. MRI was further applied to even high-risk BRCA or several family members. To enhance the detection rate, FFDM with computer-aided detection was applied to average-risk women.

Natural history of breast cancer featured by multiple detection modalities

Figure 1 shows the natural history of breast cancer defined by detection modalities including US, MA, MRI, and clinical breast examination (CBE). After the onset of the biological process, the tumour can be detected by one of the modalities in the pre-clinical screen-detectable phase. Although figure 1 shows the earliest time point is detected by MRI followed by US, MA and CBE, the sequence may not be such a case. For example, certain breast tumours with micro-calcification or breasts with fatty tissue may be easily detected by MA before US. Using the information on the sensitivity of different modalities, we can calculate the probability of staying in the preclinical detectable phase (PCDP) using different modalities. The probabilities of staying in PCDP of MA+US compared with MA, given a US study by Berg et al. (2004), were shown in figure 2. The figures of dense tumour, fatty tumour and tumours with lobular type were also depicted. In addition, the time order of detecting breast cancer by these modes can be also affected by technical quality i.e. operator characteristics. To tackle this problem, we may model all possible combinations of the four modalities and quantify each type with the proportion Pi.

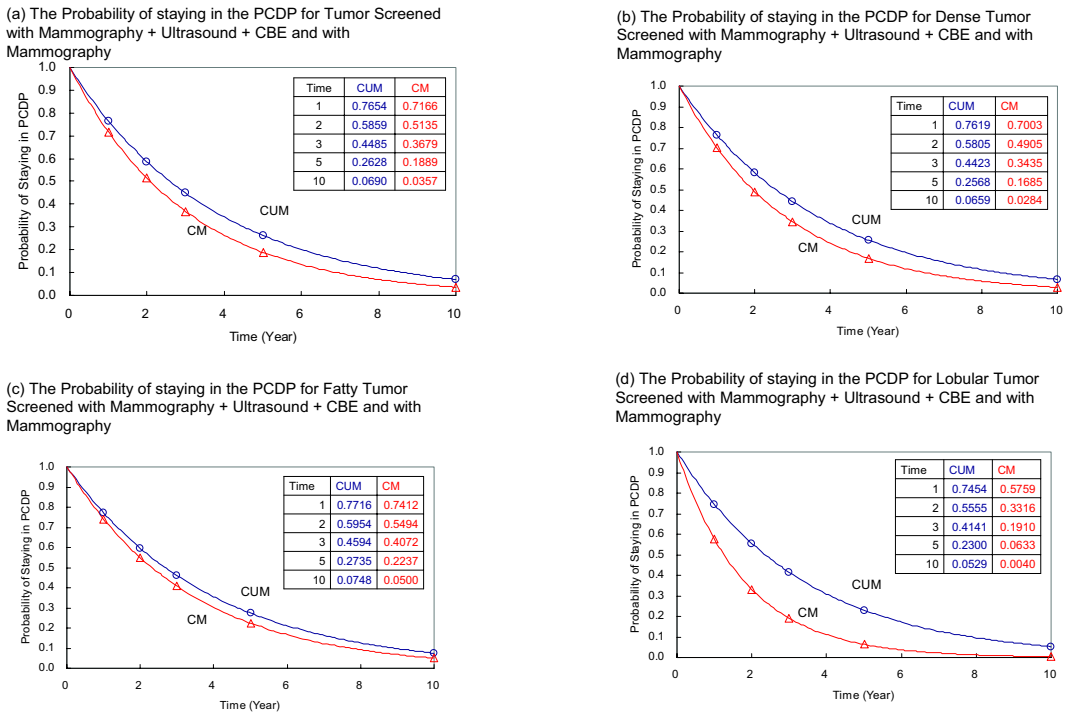


Figure 2 - The probability of staying in the PCDP for tumour screened with mammography + ultrasound + CBE and with mammography

Table 1 - The estimated results of the sojourn times between different sets of screening tools based on the stepwise method

County	Period	Partici-pants	Modality 1/Modality 2	lamda1	lamda2	Sojourn Time1	Sojourn Time2	RR of clinical cases	RR of breast cancer death
Taiwan [2]	1989-2001	105	M+C/C	0.5131	0.6127	1.9489	1.6322	0.4391	0.7273
			U+M+C/M+C	96.5106	0.3275	0.0104	3.0533	0.3333	0.6966
UK [3]	1997-2004	649	M+C/M	0.7884	0.2985	1.2685	3.3497	0.3852	0.7954
			MI+M+C/M+C	2.3220	0.5360	0.4307	1.8657	0.3704	0.8744
Nether-lands [4]	1998-2000	2020	M+C/C	1.8033	0.4957	0.5546	2.0173	0.4106	0.7402
			U+M+C/M+C	2.6597	0.9728	0.3760	1.0280	0.4957	0.7946
USA [1]	1999-2002	177	M+C/C	0.8379	0.4071	1.1934	2.4567	0.3921	0.7054
			U+M+C/M+C	2.0320	0.6143	0.4921	1.6279	0.4355	0.7656
			U+M+C/M+C(Density)	1.8349	0.5536	0.5450	1.8063	0.4173	0.7487
			U+M+C/M+C(Fatty)	2.4870	0.7460	0.4021	1.3405	0.4462	0.7756
			MI+M+C/M+C	19.0107	0.5047	0.0526	1.9814	0.3742	0.7401
			MI+M+C/M+C(Density)	13.4336	0.4477	0.0744	2.2334	0.3593	0.7288
			MI+M+C/M+C(Fatty)	63.8632	0.6296	0.0157	1.5883	0.3996	0.7580

* Lamda1 is the transition rate from the earliest screening detectable point (t_1) to the earliest detectable point with the extra screening tool in modality 1 (t_2), and Lamda2 the transition rate from t_2 to the earliest detectable point with modality 2 (t_3).

** Sojourn time 1 is the average time between t_1 and t_2, and sojourn time 2 between t_2 and t_3.

For simplification, we assume that the earliest and the latest time points detected were MRI and CBE respectively. We only make allowances for the time order changing between US and MA i.e., US followed by MA, or MA followed by US. The proportion P is used for quantifying MRI, US, MA and CBE, and 1-P for MRI, MA, US and CBE. The estimation technique has been proposed using data from literature review.

Table 1 shows the estimated results on transition rates based on the studies addressing asymptomatic breast cancers. Taking the Taiwanese study (no. 2) as an example, when comparing cancers detected by MA or CBE to CBE, the time from the earliest detectable time point to MA-detectable was estimated as 1.95 years. The time between MA-detectable and CBE-detectable was estimated as 1.6 years. When considering the modality of UA, MA and CBE combined compared to MA plus CBE combined, the first time interval was estimated at nearly no time given 100% sensitivity. However, the time lag between these two detectable points was estimated as 3 years given a comparably low sensitivity of MA plus CBE, of 62.5% of MA plus CBE. Based on a biennial screening interval in 8 years, the relative risk of being clinically detected can be obtained from a simulation technique. The 5-year survival for those breast cancer cases was projected. Results of studies from UK, Netherlands, and USA were also applied in the same way.

Effectiveness of new technologies for population-based breast cancer screening

The relative risks of being clinically detected and of breast cancer death were listed. The results show that 56% of clinically-detected cases were reduced by using the MA plus CBE regime compared to using CBE only. The reduction from using UA+MA+CBE was estimated as 67% compared to having MA plus CBE together. Using the cumulative survival by detection mode from the Swedish Two-County trial, the numbers of death from breast cancer were projected. The results suggested a 27% mortality reduction from breast cancer can be achieved with MA+CBE compared to CBE only. When the comparison was made with UA+MA+CBE to MA+CBE, the reduction was estimated as 30%.

Conclusion

Multiple detection modalities may lead to reduced mortality. However, the benefit and the selection of detection modalities are affected by biological factors including age, breast tissue type and histological type. In addition, other social factors may also affect the utilization of multiple detection modalities.

References

[1] Berg WA, Gutierrez L, NessAiver MS, Carter WB, Bhargavan M, Lewis RS, Ioffe OB. Diagnostic accuracy of mammography, clinical examination, US, and MR imaging in preoperative assessment of breast cancer. Radiology. 2004;233:830-49

[2] Hou MF, Ou-Yang F, Chuang CH, Wang JY, Lee LW, Huang YS, Huang CJ, Hsieh JS, Lai CS, Lin SD, Huang TJ. Comparison between sonography and mammography for breast cancer diagnosis in oriental women after augmentation mammaplasty. Ann Plast Surg. 2002;49:120-6

[3] Leach MO, Boggis CR, Dixon AK, Easton DF, Eeles RA, Evans DG, et al. Screening with magnetic resonance imaging and mammography of a UK population at high familial risk of breast cancer: a prospective multicentre cohort study (MARIBS). Lancet. 2005;365:1769-78.

[4] Flobbe K, Bosch AM, Kessels AG, Beets GL, Nelemans PJ, von Meyenfeldt MF, et al. The additional diagnostic value of ultrasonography in the diagnosis of breast cancer. Arch Intern Med. 2003;163(10):1194-9

Address for correspondence

Professor Tony Hsiu-Hsi Chen, Division of Biostatics, Graduate Institute of Epidemiology, Institute of Preventive Medicine, College of Public Health, National Taiwan University, Taipei, Taiwan. Room 521, No. 17, Hsu-Chow Rd, Taipei, Taiwan. Tel: +886-2-33228021; Fax: 886-2-23587707; E-mail: chenlin@ntu.edu.tw.

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Challenges in Telemedicine and eHealth: Lessons Learned from 20 Years with Telemedicine in Tromsø

Gunnar Hartvigsen[ab], Monika A. Johansen[a], Per Hasvold[ab], Johan Gustav Bellika[ab], Eirik Arsand[a], Eli Arild[a], Deede Gammon[a], Sture Pettersen[a] and Steinar Pedersen[a]

[a] *Norwegian Centre for Telemedicine, University Hospital of North Norway, Tromsø, Norway*
[b] *Department of Computer Science, Faculty of Science, University of Tromsø, Norway*

Abstract

The Norwegian Centre for Telemedicine (NST) has, over the past two decades, contributed to the development and implementation of telemedicine and ehealth services in Norway. From 2002, NST has been a WHO Collaboration Center for telemedicine. In August 1996, Norway became the first country to implement an official telemedicine fee schedule making telemedicine services reimbursable by the national health insurer. Telemedicine is widely used in Northern Norway. Since the late 1980's, the University Hospital of North-Norway has experience in the following areas: teleradiology, telepathology, teledermatology, tele-otorhinolaryngology (remote endoscopy), remote gastroscopy, tele-echocardiography, remote transmission of ECGs, telepsychiatry, teleophthalmology, teledialysis, teleemergency medicine, teleoncology, telecare, telegeriatric, teledentistry, maritime telemedicine, referrals and discharge letters, electronic delivery of laboratory results and distant teaching for healthcare personnel and patients. Based on the result achieved, the health authority in North-Norway plans to implement several large-scale telemedicine services: Teleradiology (incl. solutions for neurosurgery, orthopedic, different kinds of surgery, nuclear medicine, acute traumatic and oncology), digital communication and integration of patient data, and distant education. In addition, the following services will also be considered for large-scale implementation: teledialysis, prehospital thrombolysis, telepsychiatry, teledermatology. Last in line for implementation are: pediatric, district medical center (DMS), teleophthalmology and ear-nose-throat (ENT).

Keywords:

Telemedicine, remote consultation, teleradiology, telenephrology, telepsychiatry

Introduction

With its interdisciplinary approach to problem solving, the Norwegian Centre for Telemedicine (NST) has, over the past two decades, contributed to the implementation of telemedicine and eHealth services in Norway. Through its role as a WHO Collaboration Center for telemedicin (from 2002), NST has also influenced the implementation of telemedicine services all over the world.

The Norwegian Ministry of Health has acknowledged telemedicine as a legitimate way to deliver health services in Norway and has politically and financially supported its development. In 1993, NST was appointed national competence centre with the stated purpose of coordinating telemedicine research and national development activities.. In August 1996, Norway became the first country to implement an official telemedicine fee schedule making telemedicine services reimbursable by the national health insurer. By 2006, a wide range of information and communication technologies (ICT) have been implemented in the healthcare sector in Norway.

NST has played a leading role in investigating new telemedicine applications, along with developing and strengthening routine large-scale telemedicine services. Because of the achievements in telemedicine in Tromsø, an international masters and phd-program in telemedicine and eHealth was established at the University of Tromsø (UiT) in 2005. In 2006, the Research Council of Norway appointed Tromsø Telemedicine Laboratory, which is hosted by NST, as a center for research-based innovation in telemedicine and eHealth.

We will in this paper report on some of the results we have achieved in telemedicine in Tromsø during the last 20 years. The services will be presented in order of importance for large-scale telemedicine services.

Materials and methods

The services reported in this paper are established telemedicine services that in 2006 all were recommended for large-scale telemedicine services in North-Norway [1]. Evaluations of the presented services have mainly been published as reports at the Norwegian Centre for Telemedicine (www.telemed.no). Most reports have been written in Norwegian. A brief overview in English can be found in [2].

The Tromsø telemedicine experience

North-Norway is characterized by a scattered population and a scarcity of health service specialists, most of whom

are centered in Tromsø. This was also the starting point for telemedicine in Tromsø – to offer specialist services in the rural districts. The initial initiative for research on pilots in telemedicine was initiated by the R&D department at the Norwegian telecommunication company Telenor. The research activities were soon moved to Tromsø as the University Hospital of North Norway (UNN) has a reputation for being open to new ways of addressing the challenges of the rural region of Norway.

Telemedicine is widely used in Northern Norway. Since the late 1980's, UNN has experience in the following areas: teleradiology, telepathology, teledermatology, teleotorhinolaryngology (remote endoscopy), remote gastroscopy, tele-echocardiography, remote transmission of ECGs, telepsychiatry, teleophthalmology, telenephrology/-dialysis, teleemergency medicine, teleoncology, telecare, telegeriatric, teledentistry, maritime telemedicine, referrals and discharge letters, electronic delivery of laboratory results and distant teaching for healthcare personnel and patients.

Several hospital departments employ videoconferencing on a regular basis to conduct remote medical consultations and for training and meetings with colleagues. In recent years, the focus has changed towards off-line tele-consultations. Dermatology, ear-nose-throat (ENT) and screening for diabetic retinopathy have become routine services in the region.

In June 2005, Helse Nord (which manages the North-Norway Health Region) established an expert group (HNEG) that performed a systematic review of the tested telemedicine services that were suitable for large-scale implementation. Their evaluation [1] was based on clinical needs, cost/benefit compared with transportation of patients or health personnel, requirement to the service's functionality and user-friendliness, and requirements for relevant training.

The HNEG made its recommendations based on a study of 282 projects from NST's project portfolio and response from clinicians and researchers at UNN and UiT. After the review with use of a sorting key, 54 projects were identified for further evaluation. These were distributed among 21 different subject areas/themes. These were: acute medicine, cardiology, endocrinology, geriatric, dermatology, communication, nephrology, neurosurgery, nuclear medicine, obstetrics/gynecology, oncology, orthopedic, pathology, pediatric, caring, primary care and district medicine, psychiatry, radiology, distant education, ear-nose-throat and ophthalmology. The projects were evaluated through 7 points (type of project, relevant for evaluation, target group, need for the service, cost/benefit, requirements/premise and total rank). After an examination of the different telemedicine projects, the group decided to organize the four thematic areas into four groups in accordance with priority:

Group 1 – Must be implemented: Teleradiology (incl. solutions for neurosurgery, orthopedic, different kinds of surgery, nuclear medicine, acute traumatic and oncology), digital communication and integration of patient data and distant education.

Group 2 – Should be implemented: teledialysis, prehospital thrombosis, telepsychiatry, teledermatology.

Group 3 – may be implemented: pediatric, district medical center (DMS), teleophthalmology and ENT.

(*Group 4* – not recommended.)

The success factors for a successful implementation were: usability, user participation, adequate training, potential for research, stated requirements for Mean Time Between Failures (MTBF), and good communication between ICT personnel and clinicians.

The HNEG did not find any reasons to give priority to projects within emergency medicine (excl. prehospital thrombolysis), cardiology, endocrinology, geriatric, gynecology / obstetric, pathology, and homecare, except for the general improvement gained through the use of digital communication and distant education.

The highest ranked services (group 1 and 2) are presented in the following.

Teleradiology

Radiology has been a routine service for many years. This has been a pioneering activity, where the radiologist very early started to use and transfer digital images. The development was driven by enthusiasts at the radiology department, among them Dr. Jan Størmer, who very early saw the potential of this service. In fact, the first database solution was developed by Dr. Størmer himself. The Radiology Department at UNN (former RiTø) is linked to all hospitals in the Northern Norway Health Care Region for teleradiological consultations. UNN is used for as a reference for second opinions. In the beginning, the traditional x-ray images were digitized and transferred to UNN in digtal form. The radiologists at UNN examined the images on the screen, and reported their diagnoses.

One of the earliest remote site users was Troms Military Hospital (TMS). The hospital has no radiologist and for the most part does not take in patients in need of emergency assistance. In September 1992, teleradiology was established between TMS and UNN. X-ray pictures scanned at TMS and transferred to the radiology department, UNN. Before the teleradiology service was established, radiologists at UNN visited TMS one day a week. After the teleradiology service had been installed, TMS received immediate examination of their images. In 1995, the service was all-digital. TMS uses the teleradiology service offered by UNN more than any other hospital in the region.

Problem: To provide radiology services in Nordland, Troms and Finnmark County.

Solution: In the beginning, scanned x-ray images were transferred to UNN via ISDN lines. Later, all-digital solutions were introduced.

Equipment: An in-house developed database for administration of digitized x-ray pictures (RIS), a digtalization unit (scanner) at remote hospitals, and access to ISDN services. (Later connection to the Norwegian Health Network.)

Lessons learned:

- Radiologist at UNN became available 24 hours a day for all connected hospitals.
- Less need for travel for both patients and radiologists.
- Vendors of PACS / RIS did not follow the DICOM standard. The radiology department had to put a lot of effort into securing interoperability.
- Study: 90% of 2280 interviewed patients were very satisfied with the fact that the service offers radiology locally and that the waiting time is minimal [3].
- Effect of teleradiology: reduced need for meetings. Reduced response time from Radiology Department.
- More than 20.000 examinations were performed in 2005.

Helse Nord Expert Group recommendations [1]:

- Teleradiology is one of the success stories in telemedicine. It is used both for emergency and elective purposes, and has become increasingly important in regional clinical cooperation.
- Teleradiology has been the foundation for telemedicine services in other areas, such as neuron surgery, traumatology, orthopedic, general surgery, vascular surgery, heart surgery and oncology.
- Teleradiology is suited for large-scale operation, and the established model is a good model for other clinical areas.

Digital communication and distant education

Figures from Finnmark County (74.000 inhabitants) show a saving of about 125.000 Euro each month from using video conferences (VC) for meetings. The use of VCs started in 1989, and was routinely used from 1994. It has been important for the development of health care in the northern area. Specialists, doctors, and patients have been able to meet without traveling. In fact, most of the VC activity is in the north. Helse Finnmark alone uses 1/3 of the traffic in the Norwegian Health Network. All hospitals in North-Norway use VC. The biggest users are the outpatient clinics. The success factor is that VC has become integrated into the work of doctors, psychologists, nurses, etc.

The use of VC in education has been very important for the development health services in North-Norway. It is expected that PC-based VC solutions will improve the use even more.

In 2000, NST collaborated with the IT department and medical departments at UNN, 5 GP offices, Norwegian Health Network and Well diagnostics. Together they established a service for electronic entry of discharge letters between the hospital and GP offices. In 2004, this service was included to cover all hospitals and GP offices in Helse Nord (that were connected to the Norwegian Health Network). In addition, the GPs could send referrals and x-ray requisition to all hospitals in the area.

Helse Nord Expert Group recommendations [1]:

- Digital communication (DC) represents a high volume service that covers the whole health service. With multimedia enclosures in electronic health records (EHR), DC will be included in many demanded services (e.g., heart sound, dermatological images, etc.)
- Further education is one of the main tasks for Helse Nord. This service has a high volume, and is important in guaranteeing a high degree of expertise throughout the chain of treatment.

Teledermatology

Teleconsultation within dermatology started in 1988 between UNN and Kirkenes Hospital. In 1996, UNN established teleconsultations with Hammerfest Hospital. Both remote hospitals set up teleconsultations once a week, normally with 10-12 patients. These consultations last for 2,5 to 3 hours altogether. In addition, specialists travel from UNN to the Vadsø health care service 6 times a year. Both Kirkenes and Hammerfest hospitals have equipment for treating dermatological diseases and are able to offer their patients treatment locally.

Two teledermatology services are offered:

- Videoconference (VC) consultations: Patient and GP (or other health workers) meet with the specialist through VC. Most patients get immediate response and treatment is started (if needed). This service is in regular use between UNN and Finnmark.
- Still pictures: pictures and referral (from GP to specialist) are enclosed in emails.

In 2003, 84 of 200 GP offices in North-Norway had such equipment.

Problem: To provide dermatology services in Troms and Finnmark County.

Solution: From the beginning: Videoconference equipment at both places. Later also use of email with enclosures (images, referral letter).

Equipment: A video camera is attached to a codec and the signals are transmitted via ISDN. The codecs are H320 standard and the service uses 384 kbit/s. For email: Specialized software, digital camera, dermatoscope.

Lessons learned:

- Initial skepticism made recruitment of specialists difficult. However, simple pilots proved beneficial in spite of the limited technology.
- General practitioners who use this service are generally satisfied, and they have gained new knowledge in the field. This new knowledge has enabled them to do a better screening of patients that need to be referred to the hospital.
- In this way, both patients and accompanying persons need to travel less frequently

- With the introduction of email service instead of VC, more patients were refereed to the specialist clinic. This effect was a result of less available information.
- As many telemedicine services, the GPs become updated within dermatology and gradually took care of more cases themselves.
- Patients do not have to travel to UNN. They need only to meet with their GP.
- Equipment can be used in other settings, e.g., wound treatment (homecare), psoriasis and other chronic diseases.

Another lesson, which goes for most of the telemedicine services that have been established, is that cost saving depends on volume. The cost-effectiveness of telemedical services and electronic messages is often dependent on investment costs, the number of consultations or electronic messages per year, as well as the costs of traveling to a specialist hospital [2].

Helse Nord Expert Group recommendations [1]:

- The VC-based teledermatology solution has considerably improved dermatology competence in primary care. However, the solution is resource-demanding and the GP becomes often a ”secretary” to the specialist. Thus, the use of other groups of supporting personnel should be considered.
- The use of this service is decreasing. Possible explanations are increased level of competence in primary care and decreased enthusiasm. In addition, a dermatologist has been employed in Karasjok. He is also ambulating to Vadsø.
- VC-based teledermatology is recommended for large scale operations.

Telenephrology/teledialysis

The aim of the teledialysis service was to improve the quality of patient care by supporting patients and nurses at the remote centres (i.e., the satellite staff) with the same quality of follow-up care and support offered at UNN. With the help of telemedicine, a common workplace was created by integrating satellite staff into UNN's everyday routines. Daily communication with the satellites is established using IP-based videoconferencing.

Hemodialysis (blood- or machinedialysis) is performed 3 times per week in a health institution. The treatment takes 5 hours each time. In Norway, hemodialysis is performed at 11 hospitals with nefrologic competence. Each of these hospitals has at least one satellite. (Total 28 satelites.) Specially trained nurses manage Satelite dialysis.

Videoconferencing is used in the following areas:

1. Nursing (Day-to-day contact between nurses). In these meetings, urgent problems are discussed every day in a 15-minute session to each satellite.
2. Rounds. (Doctor and nurse). Review of all the patients and patient rounds. Ultrasound equipment is used as required. Every 14 days, alternating between Alta and Hammerfest in Northern Norway.
3. Training (In-house training once every 14 days). The satellites participate in in-house training at the University Hospital of North Norway through videoconferencing.
4. Acute problems are dealt with as needed.

Problem: To improve the quality of patient care by providing patients and nurses at the remote centers with the same quality of follow-up care and support as that received by patients and health staff at UNN.

Solution: With the help of telemedicine, a common workplace was created by integrating satellite staff into UNN's everyday routines. Daily communication with the satellites is established using IP-based VC.

Equipment: VC equipment, ultrasound equipment, digital stethoscope, dialysis software consists of: 1) A database archive which stores data from the dialysis as well as computer data, and 2) A program which keeps track of the current status of the dialysis machines to which patients are connected.

Lessons learned:

- UNN experienced improved reliability of advice given. The staff is integrated into a virtual team.
- Alta and Hammerfest experienced faster response, higher information quality, and improved safety.
- Patients experienced improved continuity in check-ups and treatment through direct contact with UNN staff (instead of through the nurse).
- Costs are saved because patients no longer need to travel to UNN for regular check-ups. Specialists travel less frequently to Alta and Hammerfest.

Helse Nord Expert Group recommendations [1]:

- Teledialysis is well established in Troms and Finnmark County, but not developed in Nordland County.
- Teledialysis should be expanded to new sites even though the volume is low.
- The first satellites were established in Hammerfest and Alta. Later a satellite was established in Kirkenes at a significantly reduced cost compared to the two first satellites. The cost of establishing telemedicine for satellite dialysis units is considered to be a small cost compared to the benefit, given the high reimbursement fees per dialysis treatment.
- The project has been very well received among the patients who to a larger extent are able to receive treatment and follow-up close to their homes.

Prehospital thrombolysis

Each year, 12-15.000 Norwegians get a stroke (cerebral infarction). Thrombolysis treatment of the cerebral infarction can, if the treatment is started within 3 hours from the symptom occurred, limit the brain damage and reduce the mortality and function failures. A treatment can be started after image diagnosis by CT or MR and clinical evaluation by a specialist [1].

Prehospital thrombolysis was started in Troms County in 2000. Community doctors and ambulance workers were trained to perform thrombolysis for acute cardiac infarction (thrombosis) based on a precise algorithm and practical procedure. ECG was sent via a mobile phone or network to a PC at UNN [1].

Studies have shown that the effect increases with increasing distance from UNN. In 2004, all ambulances and casualty clinics in North-Norway were equipped with defibrillator and ECG-equipment.

Lessons learned:

- Telemedicine can increase the availability of thrombolysis treatment through faster diagnosis. In addition, treatment can be given decentralized [4].
- For difficult cases, stroke units at the large hospitals can be contacted for advice throughVC.

Helse Nord Expert Group recommendations [1]:

- This service has a large volume and good prospects. It is recommended for large-scale services.

Telepsychiatry

Telepsychiatry has traditionally been based on videoconferencing or other means of two-way audio-visual communication. It has been used in mental health care for close to 30 years. Use in Norway has expanded steadily, particularly in North Norway [5]. Telepsychiatry includes professional supervision, education, clinical cooperation as well as actual clinical practice.

Reduced costs and improved technical solutions have increased both interest in and use of telepsychiatry. Loudspeaking telephony, videophone, phone conferencing, videoconferencing, and electronic intra- and Internet communication can overcome geographic distance between patient and therapist as well as promote cross-disciplinary collaboration, networking, and skills development. More recent developments in web-based services offer promise for expanding the reach and impact of ICT in mental health.

Lessons learned:

- VC is widely used in psychiatry, both for therapeutic and administrative purposes.
- Use of VC has considerably reduced travel costs.

Helse Nord Expert Group recommendations [1]:

- VC in psychiatry is well suited for large-scale implementation.

Results and discussion

In almost twenty years, UNN and The Norwegian Centre for Telemedicine have been in the forefront of developing telemedicine services, both regionally, nationally and internationally. Today, telemedicine is widely used in Northern Norway and is well integrated into routine health service provision. Documented areas of benefit are [2]:

- *Economic*: Travel costs, number of hospital admissions, time spent by health practitioner, and paper and postage.
- *Qualitative*: Time for other tasks, improved data quality, patients do not have to travel, health benefit where "time counts", screening of patients, competence in medical disciplines, professional confidence, access to specialists, efficient use of specialist expertise, and patient empowerment.

As of 2006, the services presented in this paper are referred to as traditional telemedicine services. The current focus of both the European Commission and others is on how to exploit ICT in efforts to support patients and the general public with tools that enable them to take better care of their own health. This increased focus on health promotion and disease prevention, along with improved collaboration between patients and health care systems, is motivated by many factors; the elderly-boom, population growth in general and growth in chronic diseases in particular, as well as an increase in lifestyle related diseases. New technologies, such as mobile terminals and wearable sensors show promise as a means for meeting these challenges. NSTs work with personalized healthcare since 2000 reflects our intention to be at the forefront of this new frontier.

References

[1] Northern Norway Regional Health Authority. Which of the tested and suitable telemedicine services should be prioritized for large scale implementation? Report from a project group appointed by the Helse Nord board 22. June 2005 in connection with the case "decentralization of specialist health service". (www.helse-nord.no) (In Norwegian)

[2] Johnsen E, Breivik E, Myrvang R, and Olsen F. Benefits from telemedicine in Norway. An examination of available documentation. HØYKOM report No. 2006:1 (ISBN 82-12-02326-5)

[3] Johansen I, and Breivik E. Is teleradiology in primary care cost effective? *Tidsskrift for Den norske lægeforening 2004*: 124 (19): 2490-2. (In Norwegian)

[4] Abelsen B, and Olsen F. Profitability analysis for three telemedicine services. Report No. 12/2004, Norut Social Science Research Ltd, Tromsø, Norway. (In Norwegian)

[5] Hanssen B, Wangberg SC and Gammon D. Use of videoconferencing within Norwegian psychiatry. *In press: Journal of Telemedicine and Telecare.*

Address for correspondence

Gunnar Hartvigsen email: gunnar.hartvigsen@telemed.no)

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

A Satellite Infrastructure for Health Early Warning in Post-Disaster Health Management

C. E. Chronaki[a], A. Berthier[b], M. M. Lleo[c], L. Esterle[a], A. Lenglet[g], F. Simon[g], L. Josseran[f], M. Lafaye[d], Y. Matsakis[h], A. Tabasco[e], L Braak[b]

[a] *Institute of Computer Science, FORTH, Heraklion, Crete, Greece,* [b]*MEDES, France ,*
[c]*University of Verona, Italy,* [d]*CNES, France*
[e]*GMV, Spain,* [f] *InVS, France,* [g]*ISCII, Spain,* [h]*Telemedicine Technologies SA, France*

Abstract

The risk of epidemics and emerging or re-emerging diseases such as avian flu, tuberculosis, malaria and other vector-borne diseases, is rising. These risks can be contained with prevention, early warning, and prompt management. Despite progress in information technology, communication is still a bottleneck for health early warning and response systems in post-disaster situations. This paper presents Satellites for Epidemiology (SAFE), a component-based interoperable architecture for health early warning that employs satellite, radio, and wireless networks, geographic information systems, integration technology, and data mining to promptly identify and respond to a disease outbreak. In a post-disaster situation, a mobile health emergency coordination center is established and integrated to public health services for health monitoring. The added-value of SAFE for post-disaster health management will be demonstrated as part of an earthquake readiness exercise regarding a typhoid fever epidemic, in the island of Crete. Advanced communication and data mining techniques in SAFE offer new tools to the "Epidemic Intelligence" and contribute to advanced preparedness and prompt response by lifting communication barriers, promoting collaboration, and reducing the isolation of affected areas.

Keywords:

health early warning, telemedicine, satellite communication, epidemic intelligence, public health surveillance

Introduction

Identification and monitoring of potential risks to human health has become an important part of the so-called "Epidemic Intelligence"; a new movement in surveillance as the population becomes less tolerant to epidemic prone diseases, making their fast detection and response an increasing priority [1]. Health early warning systems in remote and inaccessible areas or those prone to natural or man-made disasters, enabled by satellite communications, can significantly limit the risk of onset and the effects of epidemics and contribute to settling public health issues. Economically speaking, satellite communications save money by enabling rapid and coordinated response and optimal adjustment of resources when deploying an emergency plan [2].

The SAFE project aims to develop and demonstrate the added value of satellite communication services including low and high bandwidth access to the Internet, co-operative working, and geolocalisation for all phases of biological crisis including prevention, early warning, and crisis management. In this way, SAFE will contribute to a roadmap on how satellite services, by enabling or restoring access to information, will be integrated in European healthcare systems and civil protection procedures to further facilitate interaction and communication.

A practical exercise of early warning at the onset of a typhoid fever epidemic will be demonstrated on the island of Crete, Greece. The scenario, based on epidemiological monitoring after an earthquake, will validate the SAFE approach within the user community and if successful, SAFE could become part of regular earthquake readiness exercises. Furthermore, other applications of the SAFE approach are envisaged as related to early warning and management of re-emerging diseases such tuberculosis, but also to biological and radiological threats. Demonstrations of SAFE mainly in the frame of readiness exercises can be implemented all over Europe upon user-request, promoting the added value of SAFE and allowing potential end-users and stakeholders to assess the use of the SAFE system according to relevant criteria such us availability, interoperability, and security.

The rest of the paper is organized as follows: Materials and Methods (next) is dedicated to the SAFE technical architecture. It is followed by the results section, which focuses on the envisaged demonstration for typhoid fever and the evaluation methodology. In the discussion section, SAFE is placed in the context of readiness exercises and current early warning and response systems. Conclusions summarize the paper.

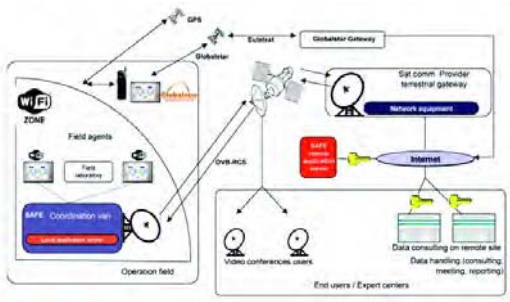


Figure 1 - The SAFE Architecture

Materials and methods

SAFE aims towards better assessment of the epidemiological risks based on real-time objective data, which is a strategic objective for public health policy and of potentially significant impact on the eHealth industry. The SAFE technical approach shown in Figure 1 will be promoted to the different actors of civil protection and epidemiological surveillance, in the context of readiness exercises. Its main components are as follows:

- *Network of expert centers* able to communicate by satellite and interoperable with other networks (e.g. gateways to Galileo[1] and GMES-RESPOND[2]);
- ***Coordination van*** equipped with DVB-RCS [5] capability and Internet access for communication with decision makers and expert centers. It has the role of the local coordination centre. Communication between the local coordinator and the mobile teams on the field is ensured via a local Wi-Fi network and Satellite phones outside the Wi-Fi coverage. Mobile teams are equipped with handheld terminals for data collection;
- *Field laboratory or lab kit* equipped with biological and biomedical equipment to identify microbial agents responsible for the threats to enable in-situ analysis;
- *SAFE information system* including a data collection sub-system and a Geographical Information System (GIS). Besides geolocalisation, the GIS will enable visualizing the evolution of the epidemic and monitoring the alarm levels associated with the different geographical regions. This system will be based on relevant standards and will be interoperable with existing information systems, i.e. national health information system, emergency response information system, epidemiological information systems, and others. It will facilitate monitoring of indicators and other measures potentially useful in modeling the transmission patterns. Dissemination of this data can be done with the Internet via satellite.

This general SAFE configuration can be adapted to the requirements of specific demonstrations or pilot studies. Satellite terminals and data collection points can operate without the coordination van. Data collection can be integrated to existing procedures with interoperability standards, cooperating with and complementing current data collection facilities.

Components of the SAFE architecture

The main components of the SAFE architecture as deployed in the post-crisis earthquake readiness exercise in Crete are: the mobile coordination center, the laboratory kit carrying analysis of biological samples, and the SAFE information system integrating the GIS, the Emergency Coordination System (EKAB), a public health information system, and a Health Observatory (HealthObs) carrying out epidemiological surveillance. A simulator will feed into the system medical cases including those of typhoid fever in a realistic fashion modeled along established readiness exercises with the involvement of WHO[3] and ECDC[4] [11,12].

Mobile coordination center

In the post-crisis demonstration scenario of SAFE, a mobile pre-hospital coordination centre will be set up following an earthquake disaster near a settlement. The coordination van will exploit satellite and radio technologies in accepting and registering rescue calls, in coordinating rescue teams, as well as in providing valuable input for health early warning.

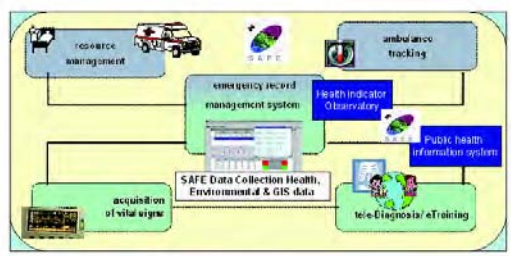


Figure 2 - SAFE adds value to the already advanced ICT infrastructure of EKAB-Crete.

The ICT infrastructure of EKAB-Crete is quite advanced [5,6]. Its components appear in Figure 2: (a) emergency record management system offering protocol-based triage for all episodes (in daily operation since 1997); (b) telemetry applications from ambulances transferring vital signs, 12 lead ECG, etc.; (c) resource management using a GIS combined with GPS-equipped ambulances for central ambulance tracking and route guidance; (d) teleconsultation subsystem, eLearning seminars/refreshing courses for different population groups including protocols for addressing BCRN (Biological, Chemical and Radionuclear Agents), biological threats, earthquakes (very frequent in the area), nuclear disasters, etc.

1 Galileo: European Satellite Navigation System: http://ec.europa.eu/dgs/energy_transport/galileo/index_en.htm
2 Global Monitoring for Emergency and Response,Humanitarian Aid http://www.esa.int/esaLP/SEM3B60DU8E_LPgmes_0.html
3 World Health Organization: http://www.who.int
4 European Center for Disease control: http://www.ecdc.eu.int/

The SAFE system adds value to the existing ICT infrastructure by making EKAB operational in minimal time and by connecting it with expert centres in other parts of the world. Existing management procedures and guidelines for health management in natural disasters that are already enforced, will be extended to demonstrate the role of SAFE and its added-value services.

The mobile coordination van will also provide all the equipment for epidemiological surveillance: (a) Local Wi-Fi network; (b) handheld terminals for mobile teams; (c) laboratory kit to perform analysis of biological or clinical samples; (d) GIS for geolocalisation of critical resources and visualization of the sanitary and epidemiological conditions; (e) links with expert centres, decision makers, or public health institutions according to the master preparedness plan.

Analysis of biological samples

A laboratory kit will be employed for the in situ-analysis of biological samples. Minimal analyses can be performed using mobile facilities (including microbial culture and optical microscope). Samples that will be analyzed for detection of the typhoid fever etiological agent may include water, blood, stools, etc. The lab equipment is connected to the SAFE information system for data collection and to microbiology experts for evaluation and interpretation of microscope images and results obtained from bacteriological culture and other biological tests.

GIS monitoring critical resources and the epidemic outbreak

The GIS application will facilitate monitoring of critical resources and visualization of the simulated evolution of the epidemic. The status of critical resources will be presented on a map to facilitate overview of the public health situation and epidemic evolution. The following aspects will be addressed in the readiness exercise: (a) real time localization on a map of mobile rescue teams and ambulances; (b) report on the status of critical resources (drinking water, excreta disposals etc.); (c) visualization of the public health situation; (d) follow-up the timeline of the simulated epidemic marking health alerts for verified and suspected clinical cases of typhoid fever.

Health indicators, data mining and health early warning

Health indicators associated with electronic health record systems linked to alerts for the onset of epidemics and appropriate use of data mining techniques will further demonstrate the value of effectively sharing and transparently processing health information. Three main information systems will be deployed and integrated in the post-crisis monitoring scenario: the SAFE information system, the EKAB emergency response information system, and a primary care/public health information system. The primary care/public health information system will record the health status of the people in the settlement. A simulator program will generate credible clinical cases for the purpose of demonstration. Finally, the HealthObs integrated environment will be connected to these systems [8,9]. HealthObjs is based on advanced data-mining and knowledge discovery operations and enables specific treatments to identify trends from the collected data and from other sources of data from external databases. HealthObjs supports discovery of interesting clinical associations and monitoring of crucial health and epidemiological indicators.

SAFE will adopt a unified approach towards collection of data from remote and heterogeneous clinical data sources. The general approach shown in Figure 3, is based on standardized medical data models and XML technology [9] as well as disease surveillance via the HealthObs integrated environment.

In essence, SAFE will demonstrate an active Electronic Health Record System, which through its link to HealthObjs raises epidemiological alerts worthy of further investigation and monitoring. HealthObs offers advanced data analysis operations towards investigation of population health dynamics. During an epidemiological investigation the query presented to the system should carry with it the health or epidemiological indicators of interest. In addition, discovery of clinical correlations (e.g. disease symptoms, or therapeutic treatment) enhances potential for evidence-based medicine since correlations, once validated, help to establish a rationale baseline for evidence-based medical decision making. Although this capability does not exist in the current system of EKAB, SAFE will demonstrate a proof of concept in the readiness exercise and possibly lead to its adoption as part of the daily practice of health surveillance in Crete.

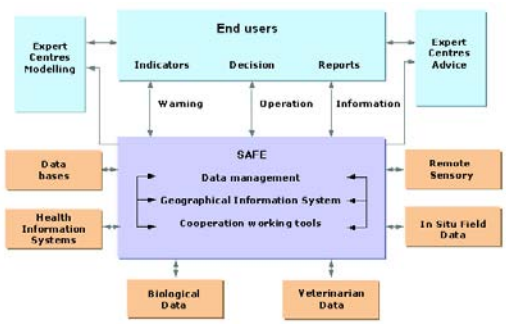


Figure 3 - Interface with external information systems.

Satellite connectivity for second opinion on simulated cases

Collaboration with expert centres and other laboratories for second opinion over the Internet will demonstrate the added value of seamless integration and in-context interpretation of simulated clinical cases of typhoid fever.

Video conferencing for coordination of mobile teams

Video-conferencing techniques using satellite technology will enable coordination of rescue teams and training of volunteers in the earthquake settlement. Teams of volunteers will be trained on site using satellite technologies before a vaccination campaign starts.

Results

The SAFE architecture will be validated in the post-crisis health monitoring section of an earthquake readiness exercise in 2007 by local authorities and WHO under a comprehensive scenario of a typhoid fever epidemic.

The proposed post-crisis monitoring scenario will focus on a simulated typhoid fever epidemic in a settlement of earthquake victims after the disaster. We assume that several critical buildings including that of the pre-hospital and emergency coordination centre of Crete (EKAB) have been destroyed. Thus, the situation calls for expert help from other regions, the set up of a mobile coordination centre, training of volunteers, and monitoring the onset of the epidemic. The emergency response system of EKAB will be restored using the capabilities of SAFE.

Network connectivity will be established using the DVB-RCS connection of the coordination van and a Wi-Fi network. The video-conference application will enable communication with expert centers, health authorities and the team members for a better coordination in support of Xenocrates, the master plan for earthquake preparedness. This application may also support the training of volunteers and possibly the education of the people in the settlement regarding health prevention or adoption of extraordinary sanitary measures. The videoconference capabilities will enable connection to expert centers and laboratories to meet the need for a second opinion on relevant cases of typhoid fever.

Epidemiological surveillance of the public health situation will be achieved with the rapid restoration and deployment of the SAFE information system. Handheld equipment will facilitate data collection in the field. It will be possible to make biological analyses in the field using the lab kit provided in the coordination van. The onset of an epidemic will be monitored using advanced data mining techniques in the integrated public health and prehospital emergency systems. Interoperability of the data collection subsystem with the national health information system and in particular public health/primary care information system and the emergency coordination center will facilitate health surveillance in HealthObs.

Although, the infrastructure at EKAB is particularly advanced, SAFE adds significant new capabilities demonstrating interoperability between traditionally diverse systems and additional functions. As the whole system will be evaluated in the context of the master plan for post-crisis management with user involvement in all levels of the hierarchy, we expect that user acceptance will be high. Furthermore, a formal user acceptance methodology will be applied to evaluate any bottlenecks that may come up due to the proposed revised master plan.

Evaluation methodology

During the various phases of the simulation, an evaluation of user acceptance will be carried out based on:

- Follow-up of the different simulation phases and the observation of the user attitudes during the simulation, and in particular at the experts centres, at the mobile pre hospital coordination centre, and during the training of the volunteers;
- Interviews carried out on representatives of stakeholders.

This evaluation approach attempts to identify the facilities and the constraints met by users in their practice during the post-crisis scenario, and to measure the suitability of the techniques and the interest of stakeholders. More generally, this study will contribute to the definition of tools and indicators to access user acceptance of health early warning systems.

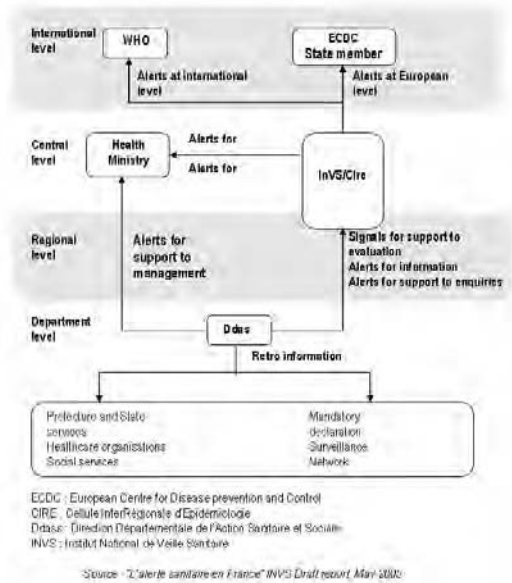


Figure 4 - Information flow in epidemic surveillance.

Discussion

At the European level, an Early Warning and Response System (EWRS) [12] has been established to alert public health authorities in Member States (MS) and the European commission (EC) on outbreaks with greater than national dimensions, where a coordinated EU action may be required. The EWRS is a telematic system focusing on infectious diseases and linking the designated authorities in MS and the EU. The system allows for immediate exchange of views on risk assessment and risk management that are crucial for timely public health action. The EWRS has already proven to be a useful tool during a number of outbreaks/incidents.

In fact, health warning systems are organized in different ways at the level of the MS. These systems include for example surveillance of indicators, health events, and animal diseases as:

- All declarations for mortality;
- Indicators of biological, chemical and physical exposures (pollution, water quality, air quality, radioactivity, aerobiological, meteorological etc);

- Indicators of infection and morbidity;
- Indicators of risky comportments.

Surveillance of events includes different types of networks:

- Networks of clinicians, civil and military;
- Networks of epidemiologists;
- Networks of public health agencies;
- Networks of civil security;
- European and international epidemiological surveillance networks (pharmacological, toxicological etc.).

Surveillance of animal diseases includes several types of networks and particularly surveillance ensured by veterinarians.

There are many organizations involved in epidemiological surveillance, from the regional to the national level. The information flow between the different actors on Figure 4 is based on the French system and provides an example of how information flow may proceed in the European context. A list of the main networks on communicable diseases at the European level can be found in [9].

SAFE will take into account the procedures and information flow reflected in the national preparedness master plan. Its focus will be on facilitating better coordination and cooperation among the actors involved e.g. civil protection and health authorities not only at the national, but also at the EU and international level. The exercise will be modeled along established exercises addressing the problems noted in their review with more effective communication tools and procedures.

Conclusions

The component-based interoperable architecture of SAFE addresses the communication and coordination needs of earthquake post-crisis health management. SAFE focuses on health early warning and monitoring the onset of epidemics emphasizing the value of satellite technologies in education, data collection, integration, processing, analysis, and notification. In particular, seamless workflows, collaboration with expert centres, geographic monitoring of rescue teams and critical resources, transparent data collection and analysis, as well as provision of health early warning will be highlighted in a scenario of typhoid fever epidemic outbreak in a settlement of earthquake victims.

SAFE is in line with the vision of ESA[5] to specify, develop, and demonstrate with support from WHO, a European health early warning system that will become a key tool for the national and European bodies in charge of epidemiological surveillance and especially ECDC.

5 European Space Agency: http://www.esa.int

Acknowledgments

The authors would like to thank all the members of the SAFE consortium for their input. The project SAFE is co-funded by the European Space Agency (ESA) and is part of the ESA telemedicine via satellite transition phase set up to pave the way for the creation of a European user-driven telemedicine via satellite program.

References

[1] Howard MJ, Brillman JC, Burkle FM, Jr. Infectious disease emergencies in disasters, Emerg. Med Clin North Am 1996, 14:413-428.

[2] World Health Organization. Earthquakes. Technical Hazard Sheet. Natural Disaster Profile 2006.

[3] Dario C. et al. Opportunities and Challenges of eHealth and Telemedicine via Satellite. European Journal of Medical Research, vol. 01, supplement 1, 2005.

[4] ETSI EN 301 790: "Digital Video Broadcasting (DVB); Interaction Channel for Satellite Distribution Systems."

[5] Tsiknakis M., Kouroubali, A., Vourvahakis, D., Orphanoudakis, SC. “Implementing a Regional Health Information Network: Impact on health care performance and the management of change”, in “Int Health Care Management - Advances in Health Care Management”, eds G. T. Savage, et al (eds) Elsevier Oxford, UK, vol 5, pp. 297-329, 2005

[6] Chiarugi F., et al: «Continuous ECG Monitoring in the Management of Pre-Hospital Health Emergencies» in Computers in Cardiology 2003;30:205-8

[7] Potamias G., Koumakis L., and Moustakis V., "Mining XML Clinical Data: The HealthObjs system", Networking and Information Systems: Information retrieval and Information mining (in press).

[8] Potamias G. and Koumakis L., "HealthObs: An Integrated System for Mining Clinical XML-formatted Data", J Quality of Life Research (JQLR), vol. 2 (1), pp. 76-82, (2004).

[9] Communicable disease networks. European Commission, DG Health and Consumer Protection. http://europa.eu.int/comm/health/ph_threats/com/comm_diseases_networks_en.htm

[10] Health Protection Agency: Exercise Common Ground: A Pandemic Influenza exercise for the European Union, Serial 5.0, March 2006 http://ec.europa.eu/health/ph_threats/com/common.pdf

[11] Health Protection Agency: Exercise New Watchman: A Pandemic Influenza exercise for the European Union, Serial 5.1, March 2006 http://ec.europa.eu/health/ph_threats/com/watchman.pdf

[12] Byrne D. Commission decision on the early warning and response system for the prevention and control of communicable diseases under Decision No 2119/98/EC of the European Parliament and the Council. Official Journal of the European Communities 2000; L21: 32–35.

Address for correspondence

Catherine Chronaki
FORTH, Institute of Computer Science
PO 1385, Heraklion, Crete, 71110 Greece
Email: chronaki@ics.forth.gr

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Remote Support for Stroke Rehabilitation: MyHeart's "Neurological Rehabilitation" Concept

Toni Giorgino[a], Paolo Tormene[a], Barbara Cattani[b], Caterina Pistarini[b], Silvana Quaglini[a]

[a] *Laboratory for Biomedical Informatics, Dipartimento di Informatica e Sistemistica, Università di Pavia, Italy.*
[b] *IRCCS Fondazione Salvatore Maugeri, Pavia, Italy.*

Abstract

Stroke is a serious neurological accident which accounts for a wide fraction of the healthcare costs in industrialized societies. Recovery from stroke and other neurological accidents usually include motor rehabilitation, maintained for several months, and logopedic training for the recovery of cognitive and speech abilities. The MyHeart consortium is addressing several aspects of cardiovascular diseases' management by combining clothes with embedded biomedical sensors and information technologies. One of the application areas is especially devoted to supporting Neurological Rehabilitation (NR). This article describes how MyHeart's Concept NR is structured and how technologies are leveraged to support both motor rehabilitation and speech/cognitive training. Information technology and garment-embedded sensors, combined, permit assisted training both at the clinic and at home, after discharge from the intensive care unit.

Keywords:

telemedicine, cerebrovascular disorders, stroke, rehabilitation, speech therapy

Introduction

Cardiovascular diseases (CVD) are among the principal causes of death and disability in industrialized countries; they represent a big social and economical problem and costs for their cure are going to increase with the growth of the mean age of population. CVDs are often correlated to a wrong lifestyle and to lack of prevention. Early symptoms detection, prompt intervention and efficient post-acute rehabilitation are necessary to properly address the problem [1].

Given the large impact of the problem and corresponding potential benefits to be gained from more efficient treatments, the *MyHeart* consortium was created. Its mission is to address various aspects of CVDs by means of information technologies (IT) and intelligent biomedical clothes (IBC). It currently includes 33 partners; the project's activities take place between December 2003 and September 2007. Research within MyHeart is partially funded by the European Union to promote advancement of the state of the art in fields such as automated remote monitoring, early symptom detection and innovative sensing technologies. It initially encompassed plans for 16 different applications of IBC to cardiovascular diseases: prevention, detection and treatment were diversely addressed among the proposals, which were named "Concepts". During the first project phase (16 months) each concept was asked to develop both technology and a relevant business model. In April 2005 each of the 16 concepts undertook an extensive evaluation phase to evaluate the corresponding potential market impact. Each concept sampled prospective acceptance of the product or service proposed by conducting standardized interviews with patients, medical experts and economic policy-makers, representing the three main stakeholders for the consortium.

At the end of the concept evaluation phase, the 16 concepts have been reshaped into four new "Product Concepts", namely:

Fitness Coach – maintenance of an healthy lifestyle
Take Care – monitoring people at risk of CVD events
Heart Failure Management – monitoring after an acute event
Neurological Rehabilitation – telemedicine and motion analysis as a support to post-event recovery.

Each Product Concept is partially independent from the others, and being a medium-sized sub-project itself, it develops the required technology and architecture. This paper provides an account of the Neurological Rehabilitation (NR) Product Concept.

Scope

Recovery from a neurological event begins in the acute unit and it is articulated into a series of intensive and extensive rehabilitation protocols which depend on the specific healthcare system in which they are inserted. Patient's needs and relative costs decrease with the progressive re-acquisition of motor and cognitive functions.

Product Concept NR's aim is to support patients in the performance of *speech* and *motor therapy*, both when they are still hospitalized, and after discharge, at home.

Eligible patients will be (1) hemiplegic patients in stable clinical condition (i.e. the problems related to the acute phase and possible complications must be over); (2) patients with mild motor and/or cognitive impairment (this criterion is essential if we want the patient to use the

system alone, or with a limited assistance from his/her caregiver); (3) aphasic patients with no other cognitive impairment that could impact on the system's use. Motor and speech problems may be present or not in the same person. The physician identifies the candidate patients and communicates the enrolment to the physiotherapy/speech therapy responsible.

Patients and their relatives are informed about the study through a meeting with the chief of the NR ward and they will be provided with a booklet with explanation about the goal and the use of the system. Their informed consent will be required.

When hospitalized, eligible patients are taught about the use of the system by the same personnel that administers the usual therapy. Organizational and liability constraints impose that IT-based rehabilitation does not replace conventional therapy hours. After hospital discharge, patients can bring the rehabilitation device at home (or long stay ward), and continue exercises with the help of their caregiver, relative or professional nurse. (The requirement of a caregiver is not imposed by the system. Often post-stroke patients are not self-sufficient.)

Patient stations are user-accessible devices which support the rehabilitation exercises (figure 1). Patients' identity is recognized via personal "check-in cards". Once inserted, the most recent exercise protocol and messages are downloaded automatically at the patient's site. For maximum accessibility, interaction with rehabilitation software, both for speech and motor therapy, takes place via a touch-screen.

Speech therapy is based on the EvoLing software developed by Dr. Hein GmbH [2]. It includes an extensive set of exercises, each promoting one ability: to recognize words, phonemes, graphemes, and pictures. Semantic and orthographic information is considered in order to build exercises with an adapting degree of difficulty.

Motor therapy is an innovative component of the system. When sitting at the patient station, patients can review an instructional movie about the exercise they are asked to perform. They are then asked with the help of the caregiver, to wear a special sensorized garment and plug it into a portable electronics box (see below). After a brief calibration phase, the motion recognition software is started; it provides real-time feedback on the progress and accuracy of exercises by means of clearly visible symbols such as colored bars and a smiling or frowning face. The movements have to be repeated until the assigned number of repetitions are performed correctly, or a timeout expires.

Motion recognition is based on strain sensor technology, directly printed on garments. Conductive elastomers (CE) are polymer-based materials which exhibit electrical conductivity [3]. Polypyrrole, for example, can be deposited in stripes on fabric [4]. After vulcanization the stripes present an electrical resistance around 10-100 k/cm, varying depending on the printed geometry but also on the stretch imposed to the fabric.

Sensing stripes of CE material are thin and can be printed by cheap industrial processes; they do not alter the mechanical properties of the material they are smeared on. After vulcanization they are stable and non-toxic. The same material can be used to realize connection wires between sensors and a portable readout electronics.

When the fabric is then used to sew appropriately-sized and shaped garments, a *sensorized garment* is obtained. For the motor rehabilitation application, Lycra was chosen as a substrate fabric, because it allows the realization of tight-fitting but comfortable shirts. Figure 2 shows an early prototype of a sensing garments. Sensing tracks are visible in black on the outer side. Current prototypes embed the sensing stripes between two layers of fabric, and therefore they are not visible. The sensor layout is presented in figure 6.

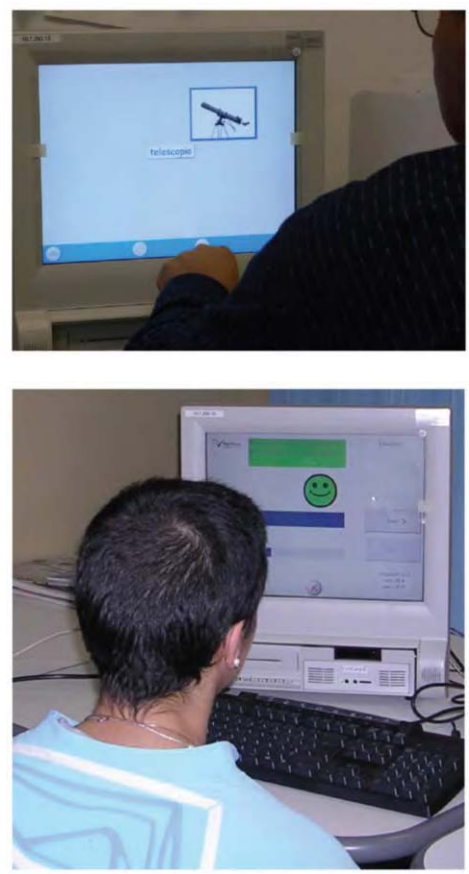


Figure 1 – Concept NR's patient station in use

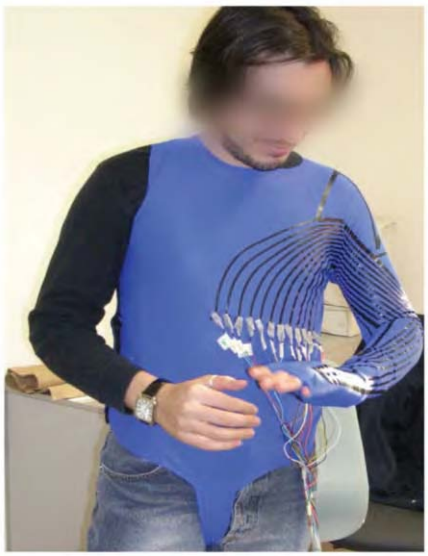


Figure 2 – Sensorized garment for motor rehabilitation. Picture of an early prototype whose CE tracks are visible.

Architecture

Figure 3 shows the telemedicine infrastructure underlying Concept NR. It is based on EvoCare, a solution adopted by Dr Hein GmbH, which is also a partner of the product concept [5]. The main components are one *server*, and a number of *therapist stations and patient stations*. The components are connected by a secure network which, depending on organizational constraints, may leverage the hospital's wired infrastructure, use wireless links, or based on public cellular network (GPRS) connections, as is the case when patient stations are used at home.

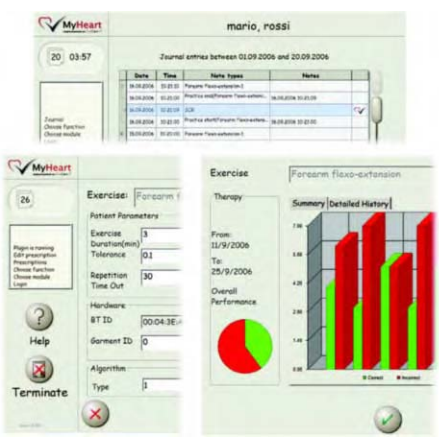


Figure 4 – Exercise prescription, configuration and results' review modes for the motor rehabilitation plug-in

Medical personnel (both physicians and physiotherapists) administer their assigned patients, prescribe exercises, and review their results using dedicated *therapists' stations*.

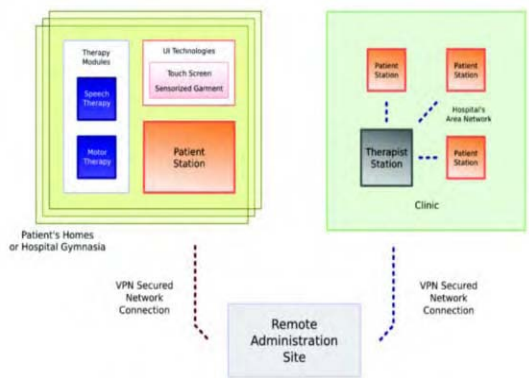


Figure 3 - Concept NR's telemedicine platform: simplified view of the architecture

The *server* acts as the repository for exercise prescriptions, results, and messages. Since session data is stored on the server, not the clients, patients can use one workstation or the other indifferently. Also, replacement of faulty workstations is straightforward.

The software architecture is based on a generic telemedicine framework which supports security and authentication, data storage and retrieval, and on-screen look-and-feel. Specific exercises, including speech therapy and motor therapy, are realized by a "plug-in" mechanism. Each plug-in implements all activities required by the specific type of exercise: it is responsible for providing configuration screens, support the execution of the actual exercise, provide appropriate feedback to the patient, and store the results. It also supports browsing and summarization of the results for the therapist to review. Figure 4 shows three views of the motor rehabilitation plugin.

CSEM (Centre Suisse d'Electronique et de Microtechnique) has realized the portable electronics required for sampling the strain sensors' resistance values (SEW, figure 5). The box, when used for the NR application, performs analog to digital conversion for 32 channels in 16 bit precision at 128 Hz, in parallel; digitized waveforms are packed and streamed in real-time over a Bluetooth connection [6]. Battery life of the current SEW prototype amounts to approximately 8 hours of continuous use.

When signals from body movement are acquired, they are processed in order to verify the correct execution of the rehabilitation exercise. Repetitions have to be counted and classified evaluating the similarity with a known "correct" or reference path. This is an important research topic by itself, because there is no direct correspondence between the fabric stretch values at various points of the garment and biomechanical parameters, like angles between limb segments. Also, strain sensors are affected by various sources of uncertainty which make the recovery of limb positions from sensor readings less than straightforward.

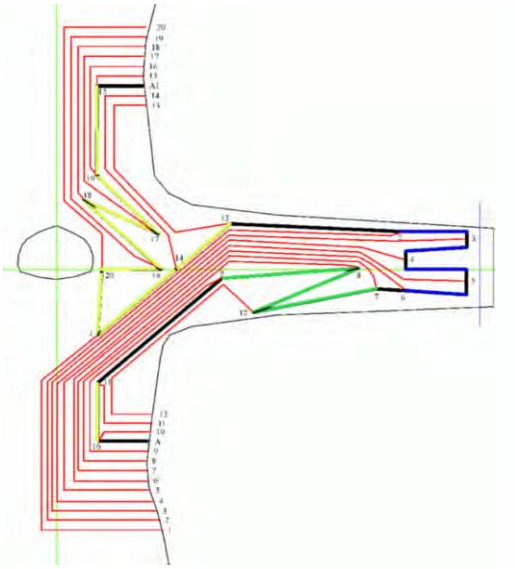


Figure 6 – CE mask in the garment. Thick lines are strain sensors; thin lines are wirelike printed connections

Classification algorithms were similarly applied for the actual posture detection. Although this strategy is partially successful, the accuracy of classification is affected by the dynamic nature of the task. On one hand, smeared sensors produce velocity-dependent artifacts; therefore, the recognition accuracy is impaired when limbs are moving. Also, one is interested in the closeness of a movement *path* to a reference, not just to detect a limited set of basic positions. For this reason, research is now ongoing to perform movement recognition on the basis of qualitative multivariate time-series analysis.

Figure 5 – Portable data acquisition electronics

Table 1 – Rehabilitation exercises under test

Lateral abduction and adduction of the arm
90° flexion of the arm in the sagittal plane
External rotation of the arm with flexed elbow
Forearm flex-extension
Forearm prono-supination
Functional: combing
Functional: eating

Results

The NR system is currently undergoing a small-scale study to assess user satisfaction and gather initial measures and feedback. It includes, as its last phase, 10 months of experience on the field, with real healthcare operators and at least 20 patients exploiting the different system functionalities. Results of the study will deal with the usability of the system (user-friendliness, lack of technical problems, comfort e.g. possible problems with the garment) and the users' satisfaction (perceived usefulness from the patients' and therapists' point of view).

Preliminary qualitative and quantitative results for speech therapy were obtained from the first four patients enrolled. One of them, recovering after a stroke accident, required only two training encounters to start using the logopedic software. He is now using the system three times a day. Although the amount of data is still limited, figure 7 shows the resulting trend: after the initial training phase carriedon with the speech therapist help (first two points), word recognition ability grows steadily.

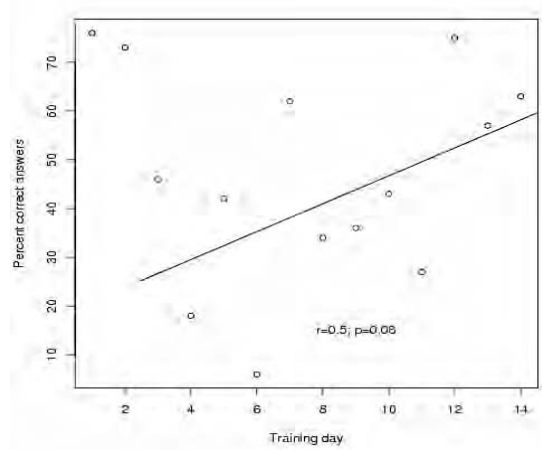


Figure 7 – Progress in speech training. During the first two days the patient was assisted. Linear regression for days > 2 indicates a positive trend (p=0.08)

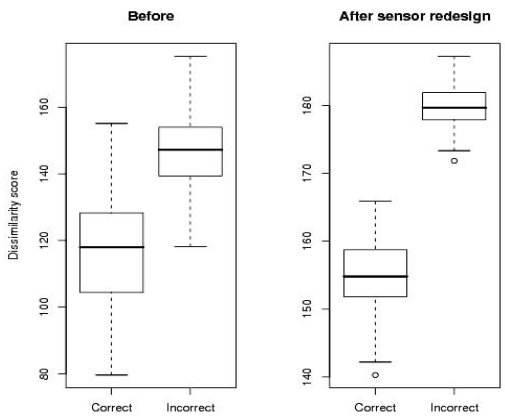


Figure 8 – Sensor layout affects recognition accuracy (exercise: sideways arm abduction)

The recognition algorithm used for the classification and real-time feedback on the exercises has been validated on several garment prototypes. The recognition is based on a "dissimilarity score" which is output by the recognition algorithm after a per-session calibration. A volunteer executed the rehabilitation exercises in sequence (table 1). Each exercise was repeated 10 times correctly, then 5 times mimicking the most frequent compensatory movements. Two distributions of "dissimilarity scores" were therefore obtained. The amount of overlap between distributions is proportional to the probability of a correct/incorrect misrecognition. Quantitative measures allowed redesigning the sensor layout in order to maximise the classifier margin (figure 8).

Conclusion

Development of the NR rehabilitation system followed an integrated approach which involved research on tissue-embeddable strain sensors, portable wireless electronics and ergonomic feedback, in order to be easy to use for patients which are often cognitively-impaired. Its most innovative component is the exercise recognition algorithms, which are able to provide feedback in real time, insensitive to several sources of noise stemming from the inherently noisy, multi-sensor approach taken.

The integrated nature of the NR system is now intended to support a multidisciplinary, telemedicine-based approach to rehabilitation. The benefits of this (and similar) IT-based aids are supported by current beliefs on rehabilitation, e.g. those expressed in the SPREAD guidelines [1]. First of all, long term rehabilitation programmes are considered beneficial for the recovery process[1]. However, their cost makes it unlikely support from insurances and national health-care systems. IT technologies are a promising solution for providing long-term rehabilitation at lower costs. At the same time, early discharge is recommended for patients with less severe degrees of disability[2]. The system allows for early discharge while still being monitored remotely.

Finally, the concept addresses a frequently-reported issue for rehabilitation professionals, i.e., lack of quantitative measures on the progress of rehabilitation. Standard functional scales, such as the FIM functional motor index, are in fact partially subjective. The diffusion of computerized monitored rehabilitation aids may increase the objectivity of practice and allow for quantitative progress evaluation and therapy comparisons.

Acknowledgments

The authors would like to thank F. Lorussi and Prof. D. De Rossi (University of Pisa); Dr. G. Maggioni and R. Mauri (Fondazione Salvatore Maugeri, Pavia); Prof. M. Stefanelli (University of Pavia); and D. Tietze and R. Setz (Dr. Hein GmbH) for essential contributions to the project. The work described in this paper was partly funded by project "MyHeart" no. IST-2002-507816 in the Sixth Framework of EU IST projects.

References

[1] The Italian Guidelines for stroke prevention. The Stroke Prevention and Educational Awareness Diffusion (SPREAD), IV edition, 2005.
[2] http://www.dr-hein.com (Accessed 25 March 2007)
[3] Schoch KF; Saunders HE. Conducting polymers, Spectrum, IEEE, vol.29, no.6, pp. 52-55, Jun 1992.
[4] Tognetti A, Lorussi F, Bartalesi R, Quaglini S, Tesconi M, Zupone G, De Rossi D. Wearable kinesthetic system for capturing and classifying upper limb gesture in post-stroke rehabilitation, J Neuroengineering Rehabil, vol. 2, no. 1, pp. 8-8, Mar 2005.
[5] Johanni M, Tietze D, Setz R, Hein A. EvoCare: a new standard in tele-therapy. Stud Health Technol Inform. 2004;108:228-34.
[6] Luprano J, Sola J, Dasen S, Koller JM, Chetelat O. Combination of body sensor networks and on-body signal processing algorithms: the practical case of MyHeart project. Wearable and Implantable Body Sensor Networks, 2006. BSN 2006. 3-5 April 2006.
[7] Giorgino T; Quaglini S; Lorussi F; De Rossi D. Experiments in the detection of upper limb posture through kinestetic strain sensors. Wearable and Implantable Body Sensor Networks, 2006. BSN 2006. 3-5 April 2006.

Address for correspondence

Email address: toni.giorgino@unipv.it

1 "A long-term rehabilitation programme promoting independence in daily activities is recommended to reduce the deterioration of the independence status obtained by the intensive or extensive rehabilitation" (Recommendation 14-73).

2 "In patients with slight-to-moderate disability following stroke, early discharge from the rehabilitation hospital is recommended […] (R. 14-61)."

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Effects of a Computerized Cardiac Teletriage Decision Support System on Nurse Performance: Results of a Controlled Human Factors Experiment Using a Mid-Fidelity Prototype

Kirsten Carroll Somoza[a], Kathryn Momtahan[b], Gitte Lindgaard[c]
[a] *Human-Oriented Technology Lab, Carleton University, Canada*
[b] *University of Ottawa Heart Institute, Canada*
[c] *Human-Oriented Technology Lab, Carleton University, Canada*

Abstract:

A gap exists in cardiac care between known best practices and the actual level of care administered. To help bridge this gap, a proof of concept interface for a PDA-based decision support system (DSS) was designed for cardiac care nurses engaged in teletriage. This interface was developed through a user-centered design process. Quality of assessment, quality of recommendations, and number of questions asked were measured. Cardiac floor nurses' assessment quality performance, but not their recommendation quality performance, improved with the DSS. Nurses asked more questions with the DSS than without it, and these additional questions were predominantly classifiable as essential or beneficial to a good assessment. The average participant satisfaction score with the DSS was above neutral.

Keywords:

Teletriage, user-centered design, user interface design, mobile computing, decision support system

Introduction

As the number of Canadian cardiac patients increases and the length of hospital stay decreases, there is a need for expert care and evaluation once patients return home from the hospital. The current services of homecare agencies, community hospitals, and family physicians are inadequate to meet the needs of this growing patient base for two main reasons:

1. inadequate access to professionals (many patients may have to wait several weeks before seeing their family physician once they leave hospital) [1] and
2. limited professional access to up-to-date, best practice guidelines for cardiac care [2].

The Nursing Coordinators (NCs) at the University of Ottawa Heart Institute (UOHI) receive more than 2000 calls per year from patients requesting help and advice for cardiology and cardiac surgery concerns. Though teletriage is significant in the care of these patients, there is no standardized protocol in place to guide the advice-giving, nor is there any formal training for staff.

The purpose of this research was to design an effective Decision Support System (DSS) to help primary healthcare professionals field these calls. The final DSS will initially be used by the NCs, but aims to assist less experienced care providers in the future. The literature suggests that challenges lie both in the practice of teletriage [3, 4, 5, 6], and in the design and development of decision support systems [7, 8, 9, 10, 11]. To the extent that an inadequate design may rest on poorly defined user tasks, one would expect improved DSSs to result if that limitation were to be overcome. This research therefore explored the use of task analysis as part of a user centered design process aiming to produce an effective DSS.

In this study, cardiac best practice algorithms are introduced via a handheld DSS. Access to these best practices is expected to improve the quality of teletriage performance [1], provided that a) the best practices are sound, and b) the DSS delivers them so that they are easily used by NCs while on the phone with a patient. This research encompassed the design and evaluation of the DSS, but did not evaluate the best practice algorithms.

The hypotheses were: (1a) that nurses' performance in assessment of a patient's problem will improve using the DSS, and (1b) that nurses' performance in recommendations for the patients will improve using the DSS; and (2) that nurses will ask fewer questions with the DSS than without it.

Since there were many aspects to this research, this article will focus primarily on the results comparing nurses' performance with and without the DSS. The design process, and the design itself, will not be discussed at length here.

Materials and methods

Development of pain scenarios and standardized patient profiles

Two Subject Matter Experts (SMEs) developed four pain scenarios, based on best practice algorithms that were determined with a larger expert team. There was one scenario for each of the following types of pain complaint: non-urgent cardiac surgery, urgent cardiac surgery, non-urgent cardiology, and urgent cardiology. The following

specific ailments were chosen as common examples of each type of pain (see Table 1).

Table 1- Pain scenarios

Scenario	Cardiac care area	Urgent or non-urgent
1: Post Pericardiotomy Syndrome (PPS)	Cardiac surgery	Non-urgent
2: Ischemic Pain	Cardiology	Urgent
3: Cardiac Tamponade	Cardiac surgery	Urgent
4: Stent Pain	Cardiology	Non-urgent

Each pain scenario was developed with a standardized patient profile whose characteristics matched the algorithm for the pain complaint in question. These simulated "patients" were created by the two SMEs, based on the algorithms and their practical experience.

Baseline test – Nurses' performance without the DSS

Because some of the NCs participated in development of the algorithms and because there are very few NCs at the UOHI, the NCs did not participate in DSS prototype testing or the baseline test. Instead, participants were cardiac care nurses from the UOHI who work with hospitalized cardiac patients.

Twelve nurses completed the baseline testing, in which their performance was assessed without the DSS. During the baseline test, nurses used only pen and paper while on the phone with the patient, as these are the tools normally available to them. Seven of these nurses subsequently completed the DSS test, which scored their performance with the DSS prototype and helped the researcher evaluate the DSS interface. During both the baseline and DSS tests, an SME acted as the patient on the other end of the phone. This SME also acted as the UOHI attendant who transfers incoming patient calls to the NCs. The study did not use real patients as participants.

This baseline provided initial data for the number of questions asked by each participant in each scenario, and for performance scores on quality of assessment and quality of recommendations. "Quality of assessment" measured nurses' ability to ask the right questions in assessing the patient's condition. "Quality of recommendations" measured their ability to give the patient correct advice based on the presenting problem in each scenario. The scoring of the quality of assessment and quality of recommendations was done independently by the two SMEs after many practice sessions and iterations of the scoring sheet. For the four scenarios tested, the average inter-rater reliability (IRR) for quality of assessment was .94, ranging from .83 to .98. For quality of recommendations, average IRR was .97, ranging from .86 to 1.

The presentation of the four pain scenarios was randomized to avoid serial order effects. In all scenarios, the SME in the role of the patient made an effort to answer the nurse's questions in an appropriate way, neither volunteering more information than requested, nor being furtive and vague. For each pain scenario, the SME referred to the corresponding standardized patient profile during the call to ensure she was giving consistent answers. Each standardized patient profile listed, in logical order determined by the algorithm, questions that the NC would likely ask with corresponding answers for that scenario. The profiles also contained answers to questions not necessarily in the optimal flow of questioning. Throughout the baseline and DSS tests, if a participant asked a question that was not on the list, the SME gave an answer consistent with the cause of pain in that scenario. She then added the new question and answer to the list, ensuring that a consistent response would be given should the same question be asked by another participant. To preserve the SME's anonymity and increase semblance to a real telephone call, the SME was in another room and communicated via telephone throughout the test.

The researcher read an introductory script before the pain scenarios began, telling the nurse that they would be acting in the role of an NC, and that their task was to assess the patient's condition and advise them accordingly. To more closely mimic real-life use, nurses were asked to hold the telephone handset to their ear during the call even though the speaker phone was in use to record the session, enabling the researcher also to hear the "patient's" responses. Participants proceeded through the phone call using a pen and paper to record whatever they wished. The participant marked the end of the call by hanging up the telephone handset.

DSS Test – Nurses' performance with the DSS prototype

Seven of the original 12 participants were available and completed the proof-of-concept DSS test, on a mid-fidelity prototype. These sessions were scheduled approximately six weeks following the baseline testing. The same SME who participated in the baseline test acted again in the role of the patients for the DSS test. The SME did not personally evaluate the interface during the DSS test.

Apparatus was similar to the baseline test with the addition of a Palm Tungsten T3 PDA including stylus, and the DSS prototype running on a Toshiba Portégé M200 tablet PC. The T3 was presented with only the software that comes installed from the manufacturer. The DSS prototype incorporated best practice algorithms for cardiac pain into an interface that supported nurses in asking the correct questions while they were assessing the patient, and assisted the gathering of patient information. With the exception of the first participant in the trial, participants did not have access to a pen and paper for the DSS test. The pen and paper were removed to encourage participants to use the prototype rather than the paper. All interaction with the tablet occurred through the included stylus.

Tablet PC interaction exercise

Each nurse completed a short interaction exercise on the Tablet PC. This was not recorded and served only to give each nurse equal opportunity to practice interacting with the device via the stylus. All participants were first-time Tablet PC users.

Practice scenario

After the first participant, it became apparent that the nurses would likely benefit from some practice on the system before being asked to use and evaluate it in a simulated working environment. Therefore all subsequent participants received the following practice scenario. The researcher walked the nurse through a sample scenario, which was based on a fictitious patient with a chest cold rather than a cardiac pain concern. Then the researcher played the role of that same patient, and allowed the nurse to click through the tool on their own while asking questions of the patient. This gave them extra practice with the stylus, and with the prototype interface elements.

Pain scenarios

The pain scenarios proceeded exactly as in the baseline test, but the participants used the prototype DSS to assist in assessing the patient. Scenario order was randomized, and participants received them in a different order than in the first session to avoid order effects. Patient and physician names were changed from the baseline test.

As participants progressed through a patient call using the DSS, they saw questions on each screen (for example, "What makes the pain worse?") and could input patient responses through checkboxes, pick lists, and freeform fields. Participants had freedom to move about within the interface, and were not instructed to read each question in sequence. At the point just before the software would throw a "flag" based on the input so far, suggesting a possible presenting issue and a recommended course of action, the participant was asked to guess what the patient's problem was and give their own recommendations. The automated assessment flag was then revealed, and the participant was asked whether they agreed or disagreed with the DSS' suggestions.

Results

Quality of assessment

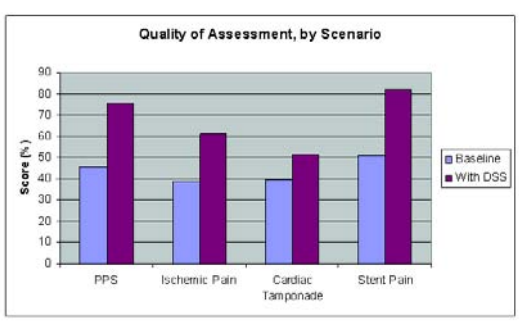


Figure 1 - Quality of assessment, by scenario

Figure 1 shows the quality of assessment scores for each scenario. While this figure suggests that performance scores improved substantially from the baseline to using the DSS in the case of Stent Pain and Post Pericardiotomy Syndrome (PPS) – the non-urgent scenarios, this improvement was less evident in the case of Ischemic Pain and Cardiac Tamponade – the urgent scenarios. A Wilcoxon Signed Ranks Test indicated that, statistically, quality of assessment did improve for every scenario with the DSS ($p < 0.05$). These results support one of the experimental assumptions, namely that that nurses' performance in assessment would improve using the DSS.

Quality of recommendations

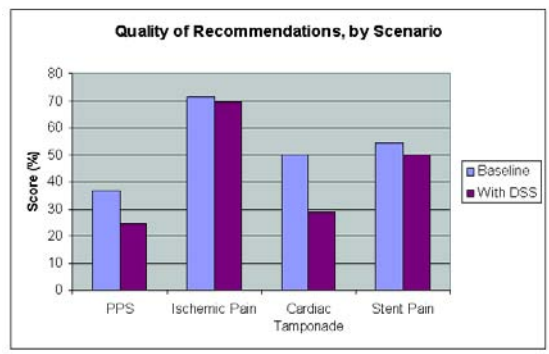


Figure 2 - Quality of recommendations, by scenario

Figure 2 suggests that recommendation performance declined at least slightly across all scenarios. This decline was most evident in the Cardiac Tamponade (urgent) and Post Pericardiotomy Syndrome (PPS) (non-urgent) scenarios. However, contrary to the predictions, the Wilcoxon Signed Ranks Test revealed no significant effect ($p > 0.05$). In seven cases out of 28, the nurse did not correctly guess the patient's problem, but they agreed with the automated DSS assessment. In these cases, recommendation quality might have improved in real life, had they seen the automated DSS flag before advising the patient – this assessment flag offered a possible condition, as well as recommended course of action.

Why was there no improvement in recommendation scores with the DSS? The nurses were provided with essential and beneficial questions to ask (contributing to the quality of assessment score), but they were not provided with the recommendations contained on the DSS. In clinical practice, telepractitioners would not be forced to advise the patient without seeing the DSS recommendations. Additionally, cardiac floor nurses would be expected to give good recommendations without assistance, so improvement may be observed more readily in a more novice user group. Nurses' lack of experience with the technology (PDAs and the tablet test platform) may have interfered with their ability to synthesize information during the phone call because a lot of their attention was devoted to trying to use the technology. Also, since this was the first time the pain scenarios were used, it is possible that these could benefit from further iteration. Finally, the floor

nurses participating in the study were not used to performing teletriage, which could have contributed to low recommendation scores both with and without the DSS.

Number of questions asked

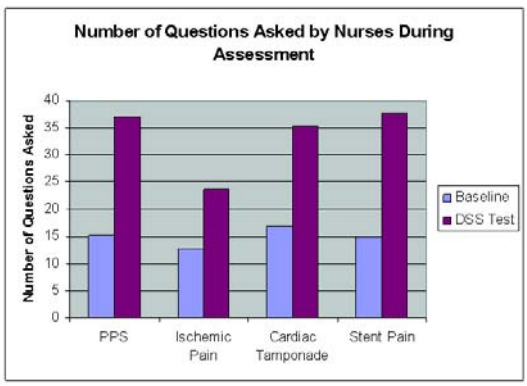


Figure 3 - Average number of questions asked by nurses during assessment

Hypothesis (2) predicted that nurses will ask fewer questions with the DSS than without it. A Wilcoxon Signed Ranks Test, comparing the number of questions asked in the baseline test and the DSS test, indicates that participants in fact asked significantly more questions with the DSS than without it ($Z= -2.31$, $p<0.05$). These results are illustrated in Figure 3 above, and they thus refute that hypothesis.

However, in addition to comparing the number of questions asked, the efficiency of questioning was also assessed. Efficiency was measured as a proportion of questions asked that were on the SME's list of essential and beneficial questions. These lists were agreed upon by two SMEs, and were iterated following the final test, before scoring was done. This iteration was deemed necessary because the nurses asked beneficial questions during the DSS test that were not on the original lists. The SME's lists of questions were not tested prior to using them for coding. Questions outside of these lists are purposely not called "bad" or extraneous questions, as it is possible that further iteration to the lists could be beneficial.

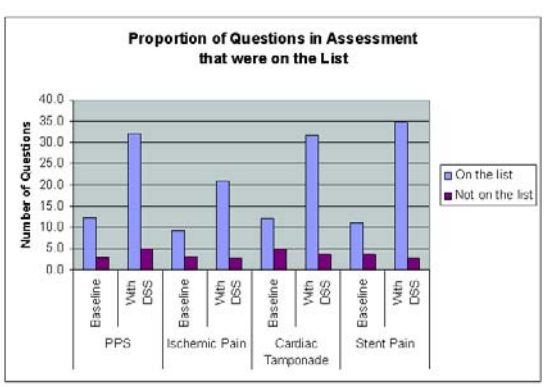


Figure 4 - Proportion of questions in assessment that were on the list of essential or beneficial questions to ask

Figure 4 shows the proportion of assessment questions asked that were on the list compared to not on the list. It suggests that most of the additional questions asked with the DSS were marked in the SMEs' list. Furthermore, the number of questions not on the list remained fairly constant across scenarios between the baseline and DSS.

A Wilcoxon Signed Ranks Test indicated that nurses asked significantly more questions on the list with the DSS than without, but there was no significant difference in the number of questions asked that were not on the list. Earlier, it was reported that assessment scores increased with DSS use. This supports the results we see here, since nurses asked more essential and beneficial assessment questions with the DSS than without it. So at the same time this figure refutes hypothesis (2), it helps confirm hypothesis (1) in that asking more essential and beneficial questions with the DSS contributed to participants' higher assessment scores.

On later discussion between the SME and an NC, it came to light that less experienced teletriage practitioners don't ask enough questions – this supports the literature [6]. In this respect, hypothesis 2 turned out to be misguided; it should have predicted an increase rather than a decrease in the number of questions asked, provided that additional questions asked could be shown to be relevant and beneficial to the final recommendation.

User satisfaction

User satisfaction was measured in a post-test questionnaire, immediately following the DSS test. Users marked their score on a 10-centimeter line scale, and these scores were translated to values between 0 and 10. Aspects that ranked most satisfying (average scores across all participants) included:

- "It was **easy to learn** to use this system" – 7.5
- "The **information was helpful** to me in completing the scenarios" – 7.5
- "**Overall, I am satisfied** with this system" – 7.5
- "I **could effectively complete the scenarios** using this system" – 7.4

Aspects that ranked least satisfying included:

- "Overall, I am satisfied with **how easy it is to use** this system" – 5.2
- "It was **easy to find the information** I needed" – 5.8
- "The **order of the screens** made sense" – 6.1

Summary

Cardiac floor nurses improved their assessment quality, but not their recommendation quality, with the DSS compared to without it. Participants asked more questions with the DSS than without it, and these additional questions were predominantly classifiable as essential or beneficial to a good assessment. This supports the literature where Wheeler found inexperienced teletriage practitioners often failed to obtain adequate patient history. Participant satisfaction with the DSS averaged above neutral.

Future research

One problem identified with the current Telepractice Record form – the paper form currently used by NCs during a patient call – was that NCs tend to fill them out incompletely. Though this is unlikely to make a difference in the quality of their advice or assessment, this tendency creates difficulties in record keeping and patient tracking. It would be interesting to study whether enabling NCs to enter patient information in the DSS affects how complete the records are when they turn them in. Since the DSS will be associated with a database collecting patient call information, the potential is there to significantly improve record keeping for patient calls.

Developing the prototype for the Tablet PC allowed us to see what users were doing during the test, and to give users a feel for stylus-based interaction with the screen. However, the form factor of the T3 handheld unit is considerably smaller, and this might therefore affect users' performance and acceptance. During the opportunity for informal feedback on the T3, none of the nurses indicated they thought it was too small to use. Despite this, testing the DSS in situ on the T3 would be advisable to discover if any interface elements are hindered by the reduction in screen size.

Finally, a computerized decision support system can only be successful if it is actually used. There will need to be a certain degree of motivation on the part of the NCs to try something new, as well as encouragement and support from the developers and researchers involved. Feedback from the NCs should be closely monitored during the implementation stages of the final DSS. Acceptance testing after the system is in place would provide helpful information for future iterations, as well as valuable insight for future nursing decision support systems. Ideally, this would incorporate longitudinal testing of participants on the DSS over a period of several weeks, rather than in a single session, to give a better impression of how performance would be affected by the DSS over time. Because the DSS will eventually be used by less experienced nurses, and because the tested group of participants did not improve their quality of recommendations with the DSS even though it helped their assessment, it is advisable that nurses receive training not only on the DSS, but also on teletriage in general before they are expected to take on a teletriage role.

Acknowledgments

Thanks to Whynne Caves and Katie Bush at the University of Ottawa Heart Institute, and to the nursing coordinators and nurses who took part in this research. Funding was provided by NSERC/ Cognos IRC Grant no. IRCPJ 234088-05, and by the Ontario Ministry of Health and Long Term Care Primary Health Care Transition Fund IRC Grant no.G03-05372.

References

[1] Schoen C, Osborn R, Huynh PT, Doty M, Zapert K, Peugh J. Taking the pulse of health care systems: Experiences of patients with health problems in six countries. Health Affairs 2005, W5(509): 509-25.

[2] Momtahan K, Burns C, Sherrard H, Labinaz, M, Mesana T. Using personal digital assistants and patient care algorithms to improve access to cardiac care best practices. Ontario Ministry of Health and Long Term Care Primary Health Transition Fund Grant No. G03-05372, 2004.

[3] Briggs JK. Telephone triage protocols for nurses. Philadelphia, PA: Lippincott, 2002.

[4] Bunn F, Byrne G, Kendall S. Telephone consultation and triage: Effects on health care use and patient satisfaction. Cochrane Database Syst Rev 2004, (3).

[5] Lowery CR. A retrospective study: Exploring selected characteristics of telenursing triage. (Doctoral dissertation, Air Force Inst of Tech Wright-Patterson AFB OH), 2001.

[6] Wheeler SQ. Telephone triage: Theory, practice, and protocol development. New York: Delmar Publishers, 1993.

[7] Braden BJ, Corritore C, & McNees P. Computerized decision support systems: Implications for practice. In U. Gerdin, M. Tallberg & P. Wainwright (Eds.), Nursing informatics : The impact of nursing knowledge on health care informatics (pp. 300-304). Amsterdam: IOS Press, 1997.

[8] Fischer S, Stewart TE, Mehta S, Wax R, Lapinsky SE. Handheld computing in medicine. J Am Med Inform Assoc 2003 Mar-Apr; 10(2): 139-49.

[9] Hastings PR. Low band telemedicine decision support system for disaster situations (Final No. 1999000221). Fort Detrick, Maryland: U.S. Army Medical Research and Material Command, 2001.

[10] Rodriguez NJ, Borges JA, Soler Y, Murillo V, Sands DZ. A usability study of physicians' interaction with PDA and laptop applications to access an electronic patient record system. Proceedings, 17th IEEE Symposium on Computer-Based Medical Systems, 2004, pp. 153-60.

[11] Rosenbloom M, Jaffe W, Adams J. Patient care efficiency and medical error reduction using PDA-based medical information. Acad Emerg Med 2001, 8(5): 587-b-588.

Address for correspondence

Kirsten M. Carroll Somoza; 130 Scenic Lane; Woodlawn, Ontario K0A 3M0; Canada

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Opportunities and Barriers for Mobile Health in New Zealand

Farhaan Mirza[a] and Tony Norris[a]

[a] *Centre for Mobile Computing, IIMS, Massey University, Auckland, New Zealand*

Abstract

Ownership and the use of mobile technologies greatly exceed those of personal desktop computer systems and countries throughout the world are beginning to understand how these technologies can enhance the delivery of healthcare (m-health). This paper reviews the opportunities and barriers for m-health and describes a study to understand its potential in New Zealand. A survey consisting of a questionnaire and in-depth interviews was used to reveal clinician and service provider attitudes to m-health. The general perception is that m-health will be an increasing component of future healthcare with many opportunities for empowering patients, delivering convenience care, and supporting carers as well as offering the potential for more effective public health and lifestyle broadcasting. Participants recognised several barriers to the acceptance and sustainability of m-health, identifying privacy of information and device form factor as major concerns.

Keywords:

mobile health, m-health, mobile phones, cell phones

Introduction

The world total of mobile phones currently stands at 2.5 billion and is set to reach 3 billion by the end of 2007 (1). New Zealand alone has 3.8 million subscribers (2) from a population of 4 million people emphasising that many users have multiple subscriptions. Evidently, ownership and use of mobile technologies greatly exceed those of personal desktop computer systems and the pervasiveness of mobile solutions (3) offers possibilities far beyond the expedient of simple communication. The next phase of smart phone development will further extend these possibilities as communication, computing, and multimedia elements (cameras, video recorders, music players) converge seamlessly in the same powerful device (4).

Countries throughout both the developed, and the developing, world are now beginning to understand how mobile technologies can enhance the delivery of healthcare (m-health) (5) and this paper describes a study of the opportunities and barriers to m-health in New Zealand. The ubiquity of mobile devices clearly offers the potential for more convenient communication between patient and physician, the transmission of clinical test results, rapid diagnosis in emergencies, and similar applications (5) but how acceptable are such interactions to the participants and what are the obstacles to their introduction? The paper discusses these issues and makes recommendations for the advancement of m-health.

M-health applications

A definition of m-health

We can define m-health as a rapidly developing area that uses small, portable, and wireless computing and communication devices to meet the information and service needs of healthcare providers and consumers.

From this definition, mobile devices are effectively transceivers that run healthcare applications over wireless (and wired) networks. The armoury of such devices is increasingly diverse ranging from (3) laptop computers, tablets, and PDAs to mobile and smart phones (4) and RFID (6) technologies. Short (< 50 m) and long-distance network protocols are similarly becoming more powerful extending their reach and reliability (7).

Mobile devices offer some obvious benefits (8) in addition to portability such as convenience (easy access), immediacy (any time) and context (location).

The providers in our definition include clinical and non-clinical practitioners whilst the consumers are the general public but in particular patients and, increasingly, their carers. This clinical/non-clinical distinction offers a useful way of classifying m-health applications.

Clinical m-health applications

Table 1 shows some typical examples of clinical m-health applications.

Table 1 - Clinical m-health applications

Web access to evidence-based databases
Medication alerts using mobile phones
E-prescribing for repeat prescriptions via mobile phones
Telemonitoring to transmit patient results to clinicians
Transmission of test results to patients via SMS messages
On-line electronic health records via computer or phone
Community nursing contact with clinical expert advice
Public health and lifestyle messages over mobile phones
Care of at-risk people, eg air-line pilots, military personnel
Emergency care for accidents, natural disasters

The critical feature of these examples is that they encompass the whole spectrum of healthcare demonstrating the central role that mobile technologies can play in wellness promotion and disease management according to:

Prevention: public health and lifestyle awareness
Monitoring: pre-disease screening and assessment
Treatment: providing efficient and effective care
Support: for patients and their carers

The seamless and integrating utility of mobile devices is also evident from these examples.

Non-clinical M-Health Applications

The smooth operation of any healthcare service also relies heavily on administrative and management functions that support the clinical imperative. Table 2 shows how some of these functions can be discharged using mobile technologies.

Table 2 - Non-clinical m-health applications

Efficient workflow via wireless communication
Rapid collection/sharing of current data via mobile phones
Optimal asset utilisation, eg hospital bed rostering
Patient or asset (eg clinical equipment) location using RFID
Patient appointment booking and alerts via wireless e-mail
Mobile phone support for patients and carers
Safety of staff checks with RFID or mobile phones/networks

This small sample reveals the ubiquity and pervasive nature of mobile technologies which make them so acceptable (but see later) as personal and lifestyle essentials.

Potential benefits of m-health

As frequently happens with the introduction and diffusion (9) of new technologies, the benefits are driven partly by the needs of the domain and partly by the capabilities and opportunities offered by the technologies themselves. This section identifies some of the benefits that, at least superficially, are the most likely to accrue from the introduction and uptake of m-health. We might label these as 'the promise of m-health'.

Healthcare drivers

Perhaps the most significant issue and driver in healthcare today is the ageing of populations in developed countries (10). For example, a frightening statistic for New Zealand (11) is that 25% (1.33 million) of the populations will be aged 65 or over in 2051 compared with 490,000 (13%) in 2004.

This trend has multiple consequences but the most apparent is the increase in the incidence of chronic diseases that demand continual rather than episodic care as patients age. As the applications in Tables 1 and 2 show, mobile technologies offer the possibility of managing non-critical care within the community (including telecare (12)) thus reducing hospitalisation, improving patients' quality of life and controlling costs.

The longevity that results from the success of modern medicine also increases the importance of preventing people from getting ill in the first place and maintaining this status through appropriate lifestyle and care. Frequently, public health interventions and campaigns have limited success because they emphasise the need for change when none is apparent and they fail to raise personal awareness. However, the pervasive nature of the mobile phone offers the opportunity for targeted marketing of the health message and its reinforcement by repetition and with incentives (13).

These examples of the potential heath benefits of applications represent two of the generic improvements, namely efficiency and effectiveness, that are sought from the introduction of new technologies. The third general benefit is increased access to services and resources and the pervasiveness of mobile technologies, particularly mobile phones, is set to have a major impact on how we think about the delivery of healthcare and the relationship between provider and consumer (14). It seems highly probable that the increased ease of communication between clinicians and patients will lead to higher workloads that will have to be managed by the involvement of intermediaries who add value for all participants. Whatever form the new services take, they can improve equality of access for disadvantaged communities.

Technology drivers

The most obvious benefit of mobility is convenience and mobile technologies offer the possibility of 'convenience medicine' more so than any other previous innovation. Instead of the patient receiving care where the medicine originates, much of the medicine will come to the patient wherever he or she is. This demand will be flamed by the fires of patient consumerism and expectation so that m-health truly forecasts a revolution in healthcare delivery.

Some of these changes may be of questionable benefit but patient empowerment and greater responsibility for their own health planning and care will be an inevitable result of more frequent involvement and demand.

We have already noted the efficiency that rapid communication can bring to service operation and the ease of data collection and exchange. This ease will in turn increase the value of data leading to improvements in their quality and longevity (15). These advances could be especially important in New Zealand which enjoys significant advantages from its national data collections. If data could be collected directly and effortlessly from people via their mobile phones then, with the appropriate management and analysis, the derived information could have an enormous impact on the quality and efficiency of service planning and operation.

A final value proposition (literally) of mobile technologies for healthcare is their apparent low cost. Certainly, unlike many medical innovations, the cost of mobile technologies to the consumer is very small and their cost-effectiveness very high. There are of course the capital and operational costs of the supporting infrastructure but such costs are shared over many markets and the expansion of generic

mobile services has produced dramatic falls in service costs; a trend that shows no signs of abating. Indeed, as the technologies become more powerful the introduction of more innovative and seamless applications will drive down costs even further. These effects will, as noted, also encourage enterprises that previously had no presence in the health sector to offer value-added products and services.

Barriers to m-health

The potential benefits of m-health should not blind implementers to the inherent barriers and challenges. From a healthcare standpoint, first and foremost amongst these problems are concerns over the privacy and security of personal healthcare information (16). Whilst these apprehensions are sometimes more perceived than real, the ethical issues surrounding the electronic storage and transmission of sensitive data and their misuse cannot be ignored. Wireless security protocols are improving rapidly but often it seems only fast enough to keep pace with the ingenuity of hackers and other intruders who wish to gain illegal access to information.

A further clinical, and perhaps more limiting, challenge to m-health is the acceptability of the technologies to both patients and healthcare practitioners. Consumer empowerment and convenience are likely to overcome patient concerns without too much resistance but clinicians are rather more traditional when it comes to alternatives to face-to-face delivery of medicine (17).

Several concerns present themselves to clinicians; the potential for automating diagnosis and clinical-related decisions concerning treatment, the multidisciplinary nature of care which means that all members of the care team must accept the technologies if treatment is to be seamless and integrated, and the impact that the wireless and mobile devices and services may have on the doctor-patient relationship (17). The involvement of commercial enterprises offering value-added services as mentioned above is an example of the last concern.

From a technology perspective, one aspect of m-health offers an interesting paradox. The acceptability of the mobile phone is based on the simplicity (some phones are no more simple to use than video recorders!) and the convenience with which they perform their main function - voice communication by telephone. However, as the technology and power of the devices progresses, their expanded functionality betrays them as examples of a disruptive technology (7, 18).

A disruptive technology is a technology that when introduced does not meet the needs of users. Thus, a desktop personal computer or a laptop meets the computing needs of users in terms of modelling (eg Excel), office applications, or web searches but mobile phones, some of which are as powerful as a 1960s mainframe, cannot satisfy these requirements. The main limitation here is of course the form factor. The demands of portability have not so far been matched by the release from the tyranny of small screens and cramped keyboards. Similarly, the reliability of mobile devices and their fault tolerance attributes do not meet the exacting requirements of the mission-critical applications found in healthcare. Even the item that makes the mobile phone portable – the battery – is a limitation. Computing technology doubles in power every 18 months or so; battery power has taken 30 years to achieve the same level of improvement.

A study of attitudes to m-health

The observations made in the previous sections are based on evidence accumulated across several disciplines and markets and many of the findings are applicable to healthcare. However, the health sector's central focus on the well-being of persons is a crucial differentiator from other sectors. The ethical priority of care often makes the health sector a slow adopter of novel technologies, particularly those that challenge long-standing practice and so it is dangerous to make too many inferences or extrapolations about the impact of mobile technologies from other areas.

With this constraint in mind, the authors carried out a survey of attitudes to m-health to determine the potential and barriers for its development and sustainability in New Zealand.

Methodology for the study

The study was a pilot investigation based on a structured questionnaire and semi-structured interviews with healthcare providers across the primary, secondary, and community divisions of healthcare delivery, and with wireless infrastructure and mobile service providers. No patients were interviewed directly in the preliminary study but the health providers reported and compared patient views together with their own responses.

A total of 18 interviews was conducted and the questionnaire was administered to a total of 34 respondents who were identified by role, personal acquaintance, referral from other participants, or by self selection by enclosing the questionnaire in the registration pack at an international conference on health informatics. These two instruments covered a range of health planners and technology strategists including CEOs, CIOs, managing directors, clinicians from all sectors, information managers, IT consultants and nurses. The health sector representatives amongst the interviewees were mainly from secondary care anticipating the finding that m-health was most evident in this area.

The questionnaire and interviews at a macro level concentrated on the m-health experience of respondents and how they saw the opportunities and barriers for its mainstream extension at local and national levels. The responses therefore sought and correlated individual and industry perspectives within and across health sector divisions as well as at delivery and policy and planning levels.

The questionnaire was structured to enquire about general m-health barriers and opportunities, integrated care, and the role of mobile health at a national level. The interviews were semi-structured, but the main focus was to learn from the interviewee about: the role of mobile technology in

New Zealand healthcare, the role of m-health across the different health sectors, the demands and need for customized mobile software, the role of m-health to achieve integrated healthcare, and lastly to discuss the privacy and security implications of m-health.

Full details of this study and a more detailed and quantitative analysis of the results will be published elsewhere (19).

Results and discussion

This section reports and discusses the results of the survey that are most relevant to the m-health benefits and barriers reviewed in the earlier part of this paper.

M-health impact on patients

The most frequently expressed view from the survey was the perceived impact that mobile technologies would have on patient empowerment. The convenience aspect, patients' greater access to information, and the more facile communication with clinicians were all seen as triggers for patients to become more involved with and responsible for their own care. Vital signs monitoring and the transmission of test results were recognised as increasingly important roles for mobile technologies and the simplicity and standard format of short message services (SMS) promoted them as backbone mobile phone services for the present and foreseeable future. SMS services could also boost the impact of public health and lifestyle messages if a suitable format or incentives (extra air time?) could be found.

M-health work to date has been mainly of a 'proof-of-concept' nature and participants were clear that much more systematic work was needed to develop m-health as a group of mainstream technologies. In particular, looking further ahead, all clinicians felt that patients should have mobile access to their electronic health records (EHRs) but many problems, operational, technical, and ethical, had to be addressed before this facility would be generally available (see later).

M-health impact on providers

Providers saw the use of mobile technologies to collect data in an electronic format as a major advance in increasing the utility of data and its value in both operational and strategic decision making. The focus will be on mechanisms for routine data collection and these should be as independent as possible of the nature of the data.

Secondary care providers were especially vocal about the value to them of mobile technologies as vehicles for communicating with colleagues, particularly members of the care team, and, obtaining access to clinical information. On a more mundane level, Auckland District Health Board had managed to reduce its missed appointments percentage from 18 to 9% over the last two years and was determined to exploit SMS to reduce this number even further.

A response, volunteered by many participants, identified a key element to the successful application of m-health technologies. This factor concerned the need for seamless use of these technologies whenever there were several stages in a communication or information sharing chain. The example of a community nurse collecting data from a patient manually or receiving instructions on paper and then transmitting the data or issuing the same directions electronically was viewed as something that would hinder rather than promote m-health. A holistic systems or business process re-engineering approach was needed to avoid such barriers to progress.

Privacy and security issues

A previous section identified privacy and security issues as potentially significant barriers to the uptake of m-health. Those who took part in the survey were keen to separate these issues. Participants felt that security was not the main issue since its governing parameters were mainly operational and technical. Privacy of information and consent to its use or to treatment (16), however, were more conceptual and philosophical matters with strong ethical overtones. Objections to the transmission of information using m-health would probably diminish over time as the technologies matured and the benefits and convenience were seen to outweigh the risks but privacy concerns were paramount and m-health (and e-health) protagonists must incorporate them in the new ways of working.

Interestingly, the privacy issue divided clinicians almost equally on the possibility of supporting EHRs on mobile devices. Everyone saw the merit of doing so (see above) but whilst some providers would alter the privacy requirements to make EHRs available, others said that they were unlikely to become available (if at all) until security improvements could guarantee privacy.

Technology issues

Some of these issues (eg SMS) have been referred to above and respondents generally had quite informed views on technical matters. The major concern was the form factor needed to ensure the portability of a mobile device and the necessary constraints it imposed on the small screen and keyboard size, particularly with mobile phones. Doctors felt that m-health applications should not require users to enter significant amounts of free-field data but rather they should be able to enter items into specified fields by accessing options from pick lists. Thus, keyboards are not the main problem but all responses pointed to the limitation of a mobile phone screen and felt that new technologies such as folding screens (20) would be needed before multimedia or web-based services could come of age.

Future projections by suppliers drew attention to voice recognition technology (clearly of some relevance to a mobile phone!) pointing to the advances that had been made over the last decade but clinicians felt that the science was yet not sufficiently reliable for a mission-critical activity such as healthcare.

A valid point made (over several years) by healthcare systems developers drew attention to the difficulty of achieving a critical mass of sales and the negative impact it had on software construction times and cost (21). Given

the pervasiveness of mobile technologies, this is potentially less of long-term problem but the short-term difficulty focused squarely on the proliferation of so-called standards and the need to port applications across multiple protocols and platforms.

Conclusions and recommendations

The survey generally painted an optimistic picture for the future and 'promise' of m-health and, whilst it was conducted in New Zealand, the implications and the results and conclusions can reasonably be extrapolated to most developed nations. The proliferation of m-health devices and technologies will continue (and is inevitable) and as applications mature we will learn more about what works well what does not. The convenience factors and the ease of communication with m-health will empower patients and when used appropriately improve both their capacity to remain well and, if they do become ill, their ability to cover quickly and enjoy greater quality of life.

A firm recommendation from this study is not to get carried away by the technology or the latest killer applications but to carry out considered and systematic research (22) alongside product and service development that extracts the principles of good practice and disseminates them to all who can deploy them.

References

[1] Taylor C. Global mobile phone connections hit 2.5bn, The Register, http://www.theregister.co.uk/2006/09/08/mobile_connections_soar/. Accessed 1 December 2006.

[2] Office of the Privacy Commissioner. Report of the Privacy Commissioner 2005/06. Wellington: Office of the Privacy Commissioner. 2006, p.10. See also http://www.privacy.org.nz/library/annual-report-of-the-privacy-commissioner-2006. Accessed 1 December 2006.

[3] Hansmann U, Merk L, Nicklous MS, Stober T. Pervasive computing: The mobile world. 2nd ed. New York: Springer; 2003.

[4] Zheng P, Ni L. Smart phone and next generation mobile computing. San Francisco: Morgan Kaufmann; 2006.

[5] Miller M. Mobile health survey – 2005. Auckland: Vodafone; 2005

[6] Glover B, Bhatt H. RFID essentials. Sebastopol: O'Reilly; 2006.

[7] Pagani M. Mobile and wireless systems beyond 3G. Hershey: IRM; 2005.

[8] Funk JL. Mobile disruption: The technologies and applications that are driving the mobile internet. New York: Wiley; 2003.

[9] Rogers EM, Rogers E. Diffusion of innovations. 5th ed. Roseburg: Free Press; 2003

[10] Lloyd-Sherlock P. Population ageing in developed and developing regions: Implications for health policy. Social Science and Medicine 2000;51(6):887-95.

[11] Pink B. Population to reach 5 million by 2041. Wellington: Statistics New Zealand, 2004. See also http://www.stats.govt.nz/NR/rdonlyres/EABFAEA8-CF19-4346-BBF5 1FB579CC98DA/0/nationalpopulationprojections 04base51mr.pdf. Accessed 1 December 2006.

[12] Norris AC. Essentials of telemedicine and telecare: Chichester: Wiley; 2002.

[13] Neville R, Greene A, Mcleod J, Tracy A. Mobile phone text messaging can help young people with asthma. British Medical Journal 2002;325:600.

[14] Martins HMG, Jones MR. What's so different about mobile information communication technologies (MICTs) for clinical work practice? A review of selected pilot studies. Health Informatics Journal 2005;11:123-134.

[15] Kerr K, Norris AC. The development of a health data quality programme. In: Al-Hakim L, ed. Information Quality Management: Theory and Applications. Hershey: Idea Publishing Group. 2006; Chapter V; pp. 94-118.

[16] Galpottage PAB and Norris AC. Patient consent principles and guidelines for e-consent: A New Zealand perspective. Health Informatics Journal 2005;11:5-18.

[17] Skulimowski AM. The challenges to the medical decision making system posed by mHealth, the IPTS Report, Progress and Business Foundation. Issue 81, 2004. See http://www.jrc.es/home/report/english/articles/vol81/ICT1E816.htm. Accessed 1 December 2006.

[18] Christensen CM. The innovator's dilemma. Harvard Business School Press; 1997.

[19] Mirza FB. Opportunities and narriers for mobile health in New Zealand. Masters thesis: Massey University. (To be published).

[20] BBC. Roll-up screens 'Moving closer'. See http://news.bbc.co.uk/2/hi/technology/5366534.stm. Accessed 1 December 2006.

[21] Burley L, Scheepers H. Emerging trends in mobile technology development: From healthcare professional to system developer. International Journal of Healthcare Technology and Management 2003;5 (3-5):179-183.

[22] Wickramasinghe N, Goldberg S. A framework for delivering m-health excellence. Proceedings of the 18th Bled E-Conference: E-Integration in Action, Bled Slovenia. June, 2005; pp. 1-21.

Address for correspondence

Prof Tony Norris, (t.norris@massey.ac.nz)
Centre for Mobile Computing
IIMS
Massey University
Auckland
New Zealand

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Applying Mobile and Pervasive Computer Technology to Enhance Coordination of Work in a Surgical Ward

Thomas Riisgaard Hansen[a] **and Jakob E. Bardram**[b]

[a] *Department of Computer Science, University of Aarhus, Denmark*
[b] *IT University of Copenhagen, Denmark*

Abstract

Collaboration, coordination, and communication are crucial in maintaining an efficient and smooth flow of work in an operating ward. This coordination, however, often comes at a high price in terms of unsuccessfully trying to get hold of people, disturbing telephone calls, looking for people, and unnecessary stress. To accommodate this situation and to increase the quality of work in operating wards, we have designed a set of pervasive computer systems which supports what we call context-mediated communication and awareness. These systems use large interactive displays, video streaming from key locations, tracking systems, and mobile devices to support social awareness and different types of communication modalities relevant to the current context. In this paper we report qualitative data from a one-year deployment of the system in a local hospital. Overall, this study shows that 75% of the participants strongly agreed that these systems had made their work easier.

Keywords:

operating room information systems, awareness, communication, user-computer interface, software design, mobile phone, handheld computers, computerized medical record

Introduction

Ensuring a smooth flow of work in a operating wards requires efficient collaboration, coordination, and communication [1,2]. A considerable amount of time and energy is spent on getting information about where people are located, what they are doing, getting information about changes in plans, and ensuring that the right staff is present inside the operating room at the right time, including the patient. Large whiteboards and a constant flow of phone calls are some of the tools that clinicians use to communicate, to stay aware on the flow of the work, and to maintain a social awareness of one another.

Information posted on whiteboards are, however, only visible in one location. For example, the whiteboard that displays today's operation program is typically only visible from the coordinating central within the operation ward. Similarly, phone calls might be a useful tool for the caller to gain knowledge and coordinate his or her work, but phone calls may also be disturbing to the recipient because they are occupied with other activities [3].

This paper describes a set of pervasive computing systems called iHospital which are designed to support the intense coordination of work in an operating ward. The goal of these systems is to address some of the shortcomings of communication and awareness systems in use today, thereby increasing the quality and efficiency of operation coordination. The paper then reports from qualitative studies of a one-year deployment of the systems which demonstrates that the systems in general made work easier on the wards by helping clinicians to reduce interruptions, locate each other, get an overview of work, coordinate work, handle changes in work, communicate in and out of the operating ward, and to reduce the amount of traffic in and out of the operating room.

Systems design

The designed system was developed in close participating with doctors and nurses over a one-year period. Several workshops and early mockup and prototypes were used to test out early ideas and get instant feedback. The final developed and deployed system was comprised of three separate but tightly integrated sub-systems: (i) a location tracking system, (ii) the AwareMedia system, and (iii) the AwarePhone system.

Location tracking

The *location tracking system* was designed to get information about the activities and whereabouts of the clinicians. The location tracking system was a zone-based tracking system that only tracks clinicians inside predefined zones, like operation rooms, the patient wards, and the recovery department. Zone-based tracking was chosen partly due to the high cost of fine-grained indoor location tracking equipment, partly because this coarse-grained location was deemed sufficient for our purpose, and partly due to privacy protection of clinicians once outside of 'interesting' zones (see [4] for a survey of tracking systems ubiquitous computing). Clinicians were only tracked inside predefined zones and no logging was performed about how much time the clinicians spent in different zones.

In the pilot study, location tracking was done by tracing Bluetooth tags and devices (e.g. mobile phones) worn by

clinicians. Location information was displayed in the AwareMedia and AwarePhone systems.

AwareMedia

AwareMedia is a full-screen, touch sensitive system running on large and medium-sized wall displays. These clients are typically deployed in the coordination central of an operating ward, inside each operating room, in different patient wards, and in the recovery ward. AwareMedia presents an overview of what is going on in relation to the daily handling of surgeries. The display shows a list of operation rooms and a list of people at work, where they are located, and what their current schedule, according to their calendar.

AwareMedia displays a detailed view of the current and scheduled activities inside each operating room. It displays the current operation, its status, the people located by the tracking system inside the room, the patient being tracked, and a complete schedule of the surgeries scheduled for this room. This information provides an overview of the status and activities inside each operating room as well as detailed and updated information about the clinicians associated with the operating ward.

The program not only provides an overview of what is going on, it also supports different types of communication. A video link from each operating room provides a passive overview of what is going on in each operating room (these video links are also called media spaces [5]). The quality is adjusted so it provides an overview of what is going on the operating room without being privacy-invading for the patient. A chat function allows clinicians to send information between the system and the operating theatres, collaborating wards or mobile devices running the AwarePhone system.

Figure 1 - The AwareMedia System running on two large displays in use at the hospital (left), The AwarePhone System displaying a list of clinicians and their location (bottom) and at the top is a clinician wearing a small tracked Bluetooth chip (top)

AwarePhone

The AwarePhone system is an application running on mobile devices and it provides a similar functionality to the AwareMedia system. The mobile device has a smart phone book where each clinician and operating room is listed. Besides presenting the name the phone book keeps an updated list of the location of clinicians, their current booking in the calendar system and a self-reported status. Next to the operating rooms is listed the current surgery in that room and its status (patient arrived, surgery just started etc). By selecting a person or an operating room, the system presents the options to call the person/room or send a message. If a message is sent to the operating room the message is presented to the people in the operating room through the AwareMedia system.

Figure 1 shows an overview of the three systems. To the left is the AwareMedia system running in the coordinating central at the hospital, to the right is a clinician tracked with a Bluetooth chip, and the bottom picture shows the AwarePhone system. The developed systems are rather complex and the technical details of the systems are presented in depth in [6, 7, 8].

Methods

The suite of systems described above was deployed in January 2006 in an operating ward in a medium-sized Danish hospital. The operating ward is a centralized ward with nine operating rooms and supports three departments (organ, orthopedic, and gynecology/obstetric surgery). Around 150 clinicians were associated with the operating department.

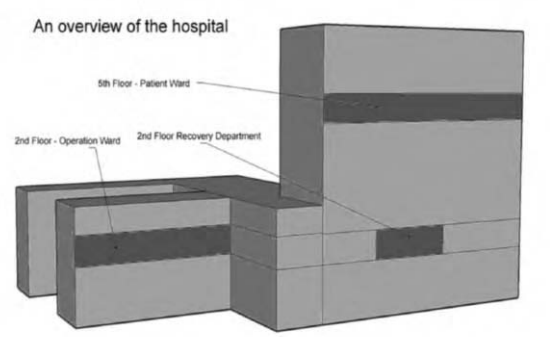


Figure 2 - A view of the hospital with the operating ward marked to the left on the second floor, the recovery department in the main building on the second floor and the patient ward at the fifth floor in the main building

For the pilot study, three operating rooms were equipped with the AwareMedia system (two orthopedic and one gynecologic), the coordinating central was equipped with a large wall-sized version of AwareMedia, and they recovery department and the patient ward for orthopedic surgery was also equipped with the system. For the tracking system 10 zones were defined and the system was able to track around 30 people (20 carrying mobile devices and 10 carrying tags). 15 mobile phones were distributed to clinicians, operation technicians, and nurses with coordination responsibilities.

Figure 2 shows a sketch of the hospital and presents the areas of the hospitals the system was installed in.

Table 1 - The Table shows ten of the forty six questions and the corresponding responses

How much do you agree with the following statements?	Strongly agree	Agree	Neutral	Slightly disagree	Disagree	Don't know	Not answered
9a. The system has lead to fewer interruptions	35%	30%	14%	0%	7%	12%	2%
9b. It has been easier to locate co-workers	42%	23%	21%	5%	5%	5%	0%
9c. I have a better overview of today's work tasks.	53%	23%	12%	5%	2%	5%	0%
9d. The system made it easier to coordinate work.	47%	21%	14%	0%	5%	5%	9%
9e. I save some steps.	49%	16%	19%	5%	7%	5%	0%
9f. It has become easier to handle changes in today's schedule.	35%	33%	23%	0%	5%	5%	0%
9g. There has been a better communication between the operating ward and the recovery department.	9%	16%	30%	0%	7%	30%	7%
9h. There has been a better communication between the operating ward and the patient ward.	23%	21%	21%	5%	7%	21%	2%
9i. The system has minimized the traffic in and out of the operating rooms.	30%	28%	14%	2%	2%	19%	5%
9j. The patient is getting a better treatment	7%	23%	37%	0%	9%	19%	5%

Results

Questionnaires

Questionnaires were given to a number of clinicians such as operation technicians, nurses, anesthesiologists, and surgeons. In total 43 questionnaires was handed out, 34 responses came from nurses and 9 from doctors and other medical personnel (approximately half the clinician working regularly in the operating ward filled out a questionnaire).

Being asked about the question, if the system in general made their work easier, 75 % strongly agreed with this statement. Only a single response stated that the system made it more difficult to use. Some other results from the questionnaires are presented in Table 1 which reveals that clinicians found that the systems helped them coordinate their work, led to fewer interruptions, helped them locate co-workers, saved them some steps, helped them handle changes in the operating schedule, helped improve communication in and out of the operating ward, and reduced traffic in and out of the operating room. The clinicians, however, reported that they did not find that the systems significantly helped improve patient treatment.

Interviews

Fourteen structured interviews were performed with doctors, nurses and supporting staff after the system had been used for three months. Each interview lasted approximately half an hour each.

The data material is huge, and in this paper we will only present very few of the statements. One of the findings highlighted by several persons was the ability to react to changes earlier than before. An operating nurse puts it this way:

[*With the system] people are starting to react earlier than before. We can see the surgery in operating room 9 is delayed so the last patient can be operated in room 4 instead. And that is a clear advantage because then the surgeon can just throw the gloves and move on to the next patient. That is a huge advantage because sometimes the last patient needed to be postponed to the next day.*

By monitoring the progress in the different operating rooms simultaneously, the coordinating nurse has a powerful tool to move patients to other rooms if the schedule is behind in one of the rooms.

Log data

The system logged all the interaction with the system. From the logged data we were able to analyze how the system was used, how many people that were tracked each day, how many messages sent, and similar data. For instance Figure 3 shows the number of surgeries created, modified, and deleted (canceled) using the system from January to September.

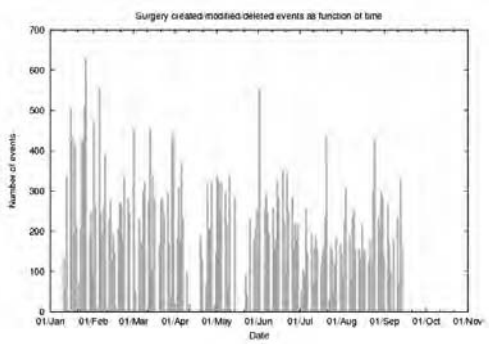


Figure 3 - The number of changes made by the AwareMedia system from January to September. A change can be a surgery being moved, a surgery being extended, the personal associated with the surgery being substituted. Some actions might result in a number of changes

Figure 3 shows that that the amount of changes made to the schedule each day is quite substantial – from 100-500 events pr. day, which correspond to approximately 10-20 changes to each operation. A small drop over time can be seen, but our analysis points to this drop being due to the clinicians getting more skilled in using the system and actually not less use. However, a larger drop can be found in the number of users being location tracked which will be further addressed in the discussion.

Discussion

Based on the study of the deployed set of systems, we will argue that there is some evidence that such mobile and pervasive computing systems may help enhance the work in an operating ward.

Efficiency

There is clear evidence that the system increased the efficiency in work at the operating ward. 67% of the clinicians agreed that the systems made coordination easier. 66% found that they could handle changes in the daily operating schedule easier and in the interviews the clinicians argued, that the systems increased the number of performed surgeries because it was easier to coordinate and gather resources to unscheduled surgeries. We are currently working on elaborating the findings by analyzing the number of surgeries carried out before and after the introduction of the systems.

The system was also designed to make communication between the operating ward and the patient ward and recovery ward easier. However, in the questionnaire only 44% and 26% respectively agreed to this. It was nevertheless in the interviews argued by the people using the system a lot[1] that the systems did enhancing cross-departmental communication and coordination. Many clinicians simply did not use this feature of the system, which also explains the large number of 'don't knows' in the questionnaire.

It is evident that the system helps clinicians locate each other and saves them many unnecessary steps. 65% agreed that it was easier to locate people and 65% agree that the systems save them some steps. Hence, critical time is saved.

Finally, the clinicians did not agree with the statement that the treatment of the individual patient had been improved based on the systems. However, the questionnaire was not handed out to people at the recovery and patient ward. At these ward they pointed out that they were able to give much more precise information to the patients and their relatives with the presented system than before. Secondly, the systems directly increase the *work quality* of the clinicians as such and hence indirectly provide better treatment of the patient.

Quality

Too much turbulence in the operating room can bring bacteria close to the surgical area and it is hence important to keep the airflow from the non-sterile areas to a minimum. Opening the door to an operating room in order to pass on a message, for example, is hence a safety hazard. It is hence interesting to note that 58% of the clinicians agreed that traffic in and out of the operating room has been minimized which clearly has improved the hygiene inside the operating rooms.

Especially, the chat messages were used as an alternative to passing messages between the operating room and the rest of the department. These messages were also mentioned as being far less interruptive than e.g. opening the door or calling the room for passing on shorter messages.

In particular from the interviews it became clear that the system could decrease errors or misunderstandings. A number of operation nurses pointed out that a number of misunderstandings previously arose simply because changes to the schedule were not passed on to all involved clinicians. As one of the operating nurses explained in an interview:

> *I remember once I was preparing for an operation. After I had finished preparing all the equipment, I was told that the operation was canceled. I had to return the equipment and start all over again, thereby delaying the flow of work. If I have had the system in that operating room it would have saved a lot of time for everybody.*

The nurses saw a significant decrease in these types of errors by using the suggested system.

1 This was typically nurses working as 'coordinators', i.e. the person in charge of coordinating the whole operating schedule for the whole operating ward. It was typically this person who was in charge of calling the patient ward and the recovery ward to ensure a smooth in- and out-flow of the patients.

Finally, we conclude that many of the interviewed clinicians found the system to make the work less stressful and overall strengthen the feeling of being ahead. Findings which are supported by the decrease in interruptions and easing of collaboration and scheduling pointed to by the questionnaires.

Challenges

While the results are promising a number of challenges exist. Interference is a highly controversial and debated area following the introduction of wireless communication technologies in hospitals. In some hospitals wireless equipment are used extensively whereas others have a complete ban on all wireless and mobile communication equipment and little research exists to guide the decisions. For our trial we got a research permit to try out mobile devices in the operation ward though with a safety distance of one meter to older medical equipment. A recently published paper in Mayo Clinic Proceedings concludes based on more than 300 tests that mobile phones do not pose the risk of interference [10].

A related problem with mobile and wireless technology is battery lifetime. While some systems can run for days, others need to be recharged regularly and managing the recharging process can be a burden.

Also privacy and local regulations pose a challenge. While the systems are designed to support quick overview and the ability to gather information by glazing at the displays while walking by, this easy access to information also pose a problem. Regulations are designed to ensure that sensitive information is only viewable by registered users. Logging of all interaction including viewing the data is often a requirement. Balancing easy to use information overview with the confidentiality of the information displayed is a clear challenge. In the trial we addressed this issue by only placing the displays in restricted areas, but even this approach might not correspond with the regulations in some countries.

Conclusion

The pressure on the health system is increasing. There is a constant demand for treating more patients with new techniques with the same amount of human and financial resources. New types of Ubiquitous Computing systems alleviate this fundamental dilemma by helping staff increase the level of treatment for the same amount of resources. Not by cutting the time spent talking to patients or increasing the overall speed, but by reducing some of the small time consuming problems encountered in everyday medical work. For example, the time spent searching for people, coordinating treatments, passing on information, being interrupted, or waiting for computers to log in [9].

In the presented set of iHospital systems we have found clear indications of how a system designed to support awareness and collaboration with the feedback from clinicians indeed can increase both the efficiency and quality of surgical treatment. The system was designed to run as a prototype for three months and currently it has been running for more than a year and clinicians call it an invaluable tool in their work. At the point of writing a new system is being developed based on the pilot study that is able to support the entire operating ward and collaborating activities.

Acknowledgments

This research has been supported by the Competence Centre ISIS Katrinebjerg. We thank the surgical staff at Horsens Sygehus who enthusiastically participated in this project as well as the management for daring this visionary project. We are also thanking Mads Søgaard, Christian Jonikeit, Martin Mogensen and Jonathan Bunde-Pedersen, who have made major contributions to the design, implementation and evaluation of the system.

References

[1] Plasters CL, Seagull FJ, Xiao Y. Coordination challenges in operating-room management: an in-depth field study. AMIA Annual Symposium Proc 2003: pp 524-8.

[2] Xiao, Y. 2005. Artifacts and collaborative work in healthcare: methodological, theoretical, and technological implications of the tangible. J. of Biomedical Informatics 38, 1 Feb. 2005: pp 26-33.

[3] Nardi BA, Whittaker S, and Bradner E. Interaction and outeraction: instant messaging in action. Proceedings of the 2000 ACM conference on Computer supported cooperative work, ACM Press, New York, NY, USA, 2000. pp 79–88.

[4] Hightower J, and Borriello G. Location systems for ubiquitous computing. IEEE Computer, 34(8): August 2001: pp 57-66.

[5] Dourish P, and Bly S. Portholes: supporting awareness in a distributed work group. In CHI '92: Proceedings of the SIGCHI conference on Human factors in computing systems, ACM Press, 1992: pp 541–547.

[6] Bardram, JE, Hansen, TR, Soegaard, M, AwareMedia – A Shared Interactive Display Supporting Social, Temporal, and Spatial Awareness in Surgery, Proc. CSCW 2006: pp 109-118.

[7] Bardram, JE, Hansen, TR, Mogensen, M, Soegaard, M, Experiences from Real-world Deployment of Context-Aware Technologies in a Hospital Environment, Proc. UbiComp, Springer Verlag 2006: pp 369-386.

[8] Hansen, TR, Bardram, JE, Soegaard, M, Moving out of the Laboratory: Deploying Pervasive Technologies in a Hospital, IEEE Pervasive Computing, 5(3), 2006: pp 24-31.

[9] Tantori M, Favela J, and Gonzalez VM, Towards the Design of Activity-aware Mobile Adaptive Applications for Hospitals, UbiHealth Workshop 2006.

[10] Tri JL, Severson RP, Hyberger LK, Hayes DL. Use of cellular telephones in the hospital environment. Mayo Clin Proc 2007;82(3):282-285, 2007.

Address for correspondence

Thomas Riisgaard Hansen
Department of Computer Science
Univeristy of Aarhus
DK-8200 Århus N, Denmark
Email: thomasr@daimi.au.dk

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Feasibility and Usability of a Home Monitoring Concept based on Mobile Phones and Near Field Communication (NFC) Technology

Jürgen Morak, Alexander Kollmann, Günter Schreier

Austrian Research Centers GmbH – ARC, Biomedical Engineering, eHealth systems, Graz, Austria

Abstract

Utilization of mobile information and communication technologies in home monitoring applications is becoming more and more common. The mobile phone, acting as a patient terminal for patients suffering from chronic diseases, provides an active link to the caregiver to transmit health status information and receive feedback. In such a concept the usability is still limited by the necessity of entering the values via the mobile phone's small keypad. The near field communication technology (NFC), a touch-based wireless interface that became available recently, may improve the usability level of such applications significantly. The focus of this paper is to describe the development of a prototype application based on this technology embedded in a home monitoring system. The feasibility and usability of this approach are evaluated and compared with concepts used in previous approaches. The high quantifier with respect to overall usability indicates that NFC may be the technology of choice for some tasks in home monitoring applications.

Keywords:

home monitoring, mobile phone, patient terminal, near field communication, usability

Introduction

Tracking health related parameters like blood pressure, heart rate, body weight, or blood glucose levels is part of many approaches in the treatment of chronic diseases like hypertension, congestive heart failure or diabetes mellitus. Such data are expected to show trends in the illness patterns and to help the physician in guiding the patient to the best possible health status by adjusting the treatment settings individually.

Using the conventional method of tracking data by hand written lists and diaries, however, often leads to poor patient compliance. Furthermore, interpreting such written records is difficult and time consuming [1].

Contemporary information and communication technology offers more appropriate home monitoring solutions. Instead of getting the patient's hand written diary once in a while in the course of consultations at the doctor's office, the patient and his/her doctor may use a home monitoring system to stay in contact.

A typical home monitoring system comprises of the following three parts:

- caregiver terminal – a PC with a secure Internet connection allows the caregivers to connect to the remote monitoring centre to view patient's data via a web - portal.
- monitoring centre - a web based system that hosts a central database to store, manage and process all data received from the patient and to show those data to authorized users. It may also offer automatic monitoring processes like continuously checking whether newly received values cross certain thresholds in order to notify caregivers.
- patient terminal (PT) - a PC or a mobile phone (MP) which enables the patient to feed the database with his/her self measurements and additional information via a secure Internet connection. The benefits of absolute mobility and high availability at low cost make the MP a very attractive PT to upload data and receive feedback.

One common way to transmit health measurements, for example blood pressure values, is to have the patients to transcribe the values from the meter's display into the MP using the keypad. Basically, this can be achieved in two different ways:

- **online** – Using the integrated micro-browser of the MP allowing the user to transmit data to the monitoring center via wireless Internet technology. The user enters the data into input-templates generated by Wireless Markup Language (WML) or the Hypertext Markup Language (HTML). Data is exchanged using the Wireless Application Protocol (WAP) [2].
- **on/offline** - values are entered by means of a stand-alone software application running on the MP (based on Java 2 Micro Edition J2ME technology) that stores the data locally and synchronizes afterwards [3] with the central database. A connection to the monitoring centre is not necessary during data entry.

However, the usability of both concepts is still limited by the necessity of dealing with the MP's small keypad and display. Fortunately, more and more health monitoring devices like blood pressure meters (BPM) already feature on-board communication interfaces. Those interfaces can be grouped by 2 different principles:

- **Wired solutions** based on universal serial bus (USB) or RS-232 cables: These interfaces are predominantly intended to communicate with a special software application running on the PC to receive and manage the data.
- **Wireless solutions** based on Bluetooth or Infrared Data Association (IrDA): They can easily be used to send measurement values to a MP which serves as a PT and facilitates automated data forwarding to the monitoring centre.

PTs based on those solutions are still not perfect. From the patient's point of view an adequate PT should feature the properties listed in the first column of table 1. These requirements can be met by technical features listed in the second column accordingly.

Table 1 – Properties of an adequate patient terminal

requirements	technical parameters and solutions
easy to learn and to use	high usability and automated methods
feedback	bidirectional communication between PT and backend system
single,integrated solution	flexibility and adaptability to various monitoring situations
privacy	data security by the means of encryption and user authentication
availability of service	on/offline data acquisition
reliability of service	error-handling, fault tolerant communication technologies
low cost	based on existing standards supported by MPs and/or PCs

All these requirements seem to be reachable by using MPs in combination with the Near Field Communication (NFC) technology. NFC technology provides an intuitive touch-based connection for automated data exchange between two devices. The present paper describes the development and initial evaluation of a home monitoring system based on NFC with the goal to assess the feasibility and usability of this method compared to concepts used in previous approaches.

Near Field Communication (NFC)

NFC is a short range communication technology for wireless peer to peer interconnections. It combines the benefits of radio frequency identification (RFID) and wireless communication technologies (Bluetooth, WLAN, IrDA) [3]. NFC is based on magnetic inductive coupling, adopted form RFID. The typical operating configuration of an RFID system is an active reader/writer device and one or more transponders as passive participants (e.g. RFID-tags or contactless smartcards). RFID technology is asymmetric in the sense that only the reader/writer device can initiate a data transmission. In contrast NFC, as the subsequent development allows bidirectional communication between two devices (similar to Bluetooth).

NFC operates in the unlicensed frequency band of 13,56 MHz and reaches a data transmission rate of up to 424 kBit/s within a short range of typically 5 to 20 centimetres. The analogue circuit for field generation and -detection and various digital components are combined in a single chip solution [4]. The NFC transmission module can operate as reader/writer unit to access tags and contactless smartcards based on ISO 14443 like MIFARE (NXP, Gratkorn, Austria) and FeliCa (Sony, Tokyo, Japan) products and additionally supports peer-to-peer communication between two devices in active and passive mode.

Basically, NFC enables data exchange just by bringing two devices close together. This is the intended purpose and the basic idea of NFC – to launch a communication session in an intuitive, easy to handle, and secure way. Literally touching NFC enabled devices, tags or contactless smartcards with a transmission device allows the user to access information and services and to run applications. Therefore, NFC is appropriate to be integrated in handheld devices like MPs and Personal Digital Assistants (PDA).

The first available MP that can be extended with NFC technology was the Nokia 3220 (Nokia Corporation, Helsinki, Finland) [5]. When equipped with a special NFC shell the MP enables users to read data from compatible RFID-tags, as well as to communicate with other NFC devices. When service instructions are provided, a build in software application (Service discovery) recognizes them and launches the appropriate MP functions like initiating a phone call, sending a text message (SMS), or connecting to a predefined web service.

Methods

Home monitoring system

We developed a home monitoring system consisting of:

Patient terminal

To enable a medical measurement device with NFC we developed an all-purpose NFC module equipped with various standardized interfaces. It consisted of a single chip NFC module (PN531, NXP, Gratkorn, Austria) [6] and an embedded microcontroller (PIC 16F88; Microchip, Chandler, AZ). We implemented this module using surface mounted device (SMD) technology to keep it small (30x35mm) and to case the integration into the UA-767PBT (A&D Company, Tokyo, Japan), a digital BPM (figure 1, a). This device offers a standardized interface and was originally connected to a Bluetooth module. Instead of the Bluetooth module we attached the NFC module and adapted the module's firmware to understand the proprietary protocol.

To fetch the measurement values from the NFC enabled BPM we utilized the Nokia 3220 which has been equipped with the special Nokia NFC shell (figure 1, b). To make the MP able to receive and forward the blood pressure value, we configured the firmware of the integrated NFC module to be compatible with the pre-installed application Service Discovery. The transmitted information consisted of:

- the measured values for systolic and diastolic blood pressure as well as the heart rate,
- a unique device identification number,
- a checksum to provide for transmission reliability,
- the web address where the data should be forwarded to by the MP.

In this configuration the BPM worked "out-of-the-box" without the need of any further settings or configurations by the user.

Figure 1 – NFC "pair": blood pressure meter UA-767PBT + NFC module inside (a) and mobile phone Nokia 3220 (b)

Monitoring center

We set up a three tier server architecture to receive, process and store the measurement values which was based on the following open source components:

- web server (Tomcat, Apache Software Foundation, Wilmington, DE) to receive data and to provide the information to the user via a web browser,
- application server (Zope 6.1, ZOPE Corporation, Fredericksburg, VA) to manage data and users (figures, statistics),
- database (Interbase 6.1, Borland, Cupertino, CA) to store the data persistently.

Caregiver terminal

The user was able to view the most recently transmitted blood pressure value or the scatter plot of all transmitted values (Figure 2) via web browser.

Evaluation of feasibility and usability

To evaluate feasibility and usability of the prototypically developed NFC based home monitoring concept we asked students in a health-care-management course to perform measurements and to transmit the values to the monitoring centre. The purpose of the survey was to quantify and compare WAP, J2ME, and NFC technology for MP based data acquisition and transmission in a controlled environment according to the following protocol:

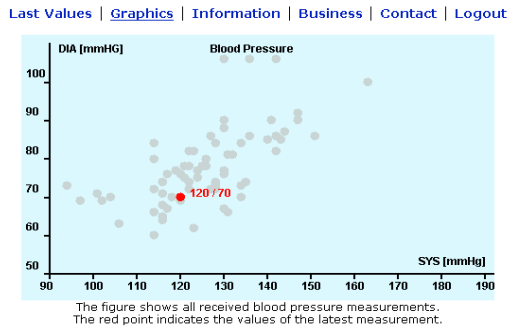


Figure 2 – View of the scatter plot of all transmitted values

- Prepare the MP for the use as PT. When WAP technology was used the web-address has to be entered and stored as a short-cut. In case a J2ME based software application was used, the software has to be downloaded and installed first. In case NFC was used, no further preparation is necessary.
- Put on the cuff and start the measurement by pushing the start button on the BPM.
- After completion of the measurement process, the displayed values have to be transferred to the MP by entering the numbers via the keypad (WAP and J2ME). When using NFC technology, the MP has to be moved close to BPM. Data transmission from BPM to the MP starts automatically and so does the software application that prepares sending the received values.
- In every case, data upload to the monitoring centre starts after pressing the "OK" button and is followed by an acknowledgement that the transmission was successful, reflecting the values that were written to the web-based database.

Each student was asked to transmit at least one dataset to the monitoring center using WAP, J2ME, and NFC. For each technology, the general user requirements as presented in table 1 were interrogated in the course of the survey. The participants were asked to weight and rate 20 predefined properties of an optimal PT concerning usability, technical feasibility, service and costs.

The requirements were weighted from 0 (not relevant) to 100 (very important) according to the individual appraisal of each student. Accumulated weight factors were normalized to a sum of 100. Thereafter the test person was asked to rate each requirement from 1 (don't agree) to 5 (strongly agree) for each method. If a certain method didn't feature a requirement it was set to 0.

The 20 requirements were merged (averaged weight and averaged rating) to the general user requirements as described in table 1. The overall quantifier for each requirement and method was estimated by multiplying the averaged weight value by the averaged rate value. Results were presented as mean +/- SD. To assess the significance of differences between the methods statistically, the

Friedman – Test was used. P < 0,05 was considered to be statistically significant.

Results

14 students (6 female, age 23,2 +/- 4,9 years) evaluated WAP, J2ME and NFC technology for the usage in MP based PT. 13 students returned duly completed questionnaires for further analyses. Overall, NFC was rated as the most suitable technology if compared to WAP and J2ME. The detailed results displayed in Table 2 indicate that general usability aspects (easy to use and learn) were weighted as most important (33% out of 100). Availability (21%) and Reliability of service (15%) were rated as second and third, respectively.

Table2 – Weighted comparison of WAP, J2ME and NFC; Bold values indicate the most suitable technology for a given requirement

requirement	weight [%]	quantifier			p-value
		WAP	J2ME	NFC	
easy to use and learn	33	99,6 +/-23,6	128,6 +/-21,0	**138,6 +/-37,9**	< 0,01
feedback	5	**16,1 +/- 4,8**	11,5 +/- 5,3	12,3 +/- 8,1	n.s.
integrated	12	35,0 +/-12,0	31,6 +/- 7,4	**38,9 +/-10,0**	n.s.
privacy	3	5,8 +/- 2,6	9,2 +/- 2,9	**10,2 +/- 2,4**	< 0,01
availabiity of service	21	59,5 +/-13,1	**80,8 +/-14,3**	78,1 +/-23,4	< 0,01
reliability in service	15	32,7 +/-11,6	48,2 +/-11,6	**58,5 +/-15,5**	< 0,01
low cost	11	**41,0 +/-13,1**	36,9 +/- 9,9	33,3 +/-12,9	n.s.
total	100	289,7 +/- 40,4	346,7 +/- 38,1	**369,9 +/- 82,8**	< 0,01

Comparing the quantifier for each method indicated that NFC was the most suitable technology in terms of general usability aspects (easy to use and learn). NFC was also found being the most suitable solution when a single, integrated system is needed, regarding privacy, as well as reliability of service, shown in table 2.

WAP was rated as the most suitable technology in terms of feedback and costs. J2ME was rated as most feasible in respect to the availability of service.

Besides the usability and feasibility test of NFC in a controlled environment the device was in use for a testing period of 11 month. Overall, more than 500 health parameters recorded by 10 different users were successfully received, mainly from Austria but also from Germany, Finland, and the U.S.

Discussion

After initial conceptual considerations on the applicability of NFC for Health Monitoring [7] the present paper provides the first report on results which had been obtained with a working prototype based on NFC technology.

Although the medical benefit of using MPs as PTs in home monitoring scenarios could already be demonstrated [8] (reduction of emergency cases and hospitalization, increase in patient therapy adherence, etc.), some patients reported a lack of usability using the conventional methods. In our experience the problems, mostly faced by elderly people, are in handling the small keypad and reading data from the display. This impression was confirmed by the mostly young test users who rated WAP technology lowest regarding general usability aspects. Moreover, using WAP technology requires an online connection to the network during data input; it is also lacking in terms of navigation and providing an intuitive user interface.

By using J2ME technology, which allows us to design much more user friendly graphical user interfaces, usability was found to be significantly better as indicated by an increase of this quantifier by more than 25% as compared to WAP (table 2). Additionally, validating the entered data using plausibility constraints is feasible for J2ME software applications resulting in lower error rates and increased reliability as compared to WAP. However, both technologies require at least basic knowledge in handling an MP which some - in particular elderly - people may not have.

Using NFC opens a new approach to simplify the health data acquisition process in home monitoring scenarios. It provides an intuitive and automated way to read out measurement values from medical devices. Compared to WAP und J2ME, NFC technology showed the highest quantifier regarding overall usability. The “bring-in-touch” paradigm of making two devices communicate provides an error-fuse way to assess the data without the need of cumbersome configurations or manual user interaction. Therefore, as compared to WAP and J2ME technology, the NFC based PT can be learned quickly without extensive training.

Another technical advantage of NFC, which was also recognized by the test users, is the fact that NFC provides an integrated solution. On the one hand existing measurement devices can be equipped with NFC modules, on the other hand NFC technology is compatible to certain parts of the existing RFID infrastructure. Therefore, RFID tags can be used to manage additional information. As an example, RFID tags could be attached to medication boxes or well-being icons. Simply touching those items with the MP would be enough to record medication intake or the well-being status.

Table 2 also indicates the main deficits of NFC technology. Because of limited availability of NFC enabled MPs the costs are relatively high at the moment. In the near future, however, NFC enabled MPs will penetrate the market leading to lower prices. The main driving force currently is the MP’s ability to act as a contactless smartcard for pay-

ment and ticketing applications. A forecast estimates that shipments of NFC enabled MPs will increase from less than 5 million in 2006 to nearly 500 million in 2011 (Figure 3).

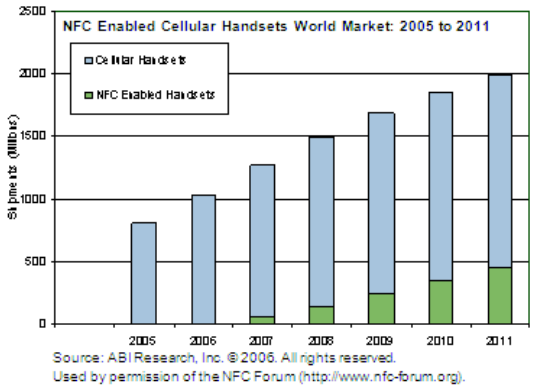


Figure 3 – Predicted shipments of NFC enabled MPs

At the moment the NFC prototype works as stand-alone-application based on the pre-installed service discovery software of the Nokia 3220 MP. Hence, only a single scenario (in our case blood pressure monitoring) is supported. To increase flexibility, we intend to develop a software-application running on the MP that will combine touch based data fetching with features provided by a J2ME application. The functions will include:

- automated fetching of measurements from medical devices and various parameters from RFID tags
- user authentication via RFID tag or RFID smartcard
- manually adding of information
- local storage of data and automated or manually synchronization with the monitoring centre
- increased data security by means of end-to-end data encryption

Additionally, we intend to redesign the NFC module to allow for integration into other medical devices like body weight scales and blood glucose meters.

Conclusion

The results of this limited survey obtained so far confirm the assumption that NFC technology does have the potential to bridge the gap between patients and the technical infrastructure in home monitoring scenarios. If similar results can be confirmed with patients in a real world home monitoring scenario, NFC technology may be established as a method that would ultimately enable more elderly and technically unskilled people to benefit from home monitoring.

References

[1] Holman H, and Lorig K. Patients as partners in managing chronic diseases. British Medical Journal 2002: vol. 320: 526 – 527.

[2] Salvador CH, Carrasco MP, de Mingo MAG, Carrero AM, Montes JM, Martín LS, Cavero MA, Lozano IF, and Monteagudo JL. Airmed-cardio: a GSM and Internet services-based system for out-of-hospital follow-up of cardiac patients. IEEE Trans Inf Technol Biomed 2005: 9: 73-85.

[3] Farmer A, Gibson O, Hayton P, Bryden K, Dudley C, Neil A, and Tarassenko L. A real-time, mobile phone-based telemedicine system to support young adults with type 1 diabetes. Inform Prim Care 2005: 13(3): 171-178.

[4] NXP. Near Field Communication. Online, last visit: Dcember 2006. available at: http://www.nxp.com/products/identification/nfc/

[5] Nokia press release, 02 November 2004. online, available at: http://press.nokia.com/PR/200411/966879_5.html

[6] Philips Semiconductors. Objective short form specification, PN531- µC based Transmission module. Revision 2.0. 2004

[7] Strömmer E, Kaartinen J, Pärkkä J, Ylisaukko-oja J, and Korhonen I. Application of Near Field Communication for Health Monitoring in Daily Life. Proceedings of the 28th IEEE EMBS Annual International Conference New York City, Aug 30-Sept 3, 2006

[8] Scherr D, Zweiker R, Kollmann A, Kastner P, Schreier G, and Fruhwald FM. Mobile phone-based surveillance of cardiac patients at home. Journal of Telemedicine and Telecare 2006: 12: 255-261

Address for correspondence

Jürgen MORAK, MSc
Reininghausstrasse 13
A-8020 Graz
P: +43 316 586570-11
F: +43 316 586570-12
E: juergen.morak@arcsmed.at

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Using Personal Digital Assistants and Patient Care Algorithms to Improve Access to Cardiac Care Best Practices

Kathryn L Momtahan[a], Catherine M Burns[b], Heather Sherrard[c], Thierry Mesana[c], Marino Labinaz[c]

[a] *The Ottawa Hospital, Ottawa, Ontario, Canada*
[b] *Systems Design Engineering, University of Waterloo, Waterloo, Ontario, Canada*
[c] *University of Ottawa Heart Institute, Ottawa, Ontario, Canada*

Abstract

In order to facilitate knowledge transfer between specialists and generalists and between experts and novices, and to promote interdisciplinary communication, there is a need to provide methods and tools for doing so. This interdisciplinary research team developed and evaluated a decision support tool (DST) on a personal digital assistant (PDA) for cardiac tele-triage/tele-consultation when the presenting problem was chest pain. The combined human factors methods of cognitive work analysis during the requirements-gathering phase and ecological interface design during the design phase were used to develop the DST. A pilot clinical trial was conducted at a quaternary cardiac care hospital over a 3-month period. During this time, the DST was used by the nine nursing coordinators who provide tele-triage/tele-consultation 24/7. This clinical trial validated the design and demonstrated its usefulness to advanced cardiac care nurses, its potential for use by nurses less experienced in cardiac care, and for its potential use in an interdisciplinary team environment.

Keywords:

clinical decision support, personal digital assistants, cardiac care best practices.

Introduction

As the worklife of healthcare professionals becomes ever more complex, there is a need to provide them with decision support tools that enhance their practice. Providing the right information at the right time in the right format is a modern challenge that healthcare researchers, educators, and clinicians are currently struggling with. At present, even though there is a trend to increase the role of clinical nurse specialists in healthcare, there are a declining number of nurses with sufficient training to perform this role [1]. In addition, there continues to be very few computerized decision support tools for advanced practice nurses [2] [3]. Every indication, however, is that this is changing. Journal articles, best practice guidelines, drug guides and other forms of decision support tools for clinicians are readily available on the world wide web. Decision support can be thought of as any resource that provides guidance for making decisions regarding clinical care. The availability of decision support tools for downloading onto personal digital assistants has seen explosive growth in the last several years, and the rate of adoption has outpaced the evaluation of such tools [4]. Strayer, Reynolds and Ebell [5] provided a comprehensive overview of the various healthcare applications available for decision support on PDAs. The majority of applications available, however, do not provide targeted summary information in order to provide information to clinicians in an efficient manner. In addition to whatever evidence may be drawn upon in the literature, there is a window of opportunity to capture the clinical expertise of experienced clinicians before the first large wave of retiring healthcare practitioners leave practice. With the regionalization of healthcare services, there is also a need for innovative ways to support teamwork across professions and between hospital care and care in the community.

The intent of this project was to demonstrate the viablity and value of implementing a cardiac decision support tool on personal digital assistants to deliver standardized care to cardiac patients using a human factors approach to the design. PDAs are particularly suited to tasks where there is an urgent need to access the decision support tool and the user is mobile.

The University of Ottawa Heart Institute nursing coordinators (NCs) receive approximately 2,000 calls a year from patients. Of these 2,000 calls, about 20% of them are from patients experiencing chest pain. Therefore, the application chosen to be designed and developed was the cardiac teleform, which was the previous paper-based documentation form that the nursing coordinators used to document calls from patients. This documentation record was used to capture patient demographic information, a description of why the patient was calling in, the pertinent clinical history, the nursing assessment, a description of the advice and recommendations that the NC gave the patient, and documentation as to whether or not the patient agreed with the advice and recommendations. The intent of the PDA-based teleform was to provide a vehicle for documenting the information exchange between the nursing coordina-

Selected for best paper award.

tors and the caller, but also for embedding decision support.

Methods

Two human factors methods were used for this project. A cognitive work analysis (CWA) approach [6] was used for the requirements-gathering/task analysis phase and the ecological interface design [7] approach was used for the design phase. Other than the cognitive work analysis (CWA) that has been done related to hemodynamic monitoring [8], there has yet to be a cognitive work analysis performed to model the work that nurses do to design computerized systems that would facilitate the nursing process.

Although a full CWA has five levels of analysis that include an analysis of complexities, an analysis of tasks, strategies analysis, social-organizational analysis, and worker competency analysis, not all five levels are appropriate to conduct for all projects. For this project, we concentrated on the strategies analysis which included in-depth interviews with all eight of the nursing coordinators (NCs) at the University of Ottawa Heart Institute. In addition to the interviews, the NCs were asked to generate examples of the types of calls that they received. Twenty-five hypothetical calls were generated.

The hospital provides patients with a number they can call 24 hours a day, 7 days a week to talk with one of the nursing coordinators. An analysis of the types of calls received was performed on the paper-based cardiac teleform documentation for a one-year period from November 17, 2002 until November 16, 2003. During this period of time, 550 calls were received from cardiology patients and 1520 calls were received from cardiac surgery patients. For the cardiology patients, chest pain calls ranked second in terms of the numbers of calls received and for cardiac surgery patients, chest pain or incision-related calls ranked second. The only calls that were more numerous in number were medication-related calls. In order to support those type of calls, commercial software was downloaded onto the nursing coordinators' PDAs.

In order to embed more explicit decision support related to chest pain calls, we were interested in developing decision trees (algorithms). A brain-storming session was held with nine subject matter experts (SMEs), three of whom had previous experience developing chest pain algorithms for tele-homecare and interactive voice response systems. Three chest pain algorithms were developed: (1) Possible ischemic pain, (2) Cardiac surgery, incision not healing well and (3) Cardiac surgery, incision healing well. The algorithms were reviewed and approved by the cardiac surgeon and cardiologist on the team.

The decision support tool that was developed was a cardiac tele-triage/tele-consultation tool called the cardiac teleform residing on a personal digital assistant (Palm Tungsten T3) with companion desktop software for downloading patient files from the memory card. Feedback on the alpha version of the PDA software was received from the NCs as well as from participants of a family physicians conference and participants of a nurse practitioner conference.

Figure 1 provides an example of how visualization tools were incorporated into the cardiac teleform. OLDCAR is an acronym for onset, location, duration, characteristics, associated symptoms/aggravating factors and relieving factors. The 'T' for 'treatment' that is usually associated with the mnemonic OLDCART was not included, as the application focused on the nursing functions of assessment, consultation and triage of patients.

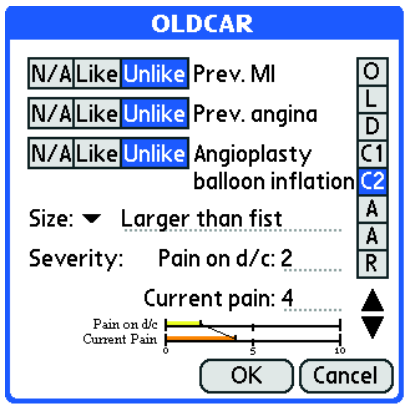


Figure 1 - Screen shot of one of the pages of the cardiac teleform decision support tool

Figure 2 is a screen shot of 1 of the 3 chest pain algorithms. The 'Cardiac surgery, incision healing well algorithm' is recommended for use if the patient's presenting problem is chest pain, they recently has cardiac surgery and the description of their pain does not sound as if it is ischemic in nature and their incision is healing well.

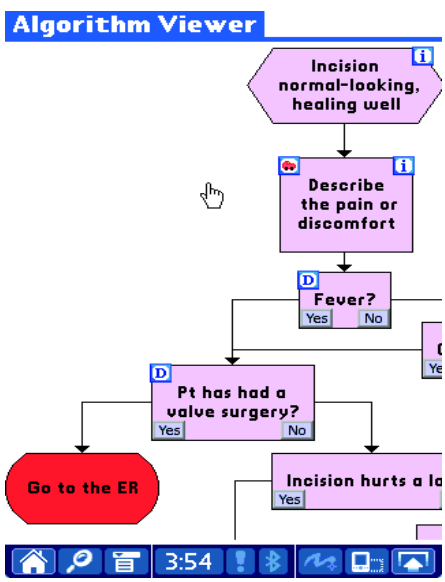


Figure 2 - Screen shot of one portion of the chest pain - incision healing well algorithm

The algorithm is meant to suggest questions that might be asked in order to help the clinician decide on a recommendation. The 'car' icon is an access point for OLDCAR data entry. The 'D' icon is an entry point to provide more information such as whether or not the patient had an aortic or mitral valve replacement, or a valve repair. The 'i' icon indicates points where more information can be obtained, such as a reminder of serious complications that can arise after cardiac surgery, the signs and symptoms of those complications, and any pertinent references.

Since a variety of pieces of information are gathered during a phone call from a patient, we provided a summary page that could be accessed at any time during the call by clicking on the 'S' icon at the top of the screen. The top of the screen also has shortcut icons to facilitate switching algorithms if you need to, going to the OLDCAR assessment page directly, or accessing the bibliography. The bibliography includes references to pertinent texts and articles but also to current related cardiac consensus guidelines from the Canadian Cardiology Society, the American Heart Association, and the American College of Cardiology.

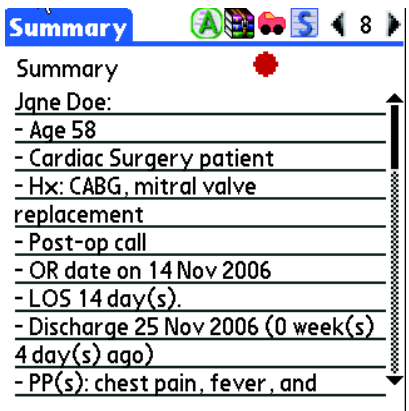


Figure 3 – Screen shot of the summary page which includes all the entered data elements from the previous teleform pages and algorithms

A 3-month pilot clinical trial was performed with the developed cardiac teleform. Outcome measures included evaluations of (a) appropriateness of the advice received as judged by the patient's physician, (b) satisfaction with how the call proceeded and the advice that was given as judged by the patient, and (c) satisfaction with using the tool as judged by the nursing coordinators A log analysis of navigation and data entry strategies was performed and all NCs participated in in-depth interviews at the end of the clinical trial.

Results

Themes from the interviews at the start of the project

Seven of the eight NCs employed by the Heart Institute at the start of the project were interviewed following a semi-structured format for between one and two hours. Although there were many areas of agreement among the NCs, the way in which the call proceeded and the types and number of questions that were asked differed. In terms of agreement, six of the 7 NCs thought that there were major differences in the way a call proceeded with a cardiology patient compared to a call from a cardiac surgery patient. Although initially the NC is assessing whether or not the patient is acutely ill (for instance, having ischemic cardiac pain, in which case the recommendation would be that they hang up the phone and call 911), the calls proceed differently if the patient recently had cardiac surgery or is being followed by cardiology. Many questions that are appropriate to surgery patients (for instance, questions related to the surgical incision) are not applicable to cardiology patients. Therefore, the first two check boxes on page 1 of the cardiac teleform have the NC choose whether the patient is a cardiac surgery patient or a cardiology patient. The subsequent screens then have items related to that subpopulation.

An example of where the call might be managed differently depending on the NC was illustrated by their answers to the following questions: "For cardiac surgery patients, do you always (or almost always) assume that the complaints have to do with the recent surgery? Would you be as likely to consider other possibilities (e.g., a new condition on its own) at the same time"? Four of the 7 NCs said that they almost always assume that the complaint has to do with the recent surgery if the patient calling was a cardiac surgery patient. The other 3 NCs said that they considered not only the recent surgery but other types of problems that might also affect cardiology patients. For this reason, in terms of the design of the teleform, a drop-down menu at the top of the teleform allowed for quick access to any of the 12 pages in the teleform and the NCs did not have to go through the teleform sequentially. The algorithms were designed so that they could be entered or exited easily from any location in the teleform and navigation within the algorithm was not constrained.

Feedback from family physicians and nurse practitioners

Although the feedback on the alpha version of the teleform received from the family physicians and the nurse practitioners at their annual conferences did not help with any design modifications, it did help validate the approach to decision support that we were taking. One family physician commented that it would be particularly useful in his clinical setting where he was the team leader for a number of nurse practitioners in a rural setting.

Results of the clinical trial

The final version of the cardiac teleform was evaluated in a 3-month pilot clinical trial. The NCs were given the choice of using the teleform on the PDA or their paper-based teleform when they received chest pain calls from patients over the 3-month period. Of the 61 chest pain calls, 46 of them were documented with the PDA software. The outcome measures used to evaluate the effectiveness of the cardiac teleform were:

- Appropriateness of the advice received as judged by the patient's physician
- Satisfaction with how the call proceeded and the advice that was given as judged by the patient
- The nursing coordinators' satisfaction level when they used the PDA teleform software on chest pain calls

An analysis of the NCs' navigation and data entry strategies was performed from the teleform log file.

Of the 61 chest pain calls, 46 (75%) of them were documented with the PDA software. The advice received by the patient from the NC was considered appropriate by the patient's physician in 59 of the 61 calls (97%). Thirty-four of the 37 patients who could be reached for feedback were satisfied (92%) with how the call proceeded with the NC and with the advice that they received. Two of the patients contacted did not answer the question. For the one patient who stated that they were not satisfied, the advice given was different from the recommended advice in the cardiac surgery algorithms in the cardiac teleform.

The median satisfaction rating of the nursing coordinators with using the PDA teleform on chest pain calls where 1 was 'not at all satisfied' and 5 was 'very satisfied', was 4.00. The navigation strategies revealed that despite the fact that power user features were built into the decision support tool to reflect the various strategies the NCs used during their calls with patients, they mainly used the cardiac teleform in a serial manner, going from page 1 to page 12 in order.

Themes from the interviews at the end of the clinical trial

Part way through the project, another NC was hired, leading to a total on 9 NCs to participate in the clinical trial. Each of the 9 NCs participated in an end-of-project semi-structured interview lasting approximately an hour. The features of the cardiac teleform that were considered the most useful and easiest to use were the pick lists, the drop down menus and the OLDCAR assessment pages.

The algorithms, which were developed with the help of the NCs and approved by them before the clinical trial, were generally liked for the same reason that the teleform and OLDCAR were liked – namely that it encouraged structure and reminded them to ask questions they might otherwise have forgotten to ask the patient. However, the algorithms were not as universally liked as the other parts of the teleform and the OLDCAR assessment pages. Any problems with using the algorithms appeared to be due to two reasons: (1) there were a small number of atypical calls related to chest pain that did not fit into any of the scenarios in the algorithm and (2) the NCs weren't using the flexibility built into the tool to exit the algorithms at any point and go to any page they wanted in the teleform from the short-cuts icons at the top of the screen.

The NCs saw the benefit of the teleform to be (a) that it provided a standard approach to patient assessment (b) that the documentation was more complete when they used the PDA teleform compared to the paper-based teleform, (c) that it produced a more professional document to send to the patient's cardiologist or cardiac surgeon and (d) that it was more legible than the paper-based teleform. Many of the comments related to improvements that could be made to the teleform were features that already existed but had not been used (for instance, the 'Go To' button at the top of each page where you could navigate to any page in the teleform from that drop down menu).

We investigated the accuracy of the NCs' perception that documentation was generally more complete with the P'DA-based teleform than it was with the paper-based teleform. We compared the documentation of the required data elements in the 46 PDA-based teleform records with a matched set of paper-based teleform records of comparable chest pain calls received immediately preceding the clinical trial. The documentation was more complete on the PDA-based record 68% of the time.

Seven of the 9 NCs thought that it was more time-consuming to use the PDA teleform than to use the paper-based teleform. There was also the fear that the information may get lost since the PDAs were stand-alone and not connected to a wireless network to be downloaded to the hospital LAN as the hospital does not currently have a fully functioning wireless network. The NCs also did not like the process of having to transfer the patient files from the memory card to the card reader and then onto the companion desktop software on the PC.

In terms of the design of the teleform and the algorithms, all of the NCs thought the DST supported their usual workflow and did not interfere or change the development of their mental models regarding the patient's condition. They all thought that the decision support tool would be particularly useful for novice nursing coordinators with the caveat that they also be encouraged to think for themselves in case the problems the patient is calling in with do not fit into the algorithms.

Discussion

The use of the combined human factors methods of cognitive work analysis and ecological interface design resulted in a decision support tool that closely matched the nursing coordinators' cognitive strategies and the mental models they develop during their tele-triage/tele-consultation sessions with patients. Feedback from the nursing coordinators and from nurse practitioners and family physicians at their annual conferences provided support for the approach being taken to develop a decision support tool that would facilitate knowledge transfer between specialists and generalists and between experts and novices while promoting interdisciplinary communication. Standardization of care based on available evidence and best practice guidelines was seen as a major benefit of such a tool, especially in a regionalized healthcare system comprised of many acute and primary care facilities and agencies. The 3-month clinical trial of the decision support tool further validated the design approach taken.

Logistically, the use of PDAs for healthcare applications are still limited by (1) the speed and robustness of the

device and (2) cumbersome data entry using graffiti or the keyboard.

Conclusion

The combination of decision support and personal digital assistants has seen explosive growth in the last several years [4] [5]. However, the majority of the applications available are generic in nature, providing information only or allowing for calculations of various sorts. The next generation of more clinically comprehensive interactive decision support tools for PDAs are only starting to emerge [9] [10]. The decision support tool reported here is another example of a next generation PDA application. The cardiac teleform application design was the cumulative result of a cognitive work analysis and ecological interface design human factors methods. The cardiac teleform and embedded algorithms for decision support provide a flexible, interactive tool which results in documentation of the nursing coordinators' tele-triage/teleconsultation session with the patient. Reference to the evidence, where it exists, to support the recommendations for questions to ask and suggested courses of action to take, are easily accessible in the tool. This type of application has the potential to be utilized in a multi-disciplinary team environment if it were to include scope of practice information for the professionals on the team. The approach taken here also provides a vehicle for capturing the expertise of specialists. Since many aspects of clinical recommendations are the result of expert opinion, this tool also demonstrates how the expertise of specialists can be captured to share with other healthcare professionals in a variety of clinical settings.

Acknowledgments

We are thankful to the NCs who have been so supportive of this project. Other supporting team members who have contributed ideas and discussion are Dan Pajek, Curtis Saunders, Dan Ho, Winnie Chen, and Jordanna Kwok. Special thanks to Whynne Caves for her clinical and PDA expertise, Yukari Enomoto for her design ideas, Vincent Ho for doing the bulk of the software development, and Katie Bush for data entry and analysis. We acknowledge the support of the Ontario Ministry of Health and Long Term Care Primary Health Care Transition Fund (Grant no.G03-05372).

References

[1] Canadian Nursing Advisory Committee. Our health, our future: Creating quality workplaces for Canadian Nurses. Final Report of the Canadian Nursing Advisory Committee to Health Canada, 2002.

[2] Lewis JA, and Sommers CO. Personal data assistants: Using new technology to enhance nursing practice. MCN Am J Matern Child Nurs 2003: 28(2): 66-73.

[3] Tooey MJ, and Mayo A. Handheld technologies in a clinical setting: State of the technology and resources. AACN Clinical Issues 2003: 14(3): 342-349.

[4] Garritty C, and El Emam K. Who's using PDAs? Estimates of PDA use by health care providers: A systematic review of surveys. J Med Internet Res. 2006: Apr-Jun: 8(2): e7.

[5] Strayer SM, Reynolds PL, and Ebell MH (eds). Handhelds in Medicine: A Practical Guide for Clinicians. Springer: New York, 2005.

[6] Vicente KJ. Cognitive Work Analysis: Toward Safe, Productive, and Healthy Computer-based Work. Mahwah, New Jersey: Erlbaum, 1999.

[7] Burns C, and Hajdukiewicz JR. Ecological Interface Design, Boca Raton, FL: CRC Press, 2004.

[8] Effken JA, Kim N-G, and Shaw RE. Making the constraints visible: testing the ecological approach to interface design. Ergonomics 1997: 40(1): 1-27.

[9] Greiver M, Drummond N, White D,Weshler J, and Moineddin R, North Toronto Primary Care Research Network (Nortren). Angina on the Palm: Randomized controlled pilot trial of Palm PDA software for referrals for cardiac testing. Can Fam Physician 2005: 51: 382-383.

[10] Rubin MA,Bateman K, Donnelly S, Stoddard GJ, Stevenson K,Gardner RM, and Samore MH. Use of a personal digital assistant for managing antibiotic prescribing for outpatient respiratory tract infections in rural communities. J Am Med Inform Assoc 2006: 13(6): 627-34.

Address for correspondence

Dr. Kathryn Momtahan, RN, PhD
Corporate Coordinator Nursing Research
and Associate Clinical Investigator
The Ottawa Hospital
Nursing Professional Practice Department
1st floor Paterson Education Centre
1053 Carling Avenue
Ottawa, Ontario, Canada K1Y 4E9
Email: kmomtahan@ottawahospital.on.ca

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Improving Systems Interoperability with Model-Driven Software Development for HealthCare

Ståle Walderhaug[a+b], Marius Mikalsen [a], Gunnar Hartvigsen [b], Erlend Stav [a], Jan Aagedal [a]

[a] *SINTEF ICT, Trondheim, Norway*
[b] *Medical Informatics and Telemedicine Group, Dept for Computer Science, University of Tromsø, Tromsø, Norway*

Abstract

An aging population and an increase in chronically ill patients demand teamwork treatment models. To support these with information systems, interoperability is a prerequisite. Model-driven software development (MDSD) with special healthcare extensions can enable reuse of components and improve conformance to international standards. In this paper, a MDSD HealthCare Framework is proposed and demonstrated for homecare services. Using the framework, information systems will improve their conformance to international standards and the interoperability with other systems.

Keywords:

Medical Informatics, home care services, information systems, systems integration

Introduction

In healthcare, the disease burden is changing from acute to chronic care, 35,000,000 people died from chronic diseases in 2005, and 60% of all deaths are due to chronic diseases [1]. New ways of providing care are being evaluated, based on teamwork treatment – demanding support from interoperable information systems. Interoperability in healthcare has been identified as an important area of research and development by many organizations, including the European Union (EU)[1], the Object Management Group (OMG)[2] and other national organizations [2]. The ability to exchange information and share services across departmental, organizational and national borders can reduce the administrative overhead and costs [3], and as a result improve the effectiveness of healthcare provided. Consequently, more patients can be treated faster with the same amount of (care) resources. A sustainable healthcare infrastructure depends upon interoperable health information services [4, 5].

The treatment and management of homecare consumers, typically elderly, chronically ill and cognitive disabled, require a coordinated effort from healthcare and social welfare services. To effectively support these care services with information systems, interoperability of core information such as patient careplan calendar and medication-list is a prerequisite.

To improve interoperabity between systems, the leading standardization bodies in healthcare information, HL7, CEN TC251 and OpenEHR, have specified standards that address systems architecture and information exchange. Although these standards have been available to the Health Information Systems (HIS) vendors for some time, they have not fully adopted them into their products. Thus, the different HIS are not interoperable, requiring the development of software adapters to be able to exchange information about the patients. There is an urgent need for a standardized interface and method to realize this information exchange.

The standardization bodies provide limited tool support to the developers of health information systems. To incorporate standard healthcare concepts in the systems' design, an operational software engineering artefact that provides both semantic and syntactic interoperability functionality [6, 7] should be available for the system architects and developers [8-10].

In 2002, the Object Management Group (OMG) introduced the Model-Driven Architecture (MDA) [11], an approach focusing on using models (e.g., UML models [12]) as first-class entities in the development of software systems. In practice, this means that the models are used directly in the implementation of an information system, either as system blueprints or as input to code generation engines that produce executable code. MDA is the most known model-driven software development (MDSD) approach, and the overall idea is to separate business functions (in Platform Independent Models - PIM) from its technological implementations (in Platform Specific Models – PSM), enabling code generation and reuse of components. The overall benefit is improved interoperability and reduced development time and cost.

Using a MDSD approach in the development of healthcare information system services could facilitate the use of standards through specification of reusable standards-based PIMs. Advanced UML mechanisms such as Profiles

1 EU Life sciences, genomics and biotechnology for health website: http://cordis.europa.eu/lifescihealth/home.html
2 Object Management Group (OMG) website: http://www.omg.org

and Patterns could be used to further extend the expressiveness of the modeling language and force the use of standardized healthcare concepts. As a result, the developed systems will increase the level of interoperability, and at the same time development and maintenance costs will decrease.

With an aging population and a rapidly increasing number of chronically ill patients [1], the need for teamwork treatment is crucial. Healthcare Information Systems (HIS) can no longer be seen as standalone systems, but need to interoperate in a health network [10]. This leads to the problem statement: *How can health information systems development be improved to ensure that systems involved in a homecare teamwork treatment infrastructure can share information in an effective and sustainable manner?*

This paper proposes a model-driven software development framework with standards-based healthcare extensions as a tool to achieve interoperability between HIS. The healthcare focus is on homecare services although the healthcare standards discussed have general applicability. The paper concludes that MDSD with the appropriate healthcare information extensions can improve software's conformance to standards and thus also the ability for caregivers to share information in teamwork treatment.

Following next is an overview of the challenges that are associated with developing such a MDSD Healthcare framework, both from a software engineering and healthcare viewpoint. Then the framework is presented along with an example from the homecare domain, before the paper concludes with a discussion of the validity of our results and directions for future work.

Immature MDSD tools and need for evaluations

In a keynote talk at the 2006 ECMDA-FA conference in Bilbao (Spain), Bran Selic (IBM) advertised for rigorous scientific studies that investigate how MDSD can improve the development process[3]. Recently, the ModelWare project[4] conducted five different scientific MDSD evaluations. A summary of the evaluations is presented in [13] and concludes that by applying MDSD, a productivity gain of 20% can be expected and the quality of the software produced would increase.

Despite these and other reports, there is a considerable skepticism in the software engineering community about the performance and usability of MDSD. The skepticism is based on three main points: 1) the UML is too generic and is conceptually too far from implementation languages making it difficult to generate efficient and fully executable code [14, 15], 2) The maturity of MDSD tools: transformation tools are not complete enough to provide return of the investment put into developing reusable UML models. E.g., the Query/View/Transformation (QVT) standard [16] by OMG does not have good tool support and 3) standards are used in different versions, some of which are not interoperable.

3 ECMDA website: http://www.ecmda-fa.org/

4 MODELWare (FP6-IP 511731) project website: http://www.modelware-ist.org

Many systems, many standards

The use of information standards to improve interoperability between information systems in the healthcare domain is not straight-forward. In a single healthcare organization, there is a plethora of information systems, each based on one or more information standards. In the context of systems development, sharing of information and services between these systems need to address the following issues: 1) Many systems (such as patient administrative systems) are dated back to the late eighties, long before the specification of today's information standards, 2) Department specific systems developed to serve one specific purpose do often not use international standards nor follow best-practice in systems architecture, 3) The information systems themselves and the information standards used are continuously being upgraded [7].

A MDSD framework for healthcare

The work presented in this paper build upon three assertions presented in the following.

Assertion 1: Model-Driven Software Development with healthcare information standards support will improve interoperability between health information systems (compared to the traditional way of developing systems)

UML allows for extensions through the use of UML Profiles. A profile defines stereotypes, tagged values and constraints that can be assigned to modeling elements in the design process. The main purpose of a profile is to extend UML's expressiveness for a certain domain, e.g. healthcare. By providing healthcare specific UML profiles and patterns as a part of a MDSD framework for healthcare, concepts defined in international healthcare information standards can be automatically built into the information systems. A healthcare profile can be used by transformation templates and code generators to explicitly implement attributes, relationships, operations and objects that provide interoperability services.

Assertion 2: Healthcare Information Standards are appropriate as reusable model-driven development artefacts.

Standards from HL7, CEN TC251 and OpenEHR make use of UML class diagrams to specify concepts and relationships. However, parts of the semantics are described textually as constraints-comments to the formal UML models. To be able to correctly incorporate these standards into model-driven development artefacts such as UML Profiles, the complete semantics of the standards must be possible to represent formally. The correctness and reusability of the models created with the UML profile will depend on the mapping between the standard and the UML profile artefacts.

Assertion 3: Healthcare information services in the homecare domain can be reused across organizations.

The usefulness of a MDSD Healthcare framework for the development of interoperable homecare services will depend on the ability to define functional and coherent information services in the domain. The services need to be reusable beyond departmental and organizational bor-

ders, preferably also national borders as some healthcare institutions have rehabilitation and treatment centers abroad, often collaborating with the local healthcare services.

Results

Using a model-driven approach such as the MDSD Healthcare Framework enables rapid development of interoperable healthcare information systems. The framework includes a set of UML profiles, models and experience reports from the homecare domain, but with generic healthcare service applicability.

Example of MDSD healthcare framework in homecare

A trivial example is provided to demonstrate how a UML Profile for healthcare can be used in the development process to achieve interoperability between information systems.

The example service is a CarePlan service where a HomeCare Center System and a General Practitioner (GP) EHR HomeCare extension can access and update the homecare patient's careplan. Both systems will need to provide a defined interface for information exchange based on the same standard. A small subset of the "CarePlan" concept in the Continuity of Care (CONTSYS) [17] standard is used for demonstration (Figure 1). A "CarePlan" is applied by one or more HealthCare Professional and addresses one or more health issues that the Subject of Care has (relation not shown).

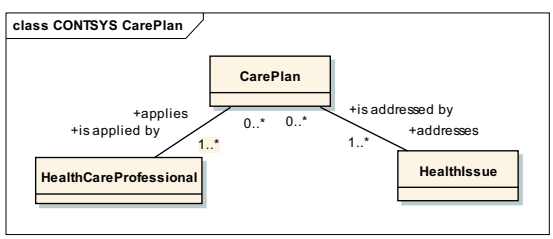


Figure 1 - A subset of the CONTSYS CarePlan concept

The goal is to develop Java based (sub-) systems that allow exchange of careplan information for the homecare patient according to the CONTSYS standard.

A simple UML profile for homecare

Based on the CONTSYS standard, the following UML extensions are specified: 1) UML Class Stereotype: SubjectOfCare: The person receiving treatment, 2) UML Class Stereotype: CarePlan: The treatment plan for one or more health issues (problem), 3) UML Class Stereotype: HealthCareProfessional: A caregiver entitled to provide care, 4) UML Association Stereotype: HealthCareProfessional_isResponsible: The healthcare professional (source element) is responsible for the target element and 5) UML Association Stereotype: SubjecfOfCare_Owns: Subject of Care (source) has owner right of the target element.

Two *tagged values* are defined: 1) Boolean: isShared: when used with a CarePlan, stating whether the careplan is shared or not and 2) Boolean: isOrganDonor: used with a SubjectOfCare to state if the person is organ donor or not.

The healthcare information systems

The two systems are being developed independently by different vendors using the same CONTSYS-based UML profile. The Care Center system platform independent model (PIM) shown in Figure 2 shows that the HomeCarePlan (stereotyped CarePlan) is related to the HomeCarePatient (owned by), the Doctor (under responsibility of) and the Visiting Nurse. All classes are stereotyped according to CONTSYS. As a result, the HomeCarePatient has a tagged value for "isOrganDonor" and the HomeCarePlan has an "isShared" tag.

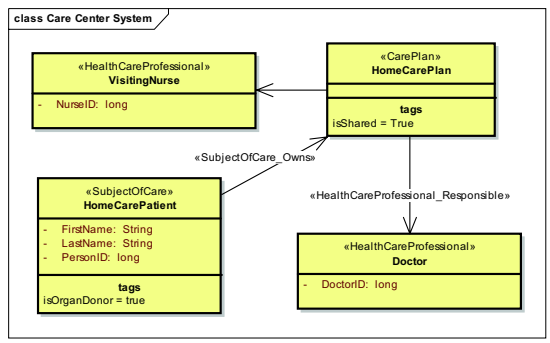


Figure 2 - The Care Center System PIM

The PIM for the GP EHR Homecare extension system (Figure 3) shows that the TreatmentPlan ("CarePlan") elements are related to one or more patient problems ("HealthIssue") according to a problem-oriented EHR [18]. This can be used to filter out treatment activities that are not related to the coordinated care of a homecare patient.

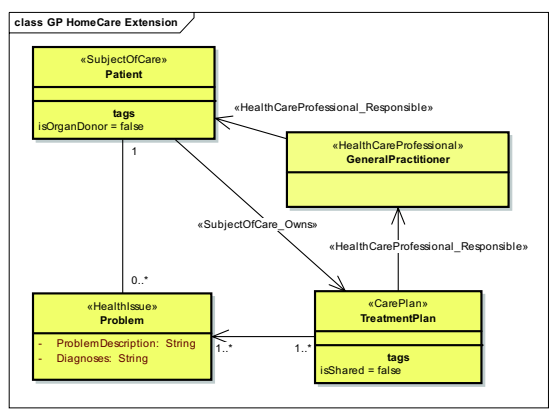


Figure 3 - The GP EHR Homecare Extension PIM

The two PIMs can be transformed to Java Platform Specific Models (PSM) using a CONTSYS-based transformation script for Java. This script utilizes the stereotypes and tagged values in the transformation process to add attributes and operations to ensure that the

required interoperability mechanisms are implemented. In this trivial example, only set and get operations for the tagged values and careplan elements are created. The Java Model for the Care Center system (Figure 4) and the GP EHR Homecare extension (Figure 5) show that during the transformation process, three operations have been created on the CarePlan-stereotyped classes. These operations, stereotyped with "CarePlan", enables exchange of CarePlan elements and retrieval of all HealthCare Professionals that are related to the CarePlan.

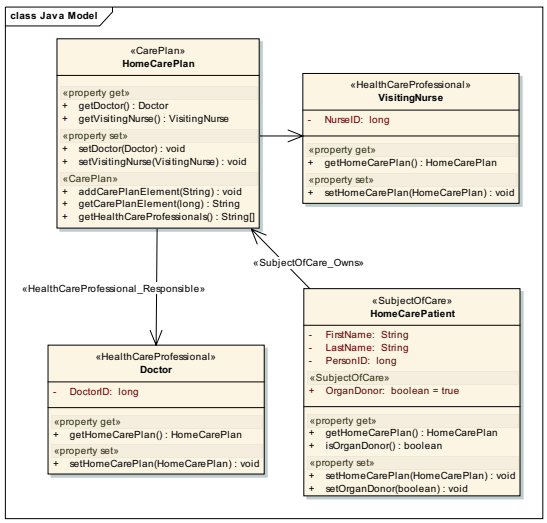


Figure 4 - Java PSM for the care center system

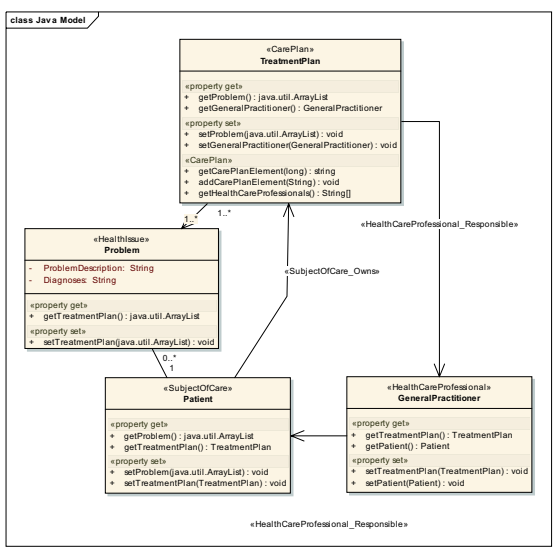


Figure 5-Java PSM for the GP EHR HomeCare Extension

From these Java PSMs, code can be generated using a standard code generation tool based on e.g. QVT [16].

To summarize: using the CONTSYS UML Profile in the design and development of the careplan service in the Care Center and GP EHR systems ensured that the services are conform to the standard and thus can exchange information correctly.

Discussion

The MDSD Healthcare framework proposed in this paper addresses the need to make information systems in the healthcare domain interoperable and sustainable. To achieve this, the framework provides tools and reusable components that incorporate international information standards into the information system design.

The effect this will have for the future healthcare information systems relies on the three assertions described in the first part: 1) the quality of artefacts artifacts produced from the framework, 2) the ability to map information standards to useful UML profiles and 3) the identification of reusable services.

The quality of the software produced by the framework will depend on the tool support and the developer. The main improvement compared to traditional software development lies in the built-in healthcare tool support, where use of healthcare UML profiles, reuse of existing platform independent models and use of code generation will reduce roundtrip time and improve the quality of the code.

The example showed a simple careplan service where a concept from CONTSYS was applied. More complex services will need more concepts, maybe from more than one standard. The MDSD framework will provide UML Profile support for the most used healthcare standards and patterns for the most recurring concepts. A modular design, in line with Beale's *archetype* concept (7), provides scalability and maintainability of the models as the standards are updated or extended. The *Archetypes* being specified in both CEN TC251 EN13606 [19] and OpenEHR, can be used by the MDSD framework as reusable models and patterns. An archetype is a model of a healthcare concept, and is represented formally using UML.

The specification of reusable services in the healthcare domain is in accordance with Service-Oriented Architecture (SOA) [20]. Many healthcare organizations are adapting SOA as the core enterprise architecture, using a message-oriented middleware with HL7 to exchange information between systems. The process of transitioning to a SOA architecture is expensive, but a fully interoperable healthcare infrastructure would reduce coordination expenses dramatically [3]. Homecare services are likely to be a part of this enterprise service architecture connected through a health network [5]. SOA-based homecare system services can enable independent development and deployment of new patient monitoring and surveillance services in the health network. A SOA based infrastructure will allow sustainable development of healthcare services.

A critical aspect when introducing new development tools and techniques is to evaluate its effect. Proper scientific methods must be applied to achieve rigor. A complete medical informatics solution should not only evaluate the artefacts isolated, but also study their effect in a real

environment [21]. The MDSD HealthCare Framework will be subject for two scientific experiments with real users in the M-Power project[5].

Future work

The framework proposed in this paper is a part of the work being done within the M-Power and Linkcare projects[6]. These projects will identify and develop reusable homecare services for the provision and coordination of homecare services. Using the first version of the HealthCare MDSD framework, some of these services will be evaluated in 2007 and 2008.

Conclusion

With an aging population and dramatic increase in chronic diseases [1], systems interoperability in the healthcare domain is of utmost importance in order to maintain the service level of today and support teamwork treatment. One way to improve interoperability is to ensure the healthcare information systems' conformance to international standards.

The Healthcare MDSD framework will incorporate standards into the development process of information systems, and as a result improve interoperability. The MDSD framework will be evaluated in two experiments in 2007 and 2008 as a part of the LinkCare and M-Power projects. These projects have a strong focus on treatment and management services for chronically ill, elderly and cognitive disabled. This will ensure the framework's relevance for the domain.

References

[1] World Health Organization. Preventing CHRONIC DISEASES -a vital investment: World Health Organization; 2005.

[2] Norwegian Ministry of Social Affairs, Norwegian Ministry of Health. Te@mwork 2007 - Electronic Interaction in the Health and Social Sector: Directorate for Health and Social Affairs; 2004.

[3] Walker J, Pan E, Johnston D, Adler-Milstein J, Bates W. D, Middleton B. The Value of Health Care Information Exchange and Interoperability. Health Tracking. 2005 January;5(10).

[4] Brailer D. Interoperability: the key to the future health care system. Health Affairs vol:Suppl Web Excl-19 2005.

[5] Beyer M, Kuhn KA, Meiler C, Jablonski S, Lenz R. Towards a flexible, process-oriented IT architecture for an integrated healthcare network. Proc. of the 2004 ACM symposium on Applied Computing. Nicosia, Cyprus: ACM Press; 2004.

5 MPOWER homepage: http://www.mpower-project.eu

6 LinkCare homepage: http://www.linkcare-bcn.org

[6] Park J. Information systems interoperability: What lies beneath? ACM transactions on information systems. 2004;22(4):595.

[7] Beale T. Archetypes: Constraint-based Domain Models for Future-Proof Information Systems. OOPSLA 2002 Workshop on behavioural semantics; 2002; Portland, Oregon, USA; 2002.

[8] Kuhn KA, Lenz R, Elstner T, Siegele H, Moll R. Experiences with a generator tool for building clinical application modules. Methods of information in medicine. 2003;42(1):37-44.

[9] Lenz R, Beyer M, Kuhn KA. Semantic integration in healthcare networks. International journal of medical informatics. 2007 Feb-Mar;76(2-3):201-7.

[10] Lenz R, Kuhn KA. Towards a continuous evolution and adaptation of information systems in healthcare. International journal of medical informatics. 004;73(1):75-89.

[11] Miller J, Mukerji J. MDA Guide Version 1.0.1: Object Management Group (OMG); 2003 2003-06-13. Report No.: omg/2003-06-01.

[12] Object Management Group (OMG). UML 2.0 Superstructure FTF Rose model containing the UML 2 metamodel: Object Management Group (OMG); 2005.

[13] Hartman A. Industrial ROI, Assessment, and Feedback - Master Document: IBM Haifa Research Lab; 2006 September 15, 2006. Report No.: D5.3-1.

[14] Fowler M. UML Modes. 2003 [cited 2006 November 14]; Martin Fowler's homepage]. Available from: http://www.martinfowler.com/bliki/UmlMode.html

[15] Mellor SJ. Executable UML: A Foundation for Model-Driven Architecture; 2002.

[16] Object Management Group (OMG). MOF QVT Final Adopted Specification: Object Management Group (OMG); 2005.

[17] CEN TC251. EN 13940-1: Health Informatics - System of Concepts to Support Continuity of Care - Part 1: Basic Concepts: European Committee for Standardization; 2006 September 26.

[18] van Bemmel JH, Musen MA. Handbook of medical informatics. 1 ed: Springer Verlag; 2002.

[19] CEN TC251. EN 13606-1-4: Health informatics - Electronic healthcare record communication Part 1-4: European Committee for Standardization; 2000 May.

[20] OASIS Open. Reference Model for Service Oriented Architecture 1.0: OASIS Open; 2006 August 2. Report No.: soa-rm-cs.

[21] Friedman CP. Where's the science in medical informatics? Journal of the American Medical Informatics Association. 1995;2(1):65-7.

Address for correspondence

Ståle Walderhaug, University of Tromsø, Department of Computer Science, Medical Informatics and Telemedicine group, 9037 TROMSØ, NORWAY. Telephone: +47 90766069, Fax: +47 77644580, email: stale.walderhaug@sintef.no

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Conformance Testing of Interoperability in Health Information Systems in Finland

Tanja Toroi[a], Anne Eerola[a], Juha Mykkänen[b]

[a] *University of Kuopio, Department of Computer Science, Finland*
[b] *University of Kuopio, Information Technology Centre, HIS Research and Development Unit, Finland*

Abstract

Conformance testing has been studied extensively but the current state of the art varies very much in different software companies in the healthcare domain. In this paper we present findings of a survey of conformance testing of interface specifications in the healthcare domain. The survey was conducted among software companies and their customers, i.e. hospital districts, in Finland. The findings of the survey show that the quality of the applications should be improved, that there is a reluctance to use external interface testing services, and that software vendors do not adequately fulfill customers' requirements in relation to testing. However, conformance testing is considered to be very important and should be improved. The main contribution of this study is the analysis of the findings and we give some recommendations for healthcare organizations, software companies, and authorities to improve the conformance testing related to the interoperability of the applications.

Keywords:

Software Validation, Quality Assurance, Health Care, Conformance Testing.

Introduction

The challenges of integration, distribution, security, and interoperability complicate the software development and testing [1]. Fortunately, the challenges can be met by means of standards, conformance testing, and the certification of systems. Conformance is defined by the ISO/IEC as the fulfillment of a product, process or service of specified requirements [2]. The requirements or criteria for conformance must be specified in the standard or official specification. After traditional testing phases (e.g. module, integration, and system testing), conformance testing is needed. In relation to interoperability, the goal of conformance testing is to assure that applications follow interface specifications so that the application integration and introduction can be done without extra adaptation or development work. Conformance testing covers testing of the interfaces between the applications, and their conformity to standards, not testing of the inner parts of the application.

Software developers are encouraged to use common standards and implement open interface specifications. If the applications have open standard-based interfaces, their interoperability is improved, introduction and integration become easier, and less local adaptation work is needed. But, in order to successfully assure software interoperability of networked applications, such as in healthcare networks, requirements for testable specifications have to be developed, too [3].

Standardization, certification, and conformance testing have been studied extensively in the telecommunication domain [4] but the practices are not well-established in other domains. Only a few studies on conformance testing in the healthcare domain [5, 6] and certification of health information systems [7, 8] have been reported. No empirical studies on the level of conformance testing in the software companies and in their customers' organizations in healthcare have been reported. However, the current level of testing has to be studied and understood before it can be improved. Therefore, we conducted a survey to find out how much and how often conformance testing is performed in relation to the interface specifications in the healthcare domain in Finland. We also wanted to find out how customers perform testing in their organizations, and how they can influence conformance of the software they acquire. In addition, we wanted to study general testing practices and test process improvement in the healthcare organizations. The results of general testing practices and test process improvement have been presented earlier [9]. In this paper the focus is on conformance testing of open interfaces.

Health information systems (HIS) aim to support a high-quality and efficient patient care [10]. The healthcare domain and HIS in Finland face big challenges which will affect testing in the near future, such as the introduction of the national electronic patient records, and many old, monolithic legacy systems which have to interoperate with new applications in the network of organizations. The other special characteristics in healthcare are high quality requirements, the safety criticalness of the systems, and non-deterministic and non-predictable processes, which often cross organization borders. Thus, efficient quality assurance and improved conformance testing processes are needed.

The rest of the paper is organized as follows. First, the research method, sample, and questionnaire of the survey are introduced. Second, the results of the survey are pre-

sented. Then, we analyze the results and make some recommendations for the further development of conformance testing. Finally, conclusions and future work are discussed.

Materials and methods

Research method

The study presented in this paper is based on empirical and constructive research. Our aim was to study conformance testing and interoperability issues using a survey research method. In addition, we wanted to make interoperability recommendations based on the survey results. A WWW-based questionnaire, with open and structured questions, was used as a survey instrument. The WWW-based questionnaire was easy and cheap to deliver to many respondents. However, we were aware of the fact that people may report higher incidences of, for example, the use of standards in self-reports than when objectively reported by outside parties. Therefore, we asked that the questionnaire be passed on, if possible, to more than one person in the same organization. We sent the participants an e-mail and asked them to respond to the survey. After a month we phoned those who had not responded and reminded them to respond.

The main research questions for healthcare application vendors and their customers were:

1. How can customers, i.e. hospital districts, influence the conformance of the software they acquire?
2. How much conformance testing do software vendors perform and how often are standards used?
3. What are the main goals of, problems with, and opinions about external interface testing services?

Sample

Conformance testing procedures concern both software vendors and their customers. Customers can require standard-based applications in calls for tenders. The expected benefits from the use of standards include better interoperability and quality of software. On the other hand, vendors can prove customers conformity of software implementations to the standards. Thus, the survey was delivered to the both groups.

The sample covered all the Finnish hospital districts, covering geographically the whole country (except _land, a self-governing territory). In addition, the sample covered those healthcare software companies which belong to a Health and Welfare IT R&D cluster in Finland. The cluster covers most of the Finnish healthcare software companies. The questionnaire was delivered to chief information officers or other leaders of each organization, and they were asked to pass it on to those responsible of testing (testers, test managers, testing engineers, etc.).

The questionnaire was sent to 47 organizations, of which 27 were software companies and 20 were hospital districts (see Table 1). Responses were obtained from 29 organizations, of which 14 were software companies and 15 hospital districts. The response rates were 52%, and 75%, respectively. Companies completed on an average of 94% of their questions (completion rate), while customers completed on an average of 91% of theirs. The companies were operating mainly in the domestic market but eight had both national and international customers.

The respondents in the software companies (henceforth *developers*) were testing engineers, designers, project managers, managers, and consultants. In the hospital districts the respondents (henceforth *customers*) were chief information officers, project managers, and designers. In addition, one senior physician responded to our survey.

The non-respondents were asked by phone why they did not respond. In most cases the reason was busyness. The other reasons were outsourcing of testing, and testing not being in focus at the moment.

Table 1 - Statistics of the sample

	Sent	**Responded**	**Resp. rate**	**Compl. rate**
Companies	27	14	52%	94%
Hospital districts	20	15	75%	91%
In total	47	29	-	-

Questionnaire

The questionnaire was developed based on the literature and our own knowledge of conformance testing theory and practice.

The questionnaire consisted of three sections: general testing practices [9], conformance testing and standards, and the organization of interface testing services. The results of conformance testing and standards, and the organization of interface testing services are discussed in this paper. The conformance testing and standards part consisted of questions on the rationale for conformance to standards, use of specifications and standards, and conformance testing methods and practices. The organization of interface testing services dealt with the use of external testing services, questions on roles and responsibilities, and comments and ideas for organizing testing services. The questions are shown on http://www.cs.uku.fi/~ttoroi /questionnaire.html.

Results

In this Section we first present the results related to conformance testing of interoperability standards. Secondly, the organization of interface testing services is presented.

Conformance testing and standards

The developers were in small (fewer than 50 employees), medium (50-250 employees) and large organizations (over 250 employees). There were 8 small, 2 medium and 4 large organizations. All the customers were from large organizations.

All the customers had understood the term "conformance testing" correctly, while two developers misunderstood questions concerning conformance testing. They had understood that the questions dealt with the testing of inner parts of the applications, and not the testing of interfaces and interoperability.

It appears that new versions of the applications are installed and conformance retesting is needed between one and four times a year for each application. However, there were significant differences in conformance testing in different software companies. Some developers' the whole system and its specifications were based on standards, while others only tested whether two particular applications could be integrated. Customers had invested in testing, and used more rigorous methods than, for example, small companies did [9].

Some comments by the respondents display a "seller's market" situation. One customer said: "From time to time there is a "take it or leave it" situation and testing does not help much. We know there are errors but we have no other choice than to buy it".

Unfortunately, testing is not an exciting and fascinating topic in software companies at the moment. As one respondent said "We had many good test practices a little while ago but do not have them any more. Our company fired the testing engineers, and does not invest in testing anymore."

The results reveal that standards are widely used in healthcare applications in Finland. All respondents replied that their applications complied with standards (except one customer who did not know whether their applications conformed to standards). The most common standards and official specifications were HL7 (Health Level 7) messages, STAKES (The National Research and Development Centre for Welfare and Health), and IEC (International Electrotechnical Commission) standards. Increasing the use of standards and official specifications was also considered to be very important.

The most often used interface specifications in the healthcare domain are natural language specifications. However, we noticed that specifications based on UML (Unified Modeling Language) and XML (Extensible Markup Language) are needed, too. In addition, one developer wanted that requirements by the authorities should be in a format that could automatically be transferred into a requirement management system.

The main sources of information concerning the standards were information provided by standards associations and other public seminars. However, information was surprisingly often acquired by rumors and surfing the Internet, too.

Developers considered recommendations by authorities to be the most desirable means to promote the use of standards. The other means were customers' requirements, laws, and commitment by management. Customers considered recommendations by authorities, common agreements, laws, and certification to be the most desirable means. One customer stated that customers also have to be more active and demand the use of standards. The respondents were divided over whether laws are a good means. Some developers claimed that no standard is implemented without legislation. Others stated that excessive legislation only delays development and innovation.

The most common problems in the use of standards were the application and interpretation of standards, national specifics which differ between countries, and the laboriousness of implementing of standards. In addition, customers ordered applications which referred to incomplete or non-official standards. Some developers stated that, when interpreting standards common solutions should be made, not tailored ones, and it is a waste of time if standards are not international. Unclear responsibilities related to standardization on a national level were also considered problematic.

Organization of interface testing services

The respondents were suspicious of an external interface testing environment. Only 9% of the responses concerning willingness to use external testing services were 'yes'; the rest were 'no' (21%), 'maybe' (38%), and 'I do not know' (32%). No respondent was willing to pay for testing services.

Respondents were divided over the responsibility questions. They responded, almost evenly, that some national actor or separate software companies should be responsible for the external testing service, but the authorities should be responsible for interoperability certification. However, certification was not considered very necessary nor useful. This view was captured by one developer who stated that "Certification does not help. Software companies must be responsible for their own products." Another developer observed that external testing of medical devices in the test labs has been organized very well, but it does not function for software code.

Respondents considered that external interface testing services may not be profitable, and may become a bottleneck. In addition, the respondents thought that external testing services cannot help in integration, and applications can be truly confirmed only by integrating them in the real local environment.

It was very surprising that respondents were not familiar with the international development of conformance testing or certification of interoperability. Only one developer was acquainted with IHE (Integrating the Healthcare Enterprise) and CCHIT (Certification Commission for Healthcare Information Technology).

Discussion

In this Section we analyze the findings of the survey and make recommendations for healthcare organizations, software companies, and authorities to improve conformance testing, and compliance to interoperability standards.

R1: Perform interoperability conformance testing more rigorously

It was no surprise that too little time was allocated for testing both within the software companies and in the customer organizations. It is worrying that even though time is not allocated to testing on either side, and customers are not satisfied with the applications they acquire, still so little is being done. If the customers do not demand more reliable and interoperable applications, developers will not invest time and effort in quality assurance and testing. Software developers and their customers should both participate in the testing process as responsible parties and improve testing in mutual understanding, as equals.

Interoperability of applications requires various conformance testing activities. Besides interoperability standards, conformance testing can also be performed against any kind of specifications, for example, if the software processes are ITIL compatible or if the applications conform to the generic requirements of electronic patient record systems. It can be concluded that more conformance testing must be performed and it must be more diversified and disciplined.

R2: Utilize open interfaces and use interface testing services

Open interfaces promote the market penetration by small software companies. However, this can be against the interest of large software companies, because they try to protect their market positions. In healthcare applications, open external interfaces between different applications and different organizations are used increasingly. Proper interface testing services or environments promote cost-effective interoperability testing. When open testing environments are used, it is not necessary for everyone to develop their own test environments. Everyone can select the counterparts of the application interfaces they need, and the correct versions, from a well-established testing service, and interoperability testing can be performed against them, simulating the real environment. This is safe, cost-effective, and saves resources. However, it requires that a neutral or official stakeholder is responsible for the development and maintenance of the testing environment. In addition, organizing and keeping the testing service in balance should be subsidized by public resources.

R3: Provide proper skills and knowledge

Increasing the use of standards was considered important. To achieve this, customers must be provided with the knowledge and skills required to demand standard-based systems. They must be educated about the benefits of standard-based interfaces, as well as certified and interoperable software products. This does not prevent free competition: it promotes standardization in widely-used interfaces and reduces local “fixes”.

Conformance testing activities varied very much in different companies. All of the respondents reported that they test the conformity of the applications to interoperability standards, but by "conformance testing" some respondents actually meant code testing or introduction testing by customers. The relation of requirements specification and process descriptions to the conformance testing were not clear. This could be noticed from responses, such as "Conformance testing is done through checking and looking at the code" or "Conformity is tested in the customers' environment". The knowledge of conformance testing must be increased through open seminars and training. What conformance testing covers and what it does not cover must be clearly stated.

R4: Enforce the recommendations by authorities

Both vendors and their customers were in favor of official recommendations. These recommendations may be related to interoperability, usability, security, and appropriateness of the applications to the customers' workflow. From the results of the survey it can be concluded that official recommendations and laws are needed to promote interoperability and quality of the applications. Therefore, recommendations by authorities have to be introduced and enforced more actively. In addition, the implementation of the recommendations must be supported, facilitated with tools, and supervised.

R5: Organize neutral testing services

On the basis of our study, there were little market possibilities for commercial interface testing services. On the one hand, the low level of interest in interface testing services is understandable since it adds one extra layer to the testing process. This adds complexity, costs, and can become a bottleneck. On the other hand, versions of the applications can be changed several times a year, and conformance retesting is needed. Thus, conformance testing must be as automatic as possible.

Studies and projects are needed, in which interface testing services are organized in a neutral manner. The organization of testing and certification should be mandated to a national institute. Companies competing with each other cannot easily establish an external testing service. In Finland, the selection of the Social Insurance Institution as a national actor for health IT and the increased coordination by the Ministry of Social Affairs and Health may alleviate the confusion related to standards and conformance testing. The increasing standardization and use of open interfaces will also help in developing external testing services.

R6: Reuse testing experiences and information

Testing experiences and information should be transferred between organizations. One feasible idea is to share testers' tacit knowledge using testing expert groups, which work independently of software vendors. This is an extension of the mentor activity performed in hospital districts, in which the introduction of the applications has been distributed between different hospital districts and the experiences from one district are utilized in other districts.

R7: Improve the testability of specifications and standards

Testing against inaccurate and ambiguous specifications is impossible. Therefore, in addition to conformance testing, the specifications and standards also have to be improved,

and UML and XML must be utilized besides natural language, to achieve accuracy and consistency [3]. Successful utilization and testing of standards requires accurate technical implementation guidelines and constraints to generic standards, such as HL7 specifications, which can be complemented with IHE integration profiles.

Conclusion

In this paper we present a survey of healthcare software companies and their customers. The objective was to survey conformance testing related to interoperability in software companies and their customer organizations in the healthcare domain, and to identify necessary improvements.

Conformance testing and quality improvement require customer demand, investments, and standard-based applications. However, it appears that there are various necessary testing improvements, including attitude problems to be solved before applications conformity to standards and interoperability can be tested. In addition, it will take quite a long time before certification of the interoperability of healthcare software becomes common in Finland. This is due to resistance to change, a fair amount of work needed to introduce standards, the immaturity of standards, and difficulty in reaching a mutual understanding between different stakeholders.

The main contribution of this study is the recommendations based on the findings of the survey. They are meant for the organizations and authorities who aim to improve conformance testing activities of interoperability between applications. Among other things, the requirements by authorities have to be enforced more actively, customers must be educated to demand standard-based applications, and projects are needed to elaborate the idea of external interface testing services.

The timing of the survey in Finland was quite favorable in relation to the national Electronic Patient Record, which increases pressure for quality assurance. On the other hand, there is a slight downtrend in software business in view. In a downtrend it is important to increase turnover and quicken the deployment, whereas quality improvement and software testing are emphasized in an uptrend.

In healthcare, many different applications have to communicate and interoperate with each other. Processes cross organization borders and are often non-deterministic. In addition, many systems are safety critical. Since the study was conducted in the healthcare domain in Finland its results can not be directly generalized without considering domain and market differences.

This study provides empirical knowledge about conformance testing in healthcare. It also discusses the viewpoints of software developers and their customers and contributes to conformance testing research.

The recommendations of this study suggest improvements to the conformance testing of interoperability. They also highlight the need for the development of testing environments and collaboration models for conformance testing. The effects of the recommendations can be surveyed using a repeated survey.

Acknowledgments

This paper is based on research in the OpenTE and SerAPI projects, funded by the National Technology Agency TEKES, software companies and hospital districts. The authors thank all the participants involved in these projects.

References

[1] Siewiorek DP, Chillarege R, and Kalbarczyk ZT. Reflections on Industry Trends and Experimental Research in Dependability. IEEE Transactions on dependable and secure computing 2004: 1(2): 109-127.

[2] ISO/IEC Guide 2: 1996 Standardization and Related Activities: General Vocabulary, 1996.

[3] Toroi T, and Eerola A. Requirements for the Testable Specifications and Test Case Derivation in Conformance Testing. In: Dasso A, and Funes A, eds. Verification, Validation and Testing in Software Engineering. USA: Idea Group Inc., 2006; pp. 118-135.

[4] ITU-T Recommendation X.290. OSI Conformance Testing Methodology and Framework for Protocol Recommendations for ITU-T Applications - General Concepts, 1996.

[5] Chronaki CE, and Chiarugi F. OpenECG: Testing conformance to CEN/EN 1064 standard. 5th European Symposium on BioMedical Engineering. Patras, Ellas. 2006.

[6] Chronaki CE, Chiarugi F, Sfakianakis S, and Zywietz C. A web service for conformance testing of ECG records to the SCP-ECG standard. Computers in Cardiology, 2005: 32: 961-964.

[7] De Moor G. Quality Labelling and Certification of Electronic Health Record systems (EHRs). EuroRec Annual Conference. Geneva Switzerland. 2006.

[8] Leavitt M. Leavitt surveys road to interoperability. Modern Healthcare online. October 30, 2006. URL: http://www.modernhealthcare.com/apps/pbcs.dll/article?AID=/20061030/FREE/61030014/1029/newsletter02

[9] Toroi T. Improving Software Test Processes. Proc. 5th International Conference on Software Methodologies, Tools and Techniques. Quebec Canada. 2006.

[10] Haux R. Health information systems - past, present, future. International Journal of Medical Informatics, 2006: 75(3-4): 268-281.

Address for correspondence

Tanja Toroi,
University of Kuopio, Department of Computer Science,
P.O.B 1627, 70211 Kuopio, Finland,
Tanja.Toroi@cs.uku.fi

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Connecting Public Health and Clinical Information Systems by Using a Standardized Methodology

Diego M. Lopez, Bernd G.M.E. Blobel

eHealth Competence Center, University of Regensburg Medical Center, Germany

Abstract

To meet the challenge for efficient, high quality and sustainable care, health systems in developed and increasingly in developing countries require extended communication and cooperation between all principals involved in citizen's care. The challenge also concerns supporting information systems, which demand interoperation with public health, bioinformatics, genomics, administrative, governmental, and other sources of data. The paper describes an architecture development methodology for modeling the integration between clinical and public health information systems that harmonizes existent standardized modeling approaches and integrates HL7 domain knowledge. An integration-architecture for information sharing between public health surveillance and clinical information systems is derived demonstrating the feasibility of the proposed methodology. Predominantly, a harmonized process for analysis, design, implementation and maintenance of semantically interoperable information systems based on formal grammars is discussed in some detail.

Keywords:

information systems, public health surveillance, semantic interoperability, architecture, unified process, HL7

Introduction

To meet the challenge for efficient, high quality and sustainable care, health systems in developed and increasingly in developing countries are moving from organization-centered to process-controlled (diseases management) health systems. This process will continue towards person-centered care (favored homecare, patient monitoring, body area networks, etc.) in the future. Such inter-organizational collaboration requires extended communication and cooperation among all principals involved in citizen's care. By that way, health systems become regionalized, nationalized or even internationalized.

Such complex collaboration requirements concern the supporting health information systems (HIS) and in consequence also their development processes. Thus, HIS need to be analyzed, planned, specified, designed, developed, deployed and maintained, bearing in mind the current/future integration with public health, bioinformatics, genomics, financial, governmental and other sources of data.

In particular, information systems serving primary, secondary, tertiary and quaternary care must be interlinked with public health information systems (PHIS). While concepts and rules defining the actual policy in clinical care settings are predefined and rather stable, different policies in an integrated environment have to be contractually harmonized through policy bridging, including procedural agreements such as clinical pathways and public health protocols; applicable, e.g., to chronic and communicable diseases.

Regarding development methodologies, medical care information systems and public health ones are often developed independently, not considering specific information requirements for information sharing. Existing development methods support specification, analysis, design, implementation, maintenance of HIS, but they normally do not consider the complexity of integrated HIS when dealing with interrelated domains such as public health.

The aim of the paper is to propose a methodology for modeling the integration between clinical and public health information systems.

Materials and methods

A key success factor for any HIS development process (requirements analysis, design, implementation, evaluation, use, and maintenance) is the definition of the system architecture in structure and behavior.

At the moment, a comprehensive development methodology meeting all requirements of future proof integrated HIS, i.e., component and service orientation, model-driven approach, flexibility, scalability, semantic interoperability, trustworthiness, etc.; doesn't exist. Such methodology can only be realized by harmonizing existing architectural approaches in software engineering and health informatics.

In software engineering, architectural frameworks (e.g., Zachman Framework [1] and ISO 10746 Open Distributed Processing – Reference Model (RM-ODP) [2]), architectural models (Model Driven Architecture (MDA)[3]), architectural styles (Service Oriented Architecture (SOA)

[4]) and development processes (The Rational Unified Process (RUP) [5]) provide some guidance for software architecture specification. In the healthcare domain, examples for architectural approaches describing an architecture development methodology are the HL7 Development Framework HDF [6], the Service Oriented Architecture for HL7 (SOA4HL7) [7] and the former CEN ENV 12967 [8], which has recently been revised and is currently under balloting. Those approaches, however, are diverse in nature touching different aspects of the system and considering different architectural paradigms.

As an alternative, the Generic Component Model (GCM) [9], developed and enhanced at the eHealth Competence Center Regensburg is a generic framework able to harmonize different architectural approaches, considering all the architectural requirements and processes mentioned. The GCM -which has meanwhile been established in some ISO standards- defines three dimensions for system modeling: the domain, the system's granularity (composition/decomposition), and the system's viewpoints as defined in RM-ODP. GCM is the only approach which allows harmonizing different architectural paradigms through 1) managing different domains thereby allowing different domain languages; 2) starting from the computation-independent business perspective and separating platform-independent (logic) from platform-specific (technology and implementation) aspects; 3) being process-controlled and service-oriented to meet the users' needs; 4) considering semantics; and 4) following ontological principles.

Results

The Architecture Development Methodology

The Architecture Development Methodology for modeling integration in HIS is described in figure 1.

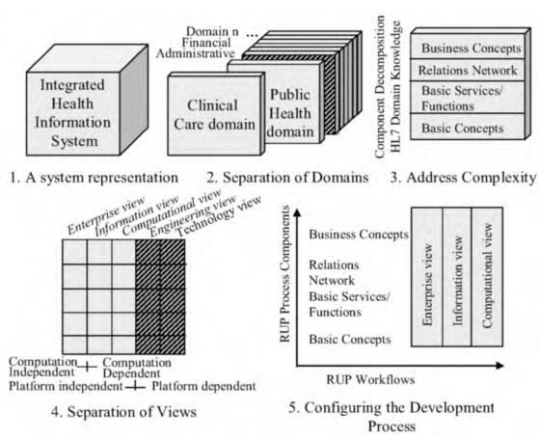


Figure 1 - The Architecture Development Methodology

The process is performed by progressively constraining the GCM. In step 1, the system to be modeled is identified. The real system under consideration is an "integrated health information system". In step 2, the GCM domain dimension is simplified by separating the domains of interest (clinical and public health domains) from other related domains (financial, administrative, security, etc). The third step in the methodology addresses the complexity dimension. All four levels of granularity defined in GCM are analyzed independently for each domain starting bottom up: basic concepts, basic functions, relation networks, and business concepts. This step intends to define the domain knowledge (reference models, reference information models, domain concepts, vocabulary) to enable semantic integration. HL7 domain knowledge was found the more comprehensive set of clinical and public health domain knowledge for the purpose of describing business concepts.

In step 4, the scope of the architecture is defined by restricting the analysis to platform-independent aspects of the system. After analyzing the enterprise architecture as well as describing the computation-independent aspects of inter-related businesses using the business viewpoint (clinical care and public health process), the platform-independent specification of the components' properties is performed considering the information and the computational viewpoint of every single component needed in the next phase of modeling. The resulting specification can automatically be transferred into platform-specific models using MDA transformations covering the engineering and the technology viewpoints.

Finally in step 5, a process guide for architecture development refining the aforementioned principles is defined. Considering that the RM-ODP specification does not prescribe any particular method for building the different viewpoints, the RUP framework is used as process guide. RUP counts as the more comprehensive development process for meeting the methodology requirements, being also configurable to integrate other development methods. An analysis of the concepts defined in the ODP standard was performed, identifying the more adequate RUP process components for inclusion into the method configuration of the Architecture Development Methodology. For each ODP perspective, a set of process components (roles, activities, artifacts ,tools) is identified in order to describe, at different levels of granularity, the HIS components. Figure 2 describes the process components configured for the Architecture Development Methodology.

Modeling the Integrated Colombian Health Information System

The Integrated Health Information System (Sistema Integrado de Información en Salud – SIIS [10]) in Colombia is used as a scenario to analyze the integration between public health and clinical information systems. The aim of SIIS is to manage virtually any kind of health information (administrative, clinical, public health) in a centralized way. In practice, SIIS deals only with administrative and financial information, however.

Health services delivery in SIIS is based on the division between personal and population- based health services, with the former controlled by public and private health

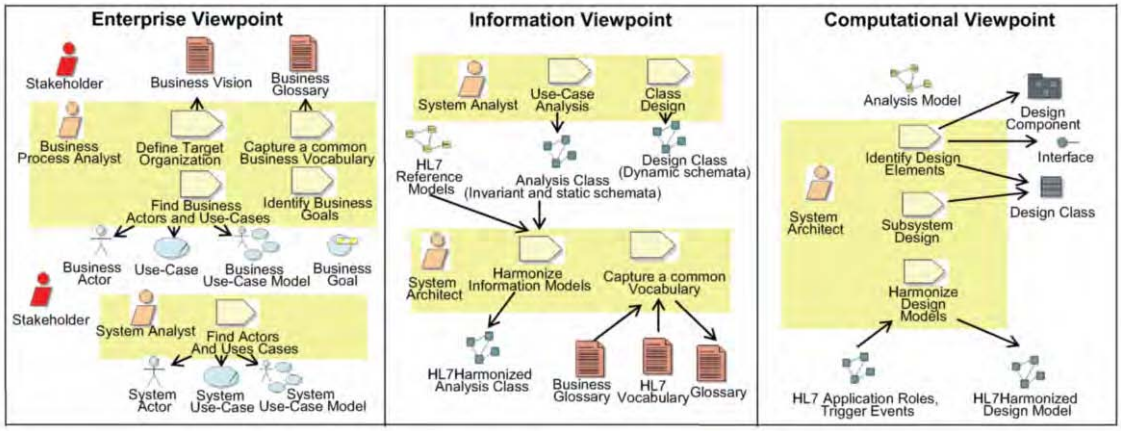


Figure 2 - Architecture Development Process details

care service providers and the latter managed by the government. This suggests the distinction of two groups of information systems: Hospital Information Systems for administrative and clinical information management and PHIS for public health information management (health surveillance, promotion and prevention). Both systems are obliged to share at least some administrative data, but at the moment they run completely independent one from the other. Occasionally they only share plain text files manually.

Figure 3 describes the proposed architecture for the Integrated National Health Information System, in which Public Health and Clinical Information Systems are integrated through an "Integration Subsystem". In the next subsections, the architecture, particularly the Integration Subsystem, is detailed according to the proposed Architecture Development Methodology.

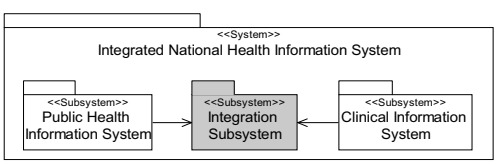


Figure 3 - The Integrated National Information System Architecture

Enterprise viewpoint

The enterprise viewpoint examines the information system and its environment in the context of the business requirements on the system, its purpose, scope and policies. The business process and the boundaries of all associated systems: Public Health and Clinical Information Systems and the Integration Subsystem are analyzed. For the sake of brevity, only two main artifacts in the methodology are detailed, the Business Vision and the Business Use Case Model.

Artifact: Business Vision

Public Health Information System. The target organization is the National Public Health Surveillance System (Sistema de Vigilancia en Salud Pública-SIVIGILA [11]). The system consists of health care institutions, protocols, norms and resources organized with the objective of supporting the systematic and ongoing collection, analysis, interpretation, delivery and evaluation of health events necessary to support health promotion and prevention programs. The system's external actors are: health services providers, health services administrators (EPS), and any other organization interested in public health information for decision making e.g. Pan-American Health Organization (PAHO):

Clinical Information Systems. The target organization is the health care service providers (IPS) network in Colombia. IPS are public and private institutions (hospitals, clinics, laboratories, primary care establishments) offering individual health care services including diagnosis, treatment, control, rehabilitation and even supporting promotion and prevention programs. The system's external actors are: Population, Ministry of Health, territorial public health authorities and EPS.

Artifact: Business Use Case Model

Integrated Public Health Information System. The business processes supported by the integrated information system are described in Figure 4.

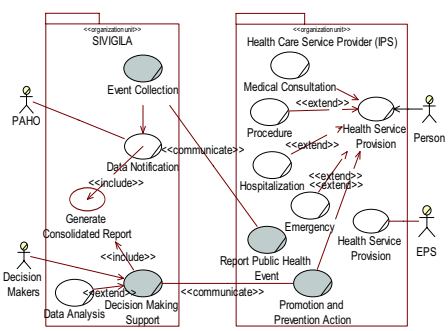


Figure 4 - The Integrated Public Health Information System Business Use Case Model

After analyzing and documenting the business process and rules of the two interacting systems; the interacting business processes are identified. The interacting use cases (colored in grey) constitute the business process identified for the Integration Subsystem.

Information viewpoint

The Information viewpoint describes the static and dynamic structure of information managed by the distributed information system to be specified, that is the "Integration Subsystem". The information components are derived from the enterprise viewpoint models and the functionality described in the national health systems policies. The process components below describe the system Analysis Class Model (RM-ODP Invariant Schemata).

Artifact: Analysis Model

Figure 5 describes the Integration Subsystem main informational components. The main entities are "Public Health Event" representing information about the SIVIGILA events, and "Person" representing the patient associated to that event.

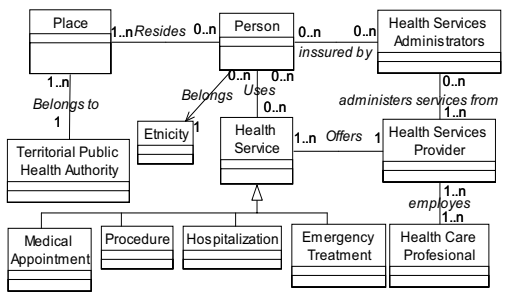


Figure 5 - The Integration Subsystem Invariant Schemata

Task: Harmonize Analysis Model

For supporting semantic integration, information models should be harmonized with reference models. Domain information models for supporting the informational components were identified in HL7 specifications. Concretely, the Public Health Reporting Domain Information Model (PORR_DM 180001) has been specialized [6]. An extract of the harmonized model depicting demographic data in the "Person" class is shown in Figure 6.

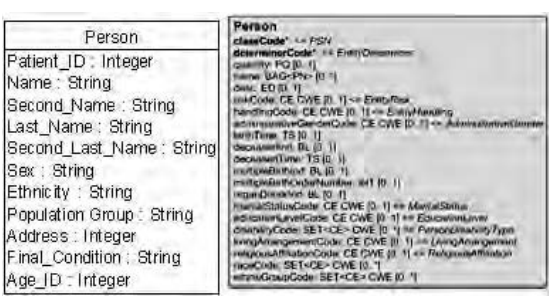


Figure 6 - Comparing the entity Person in the UML Analysis Model and the HL7 harmonized model.

Computational viewpoint

The computational view is concerned with the functional decomposition of the Integration Subsystem by specifying logical components and their interactions through interfaces. The architecture development methodology describes a top-down approach in which the computational components are derived from system use case descriptions in the enterprise viewpoint. The system Design Model is detailed below.

Work Product: Artifact Design Model

Figure 7 shows the structural diagram for the Integration Subsystem. The diagram is not exhaustive, but depicts the main components and interfaces. "Event_Look_Up", "Retrieve_Event_Data" and "Store_Event_Data" support the "Event Collection" business process presented in Figure 4. "Store_Event_Data" and "Notify Event" support the respective "Report Public Health Event" business process.

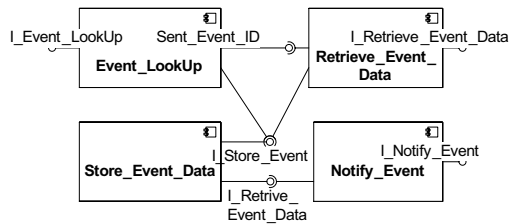


Figure 7 - The Integrated Public Health Information System Business Use Case Model

Task: Harmonize Design Models

In this stage, design components and interfaces are harmonized and refined with reference models. HL7 services, domains, application roles and CMETS serve as references for computational components and interfaces. In the scenario, the Entity Identification Service (EIS) and the Retrieve, Locate, and Update Service (RLUS) were used. As an example, figure 8 shows the operations defined for the interface "I_Event_LookUp". The semantic profile for the EIS is defined based on the "PublicHealthEntity" specified in the HL7 E_PublicHealthEntity CMET (COCT_MT840000) [6]. The interface implements the operations "Find Entities by Trait" defined in the EIS specification. The first allows for a search in the Clinical Information System for matching instances, where the attribute (traits) "ICD-10 code" is equal to one of the ICD-10 codes for notifiable diseases. Thus, a notifiable event can be retrieved from the Clinical Information System.

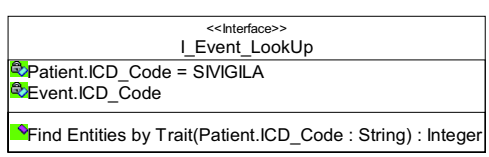


Figure 8 - "I_Event_LookUp" Interface definition

Discussion

There is not one unique methodology that can be applied to HIS architectural development; fortunately some architectural paradigms are in place to support this process. Also the resulting architectural models must include domain knowledge, which needs to be referenced when looking for semantic integration. The proposed methodology harmonizes advanced architectural approaches (RM-ODP, MDA, SOA, Component Based Development) and integrates HL7 domain knowledge into a comprehensive architecture development process.

HL7 specifications are partially reused because they are nowadays the richest source of reference to medical and public health knowledge. Furthermore, in addition to the information viewpoint, it is the only one that progressively supports enterprise and computational specifications. Nevertheless the methodology is flexible enough (through RUP configurations) to include other domain approaches, e.g. openEHR archetypes, CEN 13606 EHR Standard, CEN 12967, GPICs, etc.

The proposed methodology and the scenario architecture certainly follow very closely the HL7's SOA services specifications. Two services were reused (EIS and RLUS) as far as they are two of the three services included in the current HL7 specification [6]. Other services are under development within the HL7 SOA SIG. It is expected to integrate them into the architectural process. Policy Management, Authorization, Anonymization are some examples.

Regarding the proposed integration-architecture used to demonstrate the feasibility of the methodological approach, very important models were develop to address the necessary but not yet envisaged process of automatically interlinking the National Public Health Surveillance System in Colombia with emerging HIS.

The partially described integration architecture offers several possibilities to enhance the Colombian National Integrated Health Information System (SIIS) by providing efficient, high quality and sustainable health services. One possibility is that the architecture can be broaden and used as technical specification (platform-independent specification) to normalize all information systems developments in the country which are often autonomous projects not dealing with integration. Another possibility is to address the technical specification of the SIIS by transferring the architecture to platform-specific models using MDA, thus providing open software components and services that can be reused in local/regional developments.

Finally, the methodology and the derived architectures are an opportunity not only for Colombia but mainly for developing countries where relatively few systems are in place and resources are very short facilitating the adoption and adaptation of healthcare standards and advanced architectural approaches in use in the developed world.

Conclusion

The creation of semantically interoperable HIS including the integration of systems from other domains requires a comprehensive architectural methodology. Based on the GCM, different existing and heterogeneous architectural approaches have been adapted and partially reused in a harmonized way. The resulting architectural methodology has been deployed for integrating clinical HIS and PHIS. The feasibility of the derived architecture development process has been successfully demonstrated at the integration of clinical information systems into the National Public Health Surveillance System.

Acknowledgments

The authors are indebted to thank partners from HL7, OMG and other organizations for collegial collaboration. It's a pleasure to also thank "Dirección Departamental de Salud del Cauca" in Colombia for support and kind cooperation. This work was partially founded by the EU Programme Alban, contract E03D16197CO, and by the University of Cauca.

References

[1] Zachman J. A framework for information systems architecture. IBM Systems Journal: 1999: 38 (2-3): 454-471
[2] ISO/IEC 10746-1, 2, 3, 4. Open Distributed Processing - Reference Model. OMG, 1995-98.
[3] Object Management Group, Inc.: MDA Specification. www.omg.org/mda/
[4] Zimmermann O, Krogdahl P and Gee C. Elements of Service-Oriented Analysis and Design. IBM developerWorks. http://www-128.ibm.com/developerworks/library/ws-soad1/
[5] Kruchten P. The Rational Unified Process. An Introduction. Third Edition. Addison Wesley. 2003
[6] HL7 Inc. HL7 Version 3 Normative. Edition, 2006
[7] SOA4HL7. Healthcare Services Specification Project. http://hssp.wikispaces.com/
[8] CEN ENV 12967. "Healthcare Information System Architecture. European Committee for Standardization.
[9] Blobel B. Analysis, Design and Implementation of Secure and Interoperable Distributed Health Information Systems. Amsterdam: IOS Press, 2002.
[10] Resolución 2542 de 1998. Ministerio de la Protección Social. República de Colombia. http://www.minproteccionsocial.gov.co
[11] Decreto 3518 de 2006. Ministerio de la Protección Social. República de Colombia. http://www.minproteccionsocial.gov.co

Address for correspondence

Diego M. López, MSc. Email: diego.lopez@ehealth-cc.de

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Clinical and Translational Science Sustainability: Overcoming Integration Issues between Electronic Health Records (EHR) and Clinical Research Data Management Systems "Separate but Equal"

Robert P. DiLaura

Cleveland Clinic and Case Western Reserve University, Cleveland Ohio, USA

Abstract

The use of health information technology (HIT) is growing rapidly for patient care systems required to test, diagnose and treat patients, and to bill for these services. Today's Electronic Health Record (EHR) systems are a response to this pressure, enabling feature rich computer-assisted decisions and communication. And even though EHR benefits dramatically outweigh the costs, required investments are nonetheless significant. Continuing to invest in HIT at a revolutionary rate is unsustainable given institutional financial constraints and continuing reimbursement cuts. Future improvements must come from new treatments, test methods, drugs and devices – from research. But data management information systems for clinical research receive less funding than patient care systems, and in less coherent ways. It is easy to imagine using the high cost, patient-based EHRs for clinical research data management, and thus accelerate the speed of translating new medical discoveries into standard practice. But taking this step requires thoughtful planning to overcome significant technology, legal/regulatory, policy, process, and administrative issues.

Keywords:

medical informatics, biomedical research, clinical research, information systems,
computerized medical record systems

Introduction

Although the challenge of defining an Electronic Health Record (EHR) with multinational agreement has vexed experts around the globe, several professional associations are now reaching consensus on a basic conceptual elements.[1-3] An EHR uses computer systems to enter, store and retrieve patient medical records with accessibility by various authorized individuals. Additionally and more importantly, an EHR's extended functionality and interoperability bring significant benefits to users, healthcare providers and payers across multiple EHR systems and geographic locations (the Integrated Care EHR). Primary benefits come from faster and better decisions and reduced errors, thus improving care and saving lives. But looking more carefully at the financial picture, EHR adoption rates internationally vary based on country, care setting and provider specialization [4-5], typically anywhere between 15-85%, compared to the U.S. norm of about 25% for large hospitals[6]. While costs for nationwide implementation are well known where full adoption is nearly complete (e.g., Sweden, U.K.), other countries can only estimate the full financial requirements of such large scale projects, albeit with increasing accuracy. For the U.S., with the world's largest annual healthcare expenditures of around US$2 trillion, the cost projection for full national implementation of an ambulatory and inpatient EHR over 15 years is $114.6 billion.[7] This same report also estimates a 548% return on these investments, with savings of nearly $627.5B. Clearly the global trend to adopt nationwide EHRs is underway and will continue into the next decade. Unfortunately adoption has been slow in countries with market-based economies and privatized healthcare since financial models don't show the anticipated savings actually being accrued by the organizations that make the investments. Some large scale adopters have created new associations to encourage collaboration and financial intermediation, like the Good European Health Record project [8], or the not-for-profit company behind Australia's National E-Health Transition Authority (NEHTA), whose mission is to deliver a secure, interoperable nationwide e-health environment.[9] Such mechanisms may inevitably be required in the U.S. since projected HIT spending (Figure 1), currently less than 1% of overall healthcare cost, is expected to climb steadily (averaged estimate from CMS, IDC, Gartner, Frost & Sullivan, BCC Research, Datamonitor, Kalorama).

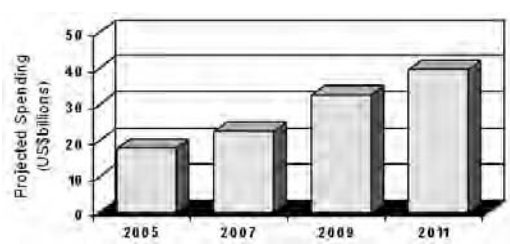


Figure 1 – Healthcare IT spending estimate for US ($billions)

U.K. national EHR implementation costs

To highlight the scale of coordination and funding required for national EHR implementation, The Lewin Group [10] summarized the U.K. National Health Service (NHS) in 2005 as:

> "…the largest organization in Europe and the third largest employer in the world, with more than 1.3 million workers. The agency is generally regarded as an efficient provider of high-quality health care to a large population. The state finances the NHS mainly through general taxes, administers the agency and makes executive decisions. Approximately 60% of U.K. primary practitioners use electronic health records. Recently, the NHS embarked on the world's largest civilian IT project, a plan to spend $11 billion on a national HIT system to replace existing local systems and paper medical records."

More recent reports have shown the size of spending in the U.K. may surpass US$20 billion before final goals are met, due partly from incorrect initial estimates and inflation, but also likely from scope creep for secondary uses of data.

Comparison with U.S. EHR funding

While the U.S. president has called for a national EHR to be available for most Americans by 2014, the necessary funding levels shown to be required in the U.K. haven't been provided thus far. And the FY2007 budget proposal recently submitted to Congress included only a request of $169M for EHR implementation, and $116M for the Office of the National Coordinator for Health Information Technology to coordinate and plan a national strategy and initiatives. The initial increase for the National Institutes of Health (NIH) in 2006, which saw a doubling of its budget over the past five years to approximately $28B, was virtually zero [11], although this was increased slightly by the NIH Reauthorization Act passed in December. But nearly two thirds of the NIH budget goes toward basic research, with only the remaining one third going to patient-oriented clinical research versus patient care; i.e., generally not for health information system infrastructure costs such as an EHR. Thus meeting the necessary expense for nationwide EHR implementation in the U.S. with such limited funding is unrealistic, with current efforts simply unsustainable.

Clinical research and its role in healthcare

Clinical research (CR) is defined by the NIH as human subject research involving an investigator that is patient-oriented including the study of – mechanisms of disease, therapeutic interventions, clinical trials (CT), and the development of new technologies. They may also involve epidemiologic and behavioral studies, outcomes research, be retrospective or prospective, utilize standard-of-care control groups or placebo comparisons, may randomize carefully selected samples of study subjects from larger groups or investigate specifically selected individuals because of their health circumstances.[12] Human subject research must follow strict ethical requirements that have been agreed to by well known international collaborations such as the World Medical Association Declaration of Helsinki, as well as regulatory requirements within, and sometimes between, national borders. Regulatory obligations typically mandate an independent Institutional Review Board (IRB) that evaluates all requests for research and ensures necessary policies and procedures are complied with. And since even large, well-known institutions can be involved with investigators who at times act inappropriately or who are not sensitive enough to study subjects special needs, especially for vulnerable populations (e.g., children, elderly, etc.), new positions and procedures have been created to independently review and endorse an institution's research enterprise and commitment to research integrity.[13]

EHRs and healthcare in general can be looked upon as vertical space in an overall strategic framework for biomedical science. The knowledge transfer that occurs from translation of scientific discovery in bench laboratories, to animal research, *in silico* research, pilot studies in humans, and all the way through large scale, multi-center, global studies into standard practice by healthcare providers was described by Sung et al [14] as the T1 translational block horizontally across these domains. Clinical research is the scientific domain prior to healthcare delivery, and upon which new standards of care depend. The need to improve the productivity and cost effectiveness of CR processes and data management systems is similar to healthcare, although on a smaller scale. The need to more tightly couple CR to healthcare, and move new discoveries forward across these domains quickly and inexpensively is also important.

Like healthcare costs, various reports [15-16] have studied the unsustainability of regulated clinical trials in particular as a subset of CR for drug and device discoveries to meet growing demands to control future healthcare problems and changing patient population demographics given the large expense and years required to get final regulatory approvals.

EHR/Clinical research integration

EHRs and CR data management systems (Figure 2) can exist as anything from independent, parallel technology platforms,

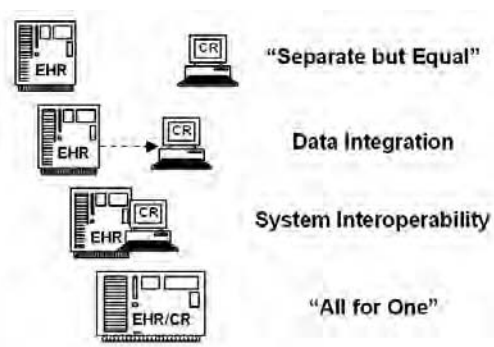


Figure 2 – EHR/CR Integration Models

to the CR system receiving a one-way data feed from the EHR, to some degree of two-way interoperability between the separate systems, to being completely integrated into

one system. This fourth alternative seems to be mostly theory at this point considering major EHR vendor products. An informal survey conducted during the 2006 HIMSS conference found only 20% with CR capabilities in current EHRs. This finding was similar to a more formal study by the American Academy of Neurology that found only 28.5% of mid-sized EHR vendors to be capable of conducting any level of CR in their current systems.[17] Usually a central IT organization governs the EHR. It also has certain budgetary or systems responsibility for CR systems, though whenever this is the case, the CR mission is almost always a very small part of the overall IT focus and funding.[18] In fact this study showed 60% of academic health centers do not even have a clear picture of how much is spent on IT for CR, and 49% have no central HIT-CR governance structure or cross-representation between these personnel. Virtually none have a strategic plan for CR. And while various attempts have been made to begin experiments to integrate EHRs with CR systems, or even to fully conduct a CR study completely within the EHR, most have been very small or ineffective due to the overwhelming issues that must be faced to make these connections work and leverage the significant investment in EHRs. Even demanding these connections through legislation in a national setting has proven futile.[19]

Issues in EHR/CR integration

As stated earlier there are many types of CR studies. All may not be appropriate to conduct directly within an EHR or via an integrated EHR/CR system, but it may be easily possible for some. The following are general categories of issues surrounding EHR/CR integration.

Technology

Most EHR products in the U.S. (other than for an OpenEHR) are proprietary technologies with vendor controlled code bases requiring certain operating platforms and network architectures. Interoperability with other information systems may be possible in some cases, but underlying data models, structure, format, vocabulary and transmission standards are certain to inhibit connectivity or data transfer without significant work.

Current FDA requirements for clinical trials of investigational drugs and devices specify that the data management system comply with system and software validation.[20-21] The steps to comply with these requirements are typically very expensive, and must be repeated whenever changes are made to the technology, which is unacceptable in a large EHR.

Legal/Regulatory

In the U.S. (similar to other countries), each patient obtaining medical treatment must be shown a Notice of Privacy Practice (NoPP) that informs them of their legal rights under the Healthcare Insurance Portability and Accountability Act (HIPAA). The notice requires "plain language", and a statement that information collected on the patients for standard care may be used for research (operational and outcomes) in addition to treatment, payment and quality purposes.[22] The use of patient data for future, unspecified research is ***not*** permitted. An assessment of the NoPP from the twelve winners of the new NIH Clinical and Translational Science Award (CTSA) for 2006 [23] indicates an average Flesch-Kincaid Grade Level of 14.3. Certainly most hospital patients are not college level readers. So even though the CTSA winners have been recognized for outstanding abilities to perform CR and to translate new science into the patient care domain, there will most likely need to be changes made in how they inform patients that research may be conducted on their data (possibly in the EHR) as normal care.

Conversely, whether a patient or not in a hospital, any person who is going to participate in prospective CR is required by regulation to be fully informed of all aspects of the study prior to making a decision to participate, and not to be coerced in any way, especially by any person in a position of authority (as is a physician-investigator when discussing a research study they are the principal investigator on with their patient). Typical informed consent forms for CR are 10-20 pages in length and can be quite complex [24] even though they should be written at less than a high-school reading level.[25]

Policy

A hallmark of an IRB is its ability to independently make ethical decisions that balance institutional interests in research against protecting human subjects involved in studies. Like criticism of the FDA because it receives money from drug companies as part of the drug approval process, the appearance of improper behavior or the potential conflict of interest is often enough to cause problems. At what point does an IRB, which may already be closely aligned with or supported by an institution, become too close and biased?

Revenues from research, whether federally funded or from industry sponsored sources generate a significant amount of income at large institutions. At what point does strict compliance with someone's interpretation of an SOP, or an unusual pattern of missing values, or delayed serious adverse event reporting require the institution to suspend or end a study (and revenue)?

How is time seeing patients as part of normal care accounted for by doctors when patients are also enrolled in a research study in the same visit? How are the costs separated, or insurance companies fully informed that the visit is for "research"?

Process

Fundamentally all CR processes are very similar, yet different enough that it has been easier to build required data management system from scratch every time a new study comes along. And since most funding for CR studies has been provided independently with little to no regard for centralized or reusable systems, investigators either received funding to build everything they needed from scratch every time, or they received little to no money for data management needs at all and thus had to cut corners in order to conduct an under-funded study in whatever way was possible within their institution. Standards for CR

(Figure 3) that improve operational consistency now exist however, especially where countries fund national EHRs.

Physicians are thus under increasing pressure to follow newly prescribed processes, especially in the case when an EHR is in place; often having 10 minutes or less to interact with patients. How can they provide the necessary additional time required to inform a patient fully about a new research study patients are eligible to participate in? Should they delegate?

Likewise, research studies have historically been statistically designed around strict visit schedules and data collection methods in order to ensure validity and significance. In many cases, these study schedules are not the standard-of-care that a physician would normally provide for their patients. So patients need

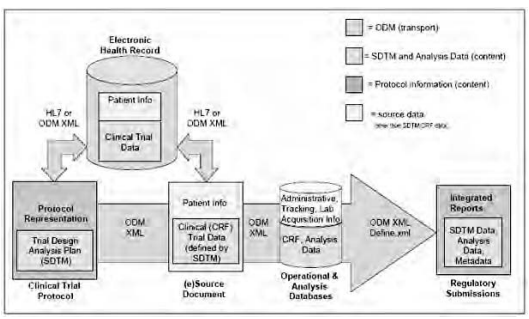


Figure 3 – Clinical Data Flow Using CDISC Std. [26]

to either come in more frequently than is normally necessary, or visit only when they are scheduled regardless of how they feel. Both situations are problematic.

Administrative

Security (i.e., confidentiality, integrity, accessibility, etc.) of a research subject's data is at least as important as a patient's data. Yet there are different rules that govern how data is managed and who can access what. Especially if CR data is collected as part of an EHR; consent forms often limit who can see study subject data to only people directly involved with the research, as opposed to anyone in the institution with general access to the EHR who is allowed to see patient records. How can this distinction be readily defined, and who will have the real-time responsibility to enter such permissions in an EHR as security administrator? Finally, healthcare data breaches are more commonplace than people want to know.[27]

EHR systems typically have well defined workflow components designed into the data collection screens for ease of use. Will new fields or special screens need to be designed for all research trials to enable complete and timely collection of relevant study variables (eCRF)? Will study subjects have the legal right to withdraw from research participation at any time for any reason without repercussion to their normal care, or to prohibit future use of their data for research purposes?

Recommendations for progress

The consensus from leading U.S. academic medical research centers at the Clinical Research Forum meeting in April 2007, as well as new dialog on this topic held at the 2007 HIMSS meeting with the EHR Vendor Association and pharmaceutical companies was that there are many issues to overcome, but none are impossible. Participation of all stakeholders is necessary to create sustainable success. There is no single "leader" or starting point to EHR/CR integration. Pick any spot relative to your unique environmental needs and just get started!

Regarding technology, if you already have an EHR, begin a dialog with your vendor on inclusion of CR into their functional design. Require this from them in the future. If you don't have an EHR, become involved in national associations that are beginning to hold these public discussions (e.g., AMIA) with mixed stakeholder groups. Require vendor involvement and compliance with nationally recognized standards. Pressure publicly elected officials regarding regulatory agencies to accelerate changes in outdated rules and guidelines in support of electronic source data collection and transmission.

Challenge legislators through efforts with national associations to harmonize the legal framework around patients and study subjects, thus lowering the burden associated with CR translational science. Rethink and revise plans and materials for patient and community awareness, and their involvement in CR including NoPP and informed consent document creation.

Form or re-energize institutional committees defining responsible research conduct, research integrity, or conflicts of interest. Establish policies for limits to IRB involvement in hospital operational research; budgeting and accounting for research activities and investigational staff effort reporting.

Investigate, then adopt or adapt a CR process model. Benchmark best practices, including sharing your work with others. And finally, continuously improve and regularly test your security model for patient and research data. It's never secure.

Conclusion

EHRs implemented on a national scale are inevitable over the next decade, but there must be significant financial support and coordinated collaboration among many stakeholders to succeed and overcome the unsustainability of sub-optimized approaches. Likewise CR data management systems have diminishing chances of existing on their own as separate source records because of the many issues identified in this paper. Integration of the EHR with CR systems, or better, enabling CR to be completely carried out within the EHR system itself, will take foresight and redesign work between central HIT staff, CR informaticists, and commercial and open source partners. In addition to the financial and quality-of-care benefits, future research will be able to leverage clinical data-CR repositories that will become available from regional and

national networks. Study subject enrollment may quickly increase and open a way to explore rare "orphan" and pediatric diseases that would have otherwise been financially or statistically unattainable.

Acknowledgments

Thanks to Drs. Anil Jain, Mary Partin, Michael Kattan and others at Cleveland Clinic; Dr. Pam Davis at CASE; Michael Lin of Mayo Clinic; Drs. Elaine Collier and Peter Highnam at the NIH/NCRR; Rob Schuff at Oregon Health Sciences University; Dr. Paul Harris at Vanderbilt; Dr. Michael Kahn at University of Colorado Health Sciences Center; and Dr. Don Detmer from AMIA for their ongoing encouragement, challenge and support.

References

[1] ISO/TR 20514:2005, Health informatics – Electronic health records – Definition, scope, and context.

[2] IOM:2003. Key capabilities of an electronic health record system. U.S. National Academy of Sciences, Institute of Medicine, 2003.

[3] The connecting for health common framework: Resources for implementing private and secure health information exchange. Markle Foundation, 2006.

[4] Detmer D, Steen E. Learning from abroad: Lessons and questions on personal health records for national policy. AARP Public Policy Institute, Research Report 2006-10, March, 2006.

[5] The future vision of electronic health records as eSource for clinical research. eClinical Forum, PhRMA EDC Task Group, 2006.

[6] Jha A, Ferris T, Donelan K, DesRoches C, Shields A, Rosenbaum S, Blumenthal D. How common are electronic health record systems in the United States? A summary of the evidence. Health Aff. 2006; 25(6):496-507.

[7] Girosi F, Meili R, Scoville R. Extrapolating evidence of health information technology savings and costs. Santa Monica: Rand, 2005.

[8] The Good European Health Record (GEHR) project [homepage on the Internet]. London:University College: Centre for Health Informatics and Multiprofessional Education. [updated 2005 Aug 22; cited 2006 Dec 4]. Available from: http://www.chime.ucl.ac.uk/work-areas/ehrs/GEHR/

[9] National e-Health Transition Authority (NEHTA) [homepage on the Internet]. No date [cited 2006 Dec 4]. Available from: http://www.nehta.gov.au/

[10] Health information technology leadership panel: Final report. The Lewin Group, March 2005.

[11] President's FY2007 budget proposal: Overview and briefing charts [homepage on the Internet]. The Henry J. Kaiser Family Foundation; c2006 [cited 2006 Dec 4]. Available from: http://www.kff.org/uninsured/upload/7472.pdf

[12] Human subjects research definition: Clinical research [homepage on the Internet]. National Institutes of Health; no date [cited 2006 Dec 4]. Available from: http://grants2.nih.gov/grants/funding/phs398/instructions/p2_human_subjects_definitions.htm

[13] Association for the Accreditation of Human Research Protection Programs (AAHRPP) [homepage on the Internet]. c2001-2007 [cited 2006 Dec 4]. Available from: http://www.aahrpp.org/www.aspx

[14] Sung N, Crowley W, Genel M, Salber P, Sandy L, Sherwood L, et al. Central challenges facing the national clinical research enterprise. JAMA. 2003;289:1278-1287.

[15] Challenge and opportunity on the Critical Path to new medical products. US Food and Drug Administration (FDA), 2004.

[16] Milne CP. US and European regulatory initiatives to improve R&D performance. Expert Opinion on Drug Discovery, 2006; 1(1):11-14.

[17] Electronic health records vendor report. American Academy of Neurology, Electronic Health Records Work Group, September, 2006.

[18] Turisco F, Keogh D, Stubbs C, Glaser J, Crowley W. Current status of integrating information technologies into the clinical research enterprise within US academic health centers: Strategic value and opportunities for investment. J Investigative Med; 2005; 53(8):425-33.

[19] Moulton B. Two years later: The impact of the EU directive. Appl Clin Trials; 2006; August 1.

[20] Computerized systems used in clinical trials (CSUCT). FDA guidance document, 1999.

[21] General principles of software validation: Final guidance for industry and FDA staff. FDA guidance, 2002.

[22] HIPAA Privacy Rule [homepage on the Internet]. Department of Health and Human Services, Office of Civil Rights; c2002 [updated 2002 Dec 3; cited 2006 Dec 4]. Available from: http://www.hipaa.state.sc.us/ocr-hip.pdf

[23] Clinical and Translational Science Award (CTSA) to transform clinical research [homepage on the Internet]. National Institutes of Health, National Center for Research Resources; no date [cited 2006 Dec 4]. Available from: http://www.ncrr.nih.gov/ncrrprog/roadmap/CTSA_9-2006.asp

[24] Flory J, Emanuel E. Interventions to improve research participants' understanding in informed consent for research: A systematic review. JAMA, 2004; 292(13): 1593-1601.

[25] Guidelines for writing informed consent documents [homepage on the Internet]. NIH Office of Human Subjects Research; no date [updated 2006 Dec 28; cited 2006 Dec 4]. Available from: http://ohsr.od.nih.gov/info/sheet6.html

[26] CDISC roadmap discussion document [homepage on the Internet]. Clinical Data Interchange Standards Consortium; no date [cited 2006 Dec 4]. Available from: http://www.cdisc.org/standards/CDISCRoadmapJan2006.pdf

[27] A chronology of data breaches [homepage on the Internet]. Privacy Rights Clearinghouse; c2005 [updated 2007 Apr 29; cited 2006 Dec 4]. Available from: http://www.privacyrights.org/ar/ChronDataBreaches.htm

Address for Correspondence

Comments welcome. Contact Dr. Robert DiLaura at:
Email: dilaurr@ccf.org.

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

The FA4CT Algorithm: A New Model and Tool for Consumers to Assess and Filter Health Information on the Internet

Gunther Eysenbach [a,b] **, Maria Thomson** [c]

[a] *Department. Of Health Policy, Management, and Evaluation, University of Toronto, Canada*
[b] *Centre for Global eHealth Innovation, University Health Network, Toronto General Hospital, Toronto, Canada*
[c] *Department of Health Studies and Gerontology, University of Waterloo, Canada*

Abstract

Background: eHealth-literate consumers, consumers able to navigate and filter credible information on the Internet, are an important cornerstone of sustainable health systems in the 21st century. Various checklists and tools for consumers to assess the quality of health information on the Internet have been proposed, but most fail to take into account the unique properties of a networked digital environment. Method: A new educational model and tool for assessing information on the Internet has been designed and pilot tested with consumers. The new proposed model replaces the "traditional" static questionnaire/checklist/ rating approach with a dynamic, process-oriented approach, which emphasizes three steps consumers should follow when navigating the Internet. FA4CT (or FACCCCT) is an acronym for these three steps: 1) Find Answers and Compare [information from different sources], 2) Check Credibility [of sources, if conflicting information is provided], 3) Check Trustworthiness (Reputation) [of sources, if conflicting information is provided]. In contrast to existing tools, the unit of evaluation is a "fact" (i.e. a health claim), rather than a webpage or website. Results: Formative evaluations and user testing suggest that the FA4CT model is a reliable, valid, and usable approach for consumers. Conclusion: The algorithm can be taught and used in educational interventions ("Internet schools" for consumers), but can also be a foundation for more sophisticated tools or portals, which automate the evaluation according to the FA4CT algorithm.

Keywords:

internet, consumer health informatics, information quality, information retrieval, education

Introduction

Searching for health information online is often said to be "one of the most popular activities on the Internet". Such sweeping (and only partially accurate) claims are mostly based on survey data, such as the Pew Internet Report, where people are questioned whether they have "ever looked online for" a certain category of information such as health, entertainment, or shopping. The Pew Internet Report 2003 [1] found that "fully 80% of adult Internet users, or about 93 million Americans, have searched for at least one of 16 major health topics online" and goes on concluding that "this makes the act of looking for health or medical information one of the most popular activities online, after email (93%) and researching a product or service before buying it (83%)."

In reality, the question "have you ever used the Internet for y" does not necessarily translate into the prevalence of day-to-day activities. To gauge these, one has to directly observe Web traffic or monitor what people are searching for. Several independent studies using these more "direct" methods to gauge online activities by tapping into the datasets from various search engines, have concluded that the actual volume of health-related searches on the Internet as a proportion of all searches conducted each day is "only" around 5% [2-5], with other areas such as entertainment, shopping, porn, research, places or business being much more popular.

In summary, survey and search data combined suggest that searching for health information is a popular, but relatively infrequent activity for most people (chronically ill people being a notable exception).

This usage pattern of health information has implications: While people may know where to go for reliable news, weather information, movie reviews, shopping, and business information, medical questions arise infrequently enough so that people not necessarily have a trusted brand names or portal in their mind. While people may be savvy and experienced enough to evaluate the credibility of a general news website or an ecommerce site, they may have insufficient experience and expertise with health websites. Consumers need to be "eHealth literate" in order to succeed in finding and filtering information. "eEhealth literacy" [6] consists of six literacy types (traditional, information, media, health, computer, and scientific literacy) which combined form the foundational skills required by consumers' to engage with electronic health information.

Several attempts have been made to create tools which can be used to educate consumers or which may assist consumers in identifying "credible" information. Most (if not all) previous tools are checklist-like instruments, designed to evaluate information on a webpage or website level.

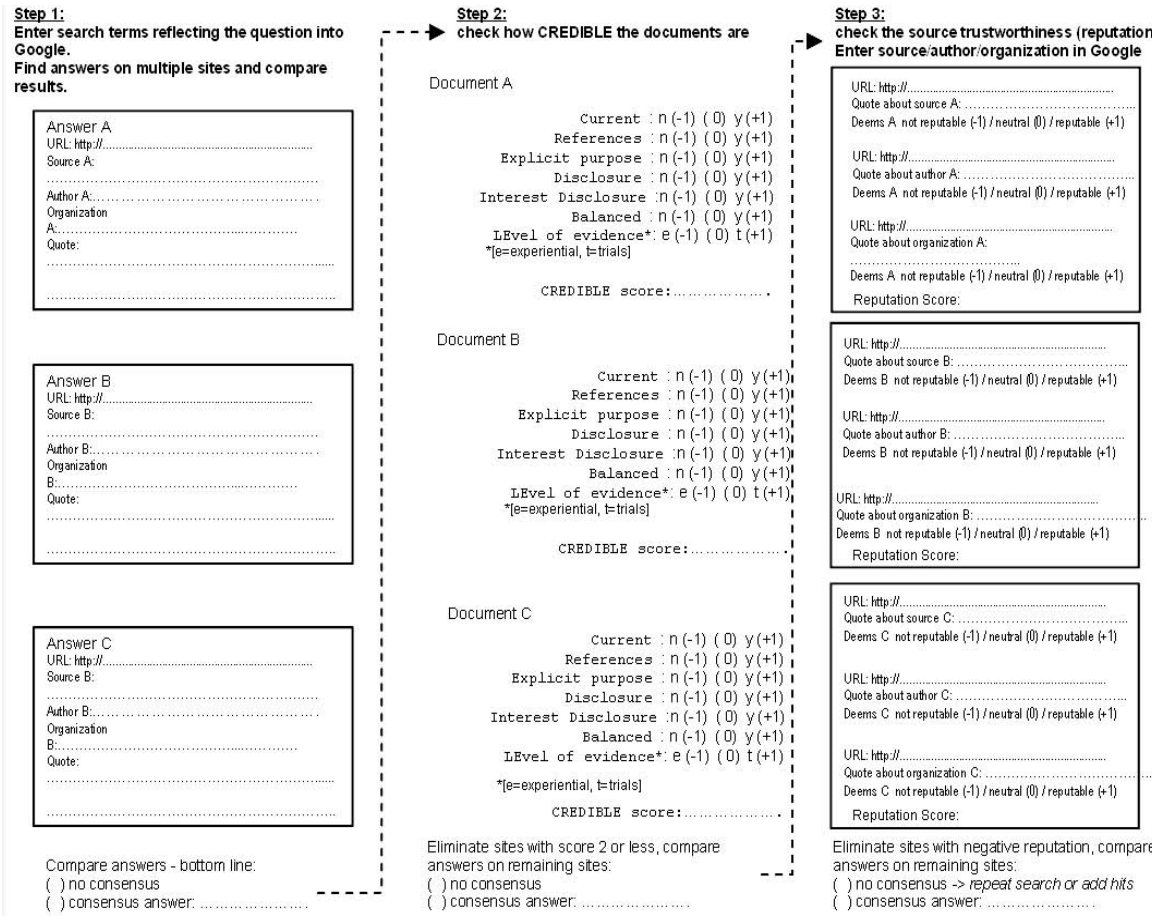


Figure 1 - The FA4CT Algorithm (Worksheet for Consumers)

A recent review of 273 instruments which can be used by patients and consumers to asses the credibility of health information has concluded that "few are likely to be practically usable by the intended audience" [7].

Many or all of today's tools are cumbersome and time-consuming checklists. They do not adequately take into account the unique features of a digital networked environment, but are still guided or influenced by our thinking about credibility in the "offline", printed world. The DISCERN instrument, developed for *printed* patient education brochures but advocated by its developers as an evaluation tool for web-based information [8;9] is a prime example. The claim of the DISCERN authors that "there's nothing radically different about information on the web" [10] illustrates a failure to recognize and to capitalize on the advantages of the Web to use the networked environment itself to assess the credibility of (health) information.

A second generation of educational tools – beyond checklists of authorship and content criteria of web documents – is needed, one that takes into account that consumers are in a *networked, digital* environment, and that credibility evaluation in this medium is a dynamic, interactive, and iterative process. Advantages of a networked environment, which should be exploited and utilized by educational and technological tools include the ability for users not to rely on only a single source, but to cross-check information on other websites and to check the credibility and reputation of the source using the Web itself. As Meola noted, rather than promoting a mechanistic way of evaluating Internet resources, a contextual approach is needed, which includes for example the possibility to corroborate information on the Web from other sources [11].

Methods

The FA4CT approach

In this paper we propose and pilot test a second generation educational model and approach which we call the FA4CT model. This educational model was originally developed in the context of an Internet school for cancer patients (I3MPACT project: Impact of Internet Instructions on Men with Prostate Cancer). FA4CT is intended for use by consumers to find and check medical facts on the Internet. In contrast to earlier approaches, FA4CT is not a checklist, but an intuitive process (or algorithm) which users are instructed to follow when assessing health information on the Web. The algorithm mimics the process expert search-

ers use for information retrieval and fact checking on the Web. For example, journalists use the technique of cross-checking facts from multiple sources to verify the credibility of their sources.

FA4CT (or FACCCCT) is an acronym for the three steps suggested in the algorithm: 1) Find Answers and Compare [information from different sources], 2) Check Credibility [of sources, if conflicting information is provided], 3) Check Trustworthiness (Reputation) [of sources, if conflicting information is provided].

The model also recognizes that consumers are usually not primarily interested in assessing the credibility of an entire "website" (or page, or document) as the unit of evaluation, but usually in the credibility of a specific health claim (fact). Thus, in order to use FA4CT, a consumer seeking information on the Internet is instructed to first formulate his factual question as clearly as possible, preferably in a way that allows a yes/no answer. He is then instructed to translate this question into search terms and to conduct an initial Google search query to locate three web sites that contain an answer to their specific medical question. The first key step (**step 1**) for making sure that the information found on the Web is "accurate" is to *compare* (cross-check) the information found on multiple websites. This is a major shift from previous approaches such as DISCERN, where checklists are used to check the credibility of the source and the information itself. In contrast, the FAC4CT algorithm suggests a source/information credibility assessment based on a checklist only as a second step, and only if there is no consensus in the three answers provided. In this case, **step 2** suggests to assess each web site using the *CREDIBLE* criteria [12]. The acronym CREDIBLE refers to Current, References, Explicit purpose, Disclosure of sponsors, Interest disclosed and no conflicts found, Balanced, and Level of Evidence. These criteria are based on empirical studies and reflect markers which have in multivariate regression models been shown to be independent predictors for accuracy [12].

Each of the seven criteria has three simple rating options "not fulfilled" (scored with -1), "neutral" (0), and "fulfilled" (+1) with a total possible credible score ranging from -7 to 7.

If after elimination of less "credible" web sites according to these criteria there is still no consensus among the remaining websites, users are in **step 3** asked to enter the name of the source into Google to check what others on the Web have to say about the source, arriving at a *reputation* score. To assess the reputation, for each web page in question the source, author or organization are entered into Google and three new sources commenting on the reputation of the source in question are identified and a quote commenting on the reputation of the source is recorded. Reputation is scored "+1" if there was an explicit statement of trustworthiness, "0" if neural or "-1" if there was an explicit statement of untrustworthiness. Figure 1 shows the algorithm as worksheet for users. In addition, a more detailed instruction sheet (not shown) is made available to users. It should be noted that the algorithm is designed for educational purposes or for implementation in automated tools assisting users. Users are not expected to go through these detailed calculations each time they check a fact, rather, they should - by applying the algorithm a few times with an instructor - develop and internalize the process on a more intuitive basis.

Formative evaluation and ROC evaluation of CREDIBLE checklist

As part of the formative evaluation of the FA4CT algorithm we had to establish 1) how many websites consumers should cross-check to arrive at a valid assessment on the accuracy of a fact, 2) what the optimal cut-off point of the CREDIBLE score from step 2 is, using a ROC (receiver-operating characteristic) curve approach.

Four questions related to a medical fact were used to pilot the FA4CT algorithm, for each question the first six websites resulting from a Google search containing the answer were assessed, resulting in a total of 24 evaluations.

The searches took place on March 3, 2006. The following are the four pilot questions used and their associated answers from gold-standard evidence based resources.

1) Do exclusively breastfed babies require vitamin D supplementation? The search terms entered were "vitamin D" "breastfeeding". The answer as derived from the CMA clinical practice guidelines developed by the Canadian Pediatric Society, which recommends that breastfeed babies be given a daily vitamin D supplement until their diet includes a reliable source or they are one year of age.
2) Does vaccination cause autism? The search terms used were "vaccination" "autism". According to the Cochrane database of systematic reviews there is no credible evidence of a link between the MMR vaccine and autism.
3) Does Echinacea cure colds? The search terms entered were "Echinacea" "colds". According to a review from the Cochrane database of systematic reviews there is no clear evidence that Echinacea prevents colds.
4) Should statins be taken for high cholesterol? This question was searched using "statins" "cholesterol". According to the recommendations of the CMA clinical practice guidelines people at high risk should be treated with the equivalent of 40 mg/d of simvastatin.

Pilot usability test with end-users

Eight participants were recruited using advertising posters distributed throughout three Toronto hospitals as well as the Consumer Health Information kiosk at the Toronto Reference Library. All participants attended an one hour session at the Centre for Global eHealth Innovation, University Health Network. The sessions were conducted from July 25-28, 2006. Computer sessions were held in a usability lab that enabled video, audio and computer recording, in addition to a one-way mirror for one observer to take notes. The session was recorded using Morae software and captured computer screen and key strokes as well as video and audio of the participant. Participants were

encouraged to speak out loud creating a narrative of their actions and decisions. Post session interviews were recorded with a simple hand held recorder.

Each computer session consisted of three tasks. Due to technical problems only five of the eight participants were included in the analysis of task two. All participants received a brief training at the beginning of the session.

Task One. This task was designed to test a "forced" step 2 of the FA4CT algorithm (regardless of whether step 1 would have triggered step 2). Participants were asked to rate three pre-selected websites using the CREDIBLE criteria. Each website provided an answer to the dichotomous question 'Do exclusively breastfed babies require vitamin D supplementation?' To retrieve these webpages the search terms "vitamin D" "breastfeeding" were entered into Google. Two of the three websites retrieved provided the 'correct' answer. This search took place on March 3, 2006. Each participant received a copy of the FA4CT algorithm and a list of the CREDIBLE criteria definitions.

To determine the reliability of the CREDIBLE criteria when used by multiple raters an inter-rater reliability score was calculated for each criteria. This measure was calculated using the Fleiss variation of the Kappa coefficient. Calculations were completed using SAS version 9.1 and the SAS MAGREE macro.

Task Two. Participants were asked to use the FA4CT tool for a health-related question of their own choice.

Both tasks were followed by a short semi structured interview and questionnaire to elicit participant feedback regarding the use of the FA4CT tool.

Results

Formative evaluation and ROC evaluation of CREDIBLE checklist

A total of 24 websites were assessed representing four different questions queried, 33% (8/24) of these web pages were determined to provide the "wrong" answer as compared to the evidence-based gold-standard, with "breastfeeding and vitamin D" having the most wrong answers at (3/6).

Although the FA4CT algorithm stipulates that step 2 (CREDIBLE evaluation) should only be carried out if there was no consensus among the first 3 websites, for this pilot evaluation a CREDIBLE score was calculated for all 24 webpages in order to determine the optimal cut-off point. The best cut-off point seems to be a threshold of 2 (sites meeting only 2 or less CREDIBLE criteria are considered not credible), where 87.5% of all web pages that contained the correct answer were correctly deemed credible and only 12.5% of web pages that contained the wrong answers were incorrectly labeled as credible.

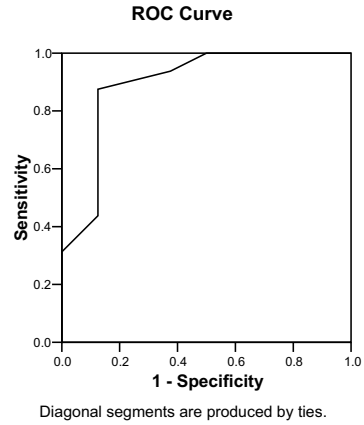


Figure 2 - ROC curve for CREDIBLE score of 24 webpages

Using 3 websites as a starting point, and a CREDIBLE score cut-off point of 2, the FA4CT algorithm performed well. Only 1/8 question sets did not result in a consensus answer at the end of the entire algorithm. Three question sets were correctly answered immediately in step 1 and a further four sets were correctly answered after employing the CREDIBLE criteria in step 2.

Usability test with end-users

The pilot user sample consisted of an equal number of males and females, with four participants aged 21-40, two participants aged 41 to 60 years and two participants over 61 years. In terms of education, three participants had completed post graduate training, four completed college or university and one completed high school. When asked about their level of confidence using the computer for the session, four reported feeling very confident, two reported that they felt confident and two reported to have some confidence. All participants had at least one year computer experience with six participants reporting more than ten years experience. All participants reported looking for health information on the internet.

Task one: The CREDIBLE criteria

All eight participants completed task one. On average the time to score one website using the CREDIBLE criteria was 10.25 minutes per page with a range of 5.1 to 18.1 min/page. Because two of the three websites provided a correct answer, it was of interest to identify whether participant criteria scores reflected this distinction. Three participants awarded final criteria scores that correctly classified all three websites therefore awarding passing scores to the two websites with the correct answer and a failing score to the website that provided an incorrect answer. Three participants correctly classified two out of three sites and two participants correctly classified only one site. The site that received the least correct classifications was the site that provided the incorrect answer, according to the gold standard.

Kappa statistics were calculated for each of the seven criteria scored on three webpages by eight raters. Generally the kappa results were satisfactory falling between k=0.52-0.79 (Table 1).

Table 1 - Kappa agreement scores from 8 consumers

CREDIBLE Criterion	Kappa Score
Current	0.62
Reference	0.62
Explicit purpose	0.79
Disclosure of sponsors	0.68
Interest disclosed and no conflicts found	0.68
Balanced	0.53
Level of **E**vidence	0.52

Task two: using the FA4CT tool

Five participants properly completed task two, using the entire FA4CT algorithm for their own health-related question. Each participant successfully asked a dichotomous medical question, completed a Google search and located three websites that answered their question. Upon evaluation of the three websites all five participants deemed a consensus answer to have been reached, i.e. no participant had to proceed to step 2 or 3. The average total time spent using the tool was 17.58 minutes with a minimum 10.1 and a maximum 27.2 minutes.

Participant search strategies may have influenced the outcome of the algorithm operation. Four out of five participants typed their medical question into Google using a full sentence complete with a question mark. Only one participant used keywords to search. This may have affected the retrieval of websites by listing first the websites that contained that specific sentence, as opposed to listing websites that contained the search key words without an exact sentence match.

Discussion

While the FA4CT approach still needs to be refined, the overarching model constitutes a major paradigm change from previous approaches. First, it is a *process*, rather than a mere checklist. While a checklist (the CREDIBLE criteria) is a part of the process, this checklist is only used as second step to eliminate less credible websites in case of lack of consensus among the first 3 sources in step 1, and is in practice rarely needed. Secondly, the approach teaches an evaluation on a fact-level, rather than a "website" or document level (though consumers could be taught to evaluate a couple of facts using FA4CT to arrive at a website/document rating). Thirdly, it takes into account the major advantages of information retrieval on the web, which is the ability to cross-check facts using different sources and to check the reputation of sources. Thus it reflects what experienced searchers do when they check the credibility of a (medical or non-medical) claim on the Internet. Consumers using the FA4CT approach are encouraged to check multiple sources and websites to arrive at an answer. They learn to eliminate clearly non-credible sources. In relatively rare instances, discordances will remain after elimination of non-credible sources, leading to the teaching point that in medicine there is often more than one answer, and sometimes even reputable sources contradict each other, which is often a sign of conflicting evidence in the literature.

Our initial experiments with the FA4CT approach have been encouraging. A caveat is that detailed instructions on how to formulate a search query in a neutral way should be part of the tool, to avoid that people enter preconceived opinions in form of full sentences into Google to find only one-sided answers. Initial findings from user testing also led to fine-tuning of the instrument, in particular related to the CREDIBLE scoring. Checking reputation (step 3) was rarely required in our initial experiments, as in most cases step 1 and step 2 already led to an accurate result, and will require further testing. Reputation checking on Google requires some more advanced search strategies, but might be facilitated by future tools specifically built for checking the reputation of sources (for some time, Google Labs offered such a feature, which is now disabled).

While the algorithm was initially designed to be used as part of educational interventions, it is conceivable that a similar algorithm based on a mix of cross-checking facts and checking of credibility markers and reputation could be part of future automated tools that help consumers to find trustworthy information on the Internet.

References

[1] Fox, S. and Fallows, D. Internet Health Resources. http://www.webcitation.org/5I2STSU61 . 2003.

[2] Eysenbach G,.Kohler C. What is the prevalence of health-related searches on the World Wide Web? Qualitative and quantitative analysis of search engine queries on the Internet. *Proc AMIA Annu Fall Symp* 2003;225-9.

[3] Eysenbach G,.Köhler C. Health-Related Searches on the Internet. *JAMA: The Journal of the American Medical Association* 2004;**291**:2946.

[4] Spink A, Yang Y, Jansen J, Nykanen P, Lorence DP, Ozmutlu S *et al*. A study of medical and health queries to web search engines. *Health Information and Libraries Journal* 2004;**21**:44-51.

[5] Pass, G., Chowdhury, A., and Torgeson, C. A Picture of Search. http://www.webcitation.org/5I2Tms4Qk INFOSCALE '06 Proc of The First International Conference on Scalable Information Systems, Ma 29-June 1, 2006, Hong Kong. 2006.

[6] Norman CD,.Skinner HA. eHealth Literacy: Essential Skills for Consumer Health in a Networked World. *Journal of Medical Internet Research* 2006;**8**(4)**:**e27.

[7] Bernstam EV, Shelton DM, Walji M, Meric-Bernstam F. Instruments to assess the quality of health information on the World Wide Web: what can our patients actually use? *Int J Med Inform* 2005;**74**:13-9.

[8] Charnock D,.Shepperd S. Learning to DISCERN online: applying an appraisal tool to health websites in a workshop setting. *Health Education Research* 2004;**19**:440-6.

[9] Shepperd S, Charnock D, Cook A. A 5-star system for rating the quality of information based on DISCERN. *Health Info.Libr.J* 2002;**19**:201-5.

[10] Shepperd S,.Charnock D. Against internet exceptionalism. *The BMJ* 2002;**324**:556-7.

[11] Meola M. Chucking the Checklist: A Contextual Approach to Teaching Undergraduates Web-Site Evaluation. *Libraries and the Academy* 2004;**4**:331-44.

[12] Eysenbach G. Infodemiology: The Epidemiology of (Mis)information. *Am J Med* 2002;**113**:763-5.

Acknowledgements

The Internet school was partly funded by the Change Foundation, Toronto (I3MPACT project).

Address for correspondence

Gunther Eysenbach MD MPH, Centre for Global eHealth Innovation, 190 Elizabeth Street, Toronto M5G2C4, Canada

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

WRAPIN : a Tool for Patient Empowerment within EHR

Michel Joubert[a], Arnaud Gaudinat[b], Célia Boyer[b], Antoine Geissbuhler[b], Marius Fieschi[a], HON Foundation Council members[b]

[a] *LERTIM, Faculté de Médecine, Université de la Méditerranée, Marseille, France*
[b] *Health On the Net Foundation (HON), Geneva, Switzerland*

Abstract

Legal and technologic trends are making medical records progressively more patient-accessible. In the near future, information technology may make it even easier to provide patients a chance to review their records. One may wonder, however, about the practical use of this technology by patients. Understanding his/her own health record will certainly be one of the main concerns of patients. WRAPIN has been designed to provide patients and citizens with trusted health information. It will help to determine the reliability of documents by checking the ideas contained against established benchmarks, and enable users to determine the relevance of a given document from a page of search results. First, we present what is, in our opinion, the most original and important patient-centred WRAPIN characteristics and functionalities. Then, we compare these characteristics with those of representatives of two main trends in information retrieval: systems based on the popularity of web sites, and on the clustering of web sites. This comparison shows that, even though patients are tempted to use popular search engines, these are not sufficiently specialized in the medical domain to help them understand their own HER.. Finally, we discuss the complexity of medical readings over the Internet and the efforts that are still required in this domain.

Keywords:

Abstracting and Indexing; Information Storage and Retrieval; Electronic Health Record; Internet; Quality Control

Introduction

In a short note Dr PH Chew related a visit to one of his patients [1]. The patient said: "Doctor, I would like to inform you that for the past two weeks my *corpus spongiosum* has not been filling adequately (…) I read on the Internet that Metoprolol® can cause impotence." Dr Chew wrote: "I realised that the internet has provided this patient with the knowledge that his recently acquired impotence was drug-induced." It is true that it is now easy to log onto a web site and obtain this kind of information about Metoprolol. We learn from this account that Internet is currently an essential source of information for patients. It can also be used by patients to understand and check diagnoses and treatments.

Legal and technologic trends are making medical records progressively more patient-accessible. In the near future, information technology may make it even easier to provide patients a chance to review their health records. Studies of technology that allows patients to access their records over the Internet have already been conducted [2, 3]. In contrast to paper records, electronic medical records should be perfectly legible. Internet-accessible records can be viewed repeatedly and in the context of rich sources of medical information available on the World-Wide-Web, potentially increasing the potency of the intervention [4].

The Health on the Net Foundation Code of Conduct (HONcode) [5, 6] for medical and health Web sites addresses one of Internet's main healthcare issues: the reliability and credibility of information. The HONcode defines a set of transparency rules to: 1) hold Web site developers to basic ethical standards in the presentation of information; and 2) help ensure readers always know the source and the purpose of the data they are reading. **MedHunt** [7] is a dedicated, medical, full-text search engine, using a global database including HONCode accredited and selected Web sites retrieved by MARVIN, the HON's robot. The global database currently includes two types of documents: 1) the **accredited Web sites** which are updated daily and reviewed manually by the HON team; and 2) the **selected health Web sites** which have been automatically retrieved from the Web by the robot MARVIN.

WRAPIN has been designed to provide patients and citizens with trusted health information [8]. WRAPIN provides an entirely new facility allowing comparisons of health/medical documents with an interconnected knowledge base, in order to discover whether the information exists in the published literature and to provide a summary of the ideas contained. It will help to determine the reliability of documents by checking the ideas contained against established benchmarks, and enable users to determine the relevance of a given document from a page of search results. This will allow patients to gather information in a trustworthy environment. Moreover, the resources referenced in the WRAPIN database are indexed with MeSH key-words. That allows to provide end-users with trusted definitions of health concepts.

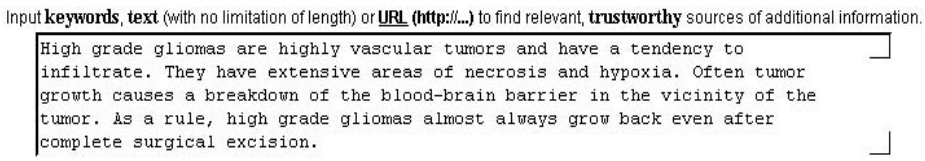


Figure 1.a. Copy and paste of a text submitted to WRAPIN for analysis.

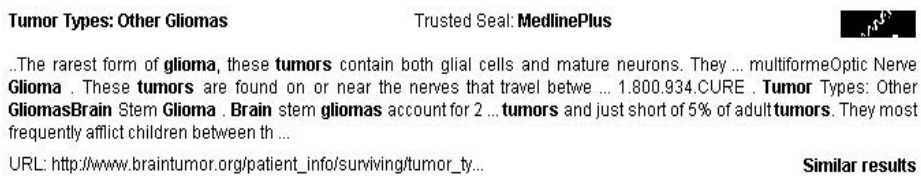


Figure 1.b. An example of a reference to a general public site related to the text submitted to WRAPIN in Figure 1.a.

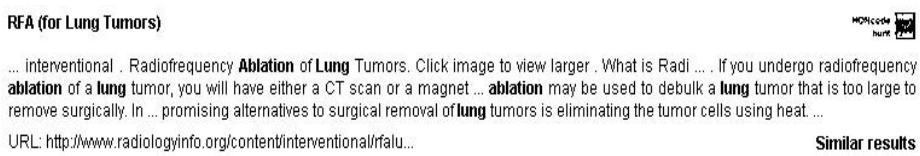


Figure 2.a - Extract of an answer to the query « ablation of a lung ».

A search engine for trustworthy health web sites

The main innovation of WRAPIN is to make available a tool to determine information reliability by automatically checking a document against matching sources from databases of known quality. Figures 1.a and 1.b respectively illustrate the submission of a text to be analysed and an extract of a related Web site of known quality. WRAPIN is fundamentally different from even the best medical search engines currently available, since an entire document may be proposed as the search query. A conventional series of search terms can also be used. A special effort has been made in the indexing process of health resources, since the extraction of MeSH terms from queries is guided by a component based on the UMLS knowledge sources [9].

Whereas other systems display search results in summary form, with bits of text extracted from matching documents, WRAPIN attempts to show how a document matches a query. To better define the results, MeSH terms (and synonyms) are highlighted in the results summaries and throughout the text of matching documents. Figure 2.a and 2.b respectively illustrate an extract from the answer to the question *ablation of a lung tumor*, and the display of a retrieved page where terms of the question are automatically highlighted.

WRAPIN is a multilingual search engine in the medical domain. If a request is expressed in a given language, French for instance, results are displayed in the same language. However, they can be adapted to another language, English for instance. Hence, a patient staying in a given foreign hospital could query WRAPIN and obtain information from data of his/her patient record expressed in his/her natural language. This is particularly significant in Europe where there exist numerous different natural languages, and where people move from one country to another, sometimes everyday, for work reasons.

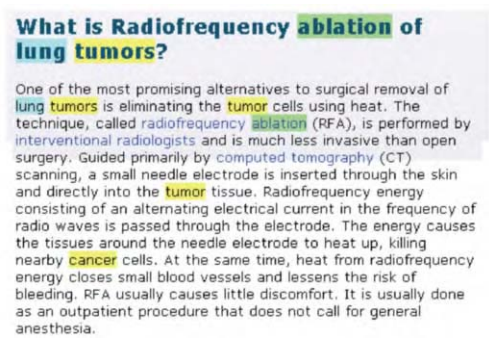


Figure 2.b - An extract of the page referenced in Figure 2.a where the terms of the query are highlighted on fly.

Another way to adapt the information retrieval process to patients' queries is to propose to rephrase queries. In the case of WRAPIN, this is done by suggesting a list of terms or relevant qualifiers, known as facets, aimed at making a query more pre- cise. This is illustrated by Figure 3.a in which the initial query is "glioblastoma". The list of terms used to rephrase a query is created dynamically with regards to the query content and is not predefined. These facets are based on the MeSH thesaurus. WRAPIN exploits other facets based on the medical domain. This kind of facet provides a useful way of helping citizens to better qualify their queries with pre-defined categories linked to a style of query. For instance, a query related to a disease activates categories such as "overview", "causes and risk factors", "screening and diagnosis", "complica-

tions", "treatment", and "prevention", as illustrated by Figure 3.b.

A last feature that WRAPIN exploits is the categorization of sources according to the complexity of the pages they publish. For instance, Medline is considered to be a complex source publishing technical information which is difficult to read by the general public.

Figure 3.a. Rephrasing the query « glioblastoma » by the means of facets.

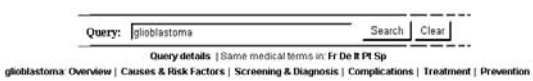


Figure 3.b - Use of pre-defined categories related to entities of the medical domain, here « disease».

A tool for patient empowerment within EHR

Figure 4 shows the general architecture of WRAPIN involving EHR. Extracts of an EHR may serve as a query to the system. Text is analysed, whatever its language (presently five European ones), by a multilingual sub-system represented in a grey rectangle. The result of this treatment is sent to the search engine which exploits an index of medical resources. Results of the query are sent to a module which displays them and suggests rewordings to render the query more accurate. If rewording is performed, a complete cycle is then processed until such time as the user is satisfied with the answers received.

On account of the above functionalities, WRAPIN constitutes an original Web search engine for the health domain. It will allow patients to involve themselves in their own care process, and to ask about and verify information related to their health care. If, for instance, a patient reads in his/her personal French EHR « … exérèse d'un gliome révélé par l'examen anatomopathologique associée à une radiothérapie … » (*…ablation of a glioma, diagnosed by a pathology exam, associated with radiotherapy…*), and if he/she wants more details in relation with this text, a copy-paste of it allows him/her to query WRAPIN and to obtain a list of trustworthy, related Web sites, as illustrated by Figure 5.

Traitement La chirurgie d'exérèse permet l'ablation de la tumeur quand celle-ci est particulièrement bien localisée. Ce traitement s'applique spécifiquement aux astrocytomes et aux craniopharyngiomes. Il existe néanmoins quelques risques de récidive dépendant de la localisation de la tumeur. Ainsi, le gliome et les tumeurs hémisphériques (situées dans chaque moitié du cerveau) sont parfois enlevées partiellement.

La radiothérapie se fait généralement complément du traitement précédent également dans le but de prévenir les récidives.

Figure 5.- French-language answer to the text « exérèse du gliome … associée à une radiothérapie » (ablation of a glioma … associated with a radiotherapy) copied and pasted from a patient EHR..

Facing the ever greater sharing of personal EHR, one may wonder what practical use patients can make of this system. Understanding one's own medical record will certainly be a major concern for patients. In Table1, we compare the WRAPIN features with those of representatives of two main trends in information retrieval: systems based on the popularity of Web sites (e.g., Google), and clustering of Web sites (e.g., Vivisimo).

Discussion

The World-Wide-Web is today probably the world's largest source of information. Among other functions, it enables patients and others to gather information about their health condition, to verify information they received and perhaps did not entirely take in, and to check the validity of their treatment. In short, it allows them to understand. WRAPIN is a tool they can use to query the health Web. One of the main features of WRAPIN is that it registers only certified quality sites. This allows it to convey reliable health information to patients and other users. Its wide-ranging features and its ease-of-use make it an unrivalled tool empowering the patient in his/her own health care process. Even if patients are tempted to use popular search engines, these are not sufficiently specialized in the medical domain to help them understand their own EHR. The comparison we made between features offered by other search engines shows that WRAPIN, as an instrument dedicated to the health domain, provides a better guarantee in terms of information quality and reliability of Web sources. Thus, thanks to its functionalities, it provides an easy-to-use search engine coupled together with EHR.

The WRAPIN search engine has been operational since 2005, having evolved from a prototype which was tested

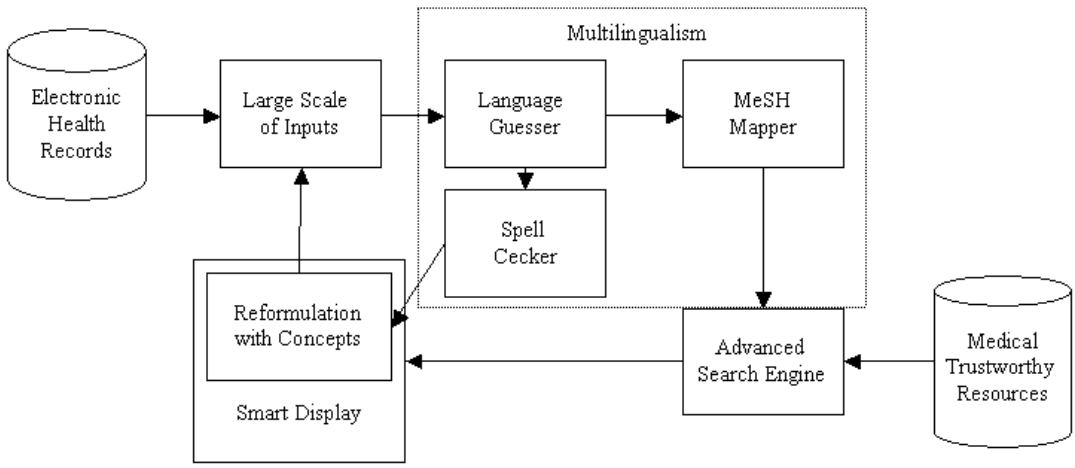


Figure 4 - EHR as an input to the WRAPIN search process.

Table 1 - Comparison of WRAPIN features with those of Google and Vivisimo.

	Google	**Vivisimo**	**WRAPIN**
Querying facilities	Limited in length	Limited in length	Copy and paste a part of a patient record
Quality of information	Related to page ranking only	Related to page ranking only	Related to trustworthy sources
Coverage/overabundance*	901,000,000	821,000,000	112,000
Cross language query**	No	No	Limited to medical terms***
Display of the results	Colorized keywords from cache or toolbar	Colorize from the toolbar	Colorize keywords on fly
Refinement/rephrasing	Limited to English	Corpus-oriented	Corpus-oriented plus a multilingual specific access to diseases
Health literacy****	English facets only for diseases Spelling suggest	No	Health information is categorized by its source Direct access to health definitions Health spelling suggest
Transparency Conflicts of interests	Commercial	Commercial	Non-governmental

Legend :*number of answers to the "health" query, Nov 26, 2006; ** e.g., the ability of a patient to query his/her health record from a foreign country; *** English, Spanish, French, Italian, Portuguese, German; **** access to easy- to-read information

and improved by evaluation outcomes [8]. The goal of this evaluation was to assess if WRAPIN really met to the user needs and particularly to know if it can help patients and individuals to judge information available on the Web. The analysis of the results made clear the important impact of WRAPIN to meeting needs expressed by great number of users. There were three types of questionnaires: technical, satisfaction and ergonomic. The evaluation was online and compared WRAPIN with another Web search engine.

There were four types of queries: free and pre-formatted questions, and free and pre-formatted URL for more than thirty medical specialties. The use of WRAPIN has, on he whole, been perceived as informative, reliable and trustworthy. Yet in a number of cases, the replies given by the system were irrelevant, but it was also true for the other engine. The rate of irrelevance/impertinence should be scientifically reckoned, and finally reduced. It seems the upfront lexical analysis of the queries, as well as the ergonomic of the reformulation leave room for improvement. In the view of the testers, WRAPIN is over-performing other engines when the replies are pertinent. By providing a synthetic and reliable reply to a query, or an assessment of a document further to the submission of an URL, WRAPIN goes beyond the other engines and brings a valuable tool to users seeking medical and health care information. It has been felt helpful that humans reviewed Web sites for appropriateness and correctness.

Usually, when citizens look for health information they look for it on the Internet with search engines like Google. It is obvious that, in order to have a more complete information related to various unclear issues of their medical health record, citizens will rush to Google again and try to clarify them. In this case, unfortunately, their search will not worry the reliability of found information. In this paper, we have presented several criteria which show that WRAPIN can be used by citizens in this context and represents a good and interesting alternative to existing non-specialized search engines. However, a more detailed evaluation should be performed in order to validate the relevance of WRAPIN for such a task and to identify more precisely appropriate functionalities needed within an EHR-oriented perspective.

To facilitate understanding of personal medical information by health consumers, the use of accessible, non-technical language is essential. Information complexity is often quantified on the basis of the complexity of words and sentence formulation, for instance, the number of words in a sentence or syllables in a word [10]. In a domain as specific as health, lexical knowledge is of paramount concern. Today, WRAPIN informs users about the *a priori* complexity of information sources. For instance, Medline is categorized as a technical source compared to Medhunt, since the presence of specific medical terms increases the level of complexity for the general public. Hence, a complexity measure based on the use of a lexicon seems promising.

Moreover, a major effort is still required if patients and non-patients are to have efficient access to health-related information. As noted by McCray in a thorough study into reading cability and the understanding of health-related documents [11], patients need to interact in a variety of health care settings, including doctors' offices, clinics, and hospitals. They also need to interact with a broad range of health-related information, including treatment instructions, patient education materials, prescriptions, bills, and insurance forms. In addition, they are being asked to take increasingly greater responsibility for their own health-care and disease management. The role of the Internet in the health-care system is just beginning to be understood. Still needed are better tools to assess and modify the comprehensibility of health materials as well as methods for improving access to information. McCray's conclusion, to which we totally subscribe, is that "health informaticians, developers of health information, and health care providers all need to work together to ensure that everyone has an equal opportunity to access, understand, and use health information".

Acknowledgments

The WRAPIN project, IST-2001-33260, has been supported by the European Commission. The authors have a special thought for Pr Jean-Raoûl Scherrer who was the originator of WRAPIN and passed away during the project. The authors wish to acknowledge all the members of the WRAPIN consortium, especially Michèle Thonnet, Anne Strauss (MISS, Mission d'Informatisation des Systèmes de Santé), Robert Baud (HUG, Hôpitaux Universitaires de Genève), Patrick Weber (NICE Computing), Pierre Uziel (XR Partner Information Systems), and Juan Bonal (Thales Information Systems). They thank Mr George Morgan who revised their English.

References

[1] Chew PH. Patient empowerment. *BMJ* 2001; 16: 1472.

[2] Cimino JJ, Patel VL, Kushniruk AW. What do patients do with access to their medical records? *Proc. Medinfo* 2001; 10: 1440–1444.

[3] Masys D, Baker D, Butros A, Cowels KE. Giving patients access to their medical record via the Internet: the PCASSO experience. *J Am Med Inform Assoc.* 2002; 9: 181–91.

[4] Ross SE, Lin CT. The Effects of Promoting Patient Access to Medical Records: A Review. *J Am Med Inform Assoc.* 2003; 10(2): 129-38.

[5] Boyer C, Selby M, Scherrer JR, Appel RD. The Health On the Net Code of Conduct for medical and health Websites. *Comput Biol Med.* 1998; 28 (5): 603-10.

[6] Boyer C, Geissbuhler A. A decade devoted to improving online health information quality. *Stud Health Technol Inform.* 2005; 116: 891-6.

[7] Baujard O, Baujard V, Aurel S, Boyer C, Appel RD. A multi-agent softbot to retrieve medical information on Internet. *Proc. Medinfo* 1998: 150-4.

[8] Gaudinat A, Ruch P, Joubert M, Uziel P, Strauss A, Thonet M, Baud R, Spahni S, Weber P, Bonal J, Boyer C, Fieschi M, Geissbuhler A. Health search engine with e-document analysis for reliable search results. *Int J Med Inf.* 2006; 75(1): 73-85.

[9] Gaudinat A, Joubert M, Aymard S, Falco L, Boyer C, Fieschi M. WRAPIN: new health search engine generation using UMLS Knowledge Sources for MeSH Term Extraction from Health Documentation. *Proc. Medinfo 2004*. IOS Press: 356-60.

[10] Kincaid JP, Fishburne RP, Rogers RL & Chissom B.S (1975). Derivation of new readability formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy Enlisted Personnel. *Research Branch Report 8* (75).

[11] McCray A. Promoting Health Literacy. *J Am Med Inform Assoc.* 2005; 12(2): 152–163.

Address for correspondence

Michel JOUBERT
LERTIM
Faculté de Médecine
27, boulevard Jean Moulin
13005 Marseille
FRANCE
e-mail: mjoubert@ap-hm.fr
URL: http://cybertim.timone.univ-mrs.fr/

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

How do Clinicians Search For and Access Biomedical Literature to Answer Clinical Questions?

Annie Y.S. Lau, Enrico Coiera

Centre for Health Informatics, University of New South Wales, Australia

Abstract

This paper presents a retrospective data analysis on how 75 clinicians searched for and accessed biomedical literature from an online information retrieval system to answer six clinical scenarios. Using likelihood ratio measures to quantify the impact of documents on a decision, and a graphical representation to model clinicians' journeys of accessing documents, this analysis reveals that clinicians did not necessarily arrive at the same answer after having accessed the same document, and that documents did not influence clinicians in the same manner. A possible explanation for these phenomena is that people experience cognitive biases during information searching which influence their decision outcome. This analysis raises the hypotheses that people experience the anchoring effect, order effects, exposure effect and reinforcement effect while searching for information and that these biases may subsequently influence the way decisions are made.

Keyword:

evidence journey, information retrieval, cognitive bias, decision making, likelihood ratio, clinicians

Introduction

Increasingly, information searching plays an important part in health consumers' decision making [1] and clinicians' practice of evidence-based medicine [2]. Decisions are improved by better access to relevant information, and searching for documents on the Web is increasingly an important source of that information [3]. Yet, studies have confirmed that clinicians do not always achieve optimal results when using information retrieval systems [4]. While much research focuses on the design of retrieval methods that identify potentially relevant documents, there has been little examination of the way retrieved documents then shape real-life decision making [5]. To develop information retrieval systems that actively support decision making, it is necessary to understand the complex process of how people search for and review information when making decisions [6].

To understand how people search for and use information to make a decision, it is important to understand the "journey" in which people undertake to arrive at the decision. Taking the definition of evidence to be "the body of observed, reported or research-derived data that has the potential to inform decision making" [7], an "evidence journey" is the process that describes an individual accessing different pieces of information retrieved from an online information retrieval system to make a decision. The notion of the evidence journey is similar to Bates' *berrypicking* metaphor, or *bit-at-a-time* retrieval, where "a query is satisfied not by a single final retrieved set, but by a series of selections of individual references and bits of information at each stage of the ever-modifying search" [8].

There is a body of literature looking at how people use information retrieved from search engines to inform their decision making. An example from the information science literature is the model of document use proposed by Wang and colleagues [9,10]. Based on a longitudinal study of academic researchers' use of documents retrieved from online databases during a research project, they proposed that document use is a decision making process and people do not necessarily use the same criteria to select, read or cite documents. Gruppen and colleagues, reporting in the medical decision making literature, suggested that information gathering and selection are more problematic than information integration and use [11]. Based on a study examining how first year house officers select information to make a diagnosis, they found that subjects selected the optimal information in only 23% of the cases but were able to use the selected information to make a diagnosis over 60% of the cases. They suggested that physicians appear to have difficulties recognising the diagnostic value of information, which results in making decisions based on diagnostically weak information.

This study analyses the evidence journeys that clinicians undertake to answer clinical questions. It graphically models the way clinicians searched for and accessed biomedical literature to make clinical decisions and identifies factors in the evidence journey that may have influenced the decision making process.

Methods

Data description

A retrospective analysis was constructed on a dataset of 75 clinicians' search and decision behaviours (44 doctors and 31 clinical nurse consultants), who answered questions for 8 real-life clinical scenarios within 80 minutes in a con-

Selected for best paper award.

trolled setting at a university computer laboratory [12]. Scenarios were presented in a random order, and subjects were asked to record their answers and level of confidence for each scenario. Subjects were then presented with the same scenarios and asked to use an Internet search engine to locate documentary evidence to support their answers.

Subjects recorded their pre- and post-search answers to each question, their confidence in these answers and their confidence in the evidence they had found using the search engine. There were four options to answer each question: yes, no, conflicting evidence and don't know. Confidence was measured by a 4 point Likert scale from "very confident" to "not confident". In addition, subjects recorded any change in answer or confidence from their pre-search response and identified which documents influenced their decision. They were asked to work through the scenarios as they would within a clinical situation and not spend more than 10 minutes on any one question.

Data from only six scenarios for which a correct answer could be identified were analysed (scenario questions are described in Table 1). The unit of measure used in the analysis is a search session, which is "the entire series of queries by a user" [13] to answer one question.

Table 1 - Clinical questions in the scenarios presented to subjects [12]

Question (scenario name)	Expected correct answer
Does current evidence support the insertion of tympanostomy tubes in child with normal hearing? (Glue ear)	No, not indicated
What is the best delivery device for inhaled medication to a child during moderate asthma attack? (Asthma)	Spacer (holding chamber)
Is there evidence for the use of nicotine replacement therapy after myocardial infarction? (MI)	No, use is contraindicated
Is there evidence for increased breast and cervical cancer risk after IVF treatment? (IVF)	No evidence of increased risk
Is there evidence for increased risk of SIDS in siblings of baby who died of SIDS? (SIDS)	Yes, there is an increased risk
What is the anaerobic organism(s) associated with osteomyelitis in diabetes? (Diabetes)	Peptrostreptococcus, Bacteroides

Document likelihood ratio

Since a document may be influencing some subjects to answer in different ways, a quantitative measure is needed to model the impact a document may have on a decision. One simple method is to associate a document with the frequency of correct and incorrect decisions made after having accessed the document. This leads to the idea of using a likelihood ratio (LR) to calculate a ratio of the frequency that accessing a document is associated with a correct answer rather than with an incorrect answer, *P(AccessedDoc|Correct) / P(AccessedDoc|Incorrect)* (Equation 1). The LR thus measures the impact a document has in influencing a subject towards a specific answer. Documents with a LR > 1 are most likely to be associated with a correct answer and a LR < 1 with an incorrect answer.

To calculate the likelihood ratio, the sensitivity and specificity of a document with respect to an answer are calculated. The *sensitivity,* or true positive rate, of a document is the frequency with which the document being accessed correlated with a correct answer being provided post-search (Equation 2). The false negative rate, the frequency with which access of a document correlated with an incorrect answer, was also calculated. The *specificity*, or true negative rate, is one minus the false negative rate (Equation 3). The sensitivity and specificity values were calculated based upon the frequency with which a document was accessed for each scenario.

$$\text{Likelihood ratio} = \frac{\text{P(Accessed Doc} \mid \text{Correct)}}{\text{P(Accessed Doc} \mid \text{Incorrect)}} = \frac{\text{Sensitivity}}{1 - \text{Specificity}} \quad (1)$$

$$\text{Sensitivity} = \frac{\text{No. of correct post-search answers where document was accessed}}{\text{Total no. of post-search correct answers}} \quad (2)$$

$$1 - \text{Specificity} = \frac{\text{No. of incorrect post-search answers where document was accessed}}{\text{Total no. of post-search incorrect answers}} \quad (3)$$

Results

Overall, subjects made 1761 searches and accessed 1873 documents across the 400 search sessions for six scenarios. In a search session, subjects took on average 404.75 seconds (standard deviation (SD): 590.824), made 4.32 (SD: 4.002) searches and accessed 4.65 (SD: 3.670) documents to complete a question.

Across the six scenarios, most subjects improved their answers after searching. There are subjects, as reported in [12], who had a wrong answer before searching and a right answer after searching, wrong-right (WR: 37%), followed by those who never answered correctly, wrong-wrong (WW: 30%), then those who answered correctly before and after searching (RR: 26%), and those who went from right to wrong (RW: 8%).

Since some documents were accessed on only a few occasions, it was not possible to calculate meaningful sensitivity and specificity measures for all documents. Thus, LR was only calculated for the subset of documents that had been accessed by at least five subjects (each document was accessed by 4.7 subjects on average). A total of 725 documents were accessed across the 6 scenarios, with a range from 78 to 138 documents per scenario. After culling, 88 documents were kept in the pool of influential documents (i.e. accessed by more than 5 subjects), with a range from 10 to 19 documents per scenario.

Did clinicians who accessed the same document arrive at the same answer?

Analysis of different evidence journeys taken by study participants reveals that subjects did not necessarily arrive at the same answer after having accessed the same document. In one scenario (MI), around 25% of WW subjects (i.e. 10 out of 39 subjects) cited the same source as WR subjects to support the opposite post-search answer. Across the six scenarios, subjects often produced different answers to questions despite having accessed the same document. The majority of frequently accessed documents were seen *both* by subjects who answered a question correctly after searching as well as those who answered incorrectly (Figure 1).

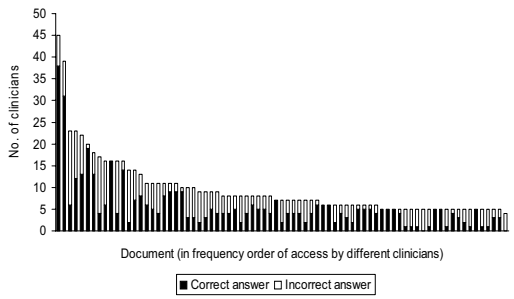


Figure 1 - Frequency of subjects answering a question correctly or incorrectly for each document accessed [14]

In addition, Figure 2 shows that documents accessed were almost equally distributed between those more likely to be associated with a correct answer (51% of accessed documents had a LR > 1) or incorrect answer (49% of accessed document had LR < 1). There was a clear variation in the likelihood that accessing different documents was associated with a subject providing a correct or incorrect post-search answer.

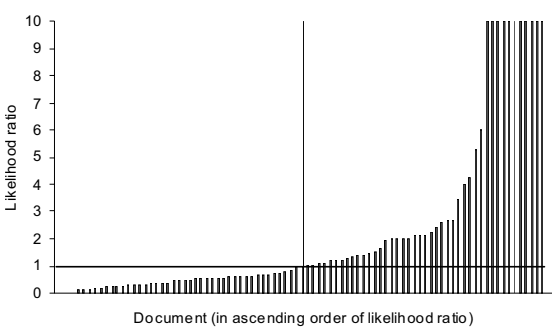


Figure 2 - Variation in influence of accessed documents in obtaining a correct post-search answer, as measured by likelihood ratio (Note: Likelihood ratio = 10 refers to documents having a likelihood ratio ≥ 10) [14]

Are there typical patterns in the evidence journey?

To better understand the way that accessing a sequence of documents might have influenced an individual in making a decision, a qualitative analysis was conducted to look for typical patterns amongst the dataset. In a search session, a *positive document* is a document with LR > 1 and is represented by a closed circle; a *negative document* is a document with LR ≤ 1, represented by an open circle; an *indeterminate document* is a document with a LR that cannot be established and it is represented with a strip-patterned circle; and each query submitted to the search engine is represented with a line.

The following examples demonstrate the evidence journeys of subjects in four different categories: RR, RW, WR and WW.

Figure 3 - Example of a subject with a right answer pre- and post-search (RR)

The subject in Figure 3 was correct before and after searching. This subject expressed being very confident in the pre-search answer. The subject made only one search and accessed one document, which is a positive document (a). One possible interpretation of this evidence journey is that the first document confirmed the subject's pre-search answer; hence, the subject stopped searching and provided the correct answer.

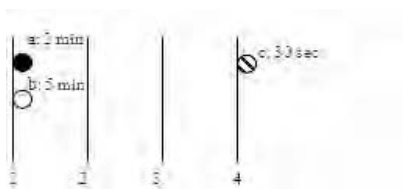


Figure 4 – Example of a subject with a right answer pre-search and a wrong answer post-search (RW)

The subject in Figure 4 gave a correct answer before searching but changed to an incorrect answer after searching. The subject made four searches. On the first search, the subject accessed two documents; the first document was a positive document (a), followed by a negative document (b). The subject then performed more searches, viewed the titles and summaries of documents retrieved on the results pages but did not access any document until the last search, which was a document with an indeterminate likelihood ratio (c). One possible interpretation is that as the subject spent more time on the negative document (b: 5 min) than the other two documents (a: 2 min, c: 30 sec); the extra time spent on the negative document may have contributed to the subject giving an incorrect answer

after searching. (Note: Time spent on a document was measured as time elapsed between the commencement of accessing the document and the subject's next action).

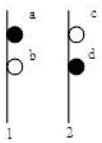


Figure 5 - Example of a subject with a wrong answer pre-search and a right answer post-search (WR)

The subject in Figure 5 gave an incorrect answer before searching, made two searches, accessed four documents and answered the question correctly after searching. The first document was a positive document (a), followed by two negative documents (b and c), and then a positive document (d). One possible interpretation is that the first and the last documents, which are positive documents, influenced the subject to change opinion and make a correct answer.

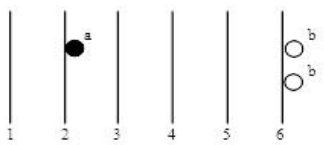


Figure 6 - Example of a subject with a wrong answer pre- and post-search (WW)

This subject in Figure 6 answered the question incorrectly before and after searching. Although the subject only accessed two documents, a positive document (a) and a negative document (b), the subject accessed the negative document twice (b). Perhaps, revisiting the negative document (b) led the subject to retain an incorrect pre-search opinion and provide an incorrect answer after searching.

Discussion

This study provides a snapshot on how clinicians searched for and accessed online biomedical literature to answer clinical questions. The likelihood ratio analysis shows that clinicians did not necessarily arrive at the same answer after having accessed the same document and that documents did not influence clinicians in the same manner. The graphical representation of evidence journeys illustrates that people take different journeys to arrive at an answer to a question and that the way documents were accessed and interpreted collectively may influence the way decisions are formed. It also reveals the possibility that the following components of an evidence journey influence the shaping of a decision outcome:

- an individual's *belief and confidence before searching,* e.g. Figure 3: RR
- the *order* of accessing documents, especially the first and last accessed documents, e.g. Figure 5: WR
- the amount of *time* spent on the documents, e.g. Figure 4: RW
- the *number of visits* the searcher made to the same document, e.g. Figure 6: WW

Each of these phenomena has been identified in the general decision making and information retrieval literature in different guises. Specifically, the literature identifies the existence of *cognitive biases* in the way people use information to make decisions [15-18]. For example, decision makers are biased by their prior beliefs and the order of information presentation when making decisions and forming impressions [19, 20].

The concept of decision biases may provide us with a theoretical model with which to understand search behaviours. Different subjects have different knowledge and levels of confidence before searching; their evidence journeys are different in search length, number and types of documents accessed; the way documents were accessed and how these documents may have interacted are possible elements that influence the way people process and use information to make decisions. Specifically, subjects may be experiencing the following cognitive biases in their evidence journeys:

- *anchoring effect*: the phenomena where one's prior belief exerts a persistent influence on the way new information is processed and subsequently affects the way beliefs are formed [21]
- *order effects:* the way in which the temporal order that information is presented or accessed affects the final judgement of an event [22]
- *exposure effect:* the phenomenon where the amount of time exposed to the information affects the final judgement of an event; e.g. [23]
- *reinforcement effect:* the phenomenon where the level of repeated exposure to information affects the final judgement of an event, which is best related to "mere exposure" discussed in [24]

Impact of these biases have been tested in a preliminary analysis that uses a Bayesian model to predict the impact of information searching on decision making [14]. The Bayesian model that predicts decision outcomes most accurately is the one that incorporates all the above-mentioned cognitive biases during information searching (without biases: 52.8% (95% CI: 47.85 to 57.59); with biases: 73.3% (95% CI: 68.71 to 77.35)).

Analysis limitations

The assumption that subjects read a document based on having accessed the document was a potential limitation in the study. Subjects may not have read documents they accessed, or only partially read them, modifying the likelihood that the document influenced them. Similarly, subjects may have been influenced by documents without accessing them, e.g., looking at the title or the abstract of the document only on the search engine results page, but not accessing the document itself.

Conclusion

This study presents a retrospective data analysis on how clinicians searched for and accessed biomedical literature to answer clinical questions. The analysis suggests that people take different journeys to answer questions, and that the way documents were accessed could contribute to different interaction effects between pieces of information, which influences the way evidence was evaluated and subsequently the decision making process. It also raises the hypotheses that people may experience cognitive biases during information searching that influence their decision making, and calls for further investigation to test these hypotheses.

Acknowledgements

The authors would like to thank Professor Johanna Westbrook, Dr Sophie Gosling and Miss Nerida Creswick for providing the data in this retrospective data analysis. Annie Lau was financially supported by the Australian Research Council during this research.

References

[1] Eysenbach G, Jadad AR. Evidence-based Patient Choice and Consumer health informatics in the Internet age. J Med Internet Res. 2001;3(2):e19.

[2] Hersh WR. Evidence-based medicine and the Internet. ACP J Club. 1996;5(4):A14-A6.

[3] Morrison JB, Pirolli P, Card SK. A taxonomic analysis of what world wide web activities significantly impact people's decisions and actions. Conference on Human Factors in Computing System; 2001; New York: ACM Press; 2001.

[4] Hersh WR, Hickman DH. How Well Do Physicians Use Electronic Information Retrieval Systems? A Framework for Investigation and Systematic Review. JAMA. 1999;280(15):1347-52.

[5] Hersh WR. Ubiquitous but unfinished: on-line information retrieval systems. Med Decis Making. 2005;25(2):147-8.

[6] Spink A, Cole C. New Directions in Cognitive Information Retrieval. 1st ed. London: Springer 2005.

[7] Ford N, Miller D, Booth A, O'Rourke A, Ralph J, Turnock E. Information retrieval for evidence-based decision making. JD. 1999;55(4):385-401.

[8] Bates MJ. The Design of Browsing and Berrypicking Techniques for the Online Search Interface. Online Rev. 1989;13(5):407-24.

[9] Wang P, Soergel D. A cognitive model of document use during a research project. Study I. document selection. J Am Soc Inf Sci. 1998;49(2):115-33.

[10] Wang P, White MD. A cognitive model of document use during a research project. Study II. Decisions at the reading and citing stages. J Am Soc Inf Sci. 1999;50(2):98-114.

[11] Gruppen LD, Wolf FM, Billi JE. Information gathering and integration as sources of error in diagnostic decision making. Med Decis Making. 1991;11(4):233-9.

[12] Westbrook JI, Coiera EW, Gosling AS. Do Online Information Retrieval Systems Help Experienced Clinicians Answer Clinical Questions? J Am Med Inform Assoc. 2005;12(3):315-21.

[13] Jansen BJ, Spink A. An Analysis of Web Documents Retrieved and Viewed. Fourth International Conference on Internet Computing; 2003; Las Vegas, Nevada; 2003. p. 65-9.

[14] Lau AYS, Coiera EW. A Bayesian model that predicts the impact of Web searching on decision making J Am Soc Inf Sci Tec. 2006;57(7):873-80.

[15] Kahneman D, Slovic P, Tversky A. Judgment under uncertainty: Heuristics and biases. New York: Cambridge University Press 1982.

[16] Elstein AS. Heuristics and Biases: Selected errors in Clinical Reasoning. Acad Med. 1999;74(7):791-4.

[17] Croskerry P. Achieving Quality in Clinical Decision Making: Cognitive Strategies and Detection of Bias. Acad Emerg Med. 2002;9(11):1184-204.

[18] Kaptchuk TJ. Effect of interpretive bias on research evidence. BMJ. 2003;326(7404):1453-5.

[19] Tversky A, Kahneman D. Judgment under Uncertainty: Heuristics and Biases. Science. 1974;185(4157):1124-31.

[20] Asch SE. Forming impressions of personality. J Abnorm Soc Psychol. 1946;41:258-90.

[21] Wickens CD, Hollands JG. Decision Making. Engineering Psychology and Human Performance. 3rd ed. Upper Saddle River, NJ: Prentice Hall 2000:293-336.

[22] Wang H, Zhang J, Johnson TR. Human belief revision and the order effect. Twenty-second Annual Conference of the Cognitive Science Society; 2000; Hillsdale, NJ: Lawrence Erlbaum; 2000.

[23] Varey CA, Kahneman D. Experiences extended across time: Evaluation of moments and episodes. J Behav Decis Making. 1992;5:169-85.

[24] Zajonc RB. Attitudinal effects of mere exposure. J Pers Soc Psychol. 1968;9(2):1-27.

Author for correspondence

Annie Lau, a.lau@unsw.edu.au
Centre for Health Informatics,
University of New South Wales,
UNSW Sydney NSW 2052, Australia

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Blogs, Wikis, and Discussion Forums: Attributes and Implications for Clinical Information Systems

Jacob B. Weiss, MS and Thomas R. Campion, Jr., BA

Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, Tennessee, USA

Abstract

Informaticians increasingly view clinical information systems as asynchronous communication systems instead of data processing tools. Outside of health care, popular web technologies like blogs, wikis, and discussion forums have proven to be platforms for effective asynchronous communication. These popular technologies have implications for improving the coordination of clinical care and social support. In order to appropriately evaluate these web-based tools for use in clinical information systems, it will be essential for the informatics community to formally identify the distinguishing attributes of these communication methodologies. The authors propose seven interpersonal and informational attributes to compare and contrast the purposes of blogs, wikis, and discussion forums. This attribute-based approach to analyzing emerging web technologies will lead to a better understanding of the design choices involved in web-based information systems. Two case studies demonstrate how informatics researchers and developers can consider these attributes in the design and evaluation of clinical information systems.

Keywords:

communication, blogs, wikis, social support, medical records

Introduction

Healthcare delivery is a social process, hence the informatics community's perspective on clinical information systems has shifted from that of data processing to an emphasis on asynchronous communication [1]. To coordinate care in a fragmented health system, information systems that promote collaboration between providers and/or patients are necessary. Emerging collaborative technologies may support the communication needs of clinical information systems.

Blogs, wikis, and discussion forums are asynchronous web-based technologies that facilitate the creation and exchange of text, pictures, and other digital media among individuals and groups. The technologies are easy to use and require no knowledge of computer programming. To create content, users typically login to content management systems, which provide straightforward ways to edit and make content available to peers via a web browser. The intended purpose of the technologies differ: blogs facilitate individual expression, wikis facilitate document creation, and discussion forums facilitate group dialog. Applications of the technologies in medicine are beginning to appear in the literature, but references to these concepts sometimes are misidentified or treated interchangeably [2]. Comparisons of the general advantages and disadvantages of the technologies exist, but little has been done to suggest methods for comparing the technologies' socio-technical characteristics [3].

For the medical informatics community to design and evaluate clinical information systems using aspects of emerging web technologies, the attributes of these tools must be distinguished. In this paper, we propose seven defining attributes of blogs, wikis, and discussion forums and assess the value of each attribute with respect to each technology. We present two case studies applying these attributes to clinical information systems: one evaluates a "problem list" in an electronic medical record, and the other explores online social support for cancer survivors. This paper is intended to stimulate discussion of these promising technologies and their attributes.

This paper is written for individuals and organizations that are developing or evaluating health communication software. For those trained in medical informatics, gaining an understanding of blogs, wikis, and discussion forums can provide insight to improve clinical information systems. For those trained outside medical informatics, viewing clinical systems through the lens of popular social software can illustrate the unique communication challenges of health care to which they can apply their knowledge and experience. With a better understanding of both collaborative software and the socio-technical underpinnings of informatics, informaticians and their industry partners can improve the design and evaluation of clinical information systems.

Analysis

In the following sections, we define blogs, wikis, and discussion forums in terms of the attributes presented in Table 1. In all three technologies, we refer to "entries" as content that is novel and discrete. "Responses" are replies or modifications to entries. We refer to creators of entries as "authors" and creators of responses as "respondents."

Selected for best paper award.

Table 1 - Seven attributes of blogs, wikis, and discussion forums

	Attribute	Blogs	Wikis	Discussion Forums
1	*Publication model*	**One-to-many**	**Many-to-many**	**Many-to-many**
2	*Display of entries*	**Reverse chronological**	**Topical**	**Chronological within topics**
3	*Response to entries*	**Append**	**Change**	**Append**
4	*Display of knowledge*	**Fragmented across entries and replies**	**Synthesized into a single document**	**Fragmented across entries and replies**
5	*Participant emphasis*	**Author > respondent**	**Author = respondent**	**Author = respondent**
6	*Visibility of author/ respondent identity*	**Strong**	**Weak**	**Strong**
7	*Tone of voice*	**Informal**	**Formal**	**Informal**

We define "identity" as a user's real name/credentials or as a persistent online pseudonym.

Blogs

A blog (an abbreviation for "web log") is a website that facilitates individual expression using a one-to-many publication model, in which a single author creates entries that any number of readers can access (attribute 1) [4]. Blog software displays entries in reverse chronological order (i.e. most recent at the top) (attribute 2). In a blog, a response is called a "comment" and consists of a text reply and respondent's name. Comments are appended to entries and do not replace the original content of the entry (attribute 3). Knowledge, the interpreted understanding of information, is fragmented across the various entries and responses; blogs lack a formal structure for the synthesis of information contributed over time (attribute 4). The blog interface places greater emphasis on the author's entry than the respondents' comments (attribute 5). For example, the interfaces for commenting often have fewer options compared to the interface for writing original entries. Comments frequently are displayed with a smaller font and sometimes are visible only by visiting a page separate from the main page of entries. Furthermore, the blog author has control over the privacy levels of the entries, and the author also has the ability to delete comments submitted by others. The identity of each individual contributor is visibly labeled for each entry and comment, except in the case where authors permit respondents to be anonymous (attribute 6). Related to the purpose of self-expression, the tone of voice in blogs is informal and spontaneous. Examples of existing health care related blogs include physicians and medical students sharing experiences and discussing clinical topics [5, 6], as well as patients updating family, friends, and/or the general public about treatment, recovery, and healthcare experiences [7].

Discussion forums

A discussion forum is a website that facilitates group dialog using a many-to-many publication model; each user can author entries that any number of readers can access (attribute 1). When an author creates an entry, called a "topic," the intent is for dialog to occur on a subject specified by the author. Users freely respond to the topics of their interest. The discussion forum interface lists topic subject lines together on one page, separating the dialog of each topic to individually accessible sub-pages. Forums also can include hierarchical categories and sub-forums akin to folders in a file system for organizing topics (attribute 2). Each topic page displays responses in chronological order (i.e. most recent at the bottom), such that new responses are appended to existing ones (attribute 3). Despite the topical and chronological arrangement, knowledge still can be fragmented across entries and responses over time (attribute 4). The typical discussion forum interface provides the same formatting for all responses in a topic so that all appear equal. Additionally, the creator of a topic and all respondents have the same abilities to include text, images, and other media when they contribute content (attribute 5). The discussion forum interface displays a user's identity alongside each topic or response he or she contributes (attribute 6). The tone of voice in forums is informal and conversational, often with the use of the first-person speaking voice (attribute 7). Discussion forums are a mature technology, and healthcare researchers have studied them extensively in terms of online social support for cancer patients and other purposes [8].

Wikis

A wiki is a website that facilitates document creation using a many-to-many publication model; any user can be an author by creating entries that any number of users can view (attribute 1). In a wiki, an entry is the creation of a page about a specific topic (attribute 2), and a response is modification to an existing entry (attribute 3). Wiki entries can link to other entries, as well as to other web resources. Because an entry's content is updated and/or replaced with new information, knowledge pertaining to a given topic is continually synthesized and displayed in a single cohesive document (attribute 4). The wiki interface provides no easily visible connection between individuals and their contributions (attribute 6). The interface also de-emphasizes any distinction between the original author and the respondents in favor of stressing the entry's content (attribute 5). Because the purpose of a wiki is to create documents, authors and respondents tend to write in a formal tone (attribute 7). Examples of healthcare wikis include Wikicancer, a social support resource for cancer patients and their families [9], and Clinfowiki, a growing

encyclopedia of medical informatics concepts sponsored by the Informatics Review [10].

Case studies

The authors present two case studies that demonstrate a systematic approach to analyzing the design of clinical information systems in terms of the attributes defined in Table 1. The first case study evaluates a "problem list" in an electronic medical record, and the second assesses online social support for cancer survivors.

"Problem list" in an electronic medical record

StarPanel, a web-based electronic medical record (EMR) developed at Vanderbilt University, contains a "problem list" for quick viewing (Figure 1) and/or updating (Figure 2) of a patient's current diagnosis, medications, and related clinical information [11]. Many aspects of the problem list resemble a wiki. All providers can create and access the problem list content (attribute 1), as well as modify the information (attribute 3). The interface presents information by clinical topic—"significant medical conditions, significant procedures, allergies and drug reactions, current medications, health maintenance, social and family history" (attribute 2) [11]. An advantage of this system attribute is that information stored by topic in discrete fields can be reused in other parts of the EMR [11]. Only the most recent version of the list is displayed; the chronology of individual updates is not immediately visible, although the user can access the previous versions of the problem list to look for changes. The interface provides a single, patient-centered source for quick access to the most recent and most relevant clinical information (attribute 4). The original author's entries are not presented differently from or emphasized more or less than each respondents' changes, although the most recent respondent is highlighted at the top of the page (attribute 5). The interface does not clearly identify each provider's contributions to the problem list (attribute 6), which may interfere with the ability of providers' to coordinate care. This example of clinical documentation features a tone of voice that is straightforward, objective, and formal (attribute 7). However, the tone of voice in the text can also be considered informal due to sentence fragments and piecemeal thoughts, which make the entry somewhat more blog-like (attribute 7).

This case study illustrates how the proposed attributes can be used explicitly to evaluate a clinical information system. For the most part, the StarPanel problem list has the attributes of a wiki. A disadvantage of the current implementation is that users cannot easily determine the individuals who contribute specific parts of the problem list. As personal health records and patient portal systems begin to give patients the ability to update portions of their own medical records, the system interface must clearly display the identity and role of the person responsible for each specific update. Rather than viewing this complication as a reason against using a wiki-based approach for the problem list, one can instead determine which of the wiki-like attributes is problematic. Simply switching to a discussion forum or blog model would not be satisfactory because this shift might alter other attributes that reduce the effectiveness of the problem list. The developers therefore can focus on redesigning the application with respect to the visibility of a contributor's identity (attribute 6) rather than redesigning the entire interface.

The attributes defined in this paper do not offer a specific design solution for each problem, but the interfaces of existing blog and discussion forum software may provide inspiration for different ways to identify the author of a specific entry. Informaticians also can study wiki-like software to learn how these applications have addressed similar challenges. The proposed attributes should be used as an objective guide to support the subjective judgment of the developers and researchers in making design choices. A possible resolution for this problem list implementation might be to display the role of the individual who entered each item by displaying each portion of the text with different background colors that represent "primary care physician," "specialist," "nurse," and "care tech."

Figure 1 - Problem list view

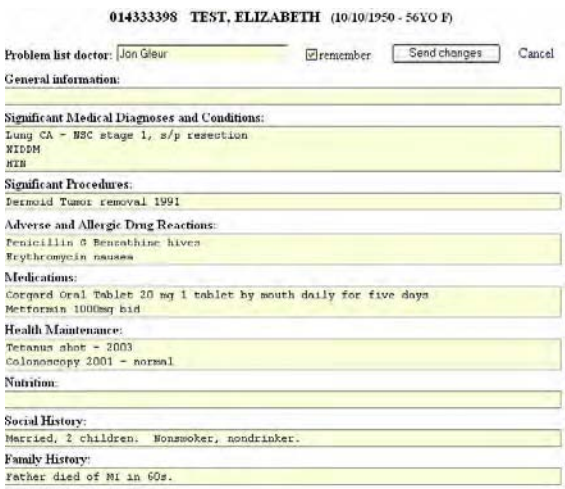


Figure 2 - Problem list edit

Case study: Online social support for cancer survivorship

An estimated 10 million individuals treated for cancer are living in the United States today [12]. Improving the qual-

ity of life of these individuals currently is a clinical challenge and priority research goal [13]. The literature includes a growing number of studies that address the use of online cancer support groups by patients and/or their informal caregivers [14, 15]. Social support is a complex concept and includes dimensions of informational support as well as emotional support [16].

Online social support interventions in health care have focused on discussion forums, newsgroups, chat rooms, and email listservs [14]. Online support groups only recently are beginning to include online communication technologies with blog and wiki functionality [9, 17]. Cancer patients and caregivers have used discussion forum software to facilitate both informational and emotional social support. Messages in online cancer support groups often include the sharing of information, encouragement, humor, and prayer [18]. Researchers have not fully considered when discussion forums might or might not be the most appropriate asynchronous collaboration tool for online social support

Independently addressing each aspect of social support might better address the advantages and disadvantages of discussion forums, blogs, and wikis. Blogs may be a powerful interface for informal, personal expression related to cancer diagnosis and survivorship (i.e. emotional support). Additionally, blogs may be a good format for providers to inform patients of timely announcements, such as new clinical trials or fundraiser events (i.e. informational support), as they present information in a reverse chronological fashion (attribute 2). A discussion forum might be appropriate for the process of discussing information that involves a variety of opinions, such as advice on certain types of alternative treatments. However, the information and knowledge created through these discussions will be fragmented in messages over time. A wiki might be an appropriate tool for creating an information resource that summarizes the collective knowledge of cancer survivors, family, friends, and providers on certain topics [17], such as what to expect during the initial period following chemotherapy and radiation treatment [19]. Adjusting the interface to clarify the identity or role of the individual who contributes each part of the document could improve the credibility of the information, as discussed previously in terms of a collaborative problem list. These examples indicate that interactive health collaboration systems may need to offer a variety of communication technologies to facilitate both the emotional and informational social support needs of cancer survivors.

Discussion

Developers of clinical applications create interfaces that inherently have attributes representative of wikis, blogs, discussion forums, or some combination of the three. These choices will fundamentally affect the way in which providers and patients use health information systems [20]. For example, if a patient, her providers, and her family members can all update a personal health record, should entries by the patient be emphasized more than the others, or should the patient, doctors, and family members all be emphasized equally in the interface design? Different sets of attributes may be appropriate for different medical or social contexts. These choices have implications for patient-provider relationships, information ownership, information filtering, social support, and other socio-technical aspects of clinical care.

By breaking down each technology into its fundamental elements, developers can select the desired properties of blogs, wikis, and forums to create interfaces that better match the goals of certain information system components. Several authors have suggested that further research is needed on how communication tools can effectively work in parallel and how they can be combined to create new technologies [3, 21-23]. For example, a “bliki” is a term that describes the combination of blog and wiki functionality [21]. The attributes proposed in this paper represent a first step toward a formal method of generating new ideas for system design by combining specific properties of emerging web collaboration software. The method could apply to health care applications as well as web-based communication applications in other domains.

This study has limitations that warrant discussion. As seen in the evaluation concerning tone of voice in the problem list case study, it is not always clear which attribute value is the most appropriate description for a given interface. Further refinements to the attribute definitions by the authors and the informatics community will be needed. The attributes are based on the literature and the authors' personal involvement in design and use of web technologies, and experimentally validating the effectiveness of these definitions on existing software applications is planned for future studies. We have excluded podcasts, newsgroups, instant messaging, and other popular technologies from the current discussion because they do not meet our inclusion criteria of primarily text-based, asynchronous web-based software. Additional analysis should extend the attributes defined in Table 1 to address these technologies as well.

Conclusion

Formally analyzing the attributes of emerging web technologies will support the creation of improved design approaches for health communication. The authors define seven attributes for researchers and developers to use in the design and evaluation of clinical information systems. Two case studies illustrate the application of the proposed attributes to improve the coordination of clinical care and social support.

Acknowledgements

The authors would like to thank Drs. Eta S. Berner, Joshua C. Denny, and Dario A. Giuse for their contributions to this paper. This work was supported by NLM T15 007450-03

References

[1] Coiera E. When conversation is better than computation. J Am Med Inform Assoc. 2000 May-Jun;7(3):277-86.

[2] Mongkolwat P, Kogan A, Koh J, Channin DS. Blogging your PACS. J Digit Imaging. 2005 Dec;18(4):326-32.

[3] Boulos MN, Maramba I, Wheeler S. Wikis, blogs and podcasts: a new generation of Web-based tools for virtual collaborative clinical practice and education. BMC medical education [electronic resource]. 2006;6:41.
[4] Mortensen T, Walker J. Blogging Thoughts: Personal Publication as an Online Research Tool. In: Morrison A, ed. *Researching ICTs in Context.* Oslo: InterMedia, University of Oslo 2002.
[5] Centor R. DB's Medical Rants. [cited 2006 November 30]. Available from: http://www.medrants.com
[6] Walker G. Over!My!Med!Body! [cited 2006 November 30]. Available from: http://www.grahamazon.com/
[7] CaringBridge: Free online service connecting family and friends. [cited 2006 November 30]. Available from: http://www.caringbridge.org/
[8] Gustafson DH, Hawkins R, Pingree S, McTavish F, Arora NK, Mendenhall J, et al. Effect of computer support on younger women with breast cancer. J Gen Intern Med. 2001 Jul;16(7):435-45.
[9] The Cancer Survivors Network, supplying information, resources, chat, stories, discussion forums and support for cancer survivors and caregivers. [cited 2006 November 30]. Available from: http://www.acscsn.org
[10] Clinfowiki. [cited 2006 November 30]. Available from: http://www.clinfowiki.org/wiki/index.php/Main_Page
[11] Denny J, Guise DA, Jirjis JN. The Vanderbilt experience with electronic health records. Sem in Colon and Rectal Surg. 2005 2005;16(2):59-68.
[12] Cancer survivorship--United States, 1971-2001. Morb Mortal Wkly Rep. 2004;53(24):526-9.
[13] Hewitt M GS, Stovall E. From Cancer Patient to Cancer Survivor: Lost in Transition. Washington, DC: The National Academies Press 2005.
[14] Eysenbach G, Powell J, Englesakis M, Rizo C, Stern A. Health related virtual communities and electronic support groups: systematic review of the effects of online peer to peer interactions. BMJ (Clinical research ed. 2004 May 15;328(7449):1166.
[15] Klemm P, Bunnell D, Cullen M, Soneji R, Gibbons P, Holecek A. Online cancer support groups: a review of the research literature. Comput Inform Nurs. 2003 May-Jun;21(3):136-42.
[16] Williams P, Barclay L, Schmied V. Defining social support in context: a necessary step in improving research, intervention, and practice. Qualitative health research. 2004 Sep;14(7):942-60.
[17] Cancer: Stories, support and information. [cited 2006 November 30]. Available from: http://www.wikicancer.org/
[18] Klemm P, Reppert K, Visich L. A nontraditional cancer support group. The Internet. Computers in nursing. 1998 Jan-Feb;16(1):31-6.
[19] Stanton AL, Ganz PA, Rowland JH, Meyerowitz BE, Krupnick JL, Sears SR. Promoting adjustment after treatment for cancer. Cancer. 2005 Dec 1;104(11 Suppl):2608-13.
[20] Wyatt JC, Wright P. Design should help use of patients' data. Lancet. 1998 Oct 24;352(9137):1375-8.
[21] Bliki. [cited 2006 November 30]. Available from: http://en.wikipedia.org/wiki/Bliki
[22] Woods L. Blogs & Wikis: Technologies for Enterprise Applications? [cited 2006 November 30]. Available from: http://gilbane.com/gilbane_report.pl/104/Blogs__Wikis_Technologies_for_Enterprise_Applications.html
[23] Herring S, Scheidt L, Bonus S, Wright E. Bridging the Gap: A Genre Analysis of Weblogs. HICSS '04: Proceedings of the Proceedings of the 37th Annual Hawaii International Conference on System Sciences (HICSS'04) - Track 4; 2004: IEEE; 2004.

Address for correspondence

Jacob Weiss
400 Eskind Biomedical Library
2209 Garland Ave.
Nashville TN 37232
USA
jacob.weiss@vanderbilt.edu

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

From Intermediation to Disintermediation and Apomediation: New Models for Consumers to Access and Assess the Credibility of Health Information in the Age of Web2.0

Gunther Eysenbach [a,b]

[a] *Department. Of Health Policy, Management, and Evaluation, University of Toronto, Canada*
[b] *Centre for Global eHealth Innovation, University Health Network, Toronto General Hospital, Toronto, Canada*

Abstract

This theoretical paper discusses the model that, as a result of the social process of disintermediation enabled by digital media, traditional intermediaries are replaced by what this author calls apomediaries, which are tools and peers standing by to guide consumers to trustworthy information, or adding credibility to information. For apomediation to be an attractive and successful model for consumers, the recipient has to reach a certain degree of maturity and autonomy. Different degrees of autonomy may explain differences in information seeking and credibility appraisal behaviours. It is hypothesized that in an apomediated environment, tools, influential peers and opinion leaders are the primary conveyors of trust and credibility. In this environment, apomediary credibility may become equally or more important than source credibility or even message credibility. It is suggested to use tools of network analysis to study the dynamics of apomediary credibility in a networked digital world. There are practical implications of the apomediation model for developers of consumer health websites which aspire to come across as "credible: Consumers need and want to be able to be co-creators of content, not merely be an audience who is broadcasted to. Web2.0 technology enables such sites. Engaging and credible Web sites are about building community and communities are built upon personal and social needs.

Keywords:

internet, consumer health informatics,
information quality, credibility

From disintermediation to "apomediation"

The debate on quality and credibility in the digital age is a result of a social process of disintermediation through digital technologies, and the health industry is no exception: Just as in many other areas of life (e.g., travel industry), information and communication technologies empower consumers and enable them to cut out the middleman or intermediary (travel agents, real estate agents, librarians, pharmacists, health professionals) to access pertinent information or services directly, whenever they need it and where they need it. For instance, on the Internet, consumers can now not only access an unprecedented amount of health information, but increasingly also personal information from their electronic health record [1]. With direct and convenient access to an abundant amount of health information on the Internet, consumers now bypass the expert intermediary and gain direct access to unfiltered information [2]. In this situation, consumers have to assume new responsibilities for assessing the credibility of the information, and intermediaries sometimes defend their role as "gatekeeper" using quality arguments.

As the role of "human" intermediaries diminishes or changes, consumers and patients are finding new ways to arrive at relevant and credible information. This can be human beings (peers) and/or technology (e.g., collaborative filtering tools).

In this paper, the author proposes to refer to these new intermediaries as "**apomediaries**", because they mediate without standing "in between" consumer and services or information. Rather, they "stand by" and provide added value from the outside, steering consumers to relevant and high-quality information without being a necessary requirement to obtain the information or service (Fig. 1). While intermediaries provide "upstream filtering", apomediaries enable and facilitate "downstream filtering" [3]. Apomediaries can help to navigate through the onslaught of information, give additional credibility cues, and provide metainformation. Examples for apomediaries are consumer ratings on Amazon or epinions, technologies like PICS or MedPICS labels and their semantic web successors [4;5] enabling the machine-processable dissemination of such ratings, collaborative filtering and recommender systems such as StumbleUpon.com, and other second generation (sometimes called Web 2.0) Internet-based services that let people collaborate and share information online in a new way - such as social networking sites, wikis, communication tools, and folksonomies.

Disintermediation not only takes place on a society level in health care and other industries, but there are also parallels to the individual emancipation process that takes place for example during puberty, when adolescents strive to become more autonomous and have the desire to reduce the influence of the intermediary (parents), with peers

(apomediaries) partly taking over the role of the former intermediary.

The disintermediation/apomediation model is useful because it allows us to analyze and discuss the implications of the disintermediation process at the societal level, for example for consumers of entire industries (ehealth), and to draw analogies to what is happening at the individual level during adolescence due to the emancipation process from traditional authorities and the use of digital media. These observations are free of judgment – it is not implied that the disintermediated / apomediated model is always better than the intermediated model. Rather, which model is "better" depends on the individual and the respective situations. In the following, the author will first discuss general implications if disintermediation takes place, and will then consider credibility implications.

Table 1 - Dichotomies in the intermediation versus apomediation model.

Intermediation Model	**Apomediation Model**
Dependency–Paternalistic System–Acute Illness–Pre-adolescent Kids–Illiterate Consumers	Autonomy–Net–Chronic Illness–Adolescents–Literate Consumers
Traditional	Digital
Centralized	Highly Networked
Managed environment	Autonomy, emancipation
Dependence on Intermediaries (physicians, parents)	Guidance by Apomediaries (peers, Web 2.0 technology)
Credibility of Authorities/Experts	Credibility of Peers
Power held by intermediaries	Empowerment of consumers/youth
Source expertise = traditional credentials (seniority, professional degrees etc)	Source expertise = first-hand experience, peers
Message credibility: professional language, message "length is strength", comprehensiveness signifies expert status	Message credibility: Understandable language, "street cred"
Top-down	Bottom-up
More formal learning	More informal learning
Static hubs	Dynamic hubs
Source credibility more important than message credibility	Message credibility and credibility of apomediaries more important than source credibility

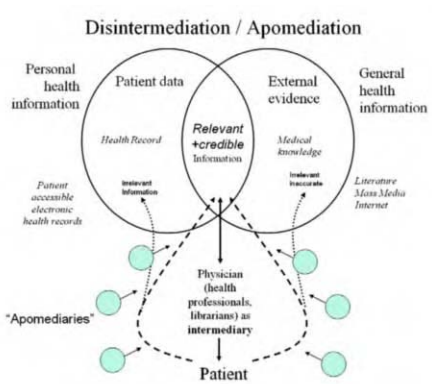


Figure 1 - Disintermediation and apomediation (circles = apomediaries assisting in "downstream filtering")

General implications of disintermediation

Decreased reliance on the intermediary

Disintermediation enabled or enhanced through technology changes the role of the recipient (consumer, patient, youth), who now has the choice to determine whether, when and how they choose to use the intermediary. The better informed the recipient is (or perceives to be), the better he knows what information or services he needs, the less likely he will need an intermediary. For example, a consumer with a chronic condition (e.g., diabetes) will have a greater knowledge and self-efficacy to critically appraise information found on the Internet than a consumer with an acute illness, and will not need an intermediary. Similarly, an older adolescent eager to learn about sexuality is less likely to rely on an intermediary such as a parent or teacher as filter than a younger child.

With increased literacy (including the ability to distinguish different types of information) and knowledge, i.e., when the receiver is knowledgeable about message content, the effects of source expertise will be attenuated, i.e., the credibility of "experts" and other authorities decreases [6], leading to an interesting positive feedback loop, where consumers learn to rely less and less on experts or intermediaries, preferring apomediation instead (Fig. 2). Again, there are parallels to what is happening during adolescence when youths learn to emancipate themselves from traditional authority figures.

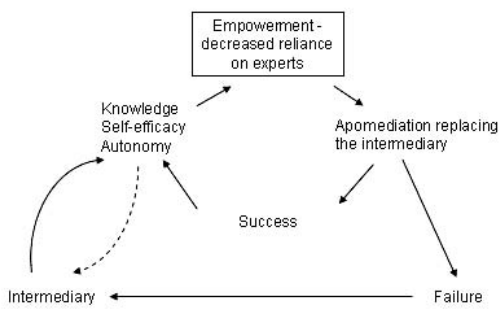


Figure 2 - Dynamic Disintermediation/Apomediation Model: Information is initially mediated and filtered by an intermediary. Once a critical threshold of knowledge, self-efficacy and autonomy is exceeded, apomediation can replace the traditional intermediary, while the recipient still has the option to choose the intermediary in case of failure.

Power shift

As a result of disintermediation, the power-relationship between recipient and intermediary changes. This may create conflicts. For instance, a significant minority of health care providers see their authority challenged, perceive a deterioration in the physician-patient relationship, and fear a negative impact on the quality of health care or health outcomes, although most embrace the shift from a paternalistic to a partnership model [2;7-12]. Parents may be equally irritated when the youths are searching for information on birth control on the Internet rather than discussing these issues with them or a physician.

Concerns have been expressed that more information does not necessarily translate into more knowledge or better

quality (self-)care, in particular as some of the quality and credibility of health information on the Internet clearly does not live up to professional standards. This view fails to recognize apomediation as an alternate mechanism to filter information.

Interestingly, the model presented above (Figure 1) also helps to explain some of the perceptions and frustrations intermediaries (health professionals, parents) often have with the disintermediation process, as they tend to see predominantly the "failures", for instance patients having found irrelevant information on the Internet.

Interpreting the message and filtering for relevance

It is interesting to note that when physicians express discontent about patients bringing stockpiles of "low quality" Internet printouts in the doctor's office, the primary complain is not so much the low quality or credibility of that information in an objective sense, but its *irrelevance* for the specific patient. In one survey 73.8% of General Practitioners said about this information that it was "accurate", and 65% of physicians even said the information was new to them, but many thought that the information often does not apply to that particular patients' condition – in the same survey, only 44.7% said the patient correctly interpreted information [13]. Another survey found that those providers who thought much of the information was often irrelevant also had a more negative view on how the patient-physician relationship was affected [11].

To what degree apomediaries actually succeed in the same way as intermediaries such as physicians or parents to filter information is largely an unexplored area. While there have been data on the "self-corrective" nature of digital media, to the effect that for example inaccurate statements on mailing-lists are corrected by peers [14], relatively little is known to what degree "apomediation" helps to filter *relevant* information and to contextualize information.

Credibility implications

The shift from intermediaries to what the author proposes to call apomediaries has implications for credibility constructs. Some credibility implications and research questions arising from the disintermediation/apomediation model are discussed in the following.

Explaining group and situational differences

Disintermediation usually means that people can, but must not use the intermediary, leading to groups of people who (or situations where people) continue to use intermediaries, trusting a more central authority, and others where people use apomediaries, relying on more decentralized, "networked" mechanisms to infer credibility (reputed, tabulated etc.). Reviewing audience factors in Web credibility research, Metzger [15] reminds us that "credibility is highly situational" and that "demographics, Internet use and reliance, and issue involvement have been found to influence perceptions of the credibility of Web-based information". The disintermediation/apomediation model hypothesizes that factors such as self-efficacy and perceived or desired autonomy (freedom from authority and capacity to make an informed decision) play a critical role in information seeking and credibility assessment behaviour. In other words, autonomous individuals are more likely to choose an apomediation model, and apomediation in turn increases autonomy. Navigation in the digital world requires and at the same time enables a high degree of *autonomy.*

As mentioned above, there are parallels between certain groups of consumers who prefer a more paternalistic patient-doctor relationship and pre-teen dependent children on one hand, and empowered consumers and teenagers striving for autonomy on the other hand. For example, a chronically ill patient is more autonomous than a traffic accident victim, and will be more likely to seek out information from peers. A pre-adolescence teenager is more dependent on parents and other authority figures than an older teenager, who deliberately seeks autonomy and questions authority – in this developmental stage, apomediaries such as peers and peer-to-peer technology gain in attractiveness and relevance.

Teenagers, people with chronic illnesses, educated people all share a common desire and capacity for autonomy, hence their credibility assessment heuristics are more on the right side of Table 1 above. On the other hand, younger kids, seniors, incapacitated people, people with acute diseases etc. are more likely to willingly submit themselves to a system in which they are dependent on intermediaries. The model proposed here theorizes that the desire for autonomy or perceived autonomy are predictors for differences in the interpretation of credibility cues (e.g., formal professional credentials versus "street cred") and differences in information seeking and verification behaviours, and explains differences between pre-Internet and Internet generation, information seeking strategies of chronic versus acute patients, pre-teens versus teens, and illiterate versus literate consumers.

What is argued here is that variables like *autonomy, self-efficacy and knowledge* (thought to be predictors for embracing disintermediation) are presumably correlated with *motivation* (issue involvement, including knowledge and personal relevance of some topic) and *ability* (e.g., cognitive abilities, literacy, time), which, according to the elaboration likelihood model (ELM) of persuasion [16], affect message processing and, thus, influence message effects and credibility judgments. The ELM theorizes that higher issue involvement (motivation) and ability (which together affect what is here called autonomy) will lead to more effortful processing of a message (central route to persuasion), while lower motivation and ability will favour a peripheral route, where environmental characteristics of the message, like the perceived credibility of the source, presentation, or the attractiveness of the source are the primary credibility cues. An extension of this model has been proposed (though not empirically tested) by Fogg and Teng, who hypothesize that people with lower motivation and ability (those who are persuaded through a peripheral route) are more likely to adopt a binary evaluation strategy (credible or not credible), whereas people with higher motivation and ability employ a spectral evaluation strategy [17]. Thus, if we accept that in

most cases autonomy is highly correlated with general motivation and ability, and if we accept that these variables also predict whether or not somebody chooses to emancipate themselves from a gatekeeper, and if we believe the predictions made by the ELM and its extensions, then the line of argument presented above becomes clear.

Boosting credibility through disintermediation

"Direct", unmediated information is often is perceived as more credible because with "greater apparent mediation comes greater opportunity to impute motives and intentions of the communicator" [18], which is one of the most cited reason for why (live) television is perceived as more credible than newspapers [15]. In other words, disintermediation has the potential to increase the credibility of information. This is particularly true the health care field, where many consumers have a "healthy" (and sometimes not completely unjustified) mistrust in a system where doctors are paid per service, and where payers are under considerable cost-pressures leading to a perceived rationing of publicly available services. In addition, many consumers view the traditional health care system as being biased against alternative medicines [19], as health care professionals are incentivized to offer expensive therapies for which they are reimbursed more generously, as opposed to therapies which are "natural" but for which they cannot charge much. Such mistrust creates the desire to bypass the intermediary, and boosts trust in information which can be received without intermediaries.

While youths will rarely rationalize their mistrust against traditional intermediaries in a similar way, questioning and mistrusting traditional authorities is a natural part of adolescence, and information mediated through traditional authorities is often perceived as biased.

Reinstating trust in the intermediary

It can also be hypothesized that transparency confirms the trustworthiness of intermediaries who "step aside" allowing and perhaps even facilitating direct access to information and transparency. In other words, once disintermediation has taken place, disintermediation has the potential to reinstate trust back to the intermediary if information provided through more direct channels proofs to confirm the information the intermediary used to provide. For example, health professionals who allow and actively encourage patients to access their own electronic health records help to reinstall patients' trust in the medical system. If however the information now obtained from other channels is perceived to contradict information from the former intermediary, then the trust-relationship will be undermined. For example, youth accessing information on issues of sexuality through the Internet will loose trust in parents and teachers if this information contradicts the information these intermediaries have provided.

Experiential credibility

While traditional wisdom from credibility research suggests that perceived "accuracy" is a hallmark for message credibility [15], it would be a mistake to assume that "accuracy" means evidence-based information based on scientific studies and that evidence-based information based on research would automatically have more credibility for consumers than anecdotes. In a focus group analysis with patients using evidence-based health information, Glenton and colleagues found that "participants described how they often made treatment decisions in a context of great pain and despair. Under such circumstances, they often had little energy to seek out written information and were sometimes too desperate to care what the research might have to say. Instead, they often gathered information about treatments through the personal anecdotes of friends and neighbors, and, in most cases, this experience-based information was considered to be more relevant than the evidence-based information" [19]. Not only is experiential information from apomediaries and peers more *relevant* for patients, it is hypothesized here that it often is also more or at least equally credible as information based on research.

Similarly. the notion of "source expertise" as being communicated primarily "through the comprehensiveness of a web site's information, its professionalism, and its sponsor's credentials" [15] is questionable in the health care context and perhaps in many other apomediated environments used by "autonomous" individuals. Here, "expertise" is not only expressed by credentials such as professional degrees and qualifications, but also first-hand experience. *Experience-based credibility* can be seen as one additional dimension of source credibility. Past research has identified that similarity in attitudes with the speaker as well as liking positively influences credibility perceptions. What might be added is that similarity of *experiences* (in the health care context: similarity of symptoms, diagnoses etc.) adds to credibility perceptions. In the context of youth, this is expressed by the term "*street cred*", which has been defined as "commanding a level of respect in an urban environment due to *experience* in or knowledge of issues affecting those environments" [20].

Applying network theory to apomediaries: credibility hubs

Apomediaries can be seen as highly complex networks of individuals and tools guiding consumers to credible information. While "networked" tools are often seen as a more equitable, democratic structure (as opposed to a system with intermediaries, who hold most of the power), network theory [21] teaches us that credibility networks are scale-free networks, where a rich-gets-richer phenomenon leads to the emergence of highly influential hubs, which in our context could be called *credibility hubs*. That is, not all apomediaries are equal, there are some apomediaries which have more influence than others. In the networked, apomediated model, some "nodes" (players or tools) become (or cease to be) credibility hubs in a more dynamic and fluid fashion than in the traditional model, where there is usually one intermediary whose credibility is influential and relatively stable.

In a "networked credibility" model with apomediaries as nodes, former intermediaries do not disappear completely, they are just one of many apomediaries, with a seemingly equal chance of becoming a "credibility hub", but in reality, they are more connected and have a better chance in ending up as a credibility hub,

For instance, a professional medical organization has a pre-existing social network which leads to other organizations linking to their website, leading that website to appear on top of Google, leading to more people linking to it, etc.

An interesting psychological phenomenon is that people attribute statements they believe to credible sources. For example, participants of an experiment who were exposed to a statement many times (and hence believed it) were more likely to attribute it to Consumer Reports (a credible source) than to the National Enquirer (a not so credible source) [22]. Such mechanisms may further increase the trustworthiness of credible sources, leading to a further rich-gets-richer phenomenon.

Conclusion

This paper discusses the idea that as a result of disintermediation, traditional intermediaries are replaced by what this author calls apomediaries, which are tools and peers standing by to guide consumers to trustworthy information, or adding credibility to information. It is hypothesized that in such an environment, tools, influential peers and opinion leaders are the primary conveyors of trust and credibility. In this environment, *apomediary credibility* may become equally or more important than source credibility or even message credibility. It is suggested to use tools of network analysis to study the dynamics of apomediaries in a networked digital world.

There are practical implications of the apomediation model for developers of digital media such as websites for consumers. Governments and other "authorities" – while certainly having credibility due to brand name recognition - do not typically do a very good job of creating "credible" Web sites – they always look and sound like government Web sites, and they lack the "edge" and the "street cred" that consumers and in particular youths are looking for to keep them engaged.

Good Web sites allow consumers to share their voices and connect with others in a safe, positive, supportive, possibly moderated, online community. Consumers need and want to be able to be co-creators of content, not merely be an audience who is broadcasted to. Engaging and credible Web sites are about building community and communities are built upon personal and social needs.

Acknowledgement

Partly funded by The John D. And Catherine T. Macarthur Foundation, full chapter to be published in the *MacArthur Series on Digital Media and Learning*

References

[1] Sittig DF. Personal health records on the internet: a snapshot of the pioneers at the end of the 20th Century. *Int J Med Inf* 2002;**65**:1-6.

[2] Eysenbach G,.Jadad AR. Evidence-based patient choice and consumer health informatics in the Internet age. *J Med Internet Res* 2001;**3**:E19.

[3] Eysenbach G,.Diepgen TL. Towards quality management of medical information on the internet: evaluation, labelling, and filtering of information. *BMJ* 1998;**317**:1496-500.

[4] Eysenbach G, Yihune G, Lampe K, Cross P, Brickley D. Quality management, certification and rating of health information on the Net with MedCERTAIN: using a medPICS/RDF/XML metadata structure for implementing eHealth ethics and creating trust globally. *J Med Internet Res* 2000;**2**:2E1.

[5] Eysenbach G,.Diepgen TL. Labeling and filtering of medical information on the Internet. *Methods Inf Med* 1999;**38**:80-8.

[6] Eastin MS. Credibility Assessments of Online Health Information: The Effects of Source Expertise and Knowledge of Content. *J Comput Mediat Communic* 2001;**6**.

[7] Anderson JG, Rainey MR, Eysenbach G. The impact of CyberHealthcare on the physician-patient relationship. *J Med Syst.* 2003;**27**:67-84.

[8] Brotherton JM, Clarke SJ, Quine S. Use of the Internet by oncology patients: its effect on the doctor-patient relationship. *Med J Aust* 2002;**177**:395.

[9] Gerber BS,.Eiser AR. The patient physician relationship in the Internet age: future prospects and the research agenda. *J Med Internet Res* 2001;**3**:E15.

[10] Hart A, Henwood F, Wyatt S. The role of the Internet in patient-practitioner relationships: findings from a qualitative research study. *J Med Internet Res* 2004;**6**:e36.

[11] Murray E, Lo B, Pollack L, Donelan K, Catania J, Lee K *et al*. The Impact of Health Information on the Internet on Health Care and the Physician-Patient Relationship: National U.S. Survey among 1.050 U.S. Physicians. *J Med Internet Res* 2003;**5**:e17.

[12] Murray E, Lo B, Pollack L, Donelan K, Catania J, White M *et al*. The impact of health information on the internet on the physician-patient relationship: patient perceptions. *Arch.Intern Med* 2003;**163**:1727-34.

[13] Wilson SM. Impact of the Internet on Primary Care Staff in Glasgow. *J Med Internet Res* 1999;**1**:e7.

[14] Esquivel A, Meric-Bernstam F, Bernstam EV. Accuracy and self correction of information received from an internet breast cancer list: content analysis. *BMJ* 2006;**332**:939-42.

[15] Metzger MJ, Flanagin A, Eyal K, Lemus DR, McCann RM. Credibility for the 21st Century: Integrating Perspectives on Source, Message, and Media Credibility in the Contemporary Media Environment. *Communication Yearbook* 2003;**27**:293-335.

[16] Petty RE, Cacioppo JT. Communication and persuasion: Central and peripheral routes to attitude change. New York: Springer, 1986

[17] Fogg, B. J. and Tseng, H. (1999). The elements of computer credibility. http://captology.stanford.edu/pdf/p80-fogg.pdf [On-line]. Archived at: http://www.webcitation.org/5Lqk5JrrA

[18] Gunther AC. Attitude extremety and trust in media. *Journalism Quarterly* 1988;**65**:279-87.

[19] Glenton C, Nilsen E, Carlsen B. Lay perceptions of evidence-based information - a qualitative evaluation of a website for back pain sufferers. *BMC Health Services Research* 2006;**6**:34.

[20] Urban Dictionary (2006). Street Cred. http://www.urbandictionary.com/define.php?term=street+cred [On-line]. Archived at: http://www.webcitation.org/5KT164TW0

[21] Barabasi AL. Linked. Cambridge, MA: Perseus, 2002

[22] Fragale AR,.Heath C. Evolving informational credentials: the (mis)attribution of believable facts to credible sources. *Pers Soc Psychol Bull* 2004;**30**:225-36.

Address for correspondence

Gunther Eysenbach MD MPH, Centre for Global eHealth Innovation, 190 Elizabeth Street, Toronto M5G2C4, Canada

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

A Mobile Phone Based Remote Patient Monitoring System For Chronic Disease Management

Mathieu Trudel[a], Joseph A Cafazzo[ac], Melinda Hamill[a], Walter Igharas[a], Kevin Tallevi[a], Peter Picton[a], Jack Lam[a], Peter G Rossos[ab], Anthony C Easty[ad], Alexander Logan[be]

[a] *Centre for Global eHealth Innovation, Toronto General Hospital, Canada*
[b] *Faculty of Medicine, University of Toronto, Canada*
[c] *Health Policy, Management and Evaluation, University of Toronto, Canada*
[d] *Institute of Biomaterials and Biomedical Engineering, University of Toronto, Canada*
[e] *Department of Medicine, Mount Sinai Hospital, Canada*

Abstract

Rising concern over the poor state of chronic disease management led to the user-informed design and development of a home tele-monitoring system. Focus groups with patients and primary care providers guided the research team towards a design that would accommodate the workflow and concerns of the healthcare providers and the low use and comfort with technology found among the patient population. The system was trialed in a before-and-after pilot study of 34 patients with diabetes and hypertension. Findings demonstrate a significant improvement in systolic and diastolic blood pressure. An RCT beginning in 2007 is being conducted to confirm these findings. It is hypothesized that this user-centred approach, utilizing focus groups, iterative design and human factors methods of evaluation, will lead to the next-generation of home tele-monitoring applications that are more intuitive, less cumbersome, and ultimately bring about greater patient compliance and better physician management.

Keywords:

blood pressure monitoring, self,
blood glucose self-Monitoring, self care, mobile phone, computers, handheld

Introduction

Chronic disease presents a growing challenge to the health and social care systems around the world. In 2005, chronic diseases accounted for 72% of the total global burden of disease in the population 30 years and older. [1]

Although many chronic conditions are preventable or can be controlled with proper treatment, evidence is mounting that current chronic disease management practices are falling short. The prevalence of type-II diabetes is expected to rise to 4.4% of the worldwide population by 2030, for a total of approximately 366 million people globally. [2] Additionally, 40-50% of type II diabetics also have hypertension, which confers a 2 to 3-fold increase in the risk of morbidity and mortality for diabetics, and is a major risk factor for both renal and cardiovascular disease (CVD). [3-5] Despite being treatable, less than 15% of these patients have reached the widely accepted goal of 130/80 mmHg. [4, 6]

This research discusses the development and testing of a novel approach to tele-monitoring for improved chronic disease management. The primary objective of this study was to determine if a hypertension tele-monitoring support system, designed to enhance patients' self-management and provide reliable information to clinicians for decision-making, will markedly improve blood pressure control in a primary care setting.

A user-centric design approach was taken to the development of a mobile phone based remote patient monitoring system for the control of hypertension in type II diabetics. The system was developed (Phase I) using affordable devices and iterative gathering of user feedback to ensure high usability and low disruption to lifestyle and workflow of patients and providers.

The system concept would start with the regular home monitoring of blood pressure (BP) by the patient. BP readings would then be automatically transmitted by a mobile phone to the central data repository. A clinical rules engine would then check the data and the patient and family physician would be notified if the readings were outside of the desired range for a period of time. Given this alerting, the system would theoretically facilitate timelier follow-up of the patient's care. Additionally, the availability of the home monitored data at patient visits may provide more information for assessing the patient's condition.

A pilot demonstration (Phase II) was undertaken to determine whether the system lowers the risk of dangerous complications for diabetic patients with high blood pressure. The study outcome was measured by assessing change in the measured blood pressure of study subjects.

Phase I - Requirements Gathering and Design

Methods

Before any development of the system began, a consultative phase was initiated to investigate the problems with the existing management of hypertension in diabetics and gather patient/physician impressions of the system concept. From these sessions, conclusions were drawn to inform iterations of the system development.

Two sets of focus groups were held. The patient set consisted of 24 type-II diabetics with hypertension, interviewed in small groups in one of four focus group sessions. The physician set consisted of 18 family physicians, also interviewed in small groups in one of four focus group sessions.

Results

Patient reaction to the concept was generally favourable, with patients showing a high regard and interest for the self-care aspects of the proposed system. Although patients were interested and intrigued by the use of a mobile device for the transmission of their BP data, they appeared to be more comfortable and familiar with the use of a mobile phone than a PDA. Surprisingly, there was little concern towards the security of the proposed electronic transmission of personal medical data. An additional significant finding from this group was that their level of Internet use was below expected. Although many had used the Internet in the past, their current use would be described as infrequent and simplistic, such as the occasional email. Many accessed the Internet through a proxy, such as a family member. These findings resulted in significant changes in the planned design of the system architecture.

The physician findings were in contrast to that of the patient group. Physicians were skeptical that patients would adhere to regular measurements of BP. Their most serious concerns were around compensation and disruption of workflow. There was concern that they would be accountable to address alerts in a timely fashion. As well, they would not be compensated for their work related to addressing alerts and reports, since it was not tied to a face-to-face patient visit. Physicians insisted that the reporting of data be tied to a patient visit. Like the patient group, physician use of computers and the Internet in the office setting was minimal and could not be seen as a means to communicate patient reports or alerts.

Design principles

Based on the findings of the study groups, the following design principles were formed:

- Personal computers are not to be used for information input, retrieval, or review by patient or physician.
- Phone and fax are the only means available to disseminate alerts and reports to physicians.
- The mobile phone or PDA is the information hub for the patient with the following considerations:
  - Data gathering will require no user-intervention.
  - User interface will consider visual and physical impairments.
  - User interface will consider limited or no previous knowledge of the operation of a mobile device.
  - Messages to the patient will be in lay terms.
  - Communication is secured via SSL connection to signed client-server applications.
  - Messaging to the patient will not be done via SMS due to usability problems. Messaging will be done via the client application, or by automated voice messages sent to the patient's home phone.

The strict adherence to these principles in order to simplify the use of the system for the end-user made the development and implementation more complex. The description of the realized system follows.

System overview

Bluetooth-enabled medical devices (BP monitor, weight scale and glucometer) are used to transmit physiological data automatically to a mobile phone. A custom application running on the mobile phone acts as a personal medical diary of the patient's test history. The application also relays data securely back to a central data repository where clinical rules are applied and alerts generated. These alerts are sent to the patient's physician by fax (in the absence of any other electronic means of communication) and to the patient by automated text and phone messages. The system is designed to remind patients by text and voice message to take their home readings if they fail to adhere to the prescribed schedule.

From the patient's point of view, the measurement of their BP is as before, with no additional steps needed for the transmission of data. The original design goal of the system was to be as unobtrusive as possible, and to require no manual intervention on the part of the patient. If desired, the patient can simply plug the mobile phone into an electric outlet to maintain battery charge and leave it unattended as it automatically relays data back to the central repository. More technically inclined patients can use the diary and reporting features of the device to review all previous measurements taken in the home, summary reports, or view a graph of historical trends in their data. Since all device interaction is optional (beyond the standard measurement procedure), no technical acumen is required of patients in order to benefit from this system.

Technical architecture

The monitoring system consists of three main components; the patient components, the back end data repository and decision support system, and the alerting and physician-

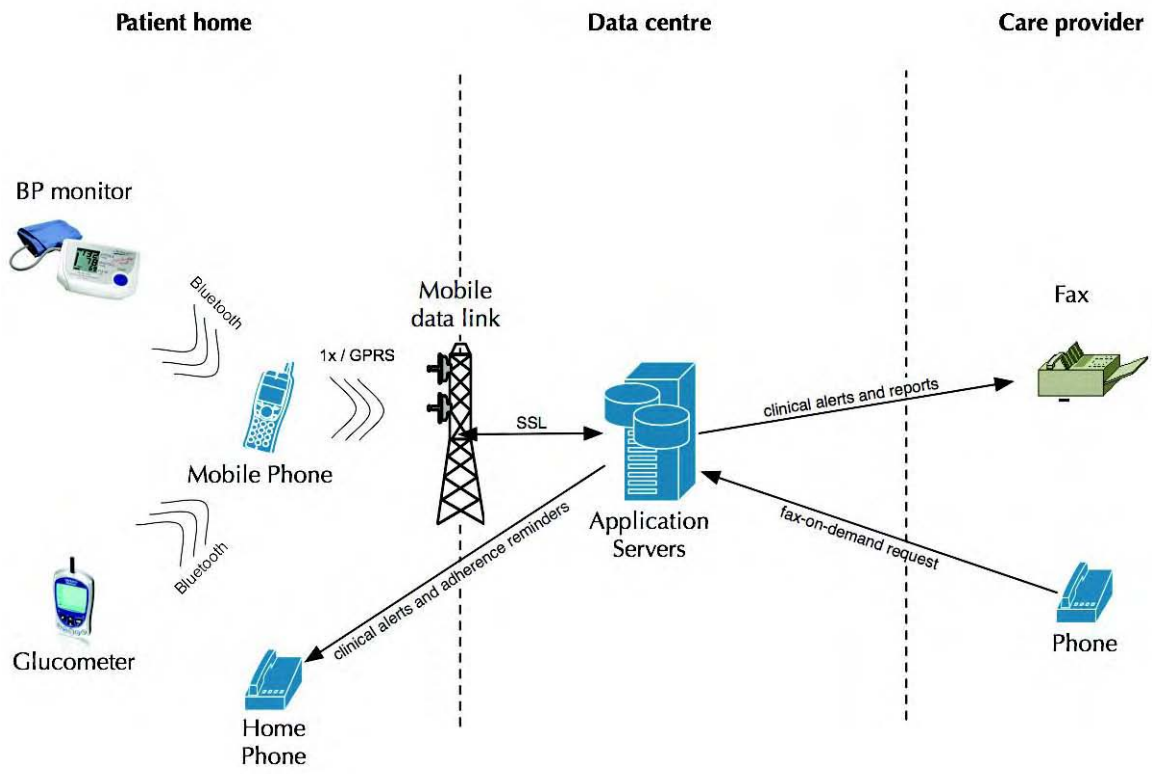


Figure 1 - Mobile Phone Based Remote Patient Monitoring System

reporting component. The data flow between these components is shown in Figure 1.

The patient component of the system comprises the medical devices themselves and the mobile phone used to relay results to the server. There is also an interface for patients to review previous readings on the device itself, in tabular, graph, and summary formats (see Figure 2). Progress messages are sent from the central server onto the mobile device after every reading, allowing for reminders and other coaching messages to be delivered automatically to the patient.

Figure 2 - BP Client Application

The back end of the system comprises a web based server application, which gathers results from client devices, a database to store the results, and a set of clinical rules, which are applied to the data as it arrives. These rules trigger events handled by the alerting and reporting component of the system.

The alerting and reporting system provides secure web, email, voice, and fax gateways into the data stored for each patient. Certain conditions (such as non-adherence to scheduled measurements) may trigger alerts to be sent to the patient via an automated telephone message to the patient's home phone number. The thresholds for these alerts can be set by the physician, on a patient-by-patient basis. Physicians can also request a fax (via an interactive telephone system) of the patient's readings to date displaying averaged readings over the past 30 days, as well as a chart of the patient's historical data measurements. These are also available online for providers who prefer to use the web for reviewing patient information.

As substantial portions of the monitoring system operate in an uncontrolled environment, measures to ensure the confidentiality of patient data were a paramount concern in the design of the system. The initial transmission of data from the measurement device to the mobile phone is done via Bluetooth, providing strong encryption of patient data during transmission, and also ensuring the authenticity of received data. Data stored on the mobile phone has no readily individually identifiable information stored with it,

as results are only tied to a specific patient at the server. This mapping associates patients to the unique numeric Bluetooth ID of the patient's medical device, and is known only to the system server. Nonetheless, in the event a device is lost or misplaced, a 'kill' message can be sent to the device from the server, causing all stored data to be permanently deleted from the device. Data is sent from the mobile device to the server via a secured SSL web connection, providing an industry standard level of confidentiality in addition to the security offered by the underlying cellular network.

Use of the interactive telephone system to request a fax report requires the physician or patient to input a valid numeric user id and password. The information is validated and authenticated against the information in the database before access is granted to the patient record. Fax-based transmission of patient results is only permitted to pre-programmed physician office fax numbers.

Phase II - Pilot study

Methods

A thirty-patient trial of diabetic patients with hypertension in the Toronto area began in December 2005 and concluded in November 2006. Patients monitored their blood pressure at home for four months using the aforementioned system. Patients had mandatory visits with their physician at baseline, two and four months. (Other visits may have taken place due to patient health factors and/or the alerting system requesting that the patient go see the doctor because of elevated hyper or hypo-tension.)

Main outcome measures of this pilot study are changes in BP compared to pre-trial baseline measurements, patient adherence, and technical and clinical feasibility of the system.

Results

Shown in Table 1 are the results. Significant results were found in all cases.

Table 1- Pre and post trial, ambulatory (24hr) and 2-week home measurement averages

	Pre (SD)	**Post** (SD)	**N**	*P*
$Systolic_{24hr}$	**143.3** (14.05)	**133.3** (12.56)	27	0.001
$Diastolic_{24hr}$	**81.4** (10.74)	**76.2** (10.70)	27	0.000
$Systolic_{2wk}$	**140.8** (16.50)	**131.8** (13.88)	32	0.001
$Diastolic_{2wk}$	**81.7** (11.70)	**78.3** (10.40)	32	0.006

Of the 34 patients that were enrolled in the study, 32 remained at the end of the four-month trial. One physician withdrew their patient, and another patient withdrew due to a confounding health problem that affected their ability to home monitor.

Discussion

The first outcome measure looked at pre and post trial ambulatory and 2-week home blood pressure averages. In both cases the results were significant. Ambulatory monitoring is currently the gold standard for measuring blood pressure, and the study patients experienced an average improvement of 10 mmHg systolic and 5 mmHg diastolic. These results show great promise as the health benefits of lowering BP in a diabetic population are well documented. [7-9] However, this was only a pilot study with 30 patients, a full clinical trial is required in order to confirm the generalizability of these results. (see "Current and Future work")

The second outcome measure was patient adherence. The success in adherence experienced in the pilot may be partially attributable to the automated phone calls to the patient's home, reminding them to continue to monitor. These phone messages were generated by the system when it was detected that the patient was not regularly monitoring their BP. The upcoming clinical trial will investigate the affect of these alerts on adherence. A comparison of the intervention group with the control arm of the study will demonstrate if adherence drops off without the adherence reminders.

Finally, the study demonstrated the technical feasibility and readiness of available and affordable technology to be applied for the purpose of home tele-monitoring (outcome measure three). Patients with the technology encountered only minor problems. Certain features of the mobile phone proved to be a challenge for some users with low dexterity or visual impairments. The small joystick for navigating on-screen menus, and the narrow power button were features that caused frustration. Additionally some patients had trouble viewing the screen when the backlight was off. The fax-back system proved to be a simple and efficacious method of providing data to the family physician, without changing their workflows significantly.

Current and future studies

An additional thirty-patient trial in the northern Ontario community of Chapleau began in July 2006 and adds glucometry in addition to home blood pressure measurement. The care model also differs in Chapleau. A nurse practitioner is the main primary care provider for diabetics in the area. Based in the local hospital, staff members have easy access to computers in the clinic, and prefer a workflow model where records and alerts are retrieved through a PC, rather than fax. A web-based application to accommodates this workflow, displays all data, reports and alerts, as well as allows patient parameters to be adjusted, such as measuring frequency, BP and glucose goals, and alert thresholds.

A randomized controlled trial (RCT) will be conducted in the greater Toronto area beginning in 2007. The control group will perform regular home monitoring of blood pressure, and the intervention group will utilize the tele-monitoring system.

Conclusions

The intent of a user-centric approach was to lead to the development of a system that was more intuitive, less cumbersome, and would ultimately bring about greater patient compliance and better physician management. The intent is to demonstrate a technology that is applicable to a wide variety of chronic diseases and conditions such as congestive heart failure and asthma and builds on the demand for new and more effective ways to improve chronic disease management. A secondary benefit of this study is the validation of the use of commodity hardware, which promises to provide significant cost savings in the delivery of the management tool.

Results from the pilot study indicate the potential efficacy of the system at reducing hypertension, which we will attempt to confirm in an RCT.

Acknowledgments

This research was funded through the Primary Care Health Transition Fund of the Ministry of Health and Long Term Care of Ontario. The clinical trial in Chapleau, Ontario is funded through Project Chapleau, a joint initiative of Bell Canada and Nortel to research the impact of advanced technology on rural and dispersed communities. The RCT is funded by the Heart and Stroke Foundation of Canada.

References

[1] Strong K, Mathers C, Leeder S, Beaglehole R. Preventing chronic diseases: how many lives can we save? Lancet. 2005;366(9496):1578-82.

[2] Wild S, Roglic G, Green A, Sicree R, King H. Global prevalence of diabetes: estimates for the year 2000 and projections for 2030. Diabetes Care. 2004;27(5):1047-53.

[3] Chobanian AV, Bakris GL, Black HR, Cushman WC, Green LA, Izzo JL, Jr., Jones DW, Materson BJ, Oparil S, Wright JT, Jr., Roccella EJ. The Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure: the JNC 7 report. Jama. 2003;289(19):2560-72.

[4] Singer GM, Izhar M, Black HR. Goal-oriented hypertension management: translating clinical trials to practice. Hypertension. 2002;40(4):464-9.

[5] Sowers JR, Epstein M, Frohlich ED. Diabetes, hypertension, and cardiovascular disease: an update. Hypertension. 2001;37(4):1053-9.

[6] Singer GM, Izhar M, Black HR. Guidelines for hypertension: are quality-assurance measures on target? Hypertension. 2004;43(2):198-202.

[7] Efficacy of atenolol and captopril in reducing risk of macrovascular and microvascular complications in type 2 diabetes: UKPDS 39. UK Prospective Diabetes Study Group. Bmj. 1998;317(7160):713-20.

[8] Effects of ramipril on cardiovascular and microvascular outcomes in people with diabetes mellitus: results of the HOPE study and MICRO-HOPE substudy. Heart Outcomes Prevention Evaluation Study Investigators. Lancet. 2000;355(9200):253-9.

[9] Estacio RO, Jeffers BW, Hiatt WR, Biggerstaff SL, Gifford N, Schrier RW. The effect of nisoldipine as compared with enalapril on cardiovascular outcomes in patients with non-insulin-dependent diabetes and hypertension. N Engl J Med. 1998;338(10):645-52.

Address for correspondence

Joseph A. Cafazzo, Centre for Global eHealth Innovation, Toronto General Hospital, University Health Network, 190 Elizabeth Street, R. Fraser Elliott Building 4S400, Toronto, Ontario, Canada M5G 2C4, joe.cafazzo@uhn.on.ca

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

How to Enhance Integrated Care towards the Personal Health Paradigm?

Bernd G.M.E. Blobel[a], Peter Pharow[a], Thomas Norgall[b]

[a] *eHealth Competence Center, University of Regensburg Medical Center, Germany*
[b] *Fraunhofer-Institute for Integrated Circuits, Erlangen, Germany*

Abstract

For improving quality and efficiency of health delivery under the well-known burdens, the health service paradigm has to change from organization-centered over process-controlled to personal health. The growing complexity of highly distributed and fully integrated healthcare settings can only be managed through an advanced architectural approach, which has to include all dimensions of personal health. Here, ICT, medicine, biomedical engineering, bioinformatics and genomics, legal and administrative aspects, terminology and ontology have to be mentioned. The Generic Component Model allows for different domains' concept representation and aggregation. Framework, requirements, methodology and process design possibilities for such a future-proof and meanwhile practically demonstrated approach are discussed in detail. The deployment of the Generic Component Model and the concept representation to biomedical engineering aspects of eHealth are touched upon as essential issues.

Keywords:

Personal health; system architecture; Generic Component Model; biomedical devices

Introduction

Healthcare systems in industrialized countries, and increasingly those in countries in transition, are faced with the challenge of ensuring efficient and high quality care independently of time, location and local resources, utilizing advanced knowledge and technologies. This challenge must be realized despite demographic developments, the growth of multi-morbidity, demands for health services and expenditures for diagnostic and therapeutic procedures, and decreasing contributions to health insurance funds. To meet this challenge, the systems have been changing from an organization-centered towards a process-controlled care paradigm, which is also called shared care, managed care or disease management. This development is combined with extended cross-organizational communication and cooperation between all of the healthcare establishments involved in patient care. This process has to be supported by deploying advanced information and communication technologies (ICT) in health, connecting primary and secondary care. Regarding the need for prevention and the integration of social care in an aging society (addressing citizens before becoming patients, and so moving the focus from healthcare to health) this process-controlled strategy is no longer sufficient. Health, nowadays provided by organizations such as hospitals, primary care offices, policlinics or medical centers as well as, has to move closer to the citizens' environment.

Observing the citizens' health status, context and conditions for providing person-centered (personalized) and dedicated health services implies the need for a new health paradigm: *personal care*, which completely integrates all of the *principals* involved in the care process. According to the definition of the Object Management Group principals are any actors in the domain in question such as persons, organizations, systems, devices, applications, components or even single objects. This does not mean that there will no loner be acute care and ambulant service, but that such services will be tailored to relevant personal care needs,.

Materials and methods

To realize patient care at any location in an individualized way, three technological paradigms have to be managed: mobile computing, pervasive computing and autonomous computing (Figure 1). Mobile computing enables the permanent accessibility of the principals involved, providing, for example, teleconsultation services. Pervasive computing allows for location-independent service provision, established as telemedicine services. For providing personalized care, services have to be flexible and cannot be rigidly predefined. Such adaptive

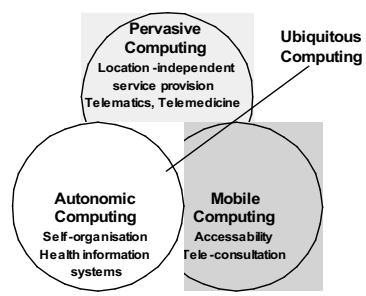


Figure 1 – Computing paradigms deployed in the personal care context (after Kirn and Müller (2005), changed [1])

health information system design, towards a self-organizing environment, draws on current challenges in the research and development for autonomous computing.

Another aspect which is characteristic for personal health (pHealth) concerns the distance between the physical and the informational world. In the traditional ICT environment, this gap is mediated through human users. Introducing advanced technologies, this gap is getting closer to the real integration of the health subject (patient) in the health system, and even becoming a part of the information system environment (Figure 2).

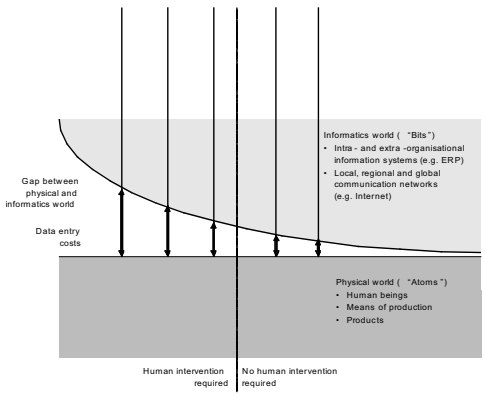


Figure 2 - Closing the gap between physical and informational world (after Kirn and Müller (2005), changed [1])

The process to be supported has to be properly described, abstracting from reality by using appropriate grammars for modeling it. A model is a partial representation of reality. It is restricted to attributes the modeler is interested in. Defining the pragmatic aspect of a model, the interest depends upon the intended audience and the reason and the purpose for modeling the reality. The resulting model is used for a certain purpose and time as a proxy for reality. Therefore, the model (which is the result of an interpretation) must be interpreted itself.

The simplification of systems through formal models can be provided in three dimensions according to the Generic Component Model [2] (Figure 3). The first level of simplification concerns the restriction to the domain of interest. Examples for such domains are the medical domain, administrative domain, technical domain, legal domain, etc. Within this domain, the system considered can be decomposed or composed for analyzing or designing it. This results in different levels of granularity or complexity, respectively, using specialization or generalization relationships. In the Generic Component Model, the following granularity levels have been derived: business concepts, relations network, basic services/functions and basic concepts. The third dimension of a generic system architecture touches different aspects of the system according to the ISO Reference Model – Open Distributed Processing [3]. Here, the business process is expressed by the Enterprise View, the informational expression of this process is expressed by the Information View and the functional aggregation of algorithms and services is expressed by the Computational View. These are described through platform independent models of the system expressing the system's logical content. Platform-specific implementation details are described by the Engineering View, and the Technology View represents technology (or implementation) aspects. The system's architecture (i.e. the system's components, their functions and relationships) is characterized through the components' concepts and their aggregations. The representation of concepts and association rules is provided by constraint models, which are derived from reference models.

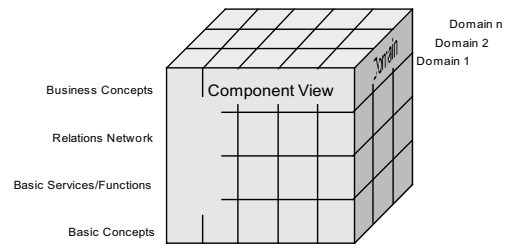


Figure 3 - The Generic Component Model [4]

The resulting pHealth information system is characterized by openness, flexibility, scalability, portability, user friendliness and user acceptance, service orientation, distribution at Internet level, being based on standards, semantic interoperability, lawfulness and trustworthiness. Organizational, contextual, rule-related or other constraining aspects of the system and its components are expressed by binding policies to the components and by ruling the component aggregation by policies .

Following the Generic Component Model approach, the pHealth information system architecture needs to combine the following paradigms: distribution; component-orientation; separation of platform-independent and platform-specific modeling, i.e. separation of logical and technological views; specification of reference and domain models at meta-level; interoperability at service level (concepts, contexts, knowledge); enterprise view driven process-controlled design; multi-tier architecture; appropriate multimedia GUIs; common terminology and ontology; unified design, development and deployment process; trustworthiness provided by appropriate security and privacy services, all as an integrated part of the design.

Results

The architectural approach of the Generic Component Model has to be applied for pHealth information systems covering health telematics aspects, telemedicine concepts, biomedical engineering solutions and bioinformatics problems in a harmonized way.

The knowledge representation is performed using meta-languages. For ensuring semantic interoperability, reference terminologies and ontologies have to be defined, using the aforementioned methods for knowledge representation in the Generic Component Model's context.

In this way, different systems deploying different modalities, belonging to different domains and different business areas, using different domain languages, can be harmonized and bound to policies. From this modeling, the specification of invocation calls and the development of XML messages can be derived as is shown in [5]. The connection of biomedical devices for patient monitoring and care is provided using the CEN ISO/IEEE 11073 standards set (which is based on ENV 13734/13735 "VITAL" and IEEE 1073-x) [6] as well as CLSI (formerly NCCLS) POCT-1A [7]. Thereby, biomedical devices can be aggregated and replaced like any other component. This is especially true for future mobile, modular, personal systems for individually caring patients. Such systems can be deployed in clinical settings and in homecare. In that way, a patient's transfer between both settings can be facilitated. Typical system components, at different levels of development and accentuation, are:

- Highly integrated sensor and human-machine interface components at the body or wearable in clothing (e.g. intelligent wireless sensors; wearable devices; PDAs);
- Components and infrastructure enabling communication between the aforementioned systems and components on the one hand and stationary systems and services on the other (e.g. Body Area Network; mobile phone/mobile network; wireless LAN; workstation with gateway function in patient's home; appropriate middleware);
- Distributed functions for sensor signal processing, state recognition and state monitoring up to person- and situation-related information and interventions offered (e.g. qualified management of emergency calls; cumulated multi-parameter records, processing and presentation using PDA or workstation);
- Information and expert systems for recognizing and managing emergency cases, for informing the patient as well as decision support for healthcare professionals (e.g. localization of principals; access to reference data; person-specific support for interpretation of data, secure access to patient's personal information/Electronic Health Record).

The system functions must realize proper escalation strategies, while the system components next to the patient must be relatively autonomous to minimize communication and maintenance effort, energy consumption and, on the other hand, communicating autonomously with external stationary system components in certain cases (exceeding of thresholds, recognition of exceptional situations, emergency cases, alerts, but also for routine communication).

Advancing the modeling approach towards the Generic Component Model methodology, biomedical device components can be designed to provide the same structure and behavior as all other system components in the sense of adaptive, self-organizing systems. In the IHE Patient Care Device Technical Framework, the eHealth technology domain is considered to have defined structural and functional requirements for semantic interoperability. Starting with an ISO OSI based layered model, the interoperability requirements have been defined by ENV 13735 Health informatics – Interoperability of patient connected medical devices. This specification, combined with others, has been moved up to ISO/IEEE 11073 "Health informatics – Point-of-care medical device communication", e.g. defining the underlying medical package model (Figure 4). The containment tree consisting of Medical Device System, Virtual Medical Device, Channel, Metric and reflecting the granularity levels, as well as its combination with the process management aspect, are consistently in line with the Generic Component Model.

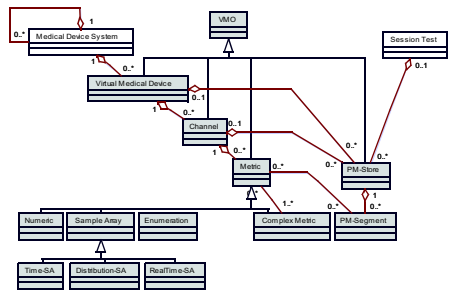


Figure 4- ISO/IEEE 11073 Medical package model

Some of the standards specifying the integration of medical devices in health information systems for enabling personal health systems, are presented in some detail below.

Standards for medical device interoperability

Unlike other clinical communication standards, the CEN ISO/IEEE 11073 family of standards provides real-time "plug-and-play" interoperability to facilitate the efficient adding and swapping of acute and continuing care devices, such as patient monitors, ventilators, infusion pumps, ECG devices, etc in critical environment settings.

To fulfill real-time requirements, a highly efficient "Medical Device Encoding" scheme is used. Conversion to alternative XML encoding reduces real-time capabilities, but enables use of XML-specific concepts and tools [8] promising dramatic cuts of development time and effort for interoperability component implementation. "Plug-and-play" practically means that all a clinician has to do is making the connection – the systems automatically detect, configure and communicate without any subsequent human interaction, maintaining both functional and semantic interoperability.

Health informatics and clinical communication standards are often generally interpreted as relating to ISO/OSI Level 7 based on the assumption of a generally available LAN-infrastructure to build upon. A prominent example for this approach is the HL7 communication standard

(appropriately named with a seven) [9]. For medical devices, interoperability explicitly implies all ISO/OSI Levels. While 11073 defines/modifies standards in ISO/OSI levels 7 – 5, it chiefly references other standards (such as 802.x, IrDA, Bluetooth, etc.) in levels 1 – 4.

For systems in homecare, and even more for mobile systems providing personal health services in dynamic environments, equivalent considerations apply. In order to enable functional interoperability using different (wired, IR and RF wireless) network technologies, CEN ISO/IEEE 11073 provides standards for internetworking in the 11073.5 branch of the 11073 standards family. In that context, the 11073 specification for "Agent Device", e.g. an infusion pump, pulse oximeter, or ventilator, and the corresponding 11073 specification for "Manager System" – a patient monitor or device manager shall be mentioned. Both are situated in different sub-networks using different network technologies. Other typical applications are wired-to-wireless transport gateways or LAN/IR access points. Based on CEN preparatory work [10], CEN ISO/IEEE 11073-60101 defines an 11073/HL7 "Observation Reporting interface" (ORI) enabling device-to-HIS-level interoperability. It is the first standard in the 11073.6 "Application Gateway" branch of the 11073 standards family which is intended to provide interoperability among different application protocols. The 11073 coding scheme has been a registered HL7 Coding Scheme since 2003, permitting its use in HL7 messages.

Body area networks

The Body Area Network (BAN) concept specifies wireless communication between several miniaturized, intelligent Body Sensor (or actuator) Units (BSU) and a single Body Central Unit (BCU) worn on the human body. It is characterized by a maximum range typical for human body dimensions, e.g. 2 meters. The BCU concentrates the BSU-originated data streams, performs intermediate storage and processing as well as communication to the outside world using standard wireless technology like DECT, WLAN or Bluetooth. The Network Access Unit (NAU) can be implemented as a medical gateway hosting an embedded web-server.

Both NAU and BCU provide standard interfaces, particularly implementing ISO/IEEE 11073 "Agent" functionality. From a communication perspective, a BAN can thus be regarded equivalent to an 11073-compliant modular medical device, implying semantic interoperability between BAN and remote professional or clinical systems.

IMEX – a micro-system perspective for interoperability

As the development of micro-sensors and micro-systems, particularly for homecare and pHealth-related applications, is progressing, the acquisition of multiple biosignals (for instance blood pressure (BP), ECG, respiration, urine flow) can be performed by means of miniaturized patient-worn equipment. Utilizing the BAN concept for collection and communication of data to enable multi-parameter monitoring, micro-sensors can also be integrated into BSU units.

The German IMEX project [11] aimed at communication between micro-systems and with external device systems, analyzing possible interfaces and their communication requirements. A Micro System Data format (MSD) was defined to enable the use of standard health telematics coding schemes for semantic core elements on the micro-system-level, minimizing the processing overhead for preparation of micro-system-generated data for external standard-based communication [12]. Thus the "semantic interoperability chain" is extended to the micro-system level. Figure 5 shows smart micro-systems as the other end of the healthcare interoperability chain.

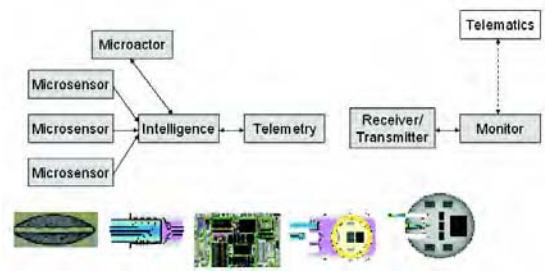


Figure 5 - Smart micro-systems as the other end of the healthcare interoperability chain

Specification and implementation process

To select the right components, aggregation rules and policies (in the broadest sense of the term policy: legal, social, organizational, ethical, functional and technical) in order to guarantee semantic interoperability, agreed constraints or profiles can be introduced similar to OMG's approach and similar to the process IHE is performing with its integration profiles.

For describing a unified process, the Rational Unified Process [13] or the HL7 Development Framework can be used. For domain-specific specializations, existing models such as HL7 Domain Message Information Models (D-MIMs), HL7 Refined Message Information Models (R-MIMs), HL7 Common Message Element Types (CMETs) [14], and also *open*EHR Archetypes [15] can be reused as demonstrated in [16]. Also other knowledge representation means can be deployed [17]. The future advanced eHealth architecture for individualized healthcare with region-crossing or European characteristics has been defined in the eHealth Action Plan of the European Commission and the EU Member States. The Electronic Health Record (EHR) is the core application of any eHealth platform; different countries are approaching this differently. The variants cover a medication file as starting point for eHealth in The Netherlands and the UK, the Sharable Record approach in Finland, right up to a comprehensive record architecture modeled and implemented within the national programme.

Discussion

Interoperability implies a number of different concepts, e.g. functional interoperability and internetworking, semantic interoperability and application gateways. Health information integration (eHealth) has established a demand for interoperability between clinical and health-care-related stakeholders, systems and processes or workflows. Domain-specific communication and interoperability standards are well established, but have to be supplemented for trans-domain use. Interoperability concepts for medical devices and for personal or mobile systems need to involve all 7 ISO/OSI reference model layers, more properly advanced to the Generic Component Model, including terminology/coding aspects.

The advanced concept of pHealth extends eHealth by the inclusion of smart sensors, body-worn mobile systems and situation-specific activation of applications and human health professionals, thus providing personalized ubiquitous health services. Body Area Networks and micro-systems are building blocks of future personalized health telematics infrastructures, and extend existing interoperability concepts. Another important eHealth pillar is the field of bioinformatics and genomics. As personal health requires personalized process models for optimal care, the underlying diagnosis and therapy has also to be individualized. This can be achieved by developing and deploying advanced bioinformatics and genomics as mentioned earlier.

The transfer to pHealth information systems with process-controlled, service-oriented, context-sensitive, semantically-interoperable information and communication architectures requires open, highly flexible individually tailored application systems for the cared for and the caring parties. Such applications cannot be pre-manufactured any more, but must be dynamically created and adapted to the actual requirements and needs. In that way, besides the well-established technology paradigms of Mobile Computing for realizing accessibility (e.g. teleconsultation) and Pervasive Computing for realizing independency of location when providing services (e.g. telemedicine), the paradigm of Autonomous Computing for realizing self-organizing systems can be introduced. The combination of the aforementioned technology paradigms leads to Ubiquitous Computing, which is bound to other paradigms and trends such as health grids. Personal health also requires an adequate legal framework and the new orientation of traditional organizational patterns.

Acknowledgement

The authors are indebted to the colleagues from HL7, ISO TC 215 and CEN TC 251 for kind support.

References

[1] Kirn S, Müller G. Ministudie zu den F&E-Perspektiven auf dem Gebiet der Gesundheitstelematik. Hohenheim, University of Hohenheim 2005.
[2] Blobel B. Analysis, Design and Implementation of Secure and Interoperable Distributed Health Information Systems. Series “Studies in Health Technology and Informatics” Vol. 89. Amsterdam: IOS Press, 2002.
[3] ISO/IEC 10746 “Information technology – Open Distributed Processing, Part 2 – Reference Model”.
[4] Blobel B. Advanced EHR architectures – promises or reality. Methods Inf Med 2006; 45: pp 95-101.
[5] Blobel B, Stassinopoulos G, Pharow P. Model-Based Design and Implementation of Secure, Interoperable EHR Systems. In: Musen MA, Friedman CP, Teich JM, eds. AMIA 2003 Symposium “Biomedical and Health Informatics: From Foundations to Applications”, American Medical Informatics Association 2003 Proceedings, Bethesda, pp. 96-100..
[6] http://www.ieee1073.org
[7] http://www.nccls.org or http://www.clsi.org
[8] http://www.w3.org
[9] http://www.hl7.org
[10] CEN SSS-HIDE (2001). Health Informatics-Strategies for harmonisation and integration of device level and enterprise-wide methodologies for communication as applied to HL7, LOINC and ENV.
[11] Becks T, Dehm J. IMEX – A New Knowledge Platform for Microsystems in Medicine. http://www.vde-mikromedizin.de
[12] http://www.iec.ch/
[13] IBM. Rational Unified Process. www.ibm.com/software/awdtools/rup/
[14] Blobel B, Engel K, Pharow P (2006) Semantic Interoperability – HL7 Version 3 Compared to Advanced Architecture Standards. Methods Inf Med 2006; 45: pp 343-353.
[15] Beale T, Heard S. Archetype Definition Language. openEHR Foundation, Ocean Informatics Australia 2004.
[16] Lopez DM, Blobel B. Connecting Public Health and Clinical Information Systems by Using a Standardized Methodology. (in this volume)
[17] Blobel B. Concept Representation in Health Informatics for Enabling Intelligent Architectures. In: Hasman A, Haux R, van der Lei J, De Clercq E, Roger-France F, eds. Ubiquity: Technology for Better Health in Aging Societies,. Series Studies in Health Technology and Informatics, Vol. 124. Amsterdam: IOS Press, 2006; pp. 285-291.

Address for correspondence

Bernd Blobel, PhD, Associate Professor
University of Regensburg Medical Center,
eHealth Competence Center
Franz-Josef-Strauss-Allee 11
93053 Regensburg, Germany
Phone: +49-941-944 6769
Email: bernd.blobel@ehealth-cc.de

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Developing Information Technology Attitude Scales for Health (ITASH)

Rod Ward[a], Katherine Pollard[a], Dr Margaret Glogowska[a], Dr Pam Moule[a]

[a] *Faculty of Health and Social Care, University of the West of England, Bristol. UK*

Abstract

This paper reports on the ongoing development and testing of a set of scales designed to elicit the attitudes of United Kingdom healthcare staff towards Information Technology. The scales were developed in the light of existing literature. Testing included a test-retest completion by over 100 staff from various disciplines in three National Health Service Trusts and a comparison with an existing scale. Exploratory principal components factor analysis identified three components, with a loading of > 3.1. This pattern of item grouping could be understood and interpreted as representing 'efficiency of care', 'education, training and development' and 'control'. The three scales comprise the Information Technology Attitude Scales for Health (ITASH). The results suggest that the developed scales together comprise a measure that can be used to establish staff attitudes towards IT use in the healthcare sector. Those undertaking research in this area might consider employing the scales to measure the factors that influence attitudes towards IT. Additionally, employers might usefully draw on the scales as they aim to support staff in IT use and embed IT systems within the healthcare workplace.

Keywords:

attitude, computers, health personnel, United Kingdom

Introduction

The introduction of Information Technology (IT) in the National Health Service (NHS) in the United Kingdom (UK) has a long history, although the use made of IT varies widely between different geographical and practice areas [1]. Since the publication of the Information for Health Strategy [2], and more recently the 'National Programme for IT' (NPfIT), the use of computer applications within the NHS has been increasing. One of the barriers to successful implementation and use of IT in the NHS is believed to be the attitude of staff to IT use, however there is limited research on this topic and no validated instruments for measuring NHS staff attitude to IT could be found in the literature.

There are several models and tools used elsewhere in the world and in other sectors. These include the Technology Acceptance Model (TAM) [3, 4] as extended by Dixon into the Information Technology Adoption Model (ITAM) [5, 6] and the Rogers' Innovation-Diffusion Model [7] as explored by Lee [8] when examining nurses' attitudes in Taiwan. All of the studies in this area show that attitude is an important predictor of the use of IT. However, none of the tools that have been employed in the UK have been validated, and those from elsewhere are not necessarily appropriate in the UK context.

Stronge and Brodt [9] over 20 years ago, studied this area in the United States of America (USA) with their Nurses' Attitudes Towards Computers (NATC) questionnaire. Others have continued this work [10-12]; however none of the resulting measures were considered appropriate, valid and reliable for use in the NHS, because of the different healthcare systems and/or the use of American phrases.

The primary aim of this study was to develop and test a scale which can be used to examine the attitudes of NHS staff towards Information Technology (IT), both in general and specifically in the workplace.

The main objectives of the research were to:

- Review existing IT attitude measurement tools
- Develop a health care specific attitude measure
- Test reliability, uni-dimensionality and validity

Methods

Development of scales

The research team collaborated in the design of the ITASH scales. Statements on issues identified from the literature [13-16] and practice were generated. Specific items were included identifying factors known to influence staff IT attitudes. The scales were incorporated into a questionnaire which also asked for demographic data, which included age group, gender, professional specialty, qualifications and current level of computer use at home and work.

The topic areas in the attitude scales included:

- Quality of care/ Communication
- Benefit to the organisation
- Confidence/anxiety/interest
- Dependability issues
- Training issues
- Professional role
- Feelings about change/control

Ten to 15 questions were written for each topic and checked for duplication and relevance. The questions were re-ordered to reduce bias and question fatigue [17]. A pool of 79 items was assembled about IT use in health care, respondents were invited to agree/disagree, using a 4 point Likert scale, as it was assumed that respondents would have an opinion. Following completion the statement values were reversed where necessary. The item pool was then used in two different ways - with NHS staff in a test-retest format to investigate reliability; and secondly with Faculty staff exploring validity.

Ethical approval was obtained from appropriate Local Research Ethics Committees.

Paper questionnaires were delivered, as online mechanisms would have biased the results by excluding those without Internet access and those with negative attitudes to IT.

Between November 2005 and January 2006, 816 questionnaires were distributed via managers/matrons/professional leads. Trust 1 (Mental Health) – 470, Trust 2 (Acute Hospital) – 200, Trust 3 (Primary Care Trust) – 146. Follow up questionnaires were sent to all respondents who included their contact details between December 2005 and March 2006.

To test for concurrent validity, copies of the questionnaire and the Stronge and Brodt [9] scale were distributed to UWE Faculty members. The questionnaire data from NHS and UWE staff were entered into SPSS for Windows (Version 13.0) and exploratory factor analysis was carried out.

Results

Response rates

Response rates are detailed in Table 1.

Table 1 - numbers of completed questionnaires

	Trust 1	**Trust 2**	**Trust 3**	**Total**
Round 1 (test)	78	35	38	151
Round 2 (re-test)	57	28	36	121

NB – Six questionnaires were returned anonymously and therefore these respondents could not be sent a second round questionnaire.

Establishment of the scales

The scales were constructed using exploratory factor analysis, a standard technique in attitude scale development [18] This process identified three components, with statement loadings >3.1, as recommended by Watson and Thompson [19]. This pattern of item grouping could be understood and interpreted as representing:

- Scale 1 'Efficiency of Care'
- Scale 2 'Education, Training and Development'
- Scale 3 'Control'

17 items were included in each of scales 1 and 2, while 14 were included in scale 3. The three scales comprise the Information Technology Attitude Scales for Health (ITASH).

The cumulative value of the responses to the statements in each scale becomes the respondent's score for that scale. For the first and second scales, the minimum score is 17, while the maximum score is 68. Scores from 17 to 34.99, 35 to 50.99 and 51 to 68 indicate negative, neutral and positive attitudes towards the use of computers in health care, respectively. For the third scale, the minimum score is 14, while the maximum score is 60. Scores from 14 to 28.99, 29 to 41.99 and 42 to 56 indicate negative, neutral and positive attitudes towards the use of computers in health care, respectively.

Reliability of the scales

To assess the stability of the scales, a test-retest administration of the scales was sent to 145 respondents, of whom 121 completed the scales a second time, over a period of 2-3 weeks. Pearson's correlation coefficients for the scores on the three scales were found to be r=.88 ($P < 0.01$), r=.75 ($P < 0.01$) and r=.83 ($P < 0.01$) respectively. The internal consistency of each scale was assessed by means of Cronbach's alpha coefficients. The coefficients obtained were =0.88 (n = 150), =0.70 (n = 150) and =0.77 (n = 150), indicating a satisfactory degree of internal consistency [17]. These results indicate that the ITASH scales are reliable.

Validity of the scales

A publicly available tool [9] appeared to measure similar constructs to one of the three scales – effectiveness of care. In total 34 members of staff of the Faculty of Health and Social Care at the University of the West of England, Bristol completed both the newly developed questionnaire and the existing scale from Stronge and Brodt [9]. The results were compared to examine the concurrent validity of the new scales.

Pearson's correlation coefficient was calculated for each of the 3 scales developed: 'Efficiency of Care', 'Education, Training and Development' and 'Control'. Scale 1 (Efficiency of Care) correlated adequately with the Stronge and Brodt [9] scale which looked at similar issues (r = .73). Scales 2 (Education, Training and Development) and 3 (Control) looked at different areas to those covered in the Stronge and Brodt [9] scale and therefore did not demonstrate correlation. It has not been possible to find measures suitable for establishing concurrent validity of the second and third scales at this time.

Analysis of scale scores and demographic factors

For each of the three scales, ANOVA was used to analyse raw scale scores on the basis of demographic variables and those relating to patterns of computer usage. The significance level for the study was set at p=0.05 [21]. 120 is accepted as a reasonable sample to provide a reliability coefficient of 0.8 with a confidence interval of +/- 0.1.[22]

Responses to each of the three scales were compared on the basis of various demographic characteristics of the respondents. The type of trust (ie community, mental health and acute) showed differences on the 'Education,

Training and Development' and 'Control' scales (F=7.23, p = 0.001; F=3.11, p=0.05 respectively). This suggests that the organisation in which the respondents worked influenced the extent to which they felt prepared for computer use and to which they felt they were able to influence their own working environment. Age revealed no differences, suggesting that age did not significantly influence attitudes toward computer usage. Gender differences were found on 'Efficiency of Care' and 'Control' scales (t = 2.92, p = 0.004; t = 3.28, p = 0.001), suggesting that men showed a significantly more positive response than women in these areas

No differences in responses to any of the scales were found based on the frequency of computer use at home. However, raw scores for each of the three scales according to frequency of PC use at work revealed significant differences (F=4.21, p=0.017; F=5.93, p=0.003; F=5.69, p=0.004 respectively). Those using the computer more frequently at work appeared to display more positive attitudes

Differences were observed for all three scales (t=2.03, p=0.044; t=2.70, p=0.008; t=2.93, p=0.004 respectively) when they were compared with the level of computer training respondents had received. Respondents with a formal computer qualification displayed more positive attitudes on each of the scales. Similar differences were found when examining experience of computer use. Respondents were asked to rate their experience from 0 (none) to 10 (extensive). Those who rated themselves as 0,1,2 and 3 were grouped as "low experience", 4,5 and 6 grouped as "medium experience" and those who rated themselves 7,8,9 and 10 as "high experience". A Kruskal-Wallis H-test of responses to each of the three scales was conducted on the basis of degree of experience. Differences were found on all three scales. (Scale 1 $K^2(2)=13.52$, $p<0.001$; Scale 2 $K^2(2)=15.63$, $p< 0.001$; Scale 3 $K^2(2)=35.99$, $p<0.001$). Those who rated their experience of computer use more highly generally showed more positive attitudes on all three scales.

Similarly, respondents were asked to rate their confidence in computer use from 0 (none) to 10 (very confident). Those who rated themselves as 0,1,2 and 3 were grouped as "low confidence", 4,5 and 6 grouped as "medium confidence" and those who rated themselves 7,8,9 and 10 as "high confidence". A Kruskal-Wallis H-test of responses to each of the three scales was conducted on the basis of degree of confidence. Differences were found on all three scales. (Scale 1 $K^2(2)=17.12$, $p<0.001$; Scale 2 $K^2(2)= 19.16$, $p< 0.001$; Scale 3 $K^2(2)=47.48$, $p<0.001$).

Discussion

Attempts were made throughout the study to reduce potential bias and accurately represent the responses of the population; however, the sample size was small when compared with the total population of the NHS in the three trusts, each of which employs around 5000 staff.

The low response rate may be due to the distribution methods. Often unit administrators/ managers left the questionnaire in staff areas for individuals to complete. As a result of the UK Data Protection Act [23] lists of names and details of individual staff were not available, which would have allowed a more targeted approach. Respondents were not randomised and were self selecting and there may be a difference between those who completed the questionnaire and those that did not, possibly reflecting their attitude towards the subject area. The respondents were from three NHS trusts, within one Strategic Health Authority therefore local initiatives and factors, such as IT training may have influenced the results

Representativeness is not claimed but the study secured a range of participants from different NHS organisations. It is not known how closely the demographics of the respondents match the make-up of the wider NHS workforce. However, a wide range of staff from various settings within the three NHS trusts participated in the study.

During attitude scale construction, the use of exploratory factor analysis is designed to reduce multiple indicators of attitudes etc. by explaining the relationships between potential scale items. It must be remembered, however, that there is always a subjective element in this process [22].

The test-retest results confirmed that the ITASH scales developed in this study correlated adequately between the first and second application and also demonstrated a level of internal consistency commonly regarded as satisfactory for attitude scales (Oppenheim, Spector) . In addition, the 'Efficiency of Care' scale showed concurrent validity with a scale [9] which has been in use in several countries for many years. Since no available measures appear to be suitable for establishing the concurrent validity of the other two scales, further work is needed to establish their validity through other means.

Nevertheless, despite their lack of established validity, the results produced by Scales 2 and 3 were consistent with the body of literature relating to healthcare staff's attitudes to IT, which show that previous experience and exposure are major determinants of attitude formation [24]. The role of the individual in an organisation and the way in which IT are introduced, along with the purpose of the introduction are also known to be relevant to attitude formation [25].

As the pace of implementation of electronic records systems and other initiatives under the NHS's NPfIT increases, the involvement of staff in the use of IT will increase. IT will become vital for both organisational objectives and individual patient care, therefore it is important to be able to identify factors which will influence staff use and evaluate the effectiveness of strategies designed to change attitudes amongst the staff.

Several interesting findings have been obtained through the administration of the scales. Those respondents who had undertaken computer training, with or without obtaining formal qualifications, showed more positive attitudes towards IT. Greater use of a PC at work was associated with more positive attitudes, although there was no such association with the amount of PC use at home. There is also some indication that the type of organisation influ-

enced the extent to which staff felt prepared for computer use. The age of respondents in this study did not have any significant effect on attitudes towards IT. However, men showed more positive attitudes than women in relation to the 'Efficiency of Care' and 'Control' scales. Attitudes on all three scales correlated positively with the respondents self reporting of their experience and confidence in computer use.

These initial results suggest that in the female dominated NHS workforce, consideration needs to be given to strategies which will influence the attitudes of staff towards IT, which may include training provision and supported exposure to computers.

The development and testing of the scale was the primary purpose of this study. The main task for the research team is now to establish the validity of Scales 2 and 3.

Conclusion and recommendations

The results suggest that the scale offers a measure that can be used to establish staff attitudes towards IT use in the healthcare sector. Those undertaking research in this area might consider employing the scale to measure the factors that influence attitudes to IT. Additionally, employers might usefully draw on the scale as they aim to support staff in IT use and embed IT systems within the healthcare workplace.

All three scales have been shown to be reliable; however only scale 1 'Efficiency of Care' has validity when compared to established instruments. Further research is needed to examine the validity of scale 2 'Education, Training and Development' and scale 3 'Control'. It would also be useful to see further research into the relationship between the different professional groups and their attitudes towards IT.

The results obtained while developing and testing the scale give a snapshot of a selection of the NHS staff in three different types of NHS trust which show fairly frequent use of IT at work, primarily for word-processing, but to a lesser extent for patient records and communication. When this is compared with the demographics of the population, including the levels of IT training and qualification, and the factors demonstrated to influence attitudes to IT, it would be useful for NHS organisations to investigate ways in which a predominately female workforce can be helped with IT use.

Acknowledgments

This study was funded by the UWE Faculty of Health and Social Care, School of Adult Nursing, Research and Staff Development Scheme.

The research team is grateful to UWE staff including Dr Jon Pollock and Dr Nikki Coghill for statistical advice and to Leigh Taylor in the Faculty Research Office for administrative support. We would also like to thank the R&D Directors, managers/ professional leads/matrons in three NHS Trusts for permission to distribute questionnaires to their staff. The contribution of all those who completed the questionnaire and attitude scales was vital in this project which could not have been completed without them.

References

[1] Brennan S. The NHS IT Project: The Biggest Computer Programme in the World... Ever! Oxford: Radcliffe Publishing Ltd.; 2005.

[2] Department of Health, Information for Health; 1998. [online] http://www.dh.gov.uk/assetRoot/04/01/17/90/04011790.pdf [accessed 01.12.06]

[3] Davis F. User acceptance of information technology: System characteristics, user perceptions and behavioural impacts. International Journal of Man-Machine Studies 1993;38: 475-487.

[4] Venkatesh V Davis F. A theoretical extension to the technology acceptance model: four longitudinal field studies. Management Science 2000;46(2) 186-204.

[5] Dixon D. The behavioral side of information technology. International Journal of Medical Informatics 1999;56(1-3): 117-123.

[6] Venkatesh V, Morris M, Davis G, Davis F. User Acceptance of Information Technology: Towards a Unified View. Management Information Science Quarterly 2003;27(3) 425-478.

[7] Rogers, E. Diffusion of Innovations (4th Ed) New York: Free Press, 1995.

[8] Lee, TT. Nurses' Adoption of Technology: Application of Rogers' Innovation-Diffusion Model. Applied Nursing Research 2004;17(4) 231-238.

[9] Stronge JH, Brodt A. Assessment of Nurses' Attitudes Toward Computerization. Computers in Nursing 1985;3(4) 154-158.

[10] Brumini G, Kovic I, Zombori D, Lulic I, PetroveckI M. Nurses' attitudes towards computers: Cross sectional questionnaire study. Croatian Medical Journal. 2005;46(1) 101-104.

[11] Kirshbaum MN. Are we ready for the Electronic Patient record? Attitudes and perceptions of staff from two NHS trust hospitals. Health Informatics Journal 2004 ;10(4) 265-276.

[12] Rosenbloom ST, Talbert D, Aronsky D. Clinicians' perceptions of clinical decision support integrated into computerized provider order entry. International Journal of Medical Informatics 2004 ;73, 433-441.

[13] Moody LE, Slocumb E, Berg B, Jackson D. Electronic health records documentation in nursing: nurses' perceptions, attitudes, and preferences. Computers, Informatics, Nursing 2004;22(6), pp. 337-44.

[14] Newton C. A study of nurses' attitudes and quality of documents in computer care planning. Nursing Standard 1995 ;9(38) 35-39.

[15] Sinclair M, Gardner J. Planning for information technology key skills in nurse education. Journal of Advanced Nursing 1999; 30(6), pp. 1441-1450.

[16] Wishart JM, Ward R. Individual differences in nurse and teacher training students' attitudes toward and use of information technology. Nurse Education Today 2002 ;22, 231-240.

[17] Oppenheim AN. Questionnaire Design, Interviewing and Attitude Measurement (2nd Ed.) Pinter: London; 1992.

[18] Rattray J, Jones MC. Essential elements of questionnaire design and development. Journal of Clinical Nursing 2007; 16, 234-243.

[19] Watson R, Thompson DR. Use of factor analysis in Journal of Advanced Nursing: literature review. Journal of Advanced Nursing 2006 ;55(3), 330-341.

[20] Bryman A, Craymer D. Quantitative data analysis with SPSS Release 10 for Windows : a guide for social scientists. Routledge, Hove; 2001.

[21] Streiner DL, Norman GR. Health Measurement Scales (2nd Ed.) Oxford: Oxford University Press; 1995.
[22] Data Protection Act 1988 [online] http://www.opsi.gov.uk/ACTS/acts1998/19980029.htm [accessed 23.11.06]
[23] Moody LE, Slocumb E, Berg B, Jackson D. Electronic health records documentation in nursing: nurses' perceptions, attitudes, and preferences. Computers, Informatics, Nursing 2004.;22(6), pp. 337-44.
[24] Doolan DF, Bates DW, James BC. The use of computers for clinical care: a case series of advanced U.S. sites. J Am Med Inform Assoc 2003; Jan-Feb;10(1):94-107.

Address for correspondence

Rod Ward, UWE Faculty of Health and Social Care,
Glenside Campus, Blackberry Hill,
Bristol BS16 1DD, UK.

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Development of Patient Centric Virtual Organizations (PCVOs) in Clinical Environment for Patient Information Management

Mohyuddin[a], W.A. Gray[a], Hazel Bailey[b], Wendy Jones[b], David Morrey[b]

[a]*Department of Computer Science, Cardiff University, Cardiff, UK*
[b]*Clinical Information Unit, Velindre NHS Trust, Cardiff, UK*

Abstract

A novel Virtual Organization framework which incorporates wireless technology support is presented in the research work. The Virtual Organization is designed for a clinical environment to provide better patient information management and enhanced collaborative working of multi-disciplinary care teams. The analysis studies the current clinical practices and looks at the general patient information resource structure currently in use for patient care. Based on this problem analysis and current requirements of the multi-disciplinary care team members, we propose a generic and sustainable Patient Centric Virtual Organization (PCVO) framework to complement the functionality of the existing infrastructure by incorporating wireless technologies support for improved patient information provision at the point of care. The preliminary results of the study identify and classify the specific point of care tasks suited to appropriate information resources needed by the care team members. This paper concentrates on the patient information management aspects brought in by incorporating wireless technologies at the point of care using patient information resources in a decentralized and distributed computing environment. This applied research is carried out in the secondary and tertiary care sector in the cancer domain. For the analysis and results of the pilot project, we have used a case study of a local NHS Cancer Hospital.

Keywords:

Patient Centric Virtual Organizations, Sustainable Patient Centric Framework, Patient Information Management, Patient Centered Care, Wireless Point of Care System

Introduction

The aim is to determine the functional requirements of multi-disciplinary care team members and provide a better infrastructure for the provision of relevant patient information at the point of care from existing information resources in an effective and efficient way by using a Virtual Organization (VO) framework. In the literature, Virtual Organizations (VOs) don't have an exact definition, as their functionality varies depending on the context of the organization, application and the way they are used. In general, VOs are formed by organizations or individuals who collaborate and work together to perform certain organizational objectives by accessing and sharing resources utilizing communication and information technologies [1], [2]. The team members in a VO might be separated by time or space. In the clinical environment, the healthcare professionals collaborating with a common objective of caring for a patient may form a VO. Some of the features which are basic for a VO can be described as: 'working for certain organizational objectives together; formation of teams for accessing and sharing resources; trust, collaboration and cooperation among team members; and usage of computer and communication technologies' [3]. In our VO, we are using wireless technologies as a means of information access and sharing among care team members due to their mobility. These technologies play an important role as a building block in a VO and serve as the backbone for all the processes within it. Looking at the nature of clinical environments with a VO context, we find a similar environment as the work practices in the clinical environment can be diverse; healthcare professionals with different skills might be caring for the same patient while working from different locations and at different times; and patient resources may be heterogeneous [4]. In our context, a VO is defined for the team members who might be dispersed physically or temporally but work to achieve a common goal using wireless technologies support in a novel way linked to traditional methods [5].

We describe how Patient Centric Virtual Organizations (PCVOs) can be developed in a clinical environment for better management of patient information using decentralized and distributed resources. PCVOs are supported by wireless technologies which play an important role in the patient treatment process at the point of care. This care is carried out at the secondary and tertiary care level by different multi-disciplinary care team members involved in the patient treatment. The care teams consist of different medical professionals having varied skills working together. In our study, the healthcare stakeholders considered are: clinicians, nurses and therapists, who access information from a variety of medical resources during a patient's treatment. Looking at the structure of secondary and tertiary care, these team members are the key players directly involved in the patient care. Since a VO provides a flexible framework, more healthcare professionals can be added or removed from the team according to patient care requirements. The analysis shows the existing information

resources used for patient treatment in the current patient care structure and reveals the issues faced by the care team members in this domain. This is used to suggest an approach of a PCVO at the clinical level which harmonizes well with the clinical settings and will utilize the existing information resources in a structured development. It will organize the patient information and support integrated working of the care team members for improved patient care. For the case study, the pilot project with anonymised data was carried out at Clinical Information Unit at Velindre Hospital, South East Wales Cancer Centre working at secondary and tertiary care level.

Motivations and problem analysis

Healthcare is a major industry and has continually evolved over time. Lack of timely information, medication errors and unavailability of centralized resources for shared information are some of the current challenges faced by hospitals [6]. Real time information at the patient bedside, which leads to better decision making support at the point of care, shows the importance of wireless devices in patient treatment. Existing computerized systems have limited capabilities as they cannot always provide appropriate information to patient care team members as required at the point of care. Current practices still involve use of manual and paper based work, which create problems such as storage of information in a suitable repository, manual data entry of the information into computerized systems, delays in system updates and mismanaged structure of patient information. One of the major concerns identified during the study was the provision of required patient information from decentralized / distributed resources to care team members at the point of care. Some of these problems can be addressed by using wireless technologies [5].

In general, the structure of resources at the secondary and tertiary care level is similar to other clinical environments. During the study, it was found that although clinical information systems contain most of the patient information required, they are not the only information resource used in patient treatment. Patient information is also stored in other information resources such as nursing notes and patient notes. Usually, at the point of care, nurses / patient notes contain most of the basic manual information about patients. So, although some of the patient notes information is stored in clinical information systems, care team members still have to refer to the manual patient notes during a patient interaction.

This highlights the need for a sustainable information infrastructure which is capable of managing the patient information resources and dealing with the patient information provision issues at the point of care. These issues are addressed by the proposed PCVO framework, which is based on a patient centric approach utilizing the power of wireless technologies at the point of care.

Related work

A VO and ad-hoc virtual team formation approach in a telecare project is discussed by Grootveld [7], in which different healthcare professionals work for the stroke service and access patient records in the emergency room for better diagnosis. It describes the stroke service which is a regional network of healthcare professionals as a VO. A VO and virtual clinic creation strategy for distributed clinical consultations is described in a case study of a prison telemedicine program which connects an academic medical centre to a prison hospital [8]. A concept of dynamic virtual collaborative healthcare teams dealing in homecare domain of cancer patients using fixed computers, mobile phones, PDAs and internet telephones for the provision of medical records is presented in enterprise project DITIS [9]. A virtual institutional infrastructure which includes different healthcare services and a linked structure of national networks is identified by a project for the integration of Cuban healthcare service market [10]. Some other papers have described work done to develop virtual communities, virtual teams and virtual systems. These cover the social context and issues of building these virtual environments in healthcare [11]–[13].

This related work has shown the benefits of virtual infrastructures and VO formation in different telecare and homecare projects, but no work has been done to analyze the potential benefits of a VO development with a Patient Centric approach at the clinical level particularly at the point of care. The potential of a VO as the patient information management infrastructure and its role at the point of care for patient information provision has not yet been investigated. Most of these environments utilized computerized systems as a communication channel, but the importance of wireless technologies in these environments at clinical level has not been utilized yet. The concept of a PCVO with wireless technologies support for patient information management as required at the point of care in clinical environment is the novel aspect presented in this paper and this area needs to be explored further.

Context analysis of existing patient information resources

This section covers details of common decentralized and distributed information resources available in an organization and through internet respectively for patient information and treatment at secondary and tertiary care level as shown in Figure 1.

Organizational resources

Clinical Information Systems most commonly provide a patient summary and the information required for patient diagnosis and treatment. This may include clinical trial information, surgery, radiotherapy, chemotherapy and home care. *Patient / Nurses Notes* are usually used by care team members for capturing the patient's vital signs, clinical workflow and bedside charting processes. Most of the information which is not available in clinical information systems is stored in patient notes. *Patient Administration*

Systems include the demographic details and patient waiting time. *Local Protocols* contain local care procedures e.g. care pathways and clinical trial protocols. In Figure 1, the highlighted area shows the Shared Information required for the purposes of patient diagnosis and treatment contained by Clinical Information Systems and Patient Notes which is obtained from Patient Administration Systems and Local Protocols [14]–[16].

Distributed resources

The distributed domain has a variety of resources for patient treatment and care like *Evidence Based Medicine (EBM) Resources, Pharmacy / Medical Information Resources and other Hospital Intranet Resources*, providing information from clinical evidence, medicine libraries, e-text books and bibliographic databases [15], [16].

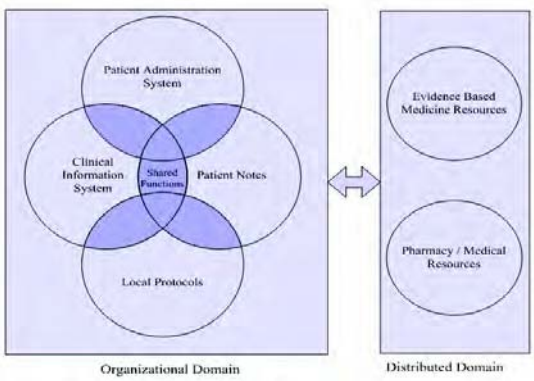


Figure 1 - Patient Information Resources Structure

Proposed approach and framework

This work presents PCVO as a Patient Centric approach at the organizational level with wireless technologies support. Looking at the building blocks of a VO, our approach is based on the functionality and tasks needed at the point of care in a clinical environment where different multi-disciplinary care team members work together in an integrated way to achieve a common goal, i.e. improved patient care. The Patient Centric approach shows that patient information is the central entity of the paradigm and a PCVO shows how each patient's data is used in different and improved ways. It aims to provide patient information at the point of care platform, which will be derived from different data sources like patient databases, other organizational resources and the internet. The PCVO framework shows only most commonly used organizational resources which are Clinical Information System and Patient Notes, because they also include the required patient information shared by the other organizational resources like Patient Administration Systems and Local Protocols, as represented by Shared Functions shown in Figure 1. Each PCVO is based on a single patient's information, so different care team members treating that patient become a part of that particular PCVO. They might be at different physical locations but will always work in a coordinated and common way for patient treatment under the same platform because the common patient information structure remains the same.

Considering the problem domain, our suggested approach presents the PCVO as an environment which works on top of existing information resources and incorporates a wireless system which is required to overcome some of the current point of care issues. It shows how a PCVO identifies and classifies the point of care tasks with respect to the information resources for care team members by using the support of a Wireless Point of Care System (WPoCS). Figure 2 shows the categorized tasks which are 'existing tasks', 'transferable tasks' and 'added tasks' or 'automated tasks' determined by classification for the appropriate information provision with respect to patient information resources. A WPoCS covering 'transferable, added or automated tasks' should play an important part in providing the needed functionality for care team members as shown in Figure 2, by capturing and providing the appropriate patient information at the point of care using PDAs.

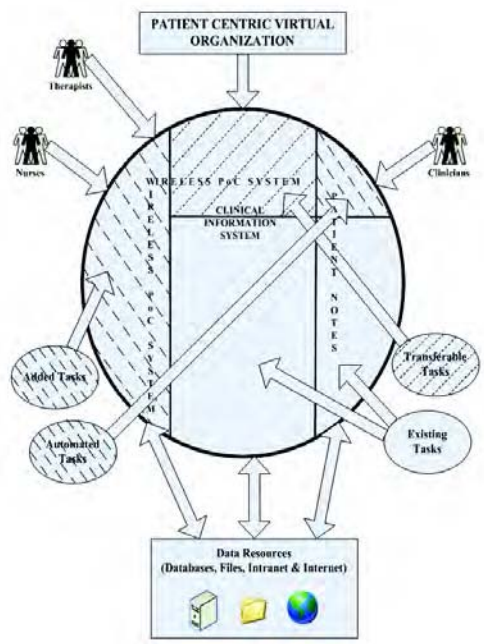


Figure 2 - Patient Centric Virtual Organization Framework

Study results: Point of care tasks classification

The results of this classification given below in the form of tables were achieved by analysis, direct observations, structured and semi-structured interviews and questionnaires conducted at various levels with different care team members, Information Analysts and IT staff in our case study conducted at the Velindre Hospital. The iterative process of this analysis and design covered the interviews, discussions and feedback from these stakeholders throughout all the phases - problem discovery, framework

approach, suitability and usefulness of task classification, development and modelling of test cases, and is still continuing in the prototype development and evaluation phases as well. The point of care tasks of the care team members are divided into three categories.

Point of care tasks suited to existing computerized clinical information system

Table 1 identifies examples of the point of care tasks which are better suited to a clinical information system. The tasks identified present a general level of usability compatible with information systems working in different environments.

Table 1 - Point of Care Tasks suited to Clinical Information System

Care Team Members	Point of Care Tasks suited to Clinical Information System
Clinicians	Data entry / order entry involving time consuming and detailed keyboard input
	Patient's detailed referrals and complete reports
	X-Ray and other large image analysis including MRI
	Looking at the case notes of the patients for detailed information
	Searching and comparing the cases for best EBM
	Searching the pharmacy resources for selection of medicines
Nurses	Data entry / order entry involving time consuming and detailed keyboard input
	Previous medical notes in dates format
	Looking at the text reports of X-Ray
Therapists	Previous medical notes in dates format
	Data entry / order entry involving time consuming and detailed keyboard input
	X-Ray and other large image analysis including MRI

Point of care tasks transferable to WPoCS

Table 2 shows examples of the suitable point of care tasks that can be transferred from Clinical Information System to a wireless system to provide easier access of relevant patient information. The framework doesn't exclude these tasks from Clinical Information System, but should provide a better interface for improved task accomplishment using wireless devices.

Table 2 - Point of Care Tasks transferable to WPoCS

Care Team Members	Point of Care Tasks transferable to WPoCS from Clinical Information System
Clinicians	Checking the laboratory test results
	Accessing summary disease history
	Accessing summary treatment history
	Checking current treatment status of the patient
	Checking current disease status of the patient
	Looking at current location status of the patient
	Checking the complications of treatment
	Checking the clinical summary of the patient
	Patient record retrieval by scanning barcode through devices
Nurses	Checking current treatment status of the patient
	Checking current disease status of the patient
	Looking at current location status of the patient
	Patient record retrieval by scanning barcode through devices
Therapists	Checking previous therapy records / past history
	Knowing the last examination of the patient, action measures taken and current status
	Checking the current location of the patient
	Patient record retrieval by scanning barcode through devices

Point of care tasks added or automated by WPoCS

Table 3 shows examples of new tasks that can be achieved using a wireless system at the point of care to assist care team members in the patient care process. It also includes the tasks that can be automated using handheld devices at the point of care instead of using a paper based approach. It must be remembered that WPoCS do not replace all the paper / patient notes but can only be utilized to automate suitable tasks which are better suited to wireless devices.

Table 3 - Point of Care Tasks added / automated by WPoCS

Care Team Members	Point of Care Tasks added / automated by WPoCS from other existing manual resources
Clinicians	Consulting reference notes / guidelines for decision making like blood pressure classification readings and average pulse rate for age-wise children and adults
	Accessing physical / vital signs like temperature, pulse rate, respiratory rate, blood pressure, body weight etc
	Looking at previous vital signs for comparison
	Recording of ward handover notes
	Accessing the trial protocols
	Accessing the treatment protocols
	Accessing the clinical pathways
	Consulting the cases for best EBM
	Consulting the right pharmacy / medical information for clinical use of medicines on the basis of symptoms
Nurse	Electronic data capture / recording of patient's vital signs like body temperature, pulse rate, respiratory rate, blood pressure, body weight etc
	Checking previous vital signs to compare
	Comparing the patient temperature and other vital signs using graphs
	Checking drugs details and timings of dosage given to patients
	Alerts and reminders for checking patient status and treatment
Nurses	Right identification of the patients and knowing their current status and details
	Step by step workflow process recording, for example, patient's check-in / position status, physical / vital signs notes etc
	Recording of ward handover notes
	Consulting actions and rationale for the specific nursing protocols
	Accessing the treatment protocols
	Right identification of the patients and knowing their current status and details
	Accessing the clinical pathways
	Accessing the trial protocols
	Accessing medical device / patient safety alerts
	Electronic bedside charting
	Pre-defined screens for data capturing
	Checking the drug prescription
Therapists	Consulting reference notes / guidelines for decision making like blood pressure classification readings and average pulse rate for age-wise children and adults
	Checking the social history / living conditions of the patient
	Knowing the last examination of the patient and action measures taken
	Accessing the treatment protocols
	Accessing the clinical pathways
	Accessing the trial protocols

Research limitations and challenges

This paper focuses on generic patient information management aspects at the point of care in the clinical environment. It concentrates on the point of care functionalities that can be achieved using organizational resources within hospital environment by a PCVO framework. We have not covered other distributed patient information resources aspects in this paper. Looking at the big picture, some of the challenges and issues must be considered if this framework is to be adopted in practice. These include the administrative,

control, management and other aspects of PCVOs. On the other hand, with regard to WPoCS working in a wireless network, security and bandwidth issues for the distributed information access must be considered. Finally, these PCVOs will not replace the current patient care infrastructure but take on suitable point of care functionalities with wireless technologies support using existing information resources in a novel and improved way.

Conclusions and future directions

This paper has discussed PCVOs development and its potential benefits in the clinical environment. The PCVOs will harness the power of wireless technologies through WPoCS, hence providing patient information management by timely information access at the point of care. The classification of care team tasks with respect to information resources shows how the PCVOs may provide a better approach for using patient information resources at the point of care without changing the current infrastructure of the existing clinical environment. The outcome of the pilot study results show there is a broad acceptance of the pilot project findings. The positive feedback from care team members received in discussions and interviews show the suitability of these tasks in addressing issues faced during their routine tasks.

So far we have designed and modelled the test cases for WPoCS. The next step is the development and implementation of a WPoCS prototype using a Pocket PC Emulator and PDA is in progress. This will determine its practical usage in the healthcare spectrum. To evaluate the pilot project we are working on qualitative measures with the healthcare stakeholders. We will also be improving and enhancing the system using feedback from care team members. The study suggests the real and tangible benefits that may be achieved with our framework. The ongoing work is likely to prove the usefulness and practical benefits to patient care in the secondary and tertiary care of cancer.

References

[1] A. Mowshowitz, "Virtual Organization," vol. 40, 1997.
[2] Y. P. Shao, M. K. Lee, and S. Y. Liao, "Virtual Organizations: The key dimensions," in Academia / Industry Working Conference on Research Challenges (AIWORC'00), 2000.
[3] M. Rittenbrunch, H. Kahler, and A. B. Cremers., "Supporting cooperation in a Virtual Organization," in Proceedings of the international conference on Information Systems, 1998.
[4] C. Bossen, "The parameters of Common Information Spaces: the heterogeneity of cooperative work at a hospital ward," in Proceedings of ACM conference of CSCW, 2002.
[5] Mohyuddin, W.A. Gray, D. Morrey, and W. Jones, "Incorporating wireless technology into Virtual Organizations supporting the work of healthcare teams," in Fourth Annual IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOMW'06), 2006.
[6] N. Wickramasinghe and S. Goldberg, "How $m=ec^2$ in healthcare," in Proceedings of International Conference on Information Technology: Coding and Computing, 2004.
[7] M. Grootveld, J. Swaal, R. Brussee, R. H. in't Veld and M. Michael-Verkerke, "Virtual Organisations in healthcare: Organisation, people and technology," Telematica Institute, Netherlands, Tech. Rep., 2004.
[8] J. W. Turner, "The integration of new communication technologies to form Virtual Organizations: A case study of a prison telemedicine program," in Pacific Medical Technology Symposium, 1998.
[9] B. Pitsillides, A. Pitsillides, G. Samaras, and M. Nicolaou, Virtual Teams: Concepts and Applications. ICFAI University Press, 2004, ch. DITIS: Virtual collaborative teams for improved home healthcare.
[10] A. C.Seror, "Integrating virtual infrastructures: A sociometry of the Cuban national healthcare system," in Academy of Management National Meetings: Democracy in a Knowledge Economy, Technology and Innovation Management Division, 2003.
[11] F. Alemi, "Virtual managed care organizations: the implications of technology-based patient management," PubMed, vol. 4, 1998.
[12] J. D. Becker, R. D. Ballentine, A. Lee, and C. Townsley, "Collaborative technology tools for virtual teaming," in Proceedings of the 5th Americas Conference on Information Systems, 1999.
[13] D. W. Forslund, R. L. Phillips, D. G. Kilman, and J. L. Cook, "Experiences with a distributed virtual patient record system," in Proceeding AMIA Annual Fall Symposium, 1996.
[14] "Information for health: An information strategy for modern NHS 1998-2005 – Executive Summary" Department of Health, UK, Tech. Rep, 1998.
[15] J. Griffiths, "Better information better health - Cancer Services Information Framework," NHS Wales, UK, Tech. Rep., 2000.
[16] "Velindre NHS Trust Annual Report: 2003-2006," Velindre NHS Trust, Cardiff, Wales, UK, Tech. Rep., 2006.

Address for correspondence

Mohyuddin, Department of Computer Science,
Queen's Buildings,
5 The Parade, Roath,
Cardiff University,
Cardiff, CF24 3AA, UK.
Email: Mohyuddin@cs.cf.ac.uk

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

SPIRS: A Framework for Content-based Image Retrieval from Large Biomedical Databases

William Hsu[a], L. Rodney Long[b], Sameer Antani[b]

[a] *Medical Imaging Informatics Group, University of California, Los Angeles, CA, USA*
[b] *National Library of Medicine, National Institutes of Health, Bethesda, MD, USA*

Abstract

With the increasing use of medical images in clinical medicine, disease research, and education, the need for methods that effectively archive, query, and retrieve these images by their content is underscored. This paper presents the implementation of a Web-based retrieval system called SPIRS (Spine Pathology & Image Retrieval System) at the U. S. National Library of Medicine that demonstrates recent developments in shape representation and retrieval from a large dataset of 17,000 digitized x-ray images of the spine and associated text records. Users can search these images by providing a sketch of the vertebral outline or selecting an example vertebral image and some relevant text parameters. Pertinent pathology on the image/sketch can be annotated and weighted to indicate importance. This hybrid text-image query yields images containing similar vertebrae along with relevant fields from associated text records, which allows users to examine the pathologies of vertebral abnormalities. Initial experiments with SPIRS have demonstrated the potential for this system, particularly on a large dataset of clinical images.

Keywords:

medical informatics applications, information storage and retrieval, content-based image retrieval, visual access methods, web-based systems

Introduction

Medical imaging is an increasingly important and versatile tool for acquiring information about a patient and disease. Image features (e.g., shape, color, or texture) have been used to diagnose a variety of conditions. Hospitals have been adopting technology such as Picture Archiving and Communication Systems (PACS) and Hospital Information Systems (HIS) to assist in the digital collection, organization, and storage of patient data. The goal of these systems is to make patient data more accessible; in reality, the amount of data that is entered and stored in these systems have created a new challenge for effective information indexing and retrieval. Historically, PACS have limited users to query by certain keywords (e.g., unique patient identifier, fields in the image header). However, these keywords often do not capture the richness of features depicted in the image itself. Over the past two decades, content-based image retrieval (CBIR) systems have been researched to address the problem of indexing and retrieving visual data in a variety of domains [1]. Rather than limiting queries to textual keywords, CBIR users provide a query sketch/image, which is then used to find similar images of the same modality, anatomical region, and disease along with the associated text records.

The goal of this work is to develop a retrieval system that implements recent developments in shape representation, efficient indexing, and similarity matching; supports whole and partial shape matching, which enables a wide range of meaningful queries to be posed; and utilizes a distributed framework, which is customizable to the needs and constraints of healthcare environments. The result, Spine Pathology & Image Retrieval System (SPIRS)[1], provides a Web-based interface for performing image retrieval on a database of digitized spine x-rays using the morphological shape of the vertebral body. A query editor enables users to pose queries by sketching a unique shape or selecting or modifying an existing shape from the database. Additional text fields enable users to supplement visual queries with other relevant data (e.g., anthropometric data, quantitative imaging parameters, patient demographics). These hybrid text-image queries may be annotated with pertinent pathologies by selecting and weighting local features to indicate importance. Query results appear in a customizable window that displays the top matching results and related patient data. SPIRS provides a working proof-of-concept that demonstrate the capability of accommodating large amounts of imaging data expected in the near future.

Background

The problem of retrieving information based on image content has been researched by various groups since the early 1990's resulting in the development of tools such as QBIC, Virage, and Blobworld [2]. CBIR in medicine has been an active area of research [3], but only a small number of proposed systems such as ASSERT [4] and IRMA [5] have been demonstrated in the clinical environment. In addition, although many large image databases exist, such

1 http://archive.nlm.nih.gov/spirs

Selected for best paper award.

as the National Cancer Imaging Archive (NCIA) or the Lung Imaging Database Consortium (LIDC) created under the aegis of the Cancer Imaging Program[2] at the U.S. National Cancer Institute (NCI), these efforts have concentrated on data collection and transmission but have left development of applications to the research community. Lack of CBIR adoption is attributed partly to the difficulty of integrating current implementations with existing healthcare systems [3]. SPIRS addresses this issue by (i) utilizing open standards to communicate among components, which can be extended to support data encryption to meet privacy regulations, and (ii) using modular open source software components. Much work has been done in the past on visual querying paradigms, as reviewed in [6]. Building upon these paradigms, SPIRS combines visual and text queries to provide users with greater flexibility in retrieving relevant results.

Biomedical database

At the U.S. National Library of Medicine (NLM), the focus of CBIR research has been to develop systems capable of performing a range of queries on large medical multimedia databases comprising various biomedical images and patient health data. Such a database in current use contains digitized spine x-rays and associated metadata from a large nationwide survey, the National Health and Nutrition Examination Survey (NHANES), conducted regularly by the National Center for Health Statistics in the United States. The goals of NHANES include estimating prevalence of selected diseases, monitoring disease trends, and studying the relationship between nutrition and health. The Lister Hill National Center for Biomedical Communications, a research division of the NLM, maintains data from the second survey, NHANES II, which was collected between 1976 and 1980 and featured over 20,000 participants. Each participant's record includes 2,000 textual data points such as health questionnaire answers, anthropometric information, and results from a physical exam. This textual data is stored in a relational database (e.g., MySQL). Supplementing the textual data is a collection of 17,000 cervical and lumbar spine x-rays that were taken from patients aged 25 – 74. These images were originally on film and subsequently digitized using a 146 dpi scanner resulting in 140 Gigabytes of data. The collection is considered valuable to radiologists, bone morphometrists and researchers interested in osteo-arthritis, and medical educators. Domain experts reviewed a sample of the data and identified 23 key biomedical features exhibited in the x-rays. Of these, anterior osteophytes, subluxation/spondylolisthesis, and disc space narrowing were determined to be frequently occurring and reliably detectable.

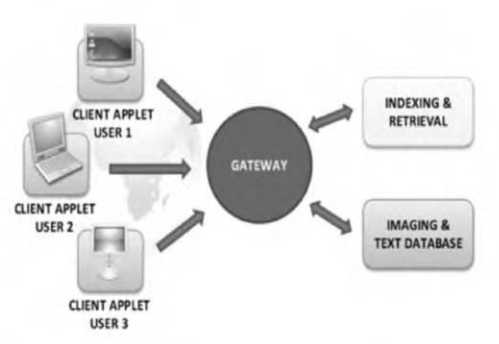


Figure 1 - The distributed architecture of SPIRS

Each of these key features may be identified by examining the boundary shape of the vertebra. However, identifying images that exhibit particular types and severities of these pathologies is extremely tedious for such a large collection. SPIRS and related tools [7], developed to analyze the images and populate the database, automate many tasks and enable retrieval of the images using sketches of the vertebral boundary with the desired feature. Furthermore, SPIRS may be used to determine what features (e.g., protrusion on the anterior edge of the cervical vertebra) are consistently associated with a certain symptom (e.g., neck discomfort) or whether a certain feature is a precursor to more serious illnesses (e.g., arthritis). Illnesses may be documented in the text of the patient record as a survey response or in the medical diagnoses.

System framework

The system's distributed architecture, shown in Figure 1, consists of four components: (i) the client applet, which provides a front-end for users to pose queries and interact with results, (ii) a gateway that acts as an mediator between client and server-side components, (iii) the indexing and retrieval server, which performs the feature representation and similarity matching, and (iv) the databases containing images and associated text data.

Indexing and retrieval

In our continuing research, we have explored and implemented various algorithms to extract, index, and retrieve shapes of vertebral bodies from x-ray images. Extraction of vertebral shapes is accomplished using manual and semi-automated algorithms such as active contour segmentation, active shape modeling, and hierarchical segmentation [8]. These algorithms have been implemented with the intention of applying them to a variety of image features; therefore, the system can be extended to a wider variety of images and text data. The resulting vertebral shape boundaries are then treated as closed polygons and represented in a variety of forms, such as Polygon Approximation, Fourier Descriptors, or geometric shape properties that help uniquely identify and characterize

2 http://imaging.cancer.gov/

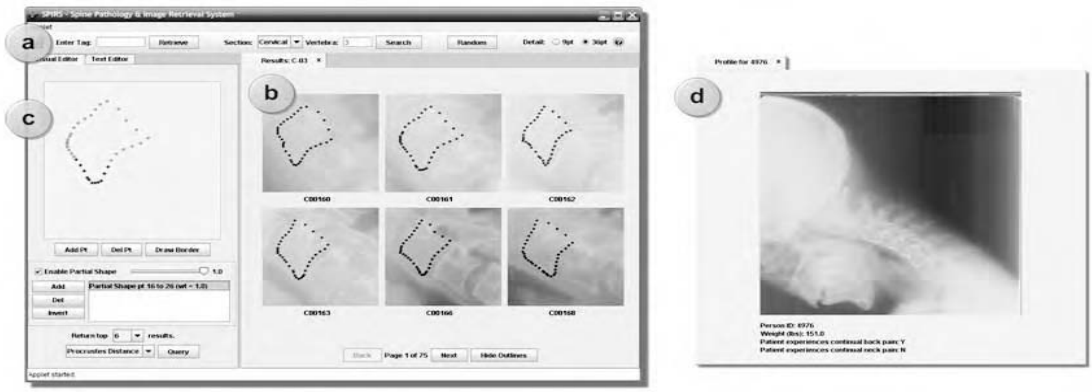


Figure 2 - A screen capture of the SPIRS client applet left) and a cropped screen capture of the overall results view tab (right)

them [9]. Traditionally, these representations are stored as a feature vector and compared individually with the query. However, for large image archives, this linear comparison approach would be impractical. SPIRS therefore uses a coordinate tree to efficiently index shapes and optimize retrieval time. Embedded with the indexing process are appropriate distance measures which identify matching shapes [10].

Client applet

The client, shown in Figure 2, is a Java applet that runs in a Web-browser and provides the user with the necessary functionality to query for particular shapes of interest and interact with the results. The applet (i) is a thin client that provides the look-and-feel of a standalone application but uses minimal computing resources, (ii) does not require installation of files onto the local drive, (iii) is easily modifiable without affecting core-CBIR functionality, and (iv) is platform independent.

Query toolbar

The query toolbar (Fig. 2a) supports query-by-example by allowing users to find vertebral shapes or spine x-ray images that match descriptors stored in the text database, such as image tag or vertebra type and number. A query vertebral shape can be selected from the returned x-ray images in the Overall View or from the selection of vertebral image crops in the Cropped View shown in Fig. 2b.

Overall view

The Overall View (Fig. 2d) displays the entire patient x-ray and allows the user to examine and select any of the segmented vertebral shapes. In addition to the unprocessed version of the scanned image, two forms of enhanced images are also available for improved visualization of subtle detail, viz., one processed with unsharp masking to enhance edges, and the other an adaptive histogram-equalized image which improves contrast. The user may use the mouse hover over and view the vertebral outlines as image overlays. A mouse click on the desired shape selects and transfers it to the query editor.

Cropped view

The Cropped View (Figure 2b) allows a user to inspect multiple vertebral shapes at once. Each displayed vertebra is cropped from the original patient x-ray using the boundary information associated with its shape. The crops are then normalized so that each vertebra is facing in the same direction, which simplifies the comparison between various vertebrae.

Query Editor

The query editor (Figure 2c) enables the user to pose queries using query-by-sketch, select text fields, and supply information pertaining to health history, anthropometric data, quantitative imaging parameters, and demographic data. The visual query component of the editor is a canvas where the user may choose to draw an entirely new shape or edit points that have been imported from an existing shape found using the query toolbar. The query editor also supports multiple partial shape query specification. This feature provides the functionality to select parts of the vertebral shape enabling the algorithms to focus on these boundary intervals, which may exhibit significant pathology. Partial shape querying is shown in the figure as the highlighted interval along the vertebral boundary. Finally, the user can configure the query execution parameters by selecting the retrieval algorithm and limiting the number of results.

Results view

The Results View offers two displays. The Cropped View (Figure 2b) shows returned matching vertebrae while the Overall View (Figure 2d) appears when a user selects one of the resulting vertebral shapes. In the Overall View, the entire x-ray of the patient is shown with the matching vertebra highlighted. Relevant text fields from the text database are displayed alongside the image and may be customized by the user. A unique feature of this system is the ability to use resulting shapes as queries. This can be considered as a form of relevance feedback through the use of iterative querying. More advanced relevance feedback methods have been explored [12] as standalone MATLAB programs and are under development for use with the Web-based system.

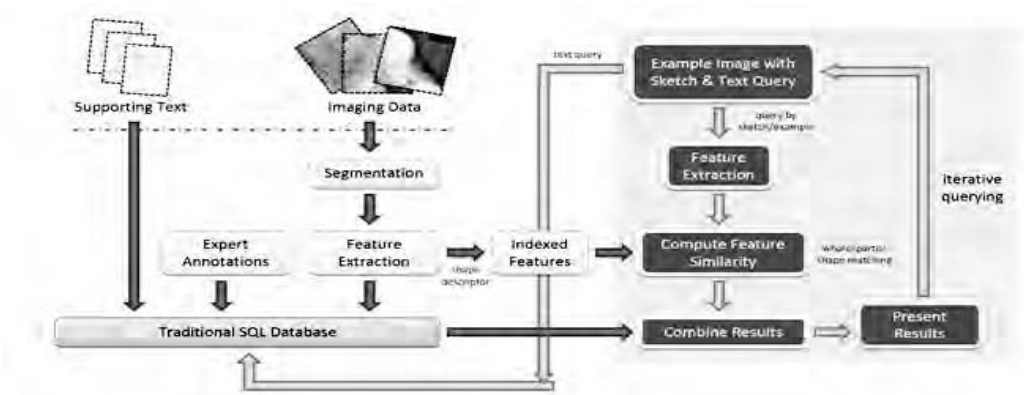


Figure 3 - The query handling process (shaded region) in the context of the overall retrieval system

Gateway

SPIRS utilizes standard Web communication interfaces to enable interaction with its components. The gateway is a Java servlet that acts as a mediator between client applets and server-side components. It manages multiple simultaneous connections (users) as separate session and queues requests to the core CBIR engine. The servlet translates text components into SQL query syntax for interacting with a MySQL database. In addition, the core architecture of the gateway makes interaction with other geographically distributed CBIR systems possible. An example is the interoperability between IRMA [5], which is a system located in Germany for CBIR of medical images using overall image intensity features, and SPIRS. While IRMA's interface provides shape query support for subtle localized pathology, it utilizes the SPIRS server located at the NIH in the USA for computing shape-similarity of spine x-rays.

Query handling

Although much CBIR research has focused on overcoming technical challenges such as developing accurate and efficient retrieval algorithms, in order for widespread CBIR adoption and use, systems need to provide users with the ability to exploit the capabilities of these algorithms in managing large biomedical multimedia databases. With this goal, we believe that SPIRS allows users to pose meaningful queries that have both clinical and research applicability.

Partial shape matching

A problem with many current CBIR approaches for medical images is that they operate on the image as a whole. Whole shape matching methods have been demonstrated to be more effective in identifying shapes with gross similarity because these algorithms operate over the entire shape. They are ineffective, however, in matching subtle shape characteristics, which may have critical pathology. It has been observed that pathologies of interest on a vertebral outline are often localized along a short interval on the boundary. For example, osteophytes are only expressed along the "corners" on the vertebral boundary as seen in the sagittal view. Partial shape matching (PSM) [7, 11] has been implemented in SPIRS allowing users to sketch or identify only the local interval of interest on the vertebral boundary. For instance, consider a user interested in identifying patients in the NHANES II data with a claw anterior osteophyte, which occurs when a spur of triangular shape rises from the vertebral rim and curves toward the adjacent disk. In the query editor, the user would highlight the subset of points that comprise the spur, weight these points greater than the other points in the shape, and execute the PSM query. Initial experiments have shown that PSM can be used to automatically classify the severity of a pathology depicted in the query shape such as whether the claw osteophyte is *slight*, *moderate*, or *severe* depending on criteria defined by domain experts [13].

Query execution

Figure 3 illustrates the process of query execution and how it integrates with the overall CBIR system. When the user formulates a hybrid text-image query, the query is separated into its visual and textual components. Relevant features and annotations are extracted from the visual component and matched against the indexed features using specific similarity computation methods. Parameters from the textual component are executed against the database containing survey and patient data. The results of these individual searches are then combined before the results are presented to the user. For example, a user who is researching the correlation between certain shape features and the degenerative spine disease spondylosis may be interested in the intersection of results between a visual query depicting, for example, a traction spur and patients who have been diagnosed with spondylosis.

Interaction with results

The standard method of reviewing the outcome of a query is through the Results View, which displays the whole patient x-ray with the matching vertebral shape highlighted and relevant text descriptions displayed alongside the image. The retrieval algorithms have been evaluated to 68% relevance (precision and recall) when querying for specific osteophyte type (claw or traction) or severity [13]. An additional 22% performance improvement is observed through use of relevance feedback methods [12, 13]. The shortcomings in the performance are linked to (i) erroneous determination of the query semantics and (ii) limitations of the shape matching algorithm. To overcome these problems, SPIRS allows users

to select a matching result as the basis of a new query. Traditionally, user interactivity has helped in minimizing similar problems with text retrieval, and user feedback has often been analyzed and employed to improve retrieval relevance. By allowing users the option to iteratively query and refine results, SPIRS implements a basic form of relevance feedback, which will be enhanced with a novel advanced weighted hierarchical feedback method using short-term memory [12] and other approaches as they are migrated from laboratory prototype routines.

Discussion

Image management and pathologically sensitive content-based image retrieval systems are increasingly necessary to interact with the growing volume of biomedical imaging data. In spite of their acknowledged importance, shortcomings in current approaches have prevented their widespread acceptance into medical research, practice, and education. We believe that a biomedical CBIR system should be easily accessible, extensible, and capable of supporting a rich set of segmentation, validation, indexing, query, retrieval, and visualization methods developed using open software and standards. It is difficult for individual systems to support unique requirements of different biomedical images. This can be addressed by providing the capability to combine this system with others (possibly geographically distant) that have complementary features. The SPIRS framework is capable of interacting with and retrieving relevant information from large databases of image and patient data using hybrid image and text query methods. It implements novel shape representation and similarity matching embedded with an index tree that allows efficient retrieval. It aims to capture query semantics through support of advanced mechanisms like multiple partial shape matching and iterative querying that provides simple yet effective relevance feedback to the system. SPIRS is built using open standards and is simultaneously developed as a service, which enables its integration with other complementary information retrieval systems. Although SPIRS focuses on shape-based queries its framework is extensible to adopt features particular to other biomedical image and data collections, e.g., its core architecture is being extended to include color, texture, and spatial location in uterine cervix images from the National Cancer Institute [13].

Future work

Future goals for the project include (i) data collection, assimilation, and validation, (ii) system feature enhancements, (iii) improved retrieval quality by learning from user feedback, and (iv) improved user interaction and visualization. Extracting and validating vertebral shapes is an ongoing process which will (at regular intervals) add to the 8,000 shapes indexed currently. In addition, three board certified radiologists are validating the segmented shapes and identifying relevant pathology on them and on the spine. Planned feature enhancements include integration of a generalized shape segmentation toolbox, which is a currently a standalone application, incorporation of additional similarity, relevance feedback, and visualization algorithms, and development of a formalized XML specification for integrating other local and global information systems (or resources) such as WebMIRS and IRMA with core SPIRS services. A multi-user comprehensive qualitative study of the interface is also planned.

Acknowledgements

This research was supported by the Intramural Research Program of the National Institutes of Health (NIH), National Library of Medicine (NLM), and Lister Hill National Center for Biomedical Communications (LHNCBC). The authors thank Dr. Alex Bui at UCLA for his valuable feedback.

References

[1] Smeulders A, Worring M, Santini S, Gupta A, Jain R. Content-based image retrieval at the end of the early years. IEEE Trans PAMI. 2000; 22(12):1349-80.

[2] Antani S, Kasturi R, Jain R. A survey on the use of pattern recognition methods for abstraction, indexing and retrieval of images and video. Pattern Rec. 2002;35(4):945-65.

[3] Muller H, Michoux N, Bandon D, Geissbuhler A. A review of content-based image retrieval systems in medical applications—clinical benefits and future directions. Int J Med Info. 2004; 73(1):1-23.

[4] Shyu CR, Brodley CE, Kak AC, Kosaka A, Aisen AM, Broderick LS. ASSERT: A physician-in-the-loop content-based retrieval system for HCRT image databases. Comp Vis & Im Understanding 1999; 75(12):111-32.

[5] Thies C, Güld MO, Fischer B, Lehmann TM. Content-based queries on the CasImage database within the IRMA framework. Lec Notes Comp Sci 2005;3491:781-92.

[6] Catarci T, Costabile MF, Levialdi S, Batini C. Visual query systems for databases: a survey. J Vis Lang & Comp. 1997;8(2):215-60.

[7] Antani S, Long LR, Thoma G. Content-based image retrieval for large biomedical image archives. Proc. 11th World Cong Med Info 2004:829-33.

[8] Long LR, Antani SK, Thoma GR. Image informatics at a national research center. Comp Med Im Graphics 2005;29:171-93.

[9] Antani SK, Natarajan M, Long JL, Long LR, Thoma GR. Developing a comprehensive system for content-based image retrieval of image and text from a national survey. Proc. SPIE Med Imaging; 2005:152-61.

[10] Qian X. Shape indexing and its optimization in medical image databases. PhD thesis, Yale University; 2005.

[11] Shou J, Antani S, Long LR, Thoma GR. Evaluating partial shape queries for pathology-based retrieval of vertebra. In: Proc. 8th World Multiconf Sys, Cyb & Info.; 2004:155-60.

[12] Xu X, Lee DJ, Antani SK, and Long LR. Relevance feedback for spine x-ray retrieval. Proc. 18th Int Symp CBMS; 2005 pp. 197-202.

[13] Thoma GR, Long LR, Antani S. Biomedical imaging research and development: knowledge from images in the medical enterprise. Tech Report. LHNCBC, NLM, NIH; Sep. 2006. Report No.: LHNCBC-TR-2006-002.

Address for correspondence

William Hsu
UCLA Medical Imaging Informatics Group
924 Westwood Blvd, Ste 420
Los Angeles, CA 90024, USA
Email: willhsu@mii.ucla.edu

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

The Quality of Reporting of Health Informatics Evaluation Studies: A Pilot Study

Jan Talmon[a], Elske Ammenwerth[b], Thom Geven[a]

[a]*Dept of Medical Informatics, Maastricht University, Maastricht*
[b]*Institute for Health Information Systems, UMIT - University for Health Sciences, Medical Informatics and Technology, Hall in Tyrol, Austria*

Abstract

The quality of reporting of evaluations in Health Informatics is considered to be open for improvement. STARE-HI provides guidance for improved reporting. This study aims at establishing the quality of reporting prior to dissemination of STARE-HI. All evaluation papers in three leading Medical Informatics Journals published in 2005 have been assessed with the aid of STARE-HI. A total of 48 papers have been analysed with respect to title, abstract, keywords and study context. Main results indicate that the type of study is seldom reported in the title, that the structured abstract contains more information than an unstructured abstract, that important keywords are often forgotten. Also important information about the system assessed is often not mentioned explicitly. Our conclusion is that there is room for improvement in the reporting of evaluations in health informatics. STARE-HI can be helpful in suggesting a structure for evaluation papers in Health Informatics and issues that should be included in such papers.

Keywords:

STARE-HI, quality control, evaluation studies, medical informatics, publishing

Introduction

Quality of reporting of studies has been a concern in medicine in the past. It has led to the development of several statements with guidelines for proper reporting of randomized controlled trials – CONSORT (1), diagnostic studies - STARD (2) etc.

A few years ago, it was recognized during a workshop on evaluation in Innsbruck that the quality of reporting of evaluation studies in health informatics was not optimal. Improvement of reporting was considered to contribute to establishing a more solid evidence base for health informatics (3).

In 2006, the authors (JT and EA) coordinated the development of STARE-HI (STAtement on the Reporting of Evaluations in Health Informatics). With input of several researchers with expertise in the evaluation of health informatics applications a draft version of STARE-HI was developed in an iterative way. Currently a draft is made available on the internet for comments (4). Based on comments from the scientific community, a final version of STARE-HI will be prepared and submitted for publication in one or more Medical Informatics scientific journals.

It has been demonstrated that CONSORT has resulted in improved reporting of RCTs (5, 6). As to be able to prove that STARE-HI has resulted in improved reporting, a baseline measurement on the completeness and quality of current reporting practices is needed. The objectives of the current study are a) to test first how well the current version of STARE-HI can be used to assess the completeness and quality of reports on evaluation of health informatics applications and b) to test whether the selected assessment approach is feasible for application in a larger baseline study on the quality of current reporting in relation to the proposed STARE-HI.

STARE-HI

The scope of STARE-HI is to provide guidelines for the reporting of evaluations in Health Informatics. We consider as the object of evaluation the health informatics application that is assessed. This assessment can cover the technical artefacts as such with its functionality, but also the effect the artefact has on the surrounding organizational environment, the working procedures and social and psychological issues.

Table 1 presents the issues that are considered relevant in STARE-HI to appear in a report on the evaluation of a health care IT application together with the results of this pilot study.

Materials and methods

For the purpose of this study, all issues of the International Journal of Medical Informatics (IJMI), Methods of Information in Medicine (MIM) and of the Journal of the American Medical Informatics Association (JAMIA) that appeared in 2005 were hand searched for articles that would fit the purpose of STARE-HI.

Table 1 - STARE-HI: Recommended items to be included in a report on the evaluation of a Health Informatics application. See (4) for a brief description of each item.

Item #	Item
1	Title
2	Abstract
3	Keywords
4	Introduction
4.1	Scientific background
4.2	Rationale for the study
4.3	Objectives of study
5	Study context
5.1	Organizational setting
5.2	System details and system in use
5.3	Study constraints
6	Methods
6.1	Study design
6.2	Frame of reference
6.3	Participants
6.4	Study flow
6.5	Outcome measures or evaluation criteria
6.6	Methods for data acquisition and measurement
6.7	Methods for data analysis
7	Results
7.1	Description of study population
7.2	Unexpected events during the study
7.3	Study findings and outcome data
7.4	Unexpected observations of outcomes
8	Discussion
8.1	Statement of principal findings
8.2	Strengths and weaknesses of the study
8.3	Results in relation to other studies
8.4	Meaning and generalisability of the study
8.5	Unanswered and new questions
9	Conclusion
10	Conflict of interest
11	References
12	Appendix

Based on title and abstract two of the authors (JT, EA) identified candidate papers independently of each other. Discrepancies among the identifications were discussed as to define more precisely the scope of STARE-HI.

Next, for all selected papers the full text version was obtained. Each paper was assessed by TG with respect to the following aspects of the STARE-HI guideline: Title (item #1 of table 1), abstract (item #2), keywords (item #3) and study context (item #5.1 and #5.2). For this pilot we wanted to restrict the analysis to a number of key items. The four items listed above were selected because a) title, abstract and keywords help in identifying relevant papers for reviews and systematic analyses and b) study context plays an important role for assessing whether a paper is applicable for the setting of the reader. In addition, as editor and reviewers for several Medical Informatics journals we had the feeling that this item in particular is often not well addressed.

In STARE-HI for each item several aspects are being specified that should preferably be included. We counted how many aspects were dealt with for the items title, abstract and keywords in the papers and compute average score across the papers.

In the title, we expect to find the type of system assessed, the study question and the study design being mentioned.

For the abstract, we assume that the following aspects are being addressed: objective, setting, participants, measures, study design, major results, limitations and conclusions.

For keywords, we looked for the term “Evaluation” or equivalents, and for descriptors for the type of system, the setting of the study, the outcome measures and the study design.

For the study context, we scored on a three point scale (0=missing, 1= partially described, some issues unclear, 2=satisfactory description) the degree in which the organisational setting was described (#5.1) and to what extent system details were reported (#5.2). In case of non-compliance of a paper with those items the identified shortcomings were recorded as to be able to provide more detailed suggestions how reporting can be improved.

In each paper we counted also (#5.2) whether the phase in the lifecycle was mentioned, whether the intended or real usage of the system was presented and whether the relation of the researchers towards the system assessed was described (e.g. developer, user or independent).

Results

Table 2 gives an overview of the number of papers that were considered eligible for inclusion in this study.

While doing the detailed analysis of the selected papers on the basis of the full text, one paper has been considered as not belonging to the scope of STARE-HI. It described an evaluation of an algorithm, although that was not clear from the abstract. Hence this paper has been excluded from further analysis. Results are reported for the remaining 48 papers.

Table 2 - overview of number of papers selected for inclusion in this study

Journal	# of publ. Papers in 2005	Selection JT	Selection EA	Initial Agreement	# of papers after consensus
IJMI – regular	27	11	7	7	11
IJMI – special issues	74	13	9	8	11 (1 excluded later)

Journal	# of publ. Papers in 2005	Selection JT	Selection EA	Initial Agree-ment	# of papers after consen-sus
MIM – regular	37	4	0	0	3
MIM – Special issues	71	4	2	1	1
JAMIA	74	23	19	16	23
Total	282	55	37	32	49

In table 3 we give an overview on the average score for each item.

Table 3 - Average scores for the various items assessed

Item	Possible range	Average
Title	0 - 3	2.2
Abstract	0 - 8	5.0
Keywords	0 - 5	2.0
Organisational setting	0 - 2	1.6
System details	0 - 2	1.6
Lifecycle, usage, and relation between evaluators and system.	0 - 3	1.2

STARE-HI suggests to include also limitations in the abstract. None of the papers addresses this in the abstract. Also other aspects are not consistently addressed as can be seen from the average score.

One of the journals does not include keywords. Papers of that journal have not been taken into account in the analysis.

It was disappointing to see that phase in the lifecycle, the usage of the system and the relation between those involved in a study and the system studied are poorly reported. In the latter category only three papers have make the relations explicit. In the other papers, the relation remains unclear or has to be deduced by careful reading of the text.

Discussion

Scope of STARE-HI.

The scope of STARE-HI is defined as: […]to provide guidelines for the reporting of evaluations in Health Informatics (4). The object of evaluation is defined as the health informatics application. However, while selecting papers that could be considered as falling within the scope of STARE-HI we also noted a number of studies that dealt with perceptions of (potential) users of a particular health informatics application or even of health informatics applications in general. We have also considered those papers that deal with potential user perceptions of particular applications, but not of health informatics applications in general. A total of 60 papers were selected by at least one of us, resulting in an initial agreement on 53% of these selected papers. After a discussion by email we obtained agreement to include 49 papers (81%). Papers we decided to exclude dealt with applications in education, a presentation of the main results of various studies, the evaluation of the content of webpages, the evaluation of methods without the real application in practice.

Still, the scope of STARE-HI is not rigid. The excluded papers could equally well be assessed along the lines of STARE-HI but some aspects may not have been applicable.

In summary we have initially included papers, based on their title and abstract, that have assessed some aspect of a health informatics application in a real user environment.

Completeness of current reporting

Title

It is clear that titles are not yet fully descriptive for the paper. In particular the study design is often missing. Only 14 of the selected papers mention the study design in the title.

Abstract

STARE-HI promotes structured abstracts. In our corpus, 37 papers have a structured abstract and 11 have not. When we compare the scores for these two groups of papers, we have an average of 5.5 for papers with structured abstracts and 3.2 for papers with an unstructured abstract. It is clear that structured abstracts provide significantly more information about the study than unstructured abstracts ($p<0.01$). It has been proposed by others to include in a structured abstract also a section on the limitations of the study (7). It is clear that our community has not yet adopted such a strategy since no abstract addresses study limitations. Other aspects that are poorly reported are the participants (46%) and outcome measures (58%).

Keywords

Also in this area reporting is far from optimal. Excluding the papers in JAMIA, which do not have keywords, on average only less than 2 of the 5 types of keywords are listed. In particular keywords addressing study design are seldom listed (only in 2 cases). Also the term evaluation (or equivalents) is listed only 3 times.

Organisational setting

We made a subjective assessment whether the organisational setting in which the study took place was described in an appropriate way. STARE-HI proves a quite extensive description of what could be included here. It does not define "mandatory" items, but presents a list of possibilities that could be addressed. We observe an average score of 1.6 (from a maximum of 2), which indicates that there is room for improvement here. We also noted which aspects were considered to be missing in the description. In particular the duration of the usage of the system is often missing.

System details

The description of the system description was assessed in a similar way as the organisational setting. The objective was to assess whether the reader would have a proper view on the functionality of the system and how it was used. We left out an assessment of the system details like name, commercially obtained, self developed etc., assessment of such details is only useful for particular type of studies. In 8 papers there were no details on the system used. Only the type was mentioned.

Lifecycle, usage and relation information

These aspects were assessed because it will allow the reader to weight the relevancy of the paper of their situation. When you consider purchasing a particular type of system, it doesn't make sense to have a look at studies that have assessed the quality of prototype systems in an experimental situation. So lifecycle information as well as usage patterns are relevant. We also think that providing information about the relation of the authors and the object of study is relevant. It makes a difference whether the evaluators are the developers of the system or an independent party. Details on lifecycle and usage were reported in about 50% of the papers. Only in 3 papers there was explicit mentioning of the relation between the evaluators and the system.

Quality of reporting

Quality of reporting is difficult to assess. Often, quality is equated with completeness. Without completeness, it is difficult to make a quality statement. When we equate quality of reporting with ease of finding information in the papers, it is our experience that papers that follow the IMRAD (Introduction, methods, results and discussion) structure are easier to analyse that papers that don't follow this structure. For a proper assessment of the relevancy of studies in Health Informatics, contextual information is of relevance as well. That aspect is not properly covered in IMRAD, since it may appear in all 4 sections.

The feasibility of STARE-HI

This is the first attempt to apply STARE-HI as a means for assessing evaluation papers in Health Informatics.

We noted that it is difficult to define precisely the scope of STARE-HI. In our set of selected articles there are many that fall within the scope of STARE-HI. However, there are a few for which it is less clear that they fit within the scope of STARE-HI. In particular for papers that evaluate not a single application or system, but assess, for example, the impact of IT in general on workflow, certain items are not applicable. Still the global objective of the various sections may still be applicable.

We applied only a few sections of STARE-HI. In particular the sections on title, abstract and keywords are clear and provide good guidance. The section on study context defines quite some aspects that may be important to report. However, when those issues are not reported, it is often difficult to assess whether they could be relevant or not. Additional guidance on how to use STARE-HI for assessment of papers is needed. Development of a uniform scoring approach is necessary.

Feasibility of the approach

The items addressed in this pilot study can be applied to assess the quality of publications. We have selected a broad scope of manuscripts that not only included the evaluation of a specific type application in a restricted (e.g. hospital) setting. We also included more survey type of papers where for example a larger group of patients has been interviewed about their perceptions and use of the internet for searching for disease specific information. When we consider the general objective of each item of STARE-HI as the guiding principle for the assessment, STARE-HI is broadly applicable. It may be difficult, however, to make statements about the quality of reporting in general when a broad spectrum of papers is taken into account.

Limitations of study

This study focussed only on a few items of STARE-HI. Results of the application of all items of STARE-HI will be reported during the meeting.

This is only a pilot study, mainly to assess the feasibility of STARE-HI as an assessment tool and to have a first indication of the completeness of reporting of evaluation studies in Health Informatics.

We included papers from only three Medical Informatics Journals that appeared in one specific year. Although one may assume that the results presented are generalisable to papers in these journals over a longer period of time, generalisation towards all papers on evaluation in Health Informatics is not yet possible. The editorial process and journal guidelines play an important role. An extensive study including more Medical Informatics journals as well as general and specialised medical journals is needed. Such an extensive study will set a baseline value for comparative studies in the future after STARE-HI has been disseminated on a larger scale and possibly adopted by several (Medical Informatics) journals.

Conclusion

There is room for improvement in the reporting of evaluation studies in health informatics. Even the sections that are often used to assess the relevancy of papers (i.e., title, abstract and keywords) score far from optimal. The fact that the completeness of structured abstracts is much better than that of unstructured abstracts is an indication that STARE-HI as a guideline for reporting on evaluation studies can contribute to quality of papers in our field.

A study covering all papers on evaluation in Health Informatics in Medical Informatics journals as well as general and specialised medical journals is indicated to have a baseline measure on the completeness and quality of publications from our domain.

References

[1] Altman DG. Better reporting of randomised controlled trials: the CONSORT statement. Bmj 1996;313(7057):570-1.

[2] Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. Standards for Reporting of Diagnostic Accuracy. Clin Chem 2003;49(1):1-6.

[3] Ammenwerth E, Brender J, Nykanen P, Prokosch HU, Rigby M, Talmon J. Visions and strategies to improve evaluation of health information systems. Reflections and lessons based on the HIS-EVAL workshop in Innsbruck. Int J Med Inform 2004;73(6):479-91.

[4] Talmon JL, Ammenwerth E. STARE-HI: STAtement on the Reporting of Evaluations in Health Informatics. http://iig.umit.at/efmi/stare-hi/Stare-HIv0.12.doc [last accessed 1 Dec 2006]

[5] Kane RL, Wang J, Garrard J. Reporting in randomized clinical trials improved after adoption of the CONSORT statement. J Clin Epidemiol 2007;60(3):241-9.

[6] Moher D, Jones A, Lepage L. Use of the CONSORT statement and quality of reports of randomized trials: a comparative before-and-after evaluation. JAMA 2001;285(15):1992-5.

[7] Haynes RB, Mulrow CD, Huth EJ, Altman DG, Gardner MJ. More informative abstracts revisited. Ann Intern Med 1990;113(1):69-76.

Address for correspondence

Jan L. Talmon
Dept of Medical Informatics
Maastricht University
PO Box 616
6200MD Maastricht
The Netherlands
Phone: +31-43-3882243
Email: Talmon@mi.unimaas.nl

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Exploring the Unintended Consequences of Computerized Physician Order Entry

Joan S. Ash[a], Dean F. Sittig[a,b], Richard Dykstra[a,b], Emily Campbell[a], Kenneth Guappone[a]

[a] *Oregon Health & Science University, Portland, OR, USA*
[b] *Kaiser Permanente Northwest, Portland, OR, USA*

Abstract

This paper summarizes the foci, activities, methods, and results of a three-year research project. Using a mixed methods approach, the Physician Order Entry Team has identified and categorized the unintended consequences of computerized physician order entry (CPOE). After analyzing 380 examples of unintended adverse consequences, the team described in detail nine major types and conducted a national survey in the U.S. to discover how hospitals recognize and deal with unintended consequences. With the assistance of a panel of experts, the team identified strategies for managing unintended adverse consequences and outlined contents of a toolkit for CPOE implementers for addressing them.

Keywords:

attitude to computers; hospital information systems; user-computer interface; physician order entry

Introduction

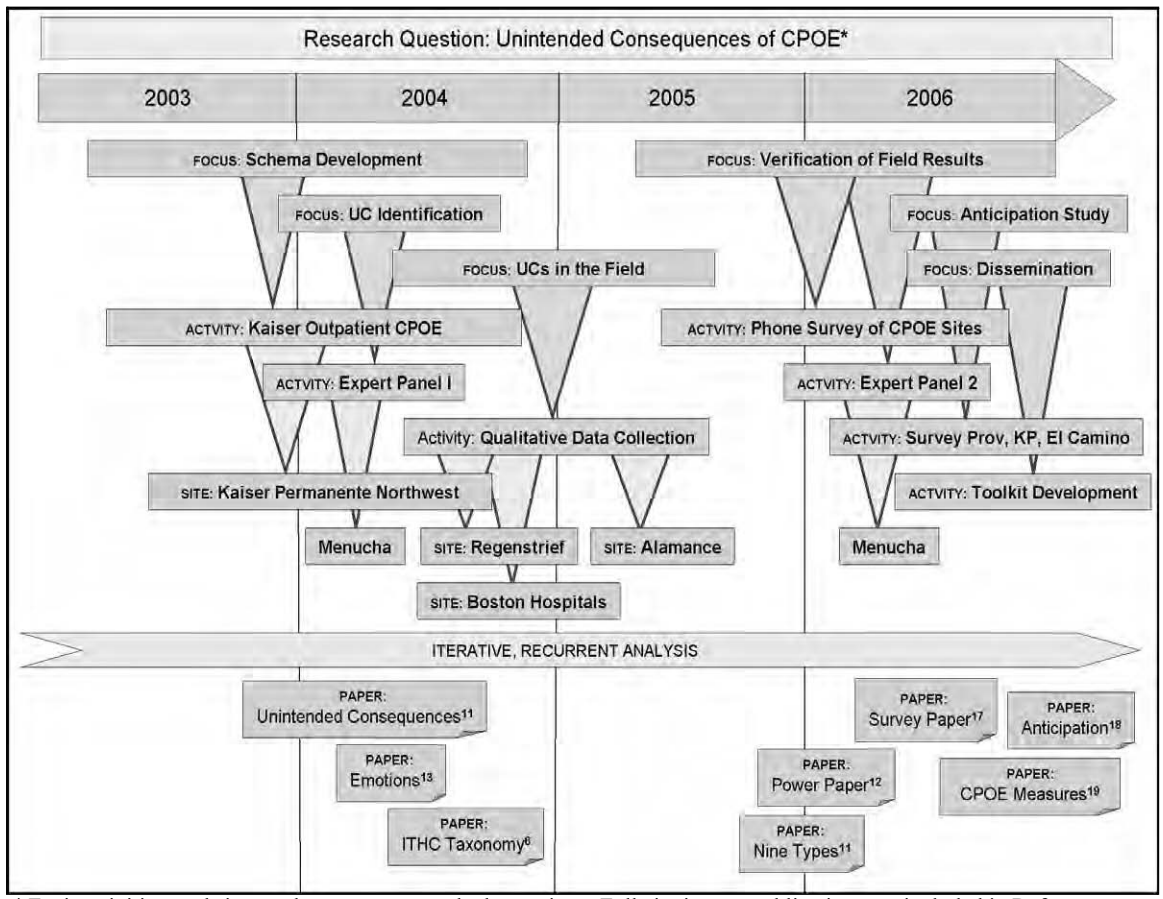
The Physician Order Entry Team (POET), a group of researchers based at Oregon Health & Science University in Portland, Oregon, U.S.A., was conducting a study of success factors for implementing computerized physician order entry (CPOE), defined as direct entry of orders into the computer by physicians or others with the same ordering privileges, when we began noticing unintended consequences (UCs), such as physicians entering orders for the wrong patient. Colleagues doing similar qualitative studies in Australia and The Netherlands were discovering these UCs as well, and a collaborative effort in 2002 produced a general description of kinds of adverse consequences caused by clinical information systems (CIS) [1]. This was a rather startling revelation at a time when CPOE was being touted as the "leap" that hospitals should take in the interest of patient safety [2] and little attention was being paid to problems caused by CPOE. With funding from the U.S. National Library of Medicine, POET has been able to conduct an in-depth study over the past three years utilizing both qualitative and quantitative methods to discover more about these UCs of CPOE. Data were gathered via two expert panel conferences, fieldwork at a total of six sites (one outpatient and five primarily inpatient), and a national telephone survey of all CPOE sites in the U.S. The aims were to identify types of UCs and strategies for preventing, managing or overcoming them, and to provide tools to help implementers address them. The following presents a summary of the research foci, methods, and results.

Methods

Sample selection

The main criterion for site selection for fieldwork was a reputation for excellence in using clinical information systems. Excellent organizations learn from their mistakes [3-4]. We found that personnel in these organizations are willing to 1) describe surprises they have experienced and managed, and 2) be observed during the order entry process. In addition, sites represented a geographic distribution, different types (e.g. teaching and community hospitals) and ownership, varying lengths of experience with CPOE, and both commercial and locally developed systems. Kaiser Permanente Northwest in Oregon, selected for excellence in outpatient CPOE, uses EpicCare (Epic Systems, Madison, WI), and is a health maintenance organization. Other sites included: Wishard Memorial Hospital, a county hospital in Indianapolis, IN using the locally developed Regenstrief system; The Brigham and Women's Hospital in Boston, MA, which uses a locally developed system; Massachusetts General Hospital, Boston, MA, which uses a newer version of the Brigham system; Faulkner Hospital, a community hospital in Boston, MA, that uses the commercial MediTech system (Westwood, MA); Alamance Regional Hospital in Burlington, NC, which uses the Eclipsys commercial system (Boca Raton, FL); El Camino Hospital in Mountain View, CA, which also uses Eclipsys; and Providence Portland, which utilizes McKesson's Horizon Expert Orders (San Francisco, CA). We received human subjects approval from each site and from the researchers' organizations. Within each site, informants for interviews were selected based on their knowledge of what had occurred during CPOE implementation and use and their representative roles (physician, nurse, pharmacist, implementer, champion, skeptic, etc.). We observed clinicians entering orders in all areas of the hospitals and clinics. Experts for the expert conferences were selected based on hands-on experience with CPOE implementation and included clinician implementers from a variety of hospital types and vendor organizations.

Selected for best paper award.



* Foci, activities, and sites as they were approached over time. Full citations to publications are included in References.

Figure 1 - PET Research Overview 2003 - 2006

Data collection

Figure 1 illustrates the progression of data gathering, analysis, and reporting of results that occurred between 2003 and 2006, starting with a transition period at Kaiser Permanente Northwest between the success factors study and the UC study. We gathered data at Kaiser on an ongoing basis between the spring of 2003 and winter of 2004, developing refined semi-structured interview and observational techniques during the transition to studying UCs. In the spring of 2004 we held a conference of invited experts at the Menucha retreat center near Portland, OR to gather stories about UCs from the experts and to gain guidance about questions to ask and what to look for in the field. We then spent three to four days at each site with four to six investigators. Observations were documented with extensive field notes. Taped oral history interviews probed the history of surprises that occurred over time at each site. Realizing that it is hard to know when consequences are indeed unintended, we developed a short survey instrument to assess what clinicians being impacted by CPOE expect, and administered it to groups at three sites. Once we had identified the major types of UCs, we developed, piloted, and administered a telephone survey instrument designed to elicit information about the nature of UCs from all U.S. hospitals reporting that they have CPOE. A second conference of experts was held in May of 2006 to help interpret our results and to plan dissemination strategies, including a set of tools that implementers could use for identifying, managing, and overcoming UCs in their hospitals.

Data analysis

We first reviewed data gathered at Kaiser [5] during the transitional period between projects to gain an overview of UCs and developed a general schema that included positive as well as negative UCs [6]. Data collected during the expert conference and at the remaining five field sites yielded 750 pages of field notes and transcripts. The POET team reached consensus on each description of what seemed to be a UC, approximately 80 from the Menucha conference and 300 from fieldwork. As we analyzed them in detail, we realized that the initial UC schema we had developed [6] was too broad and superficial and that we would need to develop a more sophisticated taxonomy. Using a card sort method [7] and grounded theory approach [8], we iteratively developed a taxonomy of nine major types of adverse UCs into which all 380 instances

fit, thus reaching saturation [9]. Using an axial coding approach [10], we then found subcategories within each of the nine. Finally, we held another Menucha conference of experts to verify that these results made sense to the experts and to plan dissemination of tools to help implementers address each kind of UC.

Results

The initial schema

The initial broad schema of consequences related to CPOE included categories of intended and unintended consequences, desirable and undesirable, direct and indirect, and "two sided" consequences that could be either desirable or undesirable depending on one's point of view [6]. This schema provided a valuable framework for fieldwork because it assured that we would not limit our foci to adverse consequences. While most informaticians are interested in undesirable consequences because they need careful management, it is also heartening to know that serendipitous, beneficial surprises occur as well.

The nine types of unintended adverse consequences related to CPOE

Using the card sort method, we developed a "short list" of nine categories based on the larger list. The categories were validated as we analyzed the full complement of UCs. We have published a paper summarizing overall results [11], as well as specific papers about changes in the power structure resulting from CPOE [12] and emotions related to CPOE [13]; a paper about impact on workflow is under review [14]. As we conducted our analysis, we discovered that a large number of consequences, over 20% of the total, emanated from issues with clinical decision support (CDS). Briefly, the categories, in descending order of frequency, are:

More / New Work Issues: Physicians find that CPOE adds to their workload by forcing them to enter required information, respond to alerts, deal with multiple passwords, and expend extra time.

Workflow Issues: Many UCs result from mismatches between the CIS and workflow and include workflow process issues, workflow and policy/procedure issues, workflow and human computer interaction issues, workflow and clinical personnel issues, and workflow and situation awareness issues.

Never Ending Demands: Because CPOE requires hardware technically advanced enough to support the clinical software, there is a continuous need for new hardware, more space in which to put this hardware, and more space on the screen to display information. In addition, maintenance of the knowledge base for decision support and training demands are ongoing.

Paper Persistence: It has long been hoped that CIS will reduce the amount of paper used to communicate and store information, but we found that this is not necessarily the case since it is useful as a temporary display interface.

Communication Issues: The CIS changes communication patterns among care providers and departments, creating an "illusion of communication," meaning that people think that just because the information went into the computer the right person will see it and act on it appropriately [15].

Emotions: As outlined in the paper by Sittig et al. [13], these systems cause intense emotions in users. Unfortunately, many of these emotions are negative and often result in reduced efficacy of system use, at least in the beginning.

New Kinds of Errors: As noted by Koppel et al. [16] and Ash et al. [1], CPOE tends to generate new kinds of errors such as juxtaposition errors, in which clinicians click on the adjacent patient name or medication from a list and inadvertently enter the wrong order.

Changes in the Power Structure: The presence of a system that enforces specific clinical practices through mandatory data entry fields changes the power structure of organizations. Often the power or autonomy of physicians is reduced, while the power of the nursing staff, information technology specialists, and administration is increased [12].

Overdependence on Technology: As hospitals become more dependent on these systems, system failures can wreak havoc when paper backup systems are not readily available.

The national survey

We had already determined that the UC categories list was both useful and easy to understand and use, so we operationalized most of the categories by asking questions about UCs in a national survey. Those surveyed were to answer these questions as yes or no, and if yes, to rate the importance of this issue from 1 (not very important) to 5 (very important) (no = 0). The questions are shown in Table 1. We surveyed the entire population of acute care hospitals listed in the 2004 HIMSS Analytics^{SM} Database as having reported that they have CPOE in place (N = 448). Since that database did not include U.S. Veterans Affairs hospitals, which we feel are important models of CPOE use, we also surveyed VA hospitals (N = 113). We conducted interviews with staff at 176 hospitals, discovering that a large number listed as having CPOE did not in fact have functioning CPOE systems. We also found that a number of hospitals have policies against doing surveys.

The survey results verified the existence of these UCs, and analysis of comments offered insight into the nature of the consequences. All types of consequences are indeed widespread. Our informants did not consider two of them, power shifts and new kinds of errors, as important as the others, however. We verified that there are positive as well as negative unintended consequences, and often the same consequence can be viewed in different ways by different people, depending on their perspectives. We can only speculate about why power shifts and new kinds of errors were not considered as important as other types: those answering the questions were generally information technology professionals who may not realize that power is

shifting in their direction. They may also believe that the new kinds of errors are not of a serious nature. A paper reporting results is under review [17].

The anticipation survey

To find out if the "unintended" or "unanticipated" consequences that we had identified were perhaps already known by others in the field and therefore actually anticipated during CPOE implementations, we designed and piloted an "anticipation survey." It was designed to determine what end user clinicians were expecting to happen in organizations that were about to implement CPOE. The questions, which were asked in person, are shown in Table 2. The survey was administered as a short interview survey to 83 clinicians at three community hospitals at common gathering spots such as the cafeteria. Results from each hospital were fed back to the implementers at those individual sites. The research team conducted a comparative analysis across sites. Briefly, end users and others affected by CPOE were usually aware that CPOE was coming and that their workflow might be disrupted for a certain period of time, but they were optimistic that in the long run it would be of benefit. Other UCs were rarely mentioned. Results are summarized in a paper under review [18].

Table 1 - Phone survey unintended consequences questions

Workflow (process)
Question 1: We have noticed in our research that when CPOE systems are in use, this alters how people do their work. Have you seen this, how important is it, and could you comment?
Communications
Question 2: Communication is really important in clinical care. Have you seen any alterations in communication patterns because of CPOE, how important are these alterations, and could you comment?
Over-dependence on technology
Question 3: As we become more dependent on technology, we've noticed that people may have a hard time when the CPOE system is not available. If your computer went down, would this be an issue for your organization, how important would it be, and could you comment?
Power
Question 4: We have noticed the balance of power may shift when CPOE is used. Have you noticed that at your organization, how important is it, and could you comment?
More work, new work
Question 5: We think of computers as labor saving devices, but we all know that sometimes they're not. Are there examples in your institution of new kinds of work that you didn't do before, how important is this, and could you comment?
New kinds of errors
Question 6: CPOE has been proposed as a solution to patient safety issues, but may have created others. Have you seen new patient safety issues with CPOE, how important are they, and could you comment?
Never ending demands of technology
Question 7: The information system typically needs a great deal of support in terms of maintenance, training, updating order sets, etc. Has this been an issue in your organization, how important is it, and could you comment?
Emotions Question 8: We have seen many emotional responses to the system. Have you seen users express strong feelings about CPOE, how important is this, and could you comment?

Table 2 - Short interview survey about anticipation of CPOE

About you:
What is your role in the organization? If clinician, continue.
• Have you heard about Computer-based Provider Order Entry being implemented here?
• If no…This is a new system that would allow the physicians to enter their patient orders directly into the computer system.
• Have you been trained on it, tested it, and/or actually used it?
• What effect do you think it will have on you?
• Do you have experience with the clinical information system now available?" How do you think the new CPOE system might compare to the current paper system? Advantages? Disadvantages?
About the organization:
• What effect do you think CPOE will have on other clinicians within the organization?
• Do you think it will be more positive or negative? In what way?
• What does this mean for patients?
• Do you think it will be more positive or negative? In what way?
• What effect will the system have on the hospital as a whole?
• Do you think it will be more positive or negative? In what way?

Affirmation and dissemination

The results of our fieldwork and the survey were disseminated to the experts prior to the Menucha conference held on May 4 and 5 of 2006. The experts affirmed our categories of UCs and recommended that we develop three tools for the UCs toolkit to be widely disseminated. These include a best practices guide for implementation to help avoid UCs, a checkup tool to help hospitals assess how well they are managing UCs, and a dashboard tool of recommended metrics that can help hospitals track their progress [19].

Discussion

In this team's earlier research on success factors for implementing CPOE, we cast a wide net because little was known at that time about factors leading to success or lack thereof. A rigorous yet open-ended grounded theory [8] approach was deemed most appropriate. As research questions became more focused, however, our qualitative methods became more structured. To investigate UCs of CPOE, we started with a broad schema of types and iteratively refined the schema by consulting experts, interviewing, and observing in the field. Once a taxonomy of types was defined, we were able to craft survey questions and administer a national telephone survey. Because such a large number of adverse UCs relate to CDS, our attention is becoming even more focused in that arena.

As the team and expert consultants develop tools for helping implementers avoid, manage, or overcome UCs, we have realized that all of the success factors identified and published in the past [20] are actually UC avoidance

mechanisms. In addition, there are other strategies for addressing UCs, and a more complete list is under development.

Conclusions and recommendations

While it is hard summarizing results of an intense three year study of UCs, we can draw some general conclusions about both methods and UCs. First, the selected methods served us well for this study. The more structured and rapid techniques such as the anticipation survey efficiently augmented other kinds of fieldwork. Second, development of a taxonomy of types and subtypes not only allowed us to craft survey questions, but was also useful in structuring an approach for addressing management of UCs. One key in prevention of UCs is to pay attention to success factors for implementing CPOE in the first place. Since so many UCs are related to CDS, it seems that a fruitful area for future research would be identification of success factors for implementation of CDS.

Acknowledgments

This work was supported by grant LM06942 and training grant ASMM10031 from the U.S. National Library of Medicine, National Institutes of Health.

References

[1] Ash JS, Berg M, Coiera E. Some unintended consequences of information technology in health care: The nature of patient care information system related errors. J Am Med Inform Assoc 2004; 11(2):104-112.

[2] The Leapfrog Group. Factsheet: Computer physician order entry. [accessed 11/21/06]; Available from: http://www.leapfroggroup.org/for_hospitals/leapfrog_safety_practices/cpoe

[3] Senge PM. The Fifth Discipline: The Art and Practice of the Learning Organization. New York: Currency Doubleday, 1990.

[4] Stavri PZ, Ash JS. Does failure breed success: Narrative analysis of stories about computerized physician order entry. Int J Med Inform 2003; 72:9-15.

[5] Ash JS, Chin HL, Sittig DF, Dykstra RH. Ambulatory computerized physician order entry implementation. AMIA Annu Symp Proc 2005:11-5.

[6] Ash JS, Sittig DF, Dykstra RH, Guappone K, Carpenter JD, Seshadri V. Categorizing the unintended sociotechnical consequences of computerized provider order entry. Int J Med Inform 2006 Jun 20; available online as preprint.

[7] Lincoln Y, Guba E. Natualistic Inquiry. Newbury Park, CA: Sage; 1985; pp. 346-51.

[8] Glaser BG, Strauss AL. The Discovery of Grounded Theory: Strategies for Qualitative Research. Chicago: Aldine, 1967.

[9] Denzin NK, Lincoln YS, eds.. Handbook of Qualitative Research, 2nd ed. Thousand Oaks CA: Sage, 2000; p. 520.

[10] Strauss A, Corbin J. Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory. Thousand Oaks CA: Sage; 1998.

[11] Campbell EM, Sittig DF, Ash JS, Guappone KP, Dykstra RH. Types of unintended consequences related to computerized provider order entry. J Am Med Inform Assoc 2006; 13(5):547-556.

[12] Ash JS, Sittig DF, Campbell E, Guappone K, Dykstra R. An unintended consequence of CPOE implementation: Shifts in power, control, and autonomy. AMIA Annu Symp Proc 2006: in press.

[13] Sittig DF, Krall M, Kaalaas-Sittig J, Ash JS. Emotional aspects of computer-based provider order entry: a qualitative study. J Am Med Inform Assoc 2005; 12(5):561-7.

[14] Campbell EM, Guappone KP, Sittig DF, Dykstra RH, Ash JS. The unintended adverse consequences of computerized provider order entry on clinical workflow. Submitted manuscript.

[15] Dykstra R. Computerized physician order entry and communication: Reciprocal impacts. AMIA Annu Symp Proc 2002:230-234.

[16] Koppel R, Metlay JP, Cohen A, Abaluck B, Localio AR, Kimmel SE, et al. Role of computerized physician order entry systems in facilitating medication errors. JAMA 2005; 293(10):1197-203.

[17] Ash JS, Sittig DF, Poon EG, Guappone K, Campbell E, Dykstra RH. The nature and extent of unintended consequences related to computerized physician order entry. Submitted manuscript.

[18] Sittig DF, Ash JS, Guappone K, Campbell E, Dykstra R. Assessing the anticipated consequences of computer-based provider order entry at three community hospitals using an expedited assessment procedure. Submitted manuscript.

[19] Sittig DF, Thomas SM, Campbell E, Kuperman GJ, Artreja A, Waitman LR, Nichol P, Leonard K, Stutman H, Ash JS. Consensus recommendations for basic monitoring and evaluation of in-patient computer-based provider order entry systems. Submitted manuscript.

[20] Ash JS, Stavri PZ, Kuperman GJ. A consensus statement on considerations for a successful CPOE implementation. J Am Med Inform Assoc 2003; 10(3):229-234.

Address for correspondence

Joan S. Ash Ph.D., Associate Professor, Oregon Health & Science University, School of Medicine, 3181 SW Sam Jackson Park Rd., Portland, OR, 97239-3098 USA. Email ash@ohsu.edu.

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Impact of Health Care Information Technology on Hospital Productivity Growth: a Survey in 17 Acute University Hospitals

Rodolphe Meyer[a,b], Patrice Degoulet[a], Louis Omnes[c]

[a]*INSERM- U729 and Hôpital Européen Georges Pompidou, Paris, France*
[b]*Centre de Traitement de l'Information Médicale des Armées, Saint-Mandé, France*
[c]*Assistance Publique Hôpitaux de Paris, Paris, France*

Abstract

The quantification of the added value of information technologies (IT) in the health sector is a major issue for decision-makers and health care professionals. This paper relates the application of an economic production function in hospitals with different integration levels of their clinical information systems (CIS).

The study concerns 17 university hospitals within the Assistance Publique Hôpitaux de Paris group that were followed from 1998 to 2005. Using an extended Cobb-Douglas production function, yearly incomes (Y) were correlated with three inputs: capital (K), labor (L) and IT expenses (T).

The results indicate that incomes are significantly and positively associated with the three input variables with elasticity coefficients: α, β, and γ of 0.81, 0.17, and 0.09 that appear to be in the range of values found in secondary and tertiary sectors. The IT elasticity coefficient (γ) is higher in the subgroup of 6 hospitals that integrate, or started to integrate, a complete CIS within the study period than in the 11 reference hospitals.

In a general production function, hospital costs appear to be positively connected to the level of IT expenses, capital and labor. Calculations in two subgroups of AP-HP hospitals divided according to the importance of the IT integration level seem to indicate that the more the clinical information system is integrated, the more its influence is positive in hospital production. The results of this first survey are sufficiently encouraging to try to refine them (better granularity) and to spread them in time (over a longer period) and space (to other hospital structures).

Keywords:

Information technologies, Production function, Cobb-Douglass, Productivity, Integration level.

Introduction and objectives

Since the 18th century, it has been common to link the growth and productivity of an enterprise to the elements, or production factors, used to generate products or services. The explanation of the economic growth by the combination of production factors led to the notion of production function, a mathematical relationship established between the production (output) and the factors put together to obtain it (inputs). A major step in this field was achieved by the American economist Paul Douglas and the mathematician Richard Cobb [1], who proposed a non-linear function linking yearn (Y), capital (K) and labor (L). The initial studies with this function undertaken in 1930 particularly concerned the industrial sector, and since then have spread to all economic sectors seeking efficiency. In 1956, Robert Solow [2] enhanced the function by introducing a new factor known as the Solow residual that is related to the technology level. In the 1970s, information technologies (IT) became tools capable of influencing capital and labor in a direct way. Since then, they have represented important investments and led to changes to workers and their organization; therefore, IT efficiency needed to be evaluated [3].

In the health care domain, the development and spread of the first hospital information systems (HIS) also began in the 1970s [4,5]. After a slow implementation and acceptance phase, their installation has accelerated due to political and economic encouragement. HIS had high strategic and economic importance, in particular after studies emphasized its possible role in improving the quality of care or reducing medical errors [6]. However, HIS constituted a new expenditure line that did not bring proof of its profitability, and this raised the problem of justifying the associated investments.

In addition, classic economic analysis models applied to hospitals must be adapted to integrate health care specificities. Economic studies in the medical information domain raise theoretical and practical issues such as the definition and measurement of hospital production, efficiency or productivity [7,8]. Yet, it is essential in the health domain to be able to estimate the impact of technology investments on hospital activities.

This study had two main goals:

1. To show that a classic econometric production function is adaptable to the health care world.
2. To compare the share of IT in the production function between two sets of hospitals split on an IT integration level basis.

Materials and methods

Hospitals

The study concerns 17 of the 38 Paris university hospitals within *Assistance Publique Hôpitaux de Paris* (AP-HP) (Table 1). These hospitals were selected according to their size (more than 350 beds) and activity (acute and short-term care). Data used in this study range from 1998 to 2005, and give a total of 136 complete annual observations. The 17 hospitals were split into two groups according to their IT integration level. Group 1 consisted of 11 hospitals having mainly administrative and ancillary department management systems (i.e., laboratory, radiology and pharmacy). Group 2 consisted of 6 hospitals that installed or had started to install an integrated clinical information system (CIS) during the study period. Hospital H12 began the installation of its CIS in 1998, H13 in 2000, and H14 to H17 throughout t 2003.

Table 1 - Sample of data from the 17 hospitals, 2005

Hospital	Nb. beds	Y(*)	K(*)	L(*)	T(*)
Group 1					
H 1	468	129.6	29.4	78.8	3.45
H 2	413	140.6	31.8	80.2	2.73
H 3	530	152.2	36.3	94.6	3.03
H 4	597	152.3	42.2	91.8	3.83
H 5	929	270.0	52.2	151.9	5.02
H 6	484	144.3	34.3	92.5	2.78
H 7	1008	214.7	40.4	157.5	5.68
H 8	602	273.4	51.5	162.5	5.28
H 9	1826	530.9	118.0	338.4	12.2
H 10	779	254.8	53.5	151.1	5.67
H 11	672	232.2	43.6	127.5	6.09
Group 2					
H 12	423	132.1	34.9	91.5	3.86
H 13	833	232.9	54.8	160.4	10.2
H 14	874	380.6	95.7	236.5	6.22
H 15	859	246.5	84.5	158.3	8.79
H 16	907	281.7	67.4	162.2	6.95
H 17	569	267.9	47.7	132.8	7.38
(*) Y = yearn, K = capital, L = labor, and T = IT in millions					

Production function

Using an aggregate Cobb-Douglas function (1) [9,10] the links between hospital production (Y) and three different inputs –capital stock (K), quantity of labor (L), information technologies (T) – and the Solow residual (A) were evaluated assuming the constant elasticity of substitution of the inputs (i.e., $\alpha+\beta+\gamma=1$) [11].

$$Y = A.K^{\alpha}.L^{\beta}.T^{\gamma} \quad (1)$$

Knowing the value of the output (Y) and of the inputs (K, L and T), the value of the α, β and γ elasticity factors were calculated to estimate the importance of each input in the explanation of the observed output.

Output and inputs

For each hospital, the value of the production **Y** was assimilated to the yearly incomes of the hospital in millions of Euros.

K regroups the financial assets owned by the structure during the study period. The capital at year *t* is defined as the accumulated assets and real estate investments of the previous years, modulated by a depreciation factor depending on the nature of the investment. This depreciation varies between 3 to 10 years depending on the French Government's guidelines in use [12]. Capital stock is defined as the capital accumulated by the hospitals from past and current investments, adjusted with depreciation. Capital stock represents the capability of hospitals' productive assets, whereas annual investments only reflect assets acquired during a particular year. Hence, productive capital stock in a hospital is measured by the following depreciation formula [13]:

$$K_t = (K_{t-1} + NK_t) - D_t \quad (2)$$

The assets of the studied year (K_t) are equal to the value of the assets of the previous year (K_{t-1}) added to the value of the new investments (NK_t) decreased by a factor of calculated markdown (D_t) [14]. In this study, the K variable excludes IT investments.

The annual measurement of labor input **L** in Euros was delivered by hospital's budget records and it involved medical as well as non-medical hospital employees during the past years excluding IT workers.

T is the aggregation of IT-specific expenses including salaries, hardware and software expenses with their depreciation (i.e., T = IT capital + IT labor) [14–20].

Statistical analysis

Data collected in our survey has been retrieved from the financial systems of the 17 hospitals and were computed via scripts using the Eviews® 5 econometric software. The results for the R2 coefficient and Durbin-Watson's statistics were given by Eviews®. These statistics have then been cross-checked with Statistica® 7.

Results

Output and inputs

The values of Y, K, L and T increased constantly since 1998 (Table 2). Between 1998 and 2005, Y increased by 47.1%, K by 69%, L by 28.4% and T by 37%.

IT expenses represented 2.8% (5.8 millions ¤) of the global expenses (205.3 millions ¤) in 2005.

Table 2 - Average output and inputs values (17 hospitals 1998–2005)

Year	Y(*)	K	L	T
1998	161.4	32.1	113.1	4.3
1999	166.4	37.2	116.1	4.6
2000	172.4	42.4	121.0	4.9
2001	182.7	48.8	125.0	5.3
2002	197.3	51.2	131.4	5.3
2003	201.8	55.2	138.1	5.6
2004	226.2	56.7	140.6	6.2
2005	237.5	54.3	145.2	5.8
(*) Y=yearn, K=capital, L=labor, T=IT (millions ¤)				

Share of Input Factors

For the 17 hospitals, the share of capital, labor and IT were respectively 0.168, 0.766 and 0.085 (Table 3). All coefficients were statistically significant ($p<0.05$). The Solow residual is 0.81. Durbin-Watson's statistic is between 1.51 and 2. The value of the R2 is highly significant (0.97, $p<0.0001$).

Table 3 – Share of factors (17 hospitals - 2005)

	Value	Std. Error	t-Statistic	p
A	0.8146	0.0430	18.935	0.0300
	0.1686	0.0509	3.3086	< 0.0001
	0.7657	0.0555	13.791	0.0216
	0.0853	0.0361	2.3643	0.0094
R2 : 0.9701			**Durbin-Watson** : 1.5772	
A = Solow residual = share of capital = share of labor = share of IT				

The value for IT is in the range of results [0.03 to 0.16] from other studies performed in secondary and tertiary economical sectors (industry and services) [14, 16–20] (Table 4).

Table 4 - Share of factors in industry and service studies

Studies [14,16-20]				R2
Hitt – 1994 [16]	0.2280	0.6860	0.0307	0.9510
Hitt – 1999 [17]	0.1300	0.7300	0.1100	0.9400
Lin – 2000 [18]	0.1240	0.7890	0.1600	0.9750
Shao – 2001[14]	0.2121	0.7040	0.0619	-
Bresnahan – 2002 [19]	0.1380	0.7530	0.0347	0.9080
Osei-Brison – 2004 [20]	0.2120	0.6630	0.0883	0.9700
= share of capital = share of labor = share of IT				

CIS integration level influence on elasticity coefficients

Table 5 compares the inputs' (K, L, and T) elasticity coefficients values in the two groups of hospitals. All elasticity coefficients remain statistically significant, and is higher in group 2 (integrated CIS) than in group 1 (0.103 vs 0.072).

Figure 1 shows that in both groups, the IT coefficients () increase from 1998 to 2005, and there is a higher slope from 2003. The difference between group 1 and group 2 has also increased since 2003.

Table 5 – Share of factors for the two groups (2005)

Group 1 (n=11) – no CIS				
	Value	Std. Error	t-Statistic	p
A	0.5985	0.0747	8.0052	0.0052
	0.1703	0.0501	3.4019	0.0010
	0.7853	0.0559	14.029	< 0.0001
	0.0723	0.0292	2.4759	0.0082
R2 : 0.9698			**Durbin-Watson** : 1.7361	
Group 2 (n=6) – integrated CIS				
	Value	**Std. Error**	**t-Statistic**	**p**
A	1.1891	0.0634	18.762	0.0057
	0.1619	0.0763	2.1218	0.0005
	0.7441	0.0837	8.8837	< 0.0001
	0.1033	0.0533	1.9369	0.0083
R2 : 0.9174			**Durbin-Watson** : 1.6822	
A = Solow residual = share of capital = share of labor = share of IT				

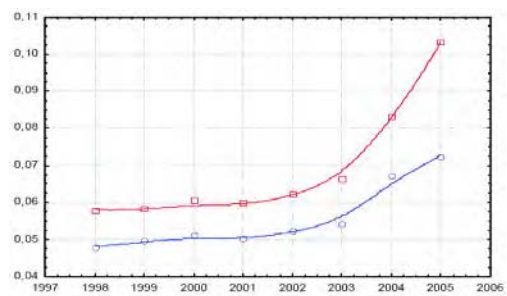


Figure 1 - Share of IT - Group 1 (circles) vs Group 2 (squares)

Discussion and conclusion

The economic approach of capital efficiency is generally represented by the informal ratio of production divided by capital expenditure [21]. The larger the ratio, the better the capital efficiency is, and thus leads to greater output. Long-term economic models, and especially growth accounting analysis [22], use the capital stock per worker ratio to study the impact of capital expenditure and capital stock increase on long-term output growth (e.g., the Robert

Solow model completed by different endogenous growth theories [23, 24]).

The common use of production functions described by Cobb and Douglas began in the 1930s, but it's only since the mid-1990s that researchers used its mathematical properties to isolate the share of IT in the production of various factories and business enterprises [15–20].

In the health care domain, numerous studies have measured the indirect earnings (quality and continuity of care, satisfaction of the users, etc.). However, very few studies have attempted to measure the direct earnings deriving from the integration of the different HIS components. The contradictory results of these studies frequently drive hospital managers to make decisions only on expected indirect benefits.

This study started with two goals: to try to use a classic econometric function on French hospitals data to isolate the share of IT in the hospital production, and to compare the share of factors between two sets of hospitals split on an IT integration level basis.

The first results show that using a function traditionally employed by economists in their analyses of the business world is possible within the health care world. In addition, the elements brought by the Cobb-Douglas production function stress the importance of the human factor in explaining the hospitals' production results. With a coefficient of 0.76, share of labor in the hospital is slightly superior to what is generally admitted in the industrial or service sector in France (0.70) [25].

If we compare the two groups, our results tend to show that the share of IT explaining the production observed is about 1.7 times more important in structures having a higher level of CIS integration than in the low-level ones. It is also important to notice that the share of IT explaining the production is measured between 0.07 (group 1) and 0.10 (group 2) and that the share of capital is measured between 0.17 (group 1) and 0.16 (group 2). When we know that IT investments represent only 11% of capital investment in the two groups, this may be a sign of the importance of IT in hospital production. It also shows that, in the end, a high level of IT integration is not necessarily more expensive than a low-level one.

The introduction of the IT input (**T)** in the production function takes into account the place of high technology in the hospital and its importance in modern health care activity. Numerous studies [18–24] emphasized their T input to smooth the differences observed with the global assets. They multiplied the IT labor value by a determined factor ranging from two to seven according to complex depreciation rules. Using the same approach in our work returned less pertinent results And therefore we didn't adopt it. The creation method of the T variable could be changed by extending it to workers and materials also closely related to the HIS-like image producing systems (i.e., scanner) and laboratory automatons that could be considered as acquisition peripherals. We limited ourselves strictly to the HIS officially assigned material and workers, having in mind that the border separating pure computer science from the techniques involving its presence is growing thinner.

A postulate known as the productivity paradox describes IT more or less like a victory that costs more than it achieves. Several authors [26, 27] analyzed this paradox in various industrial sectors to attempt to decide on its reality. The results are still contradictory. However our data seem to confirm the positive impact of IT on the studied hospital production during an eight-year period of follow-up. This global analysis needs to be completed by an individual analysis of hospital productivity before and after the introduction of their CIS since the different HIS implementation starting years in our study might have hidden some production that could appear just after the introduction of a CIS.

Since we showed that the initial production function could be enhanced using its mathematical properties, we can suppose that other constructions could be imagined that focus more specifically on one point or another (like the share of non-medical labor versus the share of the medical labor).

Information technologies are a composite mix of hardware software, knowledge, integration level, operational support and infrastructures. Our study showed us that IT investments are bringing value and contribute in positive way to hospital production. The separation of the hospitals into groups based on various other criteria (such as activity or volume) could sharpen the analysis of IT impact on hospital activity and production. Extension of this work to other hospital structures on a longer period of follow-up could also prove interesting.

On a methodological level, the econometric field is a vast area possessing a large quantity of mathematical functions that could prove useful to try on our data. These functions are either instantiations of the Cobb-Douglass function or new approaches born from mathematic research used by economists worldwide (Box-Cox, Box Tidwell, regression spline or stochastic borders analysis for example) [28, 29]. Considering the colossal increase in IT expenses and the size of IT costs in hospitals, this kind of analysis could permit hospital managers to obtain tools to aid decisions and enable the enhancement of IT governance in the health care business [30-33].

Acknowledgments

The authors would like to thank all the AP-HP staff who participated to the collection of this survey data, particularly J.P. Lefort, G. Pichaul, M.P. Ferec, D. Heudes and V. Woszto.

References

[1] Cobb C.W, Douglas P.H. A Theory of Production. *American Economic Review*, 1928 ; 18 (1) : pp. 139-65.

[2] Solow RM. A Contribution to the Theory of Economic Growth. *The Quarterly Journal of Economics*. 1956; LXX, pp.65-94.

[3] Pugh M.D, *Dashboards and scorecards : tools for creating alignment*, in The healthcare quality book, Health Administration Press Chicago, 2004 ; pp. 213-40.

[4] Van de Velde R, Degoulet P. *Clinical informations systems, a component based approach*, Spinger-Verlag New-York Inc, 2001 ; pp. 1-14.
[5] Tan J. *E-Health Care information systems*, Jossey-Bass San Francisco, 2005 ; pp. 14-16.
[6] Kohn LT, Corrigan JM, Donaldson MS, eds. *To Err Is Human: Building a Safer HealthSystem.*Institute of Medicine, Washington DC: National Academy Press; 1999.
[7] Koc C. The productivity of health care and health production functions. *Health Econ*. 2004; 13(8): 739-47.
[8] Hollingsworth B, Dawson PJ, Maniadakis N. Efficiency measurement of health care: a review of non-parametric methods and applications. *Health Care Manag Sci*. 1999; 2(3): 161-72
[9] Greene W. *Économétrie*, Pearson éducation France, 2005 ; pp. 7-10,19-38,484-487.
[10] Stadelmann D. *Fonction de production – illustration des propriétés pour l'analyse économique.* Séminaire de l'université de Fribourg, 2003-2004.
[11] Solow RM. A contribution to the theory of economic growth. *The Quarterly Journal of Economics*. 1956 ; LXX, pp. 65-94.
[12] Guide de la comptabilité analytique, AP-HP - direction économique et financi_re - département contrôle de gestion - service analyse de gestion. Version 3 du 1er mars 2006.
[13] Gapenski C. *Healthcare Finance : an introduction to accounting and financial management*, Health Administration Press Chicago, 2005 ; pp. 95.
[14] Shao B, Lin W. Measuring the value of information technology in technical efficiency with stochastic production frontiers, *Information and Software Technology*, 2001 ; 43 : pp. 447-456.
[15] Menon N, Lee B, Eldenburg L. Productivity of information systems in the healthcare industry, *Information Systems Research,* 2000 ; 11 (1) : pp. 83-92.
[16] Hitt L, Brynjolfsson E. *Creating value and destroying profit ? Three measures of information technology's contributions*. MIT Sloan school papers, december 1994.
[17] Hitt LM, Snir EM, *The role of information technology in modern production: complement or substitute to other inputs?* University of pennsylvania papers, January 1999.
[18] Lin T, Shao B. Relative sizes of information technology investments and productive efficiency: Their linkage and empirical evidence. Journal of the Association for Information Systems, 2000; 1(7): pp. 1-35.
[19] Bresnahan TF, Brynjolfsson E, Hitt LM. Information technology, Workplace Organization, and the demand for skilled labor: Firm-level evidence. Quarterly Journal of Economics, *The Quarterly Journal of Economics*, 2002; 117(1): pp. 339-376.
[20] Osei-Brison KM, Ko M. Exploring the relationship between information technology investments and firm performance using regression splines analysis. *Information & management*, 2004 ; 42 : pp. 1-13.
[21] Brinkerhoff RO, Dressler DE. *Productivity measurement: a guide for managers and evaluators*, Newbury Park (Calif.): Sage. 1990; pp. 16.
[22] Baier SL, Dwyer GP, Tamura R. *How Important Are Capital and Total Factor Productivity for Economic Growth?* Federal Reserve Bank of Atlanta Working Paper No. 2002-02, Mar 2005.
[23] Barro RJ. Notes on Growth Accounting. *National Bureau of Economic Research Working Paper No.W6654*. Jul 1998.
[24] Romer PM. The Origins of Endogenous Growth. *Journal of Economic Perspectives*. 1994; 8(1): 3-22.
[25] Bernier B, Simon Y. Initiation à la macroéconomie. Dunod : Paris, 2001 ; pp. 507-47.
[26] Brynjolfsson E. *The productivity paradox of information technology : Review and assessment*. MIT Sloan school papers, december 1991.
[27] Brynjolfsson E. The Productivity Paradox of Information Technology, *Communication of ACM*, 1993 ; 36 : pp. 67-77.
[28] Osei-Brison KM, Ko M. The productivity impact of information technology in the healthcare industry: an empirical study using a regression spline-based approach. *Information & Softwar*e *Technology*, 2004 ; 46 : pp. 65-73.
[29] Lin T, Shao B. The business value of information technology and inputs substitution: the productivity paradox revisited. *Decision Support Systems*, article in Press, 2005.
[30] Georgel F. *IT gouvernance: maîtrise d'un syst_me d'information*. Dunod, Paris, 2005; pp. 95-105.
[31] Fahy M, Roche J, Weiner A. *Beyond governance: creating corporate value through performance, conformance and responsibility.* Willey & Sons ltd, Chichester England, 2004; pp. 131-61.
[32] Bleom J, VanDoorn M, Mittal P. *Making IT governance work in a Sarbanes-Oxley world.* Willey & Sons Inc, Oboken, USA, 2006; pp. 46-90, 153-157.
[33] Frist B, Clinton H. How to Heal Health Care. *Washington Post.* Wednesday, August 25, 2004; Page A17.
[34] http://www.washingtonpost.com/wp-dyn/articles/A30277-2004Aug24.html]

Address for Correspondence

Dr Rodolphe Meyer, MD, MSc
Centre de Traitement de l'Information Médicale des Armées.
BP 130 – 00481 Armées
rodolphe.meyer@santarm.fr

Pr Patrice Degoulet, MD, PhD
Département d'Informatique Hospitali_re - HEGP
20, rue Leblanc - 75015 Paris
patrice.degoulet@egp.aphp.fr

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Risk Management and Measuring Productivity with POAS - Point of Act System

Masanori Akiyama [ab], Tatsuya Kondo[b]

[a] *Sloan School of Management, Massachusetts Institute of Technology, Cambridge, MA, USA*
[b] *International Medical Center of Japan, Tokyo, Japan*

Abstract

The concept of our system is not only to manage material flows, but also to provide an integrated management resource, a means of correcting errors in medical treatment, and applications to EBM through the data mining of medical records. Prior to the development of this system, electronic processing systems in hospitals did a poor job of accurately grasping medical practice and medical material flows. With POAS (Point of Act System), hospital managers can solve the so-called, "man, money, material, and information" issues inherent in the costs of healthcare. The POAS system synchronizes with each department system, from finance and accounting, to pharmacy, to imaging, and allows information exchange. We can manage Man, Material, Money and Information completely by this system. Our analysis has shown that this system has a remarkable investment effect – saving over four million dollars per year – through cost savings in logistics and business process efficiencies. In addition, the quality of care has been improved dramatically while error rates have been reduced – nearly to zero in some cases.

Keywords:

POAS (point of act system), hospital management, ERP (enterprise resource Planning), financial management, risk management

Introduction

There has been a tendency in medical care to give low priority to management and the improvement of efficiency; medicine has been regarded as a sacred area exempt from such moves. However, in September 2001 the Ministry of Health, Labor and Welfare made public a draft plan of medical system reform because of the need to seriously review the country's medical services. This was brought about both by the harsh economic conditions existing after the collapse of the asset-inflated bubble economy in the early 1990s and the aging of society accompanied by declines in the birthrate. The plan, which not only visualizes reform of the medical insurance system but also pictures an ideal system of medical care for the future, is a comprehensive draft for institutional reform. In concrete terms, the plan calls on medical professionals to respect their patients' points of view and allow patients to take responsibility for decisions regarding their own care, to improve the environment within which information is supplied, to provide high-quality, efficient medical care, improve the quality of medical service, regional medical care security and to introduce the use of information systems in providing medical services. The point of these suggestions is to foster respect for the options chosen by patients, to provide the information necessary for informed decision making, to establish a system that provides high quality, efficient medical service and to build a foundation for public confidence. Because of these proposals, economic efficiency in medical care is becoming an important public issue. In this context, information technology (IT) can serve as a helpful tool. When the improvement of efficiency is stressed, the quality of medical care may tend to be sacrificed.We have developed a system that, utilizing IT, can accurately calculate costs, in a bid to maintaining a balance between efficiency and quality. At the same time, the system can also be used as a yardstick for the measurement and improvement of efficiency.

Materials and methods

Points that need to be addressed

The traditional hospital information system (HIS), built by connecting order entries and the medical clerical system, takes in information about orders and outputs medical payment requests via a medical accounting system, which is actually a payment system. However, this kind of system has the following problems: 1) Although physicians are supposed to enter correct payment information, the information is often incomplete (occurrence of uncollected balance). 2) The data terminals within divisions and those at the HIS are not integrated. As a result, duplicate entries are required, resulting in unnecessary extra work. 3) While data held in the HIS can be sent to the medical financial system, divisional data necessary for payment cannot be entered due to inconsistencies in the master system. 4) It is difficult or impossible to search the information held by the medical financial or divisional systems via the order systems. 5) A most important problem is that the existing systems have been used primarily for preparing medical payment requests. As a result, data on clinical activities, which have nothing to do with medical insurance, are not received (and could not be handled anyway) by the existing systems.

In these circumstances, when certain expenses are not covered by medical insurance, it has not been possible to make accurate assessments of expenses for materials and person-

nel through cost calculations based on the data held in the medical financial systems.

Outline of the system

To deal with these problems, we have designed a three-tier model. The middle-tier application server is located at the center. Each divisional system manages data that has resulted from that division and its clinical work processes[1,2]. Each division manages and preserves detailed data, including its reports, and provides only the "outlines" of the data to the application server. Thus, the actual data are not sent to or preserved in the application server. Since only outlines of data are held in the central application server, the volume of data stored there will not increase dramatically. Each client communicates with the others via the application server, and a graphic user interface (GUI) is provided for each occupational category.

Calculating medical care costs

Calculating medical care costs, which had posed difficulties that needed to be resolved, has now become possible. POAS, which stands for the Point of Act System, is a design feature of this comprehensive medical information system. Its characteristics are as follows. 1) Information on all medical activities is collected as detailed data at major "action" points, from the time orders are issued on through to their implementation. 2) The system is organically linked to various medical devices, such as medical diagnostic instruments, X-Ray equipment and equipment in the pharmaceutical division. It records information about medical activities, and their results, in a general-purpose database in various forms such as images, numerical values and text. 3) It uses a general-purpose data description method that enables flexible incorporation in response to advances in IT technologies. 4) It has a data warehouse structure, which collects and permits the analysis of detailed data at the level of individual medical activities. 5) It helps prevent medical errors – including mistakes at the stage of implementation – by making it possible to cross-check such data as patient identification, ongoing medical activities, medicines to be used and what personnel carry out the medical activities, each time an activity is executed. 6) It can be used to calculate profits and costs, based on orders. It will total them by medical fees, sectors or patients. These figures can be utilized as management information.

Mechanisms for data collections and structure of order item

Data on medical activities at the points of action listed below can be collected centrally by direct connections to the order systems and the medical equipment in each division. Order is input, received, changed or cancelled, implemented (contact is made with the accounting section), and completed. Necessary units of data recorded by the system, based on the idea of 6Ws and 2Hs , are as follows: Who-the implementer (the person who placed an order, or the person who carried it out); to Whom - the patient; How - medical activities and changes in them; What - materials used (pharmaceuticals, medical materials and others); How Much - amount of materials used and the number of applications; For What - name of the disease subject to these medical activities; When - date when the order was placed, when it was implemented, and when it was discontinued; and Where - place of implementation (department, hospital ward, and equipment used). We have made it possible to calculate the costs related to each type of disease by entering the name of the disease along with each order.

Results

Operational track records

The underlying concept of this system is POAS, which enables records of "who did what to whom, where, when, using what, and for what reason" [1]. In short, real-time input becomes possible at the point of action. Logs, including inventories, are created. It becomes possible to reduce to a minimum the difference between expenses from medical activities and the amount claimed as lost by adopting the "accrued basis of corporate accounting" concept. In short, the management of divisions and their work, using a corporate financial/accounting system, has become possible by identifying the divisions that are incurring losses. The system operates continuously at the International Medical Center of Japan, handling 100 transactions per second, or more than 360,000 transactions per hour. It has been in continuous operation for four years.

Linking the hospital information system (POAS) with the management information system

The hospital information system concerning diagnosis and treatment (POAS) and the management information system, centered on accounting, are separate systems. Data collected as described above are compiled at midnight each day in the clinical database and then sent to the management information system. It calculates all costs in the early morning, using batch processing. As a result, management information from the previous day is available by 6:00 a.m.

Positive management analysis

1. The use of POAS makes possible management analyses based on objective data. The following kinds of analyses can be performed.
2. Profit-and-loss calculations for medical treatment departments/divisions
3. Profit-and-loss calculation by patient category
4. Calculating cost by disease
5. Profit-and-loss calculation by physician

Risk management

Characteristics

The difference between POAS and conventional systems is that POAS is not based on orders but on actions. Essentially, traditional systems were expanded versions of the medical accounting systems that were brought into nurse stations and outpatient departments. This means they were only capable of processing orders by day. As a result, these systems can cause time lags of anywhere between 10 minutes and up to several hours, posing a major problem for the medical workplace. To shorten the time-lag to meet the requirements of medical workers at the patient's bedside - about 2 to 3 seconds- data granularity must be based on

single vials. It is important to recognize at the outset that the Medical Affairs Section and the sections responsible for executing actual medical actions require different data granularities. If a system's granularity were to be based on single items to begin with, its data could be easily compiled to derive the data required by the Medical Affairs Section as well. This is the reason why conventional systems have not been useful for improving productivity, gathering clinical data or improving management efficiency. While manufacturers conduct production control on their drugs and medical supplies by single types, by the time these products reach the hospital through the wholesaler, they are batched together into units of boxes or purchase orders. As a result, conventional material flow systems process these items by the shipping slip and not by single types. Even if these products were checked by shipping slip or per day, once an accident occurs, it would be too late to prepare electronic medical charts. To prevent accidents, these products must be controlled as single items from the outset. When the shipment is received, POAS controls these items as single types, not by shipping slips. This helps prevent accidents since it allows hospital operators to implement the same level of quality control as the manufacturers.

Effects

According to a survey of injection prescriptions previously conducted at the International Medical Center, changes were ordered for 20% of these prescriptions at one time or another. Since then, the average hospital stay has been halved to 15 days, and we used POAS to calculate the rate of injection instruction changes for a one-year period ending October 2004.

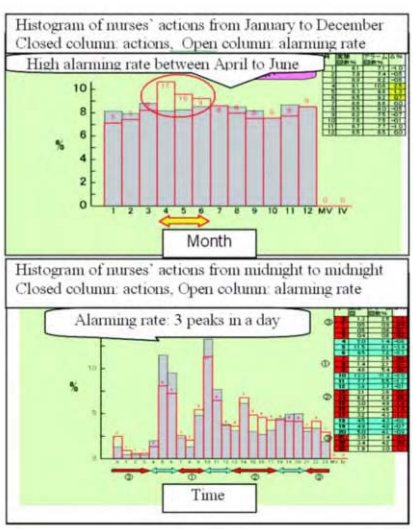


Figure 2 - Basic analysis: the frequency distribution of variables

Changes were ordered for 24% of the ﾒ injection prescription ﾄ between the time they were issued and the ﾒinjection mixingﾒ step. In addition, changes were ordered for 15% of the instructions after "injection mixing" (Fig.1).This shows that changes were ordered for a total of about 40% of the instructions. These changes should have doubled the amount of work for nurses and pharmacists, but their actual workloads did not increase. There was a reduction in nurse overtime and the number of accidents fell to zero. Similar improvements were seen at the Morioka Red Cross Hospital after they began using POAS. This was because automation eliminated tasks such as filling out and transferring slips, which previously took up most of the nurses' time.

Figure 1 - The effects of making injection action entries (calculated from performance data)

The figure 2 shows a map of nurses' actions from midnight to midnight. POAS also records nursing and care procedures. The logged records are 400 thousands / month, then about 80 million logs and 18 million records accumulated over two years. We can see that a variety of workloads are concentrated in the 9:00 AM to noon timeframe. This is because the morning shift starts at 8:30 for types of work. Most of the important medical actions are carried out before noon and 40% of the prescription instruction changes also procedures during this time. This is a hazard-prone timeframe that produces the most accidents and incurs the most wasted.

Injection accidents are most likely to result in personal injury. Therefore, here we analyze the causes of injection alarms. The horizontal axis shows the total number of injections performed and the vertical axis shows the alarm rates (Fig.3). Each point corresponds to a single sample with duration of 30 minutes each. The values are for the entire hospital for a period of one year. It shows that alarm rates were lower during time segments in which a large number of injections were performed and were higher during time segments in which fewer injections were performed. This indicates that accidents do not necessarily occur because nurses are busy.

Fig.4 shows the number of errors and error rates in 30-minute increments. You can see that the alarm rate increases immediately after a shift change. Additionally, you can see that the execution of instructions that were specified for 6:00 a.m. were scattered over several hours between 4:30 and 7:30. Conventional electronic medical chart systems will show these as being administered at 6:00 a.m. and there will be no way of getting a picture of the actual situation.

With the POAS system, users can not only track this information, but also analyze the effectiveness with pharmacokinetics and blood kinetics, as well as efficacy for different administration times.

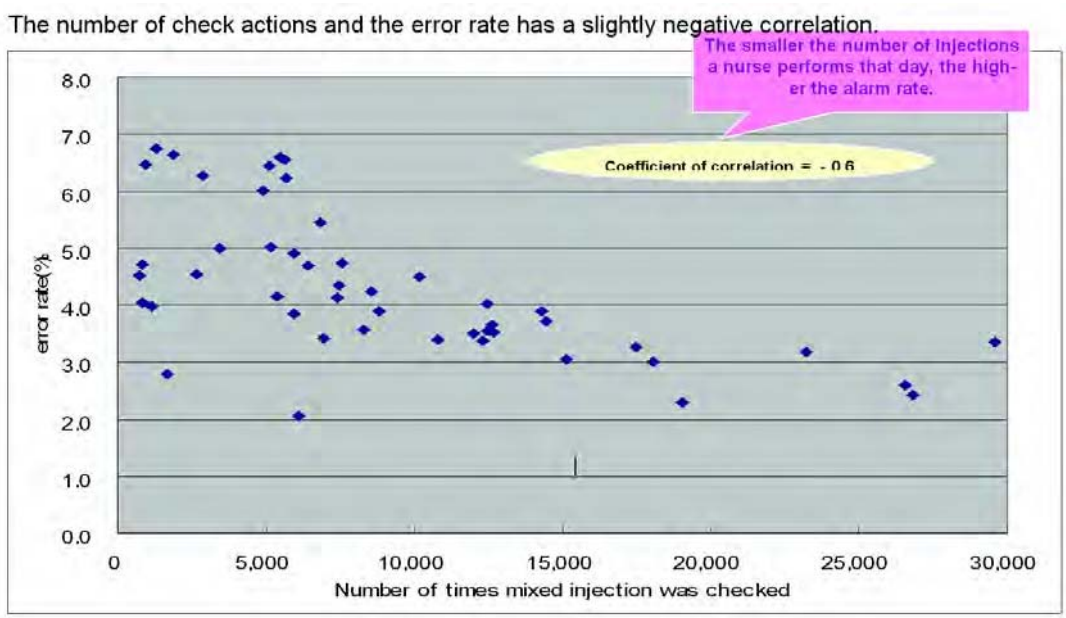
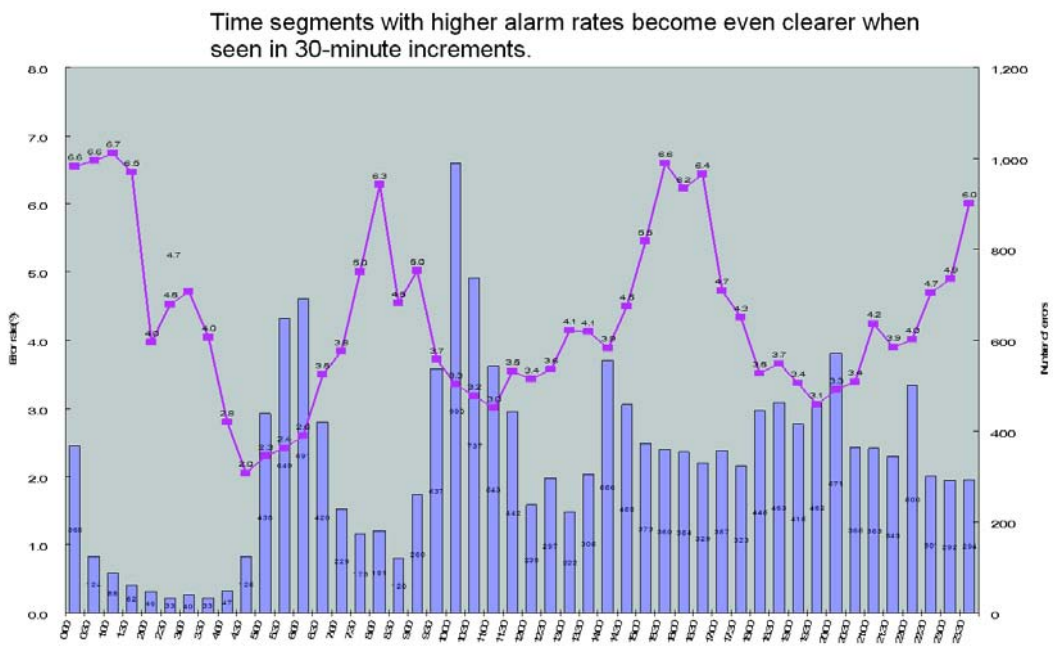


Figure 3 - Comparison of the number of times mixed injections were checked and error rate (%)



The height of the curve becomes progressively lower for the second, third and fourth injections and the dispersion become more evident. This is likely due to the fact that the first injection is started at around 6:00 a.m. and the second and later injections are IV replacements performed in response to nurse calls. The important point here is that the timeframe before 10:00 a.m. is potentially extremely hazardous. This is the timeframe during which powerful drugs are used the most often, with a corresponding spike in the number of incident and accident reports.

Discussion

Conventional hospital information systems are typically linked to other divisional information systems through the order entry system vendor [3]. However, such centralized

systems do not necessarily conform to the actual state of affairs within which the front line of medical care functions, led by the separate divisions. Linkage between divisional systems has become easier, thanks to recent distributed object technologies that have made it possible to design an order entry system centering on the divisions. With this system, on-site units not only output images and medical payment requests, but also record "where and when who did what to whom, using what, and for what reason." In short, it becomes possibleto enter the sources of events which generate logs that are useful for healthcare management.

As a result, it also becomes possible to assess each admission. Appraisal and review of clinical paths themselves is also made possible. We have found that there is no point in comparing the profits and losses of medical treatment departments and wards uniformly, since actual analyses of profit-and-loss indicated that each medical treatment department/ward has different characteristics. Therefore, we found it better to focus our efforts at improvement using comparisons of time series data from the same medical treatment department/ ward [4]. Losses are inevitable at some medical treatment departments and wards for institutional and political reasons. In this context, we concluded that money-losing divisions should be asked to make efforts to reduce their losses by setting goals, rather than to attempt to turn uniformly profitable.

This system not only aims to provide risk and logistics management, but also comprehensive oversight of management resources, means of preventing medical errors and the application of medical execution records to evidence based medicine (EBM). The system can also conform to a package payment system. Conventional systems could manage logistics in the central materials division, but it was difficult to manage materials accurately at the point of consumption in each division and department. The POAS system incorporates an online barcode check via newly developed portable terminals. It firmly establishes an efficient business system that records the relationships between materials use and work – which had not previously been recognized – and it eliminates waste. Simply stated, POAS enables users to relate medical activities to materials accurately, even though they are not listed on medical payment requests, and to confirm the real-time movements of materials and patients, after eliminating duplicate inputs and reducing clinical labor. Material on data that accrues within each division's system is transferred simultaneously to the management control system.

In conventional systems, prescription changes made by a physician would take several minutes to several hours to be reflected on terminals used by nurses or the pharmacy. Hospitals have experienced accidents even when barcodes were scanned during a procedure because no alarm sounded at the time. In POAS, all the information including prescription changes made by the physician, as well as confirmation points for nurses and pharmacists are on a shared database, so the information is reflected in 2 to 3 seconds. This eliminates injection accidents. The displays also reflect the use of a shared database. While displays for physicians, nurses, pharmacists and co-medical staff all have a different appearance, any shared data content is queried from the same database. POAS uses a single set of data for controlling processes. This is because data duplication results in inconsistencies and accidents.

You can imagine that an accident occurs. Is the physician responsible for the accident? The relation between a sequence of processes is crucial for preventing accidents. It would be impossible to reproduce the situation in conventional systems, but POAS makes this possible. By analyzing the actions of the person in question as well as those of the people connected with this person, hospital operators can establish systems and organizations that are robust in terms of accident prevention.

Conclusion

Since our focus was to develop a system based on data capture at the point of action throughout the hospital, this system is designed to be able to unitarily analyze data for better healthcare management. Our analysis has shown that the cost savings effect alone is over four million dollars per year. The quality of care and improved outcomes has shown equally significant improvement.

References

[1] Akiyama M. Migration of the Japanese healthcare enterprise from a financial to integrated management: strategy and architecture, Medinfo. 2001: 10(Pt 1): 715-718.

[2] Akiyama M. A Medical Information System as ERP (Enterprise Resource Planning) for the Hospital Management Medinfo. 2004: 11: 1502.

[3] 7PKC Corporation. A Problem Oriented Approach to the Computerized Patient Record. 1998. http://www.pkc.com

[4] Kaplan, Norton. The strategy-Focused Organization: How Balanced Scorecard Companies Thrive in the new Business Environment. Boston, MA: Harvard Business School Press, 2000.

Address for correspondence

Masanori Akiyama M.D., Ph.D
Massachusetts Institute of Technology Sloan School of Management,
3 Cambridge Center,
Room NE20-336, 02142-1347,
Cambridge, MA,USA
Email: poas@mit.edu

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Connecting the Dots: Creation of an Electronic Regional Infection Control Network

Abel N. Kho[a,b], Paul Dexter[b], Larry Lemmon[b], Dauna Carey[b], Heather Woodward-Hagg[c], Steve Hare[b,d], and Brad Doebbeling[b,c, d]

[a]*Northwestern University, Chicago, IL, USA,* [b]*Regenstrief Institute, Inc, Indianápolis, IN, USA*
[c]*Indiana University, Purdue University, Indianápolis, IN, USA*
[d]*VA HSR&D Center on Implementing Evidence-based Practice, IN, USA*

Abstract

The prevalence of Methicillin-resistant Staphylococcus aureus (MRSA) and Vancomycin-resistant Enterococci (VRE) continues to increase dramatically worldwide. Successful programs to reduce infection rates of resistant organisms require regional or national compliance with strict infection control measures and feedback on implementation and reduced rates. We partnered with local infection control professionals (ICPs) and leveraged our existing electronic network to create a comprehensive citywide network to track and uniformly respond to patients admitted with a history of MRSA or VRE. We successfully standardized and included electronic data from four out of six of the major healthcare systems within Indianapolis. We created tailored abstracts to deliver key infection control data to ICPs when a MRSA patient is admitted to a participating hospital. We created web-based data entry forms for ICPs to modify and enter new infection control data. This paper describes our design and initial implementation of a working electronic regional infection control network.

Keywords:

computerized medical tecord dystems, infection control, MRSA, VRE, antibiotic resistance

Introduction

Rates of colonization or infections with drug-resistant organisms are steadily increasing [1]. Patients infected with MRSA suffer increased morbidity and mortality and have significantly increased lengths of stay and health care costs [2]. Delay in correctly detecting these infectious patients and implementing measures to prevent transmission represents a danger to patients and providers who come in direct contact with these patients without appropriate precautions [3].

Some countries, notably Denmark and The Netherlands have successfully controlled rates of MRSA by instituting nationwide policies for prompt barrier isolation of MRSA positive patients and controlling antibiotic prescribing practices [4,5]. Within the United States, a network of hospitals demonstrated that sharing best practices between multiple hospitals significantly reduced rates of nosocomial infection and colonization [6]. Evans et al. at LDS hospital in Utah successfully created a system-wide surveillance system for MRSA patients and demonstrated the itinerant nature of MRSA patients [7]. Samore et al. successfully applied electronic clinical decision support to improve appropriate antimicrobial use across two states [8]. However, such success is the exception, and in most countries, including the United States, MRSA rates continue to steadily increase.

Indianapolis is well positioned to merge the benefits of inter-institutional collaboration and the efficiency of electronic data exchange and build upon the published experience. Investigators at the Regenstrief Institute created the Indiana Network for Patient Care (INPC) in 1994 with the goal of improving the care of patients [9]. The INPC is an *operational* community wide electronic medical record. The system currently includes data from 17 hospitals in five health systems, the Marion County Health Department (MCHD) and various physician practices.

Here we describe our rationale, design and initial implementation of an electronic infection control network paired with a standardized infection control approach to reduce the transmission of MRSA across a city.

Materials and methods

Background

Creating a citywide infection control network involved a concerted effort to build consensus among various stakeholders, and depended on a number of prior developments and studies.

In 1994, researchers at the Regenstrief Institute created the INPC to improve the care of patients across multiple institutions. Today, the INPC connects almost all major health care institutions within Indianapolis spanning over 95% of all inpatient care. When a patient registers at an INPC institution for emergency department care or admission to the hospital, a paper abstract prints which includes key clinical data on that patient from any INPC institution. The

core set of data currently received from all participants includes demographics, laboratory data, ED and inpatient encounter data including chief complaint, coded diagnoses and coded procedures [9].

Prior work within one of the participating INPC institutions highlighted the potential for the timely delivery of accurate data on MRSA history to improve the care and efficiency of patients [10]. In this study, computerized reminders to barrier isolate patients with a history of MRSA increased barrier isolation rates from 33% to 91%. Despite this success, rates of nosocomial infection did not significantly change. One suspected reason for this apparent paradox was the known itinerant nature of patients. A previous trial linking patient registrations across multiple INPC institutions demonstrated that up to 60% of all patients within the Indianapolis area are treated at two or more healthcare institutions within a single year [11]. Currently, information on positive cultures from prior institutions is not readily available. Based on our experience and models of community and hospital spread of MRSA we concluded that effective infection control requires efforts beyond the reach of a single institution [12,13].

Scope of the problem

To quantify the extent of this problem of itinerant patients spreading MRSA within our population, we conducted a retrospective analysis to determine how often patients with a history of MRSA known at one institution presented to an institution where they were not known to be MRSA positive. We obtained lists of patients with a prior history of infection or colonization with MRSA from the infection control departments of three health care systems within central Indianapolis, as of January 2006. The INPC has a tool to link patients with different hospital numbers (within or across systems) who are the same. It uses a number of patient attributes including name, birthdate, and gender to identify matches. Studies conducted on a test population resulted in a 92% sensitivity and 100% specificity for true matches [14]. We matched the common global identifiers from each institution's MRSA list to those on the MRSA lists of the other two in order to determine the subsets of patients held in common on each list, and to determine patients that were only on the list at one institution.

Within a subset of three centrally located INPC institutions we identified 8,895 patients with a history of MRSA based on the infection control lists against the registration lists of patients presenting to three participating institutions. Over a single year, these patients accounted for 4.4% (5,244) of the estimated 120,000 admissions for these three institutions. More concerningly, a subset of MRSA patients not known to have a history of MRSA (i.e. never showed up on the receiving institutions MRSA lists) generated 594 admissions with an average length of stay of 7.2 days. This accounted for an additional 11% of total admissions of MRSA patients over one year.

Addressing the problem: building consensus

With these results in hand, we approached the members of the INPC steering committee and successfully garnered support. We approached the individual hospital ICPs to raise the possibility of an electronic information sharing network for infection control data, and to elicit feedback, we conducted a survey of ICPs to determine their information needs and preferences to guide our subsequent design. Eight ICPs representing four INPC institutions responded to our survey. All respondents preferred electronic notification and web access for both storage and retrieval of infection control data.

With support from both the INPC steering committee, the local ICPs, and the approval of multiple institutional IRBs, we successfully competed for national funding to support further development work.

Through weekly teleconferences the researchers met with ICPs representing all participating institutions and together crafted a standard document and goals. An informatician was present to determine how standard practice could be captured and delivered electronically to ICPs.

We created a secure website using a commercially available software package [15] to coordinate the many participants and provide a common ground for the exchange of documents.

Partnering for organizational change

A key component of our design acknowledges that simply knowing a patient's MRSA status does not necessarily translate into action. We realized that for long term success, we needed to implement organizational change and drive implementation from the lowest level possible from within the frontline staff of an organization.

Lean is a quality improvement methodology based in Systems and Industrial Engineering techniques. These techniques have been shown to be highly effective for continuous quality improvement within manufacturing environments and there is evidence to suggest that appropriately developed and optimized Six Sigma and Lean techniques are effective within healthcare settings[16,17,18]. We partnered with faculty from the Regenstrief Center for Healthcare Engineering and the Purdue School of Engineering and Technology at IUPUI to identify and remove operational barriers present within the processes used to respond to the MRSA positive patient across all institutions. We focused on three processes: hand hygiene, barrier isolation compliance, and active surveillance cultures.

At each institution we identified a Project Team and a Process Owner for each of the three processes. We worked within the current quality improvement framework (e.g. Six Sigma, Model for Improvement) for each institution to clearly define the problem, the project goal and scope, and define metrics to measure success for each Institution's Project Team. Additionally, we worked with front line staff teams in application of Lean tools to identify and remove operational barriers within each of the three processes.

Results

Design and implementation

Based on our weekly meetings, and review of the data collected individually by ICPs at each institution we created seven standardized vocabulary terms to document potential infection sites, and study status (Table 1). To preserve

as much of the original descriptive data stored by ICPs we opted to maintain culture site as free text data. All observations are date and time stamped at the time of data entry, and stored in the individual institutions standardized medical record file.

Table 1 – Standardized infection control terms

Term	Type
MRSA Culture Site	Free text
MRSA Culture Infection Status	Coded
VRE Culture Site	Free text
VRE Culture Infection Status	Coded
Study Action Status	Coded
Barrier Isolation Status	Coded
Infection Control Comments	Free Text

To date we have standardized the data for 12,253 patients with a known history of MRSA, spanning four out of the six major health care institutions within Indianapolis. When a patient with a history of MRSA is admitted to an INPC hospital, we can deliver a printed or electronic abstract containing the standardized infection control data securely to the institutions ICP and admitting office (Figure 1).

We created web-based data entry forms for ICPs to update or make changes to the patient data, and document compliance with our standardized treatment protocol (Figure 2). We included a section for ICPs to write free text comments for delivery on subsequent abstracts.

We have standardized the infection control data and assembled implementation teams of frontline staff at three out of the six major healthcare systems in Indianapolis. We achieved over 90% adherence with these three measures in 2 ICUs at one participating institution, with a resulting elimination of MRSA transmission over the first 60 days of the program.

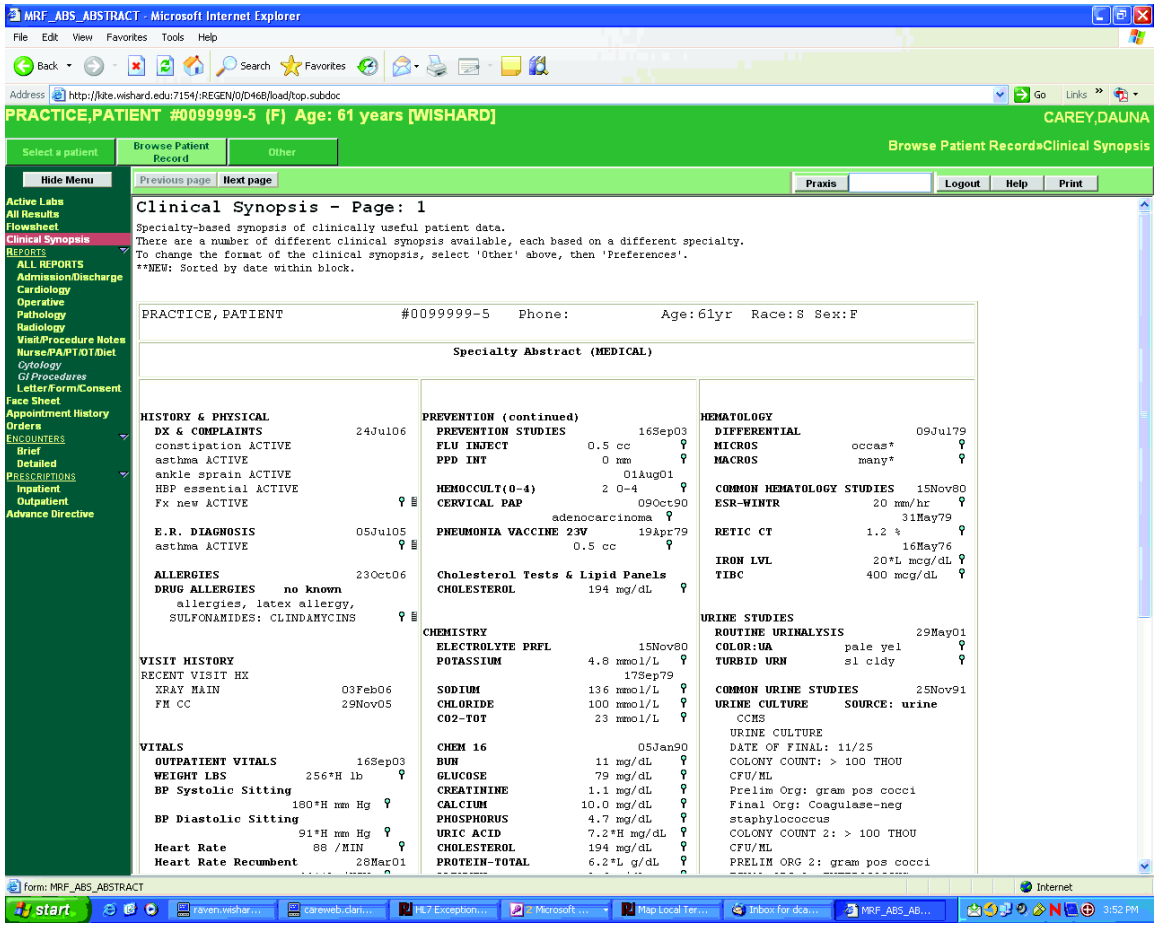


Figure 1 - Screenshot of sample abstract delivered to ICP when a patient admission

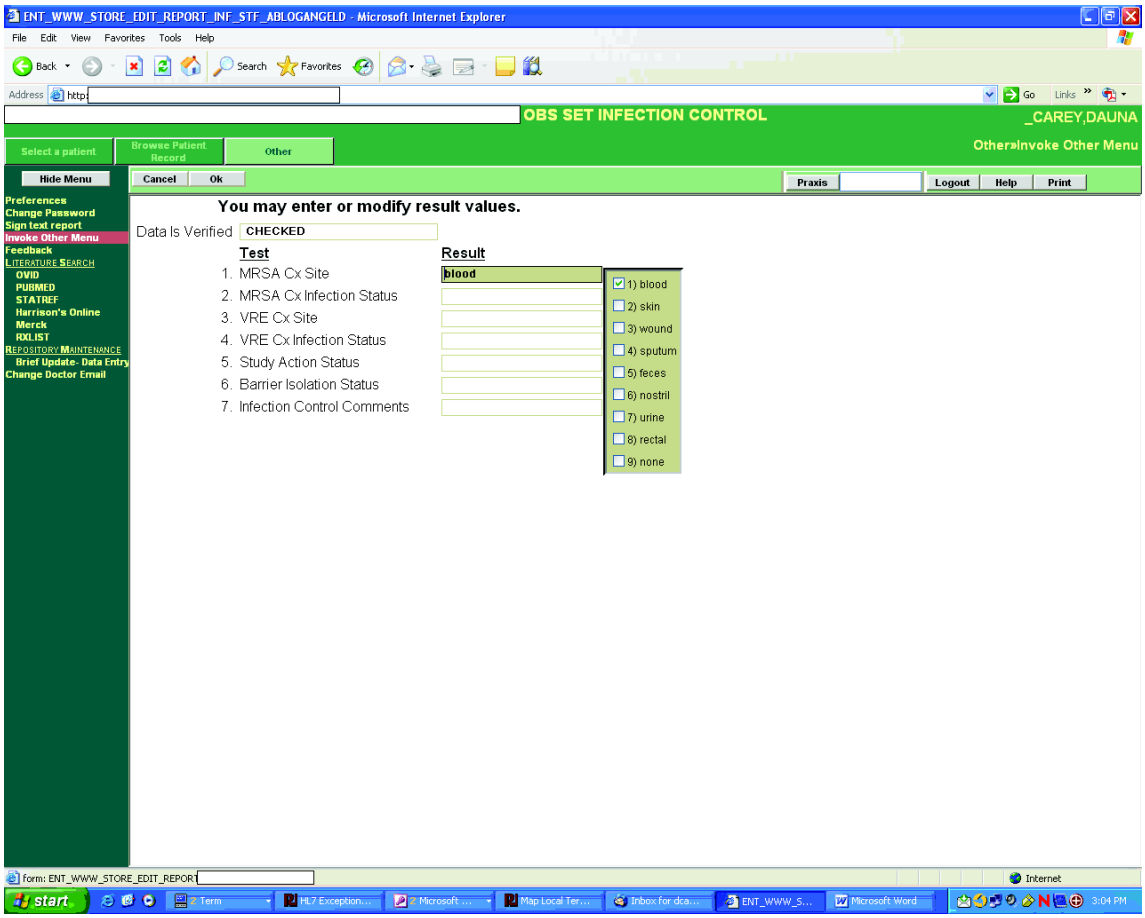


Figure 2 - Screenshot of sample webform for ICP to update patient data and document

Discussion

Common sense and the supporting literature demonstrate that the spread of bacterial resistance is a regional problem [7,12,13]. Patients colonized or infected with drug resistant bacteria travel without restriction and may arrive at another treating institution without warning labels. Our work focuses on delivering the latest available data from the whole community to guide the actions of the front line infection control staff and support implementation of evidence-based measures to control spread.

Our electronic network depends on years of "pipe laying" and builds on prior studies at our institution and others [6,7,8,9,10,11,14]. We successfully used our information sharing network to quantify the problem of crossover of MRSA patients between institutions within our network. We subsequently created a "closed" electronic network to track basic infection control information on any patients with a known history of MRSA upon admission to participating institutions. Coupled with coordination of best practices and standardization of infection control methods, we can both identify, and act to prevent the spread of infection. By combining the efficiency of electronic information exchange with continuous quality improvement methodology we hope to achieve and sustain the effort required to reduce drug resistant infections. Our aim is not 100% success at each opportunity, but rather, to reduce the rate of subsequent infections from each case to less than one. We believe that this approach will follow the model laid forth elsewhere and achieve similar reductions in resistance rates.

Our study is a work in progress, and we continue to incorporate feedback from the frontline ICPs. Our system does not currently integrate with the pre-existing institutional MRSA/VRE tracking systems, and this limitation may create additional work for ICPs and prevent ICPs from generating their own reports. We do not currently have the means to consistently deliver antibiotic susceptibility patterns from all institutions. Our current development work relies on grant support, although we believe that successful reduction in infection rates will more than justify costs.

Future work will focus on quantifying the frequency which our system correctly identifies patients with a prior history of MRSA/VRE, and the effect on ICP compliance with three quality improvement processes (barrier isolation, hand hygiene, and active surveillance). We will ultimately

measure our success by a citywide reduction in infection rates and their related adverse outcomes.

Acknowledgments

The authors are grateful to the valuable contributions of our MRSA Reduction Research Working Group: Laurie Fish, Marie Commiskey, Loretta Marsh, Claire Roembke, Doug Webb, Shawn Hoke, Stephen Wilson, Glenn Bingle, and Jamie Workman-Germann without whom this work would not be possible.

This work was supported by an Agency for Healthcare Research and Quality Contract HHSA290200600013I.

References

[1] National Nosocomial Infections Surveillance (NNIS) System Report, data summary from January 1992 through June 2003, issued August 2003. American Journal of Infection Control. 2003; 31:481-498.

[2] Abramson M, Sexton D. Nosocomial methicillin-resistant and methicillin-susceptible Staphylococcus aureus primary bacteremia: at what costs? Infection Control and Hospital Epidemiology. 1999; 20:408-411.

[3] Ward MM, Diekema DJ, Yankey J, Vaughn TE, Bootsmiller BJ, Pendergast JF, Doebbeling BN. Implementation of strategies to prevent and control the emergence and spread of antimicrobial-resistant microorganisms in U.S. hospitals. Infection Control and Hospital Epidemiology 26:21-30, 2005.

[4] Voss A, Doebbeling BN. The worldwide prevalence of methicillin-resistant Staphylococcus aureus. International Journal of Antimicrobial Agents 1995;5:101-6.

[5] Vandenbroucke-Grauls CM. Methicillin-resistant Staphylococcus aureus control in hospitals: the Dutch experience. Infection Control and Hospital Epidemiology 1996;17:512-13.

[6] Kaye KS, Engemann JJ, Fulmer EM, Clark CC, Noga EM, Sexton DJ. Favorable Impact of an Infection Control Network on Nosocomial Infection Rates in Community Hospitals. Infection Control and Hospital Epidemiology. 2006; 27:228-232.

[7] Evans RS, Lloyd JF, Abouzelof RH, Taylor CW, Anderson VR, Samore MH. System-wide Surveillance for Clinical Encounters by Patients Previously Identified with MRSA and VRE. Medinfo 2004:212-216.

[8] Samore M, Bateman K, Alder S, et al. Clinical Decision Support and Appropriateness of Antimicrobial Prescribing: A Randomized Trial. JAMA 2005;294(18):2305-14.

[9] McDonald CJ, Overhage JM, Barnes M, et al. The Indiana Network for Patient Care: a working local health information infrastructure. An example of a working infrastructure collaboration that links data from five health systems and hundreds of millions of entries. Health Affairs. 2005; 24(5):1214-1220.

[10] Kho AN, Dexter PR, Warvel J, Commiskey M, Wilson SJ, McDonald CJ. Computerized Reminders to Improve Isolation Rates of Patients with Drug-Resistant Infections: Design and Preliminary Results. AMIA Annu Symp Proc. 2005:390-4.

[11] Finnell JT, Overhage JM, Dexter PR, Perkins SM, Lane KA, McDonald CJ. Community clinical data exchange for emergency medicine patients. AMIA ... Annual Symposium Proceedings/AMIA Symposium. 2003:235-238.

[12] Cooper BS, Medley GF, Stone SP, et al. Methicillin-resistant Staphylococcus aureus in hospitals and the community: Stealth dynamics and control catastrophes. Proceedings of the National Academy of Sciences. 2004; 101(27):10223-10228.

[13] Smith DL, Levin SA, Laxminarayan R. Strategic interactions in multi-institutional epidemics of antibiotic resistance. Proceedings of the National Academy of Sciences 2005;102(8):3153-8.

[14] Grannis S, Overhage JM, McDonald CJ. Analysis of Identifier Performance Using a Deterministic Linkage Algorithm. Proceedings of the American Medical Informatics Association 2002:305-9.

[15] Basecamp available at http://www.basecamphq.com/ Last accessed 12/1/06.

[16] Johnson K. Quality Progress. Quality Progress 2004;46-55

[17] Seshadri S. Strategic Finance. Strategic Finance 2001;40-46.

[18] Simmons-Trau D, Cenek P, Counterman J, Hockenbury D, Litwiller L. Reducing VAP with 6 Sigma. Nursing Management 2004; 35(6);41-45.

Address for correspondence

Abel Kho MD, MS
Northwestern University
Division of General Internal Medicine
Regenstrief Affiliated Scientist
676 N St Clair St Suite 200
Chicago, IL 60611 USA
abel.kho@nmff.org
(312) 695-1917

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Dealing with Ecological Fallacy in Preparations for Influenza Pandemics: Use of a Flexible Environment for Adaptation of Simulations to Household Structures in Local Contexts

Toomas Timpka[a,b], Magus Morin[c], Johan Jenvald[c], Elin Gursky[d], Henrik Eriksson [a]

[a] *Department of Computer and Information Science, Linköping University, Linköping, Sweden*
[b] *Department of Social Medicine and Public Health, Linköping University, Linköping, Sweden*
[c] *VSL Research Labs, Linköping, Sweden*
[d] *ANSER Analytic Services Inc., Arlington, USA*

Abstract

Development of strategies for mitigating the severity of a new influenza pandemic is a global public health priority. The aim of this study is to examine effects on simulation outcomes caused by variations in local socio-demographic data. We used a spatially explicit geo-physical model of a virtual city as a baseline and employed an ontology-modeling tool to construct alternative population distributions and household structures. We found that adjustment for the case when single parents in practice were cohabiting led to a higher reproduction rate than that observed for a population with the highest formally recorded share of households with >2 children. When antivirals sufficient to protect 10 percent of the population were administered to schoolchildren, a preliminary effect on the reproduction rate was observed. This effect was eliminated when the household structure was adjusted for cohabiting single parents. Nations have been encouraged to develop estimates of morbidity and mortality during a possible pandemic outbreak. In order to deal with ecological fallacy, the present results suggest that this recommendation can be extended also to local communities.

Keywords:

public health informatics, influenza, epidemiological simulations, basic reproduction rate, health policy

Introduction

Recent human deaths due to infection by highly pathogenic (H5N1) avian influenza A virus have highlighted the possibility of a devastating pandemic similar to that of 1917-1918, [1]. The development of strategies for mitigating the severity of a new influenza pandemic is therefore a global public health priority. Even the concern that a global influenza pandemic may cause millions of fatalities is a public health issue on its own. In planning how communities should react to a potential pandemic, a reasonable balance must be struck between the rights and obligations of the individual and the protection of the population at large. To support such plans, national-scale epidemic simulations to examine intervention have been undertaken, e.g. using Great Britain and the United States as examples [2,3]. In practice, however, much of the planning of interventions and implementation of response actions have to be carried out at local and regional levels in order to deal with ecological fallacy. Moreover, the validity of policy effectiveness estimates will change if the characteristics of the pandemic strain differ substantially from those seen in past pandemics [4]. It is therefore necessary to allow a high level of flexibility in simulation models to be used at the local level during a factual pandemic outbreak.

We have previously reported general system requirements [5] and from the management of social geographic issues [6] during the planning of pandemic simulations in a local public health context. Based on these experiences and preliminary comparative simulations of influenza outbreaks in local communities in Europe [7], the aim of this study is to examine effects on simulation outcomes caused by variations in local socio-demographic data. Specifically, the geo-physical structure of a virtual city constructed on the basis of data from Linköping (population 135.800), Sweden, is used to study alternative household distributions and their influences on intervention planning. Studies of the morbidity patterns during the pandemic outbreak 1917-1918 have shown that the household structure is an important determinant of the strength of the epidemic [8]. One issue associated with modern census records is that they report households to which individuals formally are associated. In some cases, however, people factually stay at other locations than their formal residence, e.g. students and unmarried couples. The problem for pandemic simulations is thus that people are factually living together even though they are recorded as staying separately in the official registries. There is, thus, a need to study the impact from these variations on estimates used a basis for local policy development.

The basic reproductive rate, R_0, is defined as the expected number of secondary infections arising from a single individual during his or her entire infectious period, in a population of susceptible persons. R_0 is often used as a threshold parameter that predicts whether an infection will spread [9]. If $R_0 > 1$, the infectious agent is theoretically

able to invade the susceptible population. Accordingly, this rate was used as the main outcome measure in the present study.

Methods

We used a spatially explicit geo-physical model of a virtual city based on Linköping, Sweden (Table 1) as a baseline and employed an ontology-modeling tool [10] to construct alternative population distributions and structures. A mixing group design was used for the simulations. The model includes mixing groups for households, neighborhoods, schools, daycare centers, and playgroups. Children are assumed to attend local elementary and middle schools (n=45), whereas adolescents are randomly assigned to city high schools (n=5). One way of understanding this explorative experiment is to view it as moving populations with different household and family compositions to the geophysical structure of a specific city. R_0 was estimated by randomly selecting and infecting persons (n=1,000) from a fully susceptible population (n=135,800) and counting the number of secondary cases. The average number of secondary cases represents R_0. Our estimate of R_0 is the mean over 100 populations. Influenza transmission is represented by the number of infected individuals that are expected to pass through two phases. The first phase is an incubation period when they are not infectious and the second phase is the infectious period. The lengths of the incubation and infectious periods follow empirical probability distributions, with mean lengths of 1.9 and 4.1 days, respectively [3,11]. During the infectious period, the infected person may or may not develop influenza symptoms. The probability that a person will be symptomatic given that person has been infected is 0.67. An asymptomatic infection is assumed to be 50 percent as infectious as a symptomatic infection. In the model, the probability that persons with symptoms withdraw to the household is 0.67, exposing only the other members of their household. People who do not withdraw continue circulating in their mixing groups.

Table 1 – Age distribution for the population of the virtual city (n=135.800)

Category	*n*
Children, 0–6	8.800
Children, 7–18	28.000
Adults, 19–65	85.000
Elderly, 66–	14.0 00

In the first test case, we compared the effects of different household distributions on the disease spread during an influenza outbreak. The distribution of household size (number of persons) was varied according to data from Statistics Sweden, USA standard data used for mixing group simulations, and local U.S. Census data on general demographic characteristics (DP-1) and households and families (QT-10) from the Census 2000 Summary File 2 (SF 2) 100-Percent Data for Utah and the District of Columbia (D.C.). The Utah and D.C. locations were selected because they represent U.S. extreme cases. For example, D.C. has the highest rate of single-person households (n=1; 43.8%). It can also be noted that more than 8% of the households in Utah contained more than three children, and >12% of the households in both D.C. and Utah contained more than two adults aged 19-65 years. These census data were the basis for the household model in the ontology-modeling tool. We also developed two alternative household structures for the Swedish reference populations. First, we moved in single adults to live with all single parents. For the second alternative structure, we pared all single adults to 2-person households.

Table 2 shows the distribution (in percent) of household types for the populations used to populate the virtual city in the study. The reference distribution is based on the Swedish national average, while the US standard (U.S. Std.) has previously been used for mixing group simulations [12].

Table 2 – Distribution (percent) of household types for study populations

Household type	Population			
	Refer -ence	U.S. Std.	D.C.	Utah
1 person	41	33	44	18
Adult	*32*	*24*	*34*	*12*
Elderly	*9*	*9*	*10*	*6*
2 persons	32	34	23	16
Adults	*17*	*17*	*12*	*11*
Elderly	*1*	*3*	*1*	*1*
Adult, elderly	*10*	*14*	*2*	*1*
Adult, child	*4*	*0*	*8*	*3*
3 persons	10	13	16	20
Adults, child	*8*	*12*	*3*	*9*
Adult, 2 children	*2*	*0*	*4*	*2*
3 adults	*0*	*0*	*6*	*6*
Other	*0*	*0*	*3*	*3*
4 persons	12	10	10	20
Adults, 2 children	*11*	*10*	*2*	*9*
Adult, 3 children	*1*	*0*	*2*	*2*
4 adults	*0*	*0*	*4*	*6*
Other	*0*	*0*	*2*	*3*
5 persons	4	7	4	13
Adults, 3 children	*4*	*7*	*1*	*6*
5 adults	*0*	*0*	*2*	*4*
Other	*0*	*0*	*1*	*3*
Other	1	3	3	13

In the second test case, we investigated the influence from different household structures on estimates of intervention outcomes. In a first analysis, we simulated a policy intervention restricting interactions in the community. Closing the five high schools in the virtual city was used as an example. In the second analysis, we examined three different strategies to distribute antiviral medication in preventive doses to groups being candidates for prioritization on social or medical grounds. Preliminary studies in household contexts have suggested a satisfactory effect (80%) on reducing infectiousness from Oseltamivir [13]. Because these studies have not been randomized, we assumed a 60% antiviral efficacy from the preventive medication. In addition, the length of the infectious period was assumed to be one day shorter. The first strategy was to treat adults working at vital societal positions, e.g. in the police force, health care, and rescue services. The second strategy was to treat school children. The cost of influenza-related hospitalizations in children has recently been reported to be considerably higher than previously estimated. [14]. The third strategy was to treat the elderly. Excess hospitalization estimates among adults aged 65 years and high-risk 50-64 year olds during non-pandemic influenza seasons suggest that these groups should have priority for medical preventive interventions [15]. The latter strategies were studied with regard to R_0 in two resources contexts, having either 5.000 or 15.000 individual treatment packages available.

Results

Household structures

Computations of R_0 based on crude registry data displayed as expected the highest rate for the population with the highest number of large households with children (Utah; R_0 3.05) and the lowest for the population most single adult households (D.C.; R_0 1.49). The reference population (R_0 2.23) and the USA standard (R_0 2.45) showed similar rates. Table 3 displays the mean R_0 and standard deviation (*S.D.*) for the virtual city by the alternative populations. The Swedish census data are used as reference.

Table 3 – Mean R_0 for the virtual city by alternative populations

Population	***R₀***	***S.D.***
Reference	2.23	0.24
U.S. Std.	2.45	0.26
D.C.	1.49	0.16
Utah	3.05	0.26

However, when the reference population was adjusted for the instance when single parents formally recorded as staying alone but factually stayed together with another single adult, the R_0 reached a level (3.10) higher than that observed for the Utah population.

Table 4 – Effects on R_0 from differences in reported and possible factual household structures

Population	***R₀***	***S.D.***
Reference	2.23	0.24
Single parents cohabiting	3.10	0.31
Singles cohabiting	2.25	0.20

Table 4 shows the effects on R_0 from differences in reported and possible factual household structures, where the Swedish census data are used as reference. Adjustments are displayed both for the case when all single parents are cohabiting with another single adult, and when all single adults are cohabiting with another single adult. In contrast to the case of single parents, moving all single individuals without children together into 2-person households did thus not significantly influence R_0.

Interventions

Table 5 displays the mean R_0 for the virtual city, with high schools closed, by alternative populations and change in R_0 (R_0). In this scenario the Swedish household structure used as reference. The reference structure is also displayed adjusted for the hypothetical case when all single parents are cohabiting with a single adult. The most significant effect from closing high schools in the virtual city was observed when it was populated with the reference population adjusted for single parents cohabiting with single adults (Ro –1.64). In fact, in the city with closed high schools, the differences in R_0 between the reference population and the adjusted reference population were almost eliminated.

Table 5 – Mean R_0 for the virtual city, with high schools closed, by alternative household structures.

Population	***R₀***	***S.D.***	***R₀***
Reference	1.32	0.094	-0.91
Reference adjusted	1.46	0.099	-1.64
U.S. Std.	1.49	0.11	-0.96
D.C.	0.99	0.067	-0.50

Table 6 shows the mean R_0 for the virtual city with the reference household structure displayed by the three preventive strategies (covering adults in vital societal positions, school children, and high-risk elderly, respectively) and two resource conditions (5.000 or 15.000 longitudinal doses) for providing antiviral protection. The R_0 for the strategy targeting school children is displayed also with the household structure adjusted for cohabiting single parents.

The effect on R_0 was low for all strategies in the resource situation when 5.000 longitudinal dosages of antiviral

medication were administered. When 15.000 dosages were available, the only significant effect on the reproduction rate (R_0 reduction -0.49, 22%) was observed when the antiviral protection was administered to schoolchildren. However, when the adjusted household structure was used, the reduction in R_0 was similar in magnitude (R_0 -0.56; 18%) and did not alone lead to a satisfactory level of R_0 (2.54 compared to a desired level <1.80).

Table 6 – Mean R_0 for the virtual city by three strategies and two resource conditions for supply of antiviral protection.

Target group	*n*	R_0	S.D.	R_0
Adults	5.000	2.15	0.25	-0.08
Children	5.000	2.05	0.19	-0.18
Children adjusted	5.000	2.91	0.13	-0.19
Elderly	5.000	2.21	0.22	-0.02
Adults	15.000	2.18	0.21	-0.05
Children	15.000	1.74	0.16	-0.49
Children adjusted	15.000	2.54	0.22	-0.56
Elderly	5.000	2.20	0.21	-0.03
Reference		2.23	0.24	

Discussion

The aim of this study was to examine deviations in estimates of local morbidity during an influenza pandemic caused by variations in socio-demographic data. In specific, the geo-physical structure of a virtual city was used to study alternative household distributions and their possible influences on estimated reproduction rates. In the first test case, we compared the effects of different household distributions on the disease spread at the initial stages of an outbreak. We found that adjustment for the case when single parents were cohabiting led to a higher reproduction rate than that observed for a population with the highest formally recorded share of households with >2 children. The impact of possible differences between formal and factual household structures has relevance for local planning, when the intention is to suppress a pandemic before it is established in the local community, i.e. to fight the infection at the sites where it is most likely to disseminate at a large scale. If the R_0 is assumed to be relatively low in a community, it may be considered sufficient to protect high-risk groups and suggest general measures for reduction of person-to-person virus transmission, e.g. by increased personal hygiene. However, if this assumption is wrong, public policy interventions, such as closing schools, may be erroneously delayed with possibly fatal consequences.

In the second test case, we investigated the influence from household structures on estimates of intervention outcomes. The most significant effect from the policy intervention 'closing high schools' in the virtual city was observed when it was populated with the reference population adjusted for cohabiting single parents. In this setting, the differences in R_0 between the reference population and the adjusted reference population were almost eliminated. Regarding interventions involving antiviral medication, the effect on general disease transmission was low for all strategies in the resource context when 5000 longitudinal dosages were distributed to the virtual population of 135.800. When 15000 dosages were available, the only satisfactory effect on the reproduction rate was observed when the antivirals were administered to protect schoolchildren, but this effect was reduced when the household structure was adjusted for cohabiting single parents. These results suggest that it is important clarify the purpose of the intervention; containment of the epidemic or protection of high-risk groups. Also, these results highlight the fact that there are three broad areas in which the ethical issues associated with handling public health emergencies must be addressed,, i.e. rationing, restrictions and responsibilities [16]. In light of a pandemic influenza threat, some experiences from seasonal influenza may apply to the pandemic setting, but many preparatory decisions must be based on estimates. In 2006, two out of three nations prioritize health care workers at the highest level for antiviral distribution [1]. The justification for this prioritization is that health care workers are at increased risk of exposure and, therefore, for acquiring infection and/or transmitting it to patients. Additionally, health care workers perform vital services to the community. What ethical principles should guide rationing-based decisions in local communities, what data should inform these decisions, and how should rationing be implemented based on these data are all important questions.

Geo-physically explicit models of local communities seldom have been used for comparative experiments in the area of influenza response planning. However, detailed computational models of smallpox transmission and control based on agent representations or social networks have been reported [17,18]. The agent-based models explicitly represent an "artificial society" of individual human beings and physical structures, each implemented as a distinct object in the software. The agents interact locally with one another in code-represented physical places, e.g. homes, workplaces, schools, and hospitals. Although being more intuitive than the mixing-group method used for the present simulations, a limitation of agent-based model is the trade-off between inclusion of a large level of detail and the complexity of the resultant model. A secondary, and more important, consequence of the complexity is that the use of the results in policy development should be made with great caution.

The starting point for this study was the necessity for dealing with ecological fallacy in the planning of local responses to pandemic influenza: It must be acknowledged that what may be observed at global or national levels may

not always be applicable in the local setting. Nations have been encouraged to develop estimates of morbidity and mortality during a possible pandemic outbreak. These results suggest that this recommendation can also be extended to local communities.

Acknowledgments

This work was supported by the Swedish Emergency Man-agement Agency (KBM) under contract 0700/2004. Einar Holm and Magnus Strömgren, Umeå University, Sweden, James Nyce, Ball State University, USA, and Britt Åkerlind and Kristina Kinnunen from Linköping University, Sweden have contributed valuable comments to the analyses presented in this report.

References

[1] Uscher-Pines L, Omer SB, Barnett DJ, Burke TA, Balicer RD. Priority setting for pandemic influenza: an analysis of national preparedness plans. PLoS Med. 2006 Oct; 3(10):e436.

[2] Ferguson NM, Cummings DA, Fraser C, Cajka JC, Cooley PC, Burke DS. Strategies for mitigating an influenza pandemic. Nature. 2006 Jul 27;442(7101): 448-52. Epub 2006 Apr 26.

[3] Germann TC, Kadau K, Longini IM Jr, Macken CA. Mitigation strategies for pandemic influenza in the United States. Proc Natl Acad Sci U S A. 2006 Apr 11; 103(15):5935-40. Epub 2006 Apr 3.

[4] Mounier-Jack S, Coker RJ. How prepared is Europe for pandemic influenza? Analysis of national plans. Lancet. 2006 Apr 29;367(9520):1405-11.

[5] Timpka T, Morin M, Jenvald J, Eriksson H, Gursky E. Towards a simulation environment for modeling of local influenza outbreaks. AMIA Annu Symp Proc. 2005;:729-33.

[6] Holm E, Timpka T. A discrete time-space geography for epidemiology: from mixing groups to pockets of local order in pandemic simulations. MedInfo 2007;

[7] Morin M, Jenvald J, Eriksson H, Timpka T. Simulating Influenza Outbreaks in Local Communities. In Fritzon P, ed. Proceedings of Biomedical Simulation Conference '05 (BioMedSim '05). Amsterdam: IOS Press, 2005.

[8] Mamelund SE. A socially neutral disease? Individual social class, household wealth and mortality from Spanish influenza in two socially contrasting parishes in Kristiania 1918-19. Soc Sci Med. 2006 Feb;62(4):923-40.

[9] Heffernan JM, Smith RJ, Wahl LM. Perspectives on the basic reproductive ratio. J R Soc Interface. 2005 Sep 22 ;2(4):281-93.

[10] Eriksson H, Morin M, Jenvald J, Gursky E, Holm E, Timpka T. Ontology-based modeling of pandemic simulation scenarios. MedInfo 2007;.

[11] Halloran M., Longini IM, Cowart DM, Nizam A. Community trials of vaccination and the epidemic prevention potential. Vaccine 2002;20:3254–3262.

[12] Longini IM Jr, Halloran ME, Nizam A, Yang Y. Containing pandemic influenza with antiviral agents.Am J Epidemiol. 2004 Apr 1;159(7):623-33.

[13] Halloran ME, Hayden FG, Yang Y, Longini IM Jr, Monto AS. Antiviral Effects on Influenza Viral Transmission and Pathogenicity: Observations from Household-based Trials. Am J Epidemiol. 2006 Nov 6; [Epub ahead of print]

[14] Keren R, Zaoutis TE, Saddlemire S, Luan XQ, Coffin SE. Direct medical cost of influenza-related hospitalizations in children. Pediatrics. 2006 Nov;118(5):e1321-7.

[15] Mullooly JP, Bridges CB, Thompson WW, Chen J, Weintraub E, Jackson LA, Black S, Shay DK; Vaccine Safety Datalink Adult Working Group. Influenza- and RSV-associated hospitalizations among adults. Vaccine. 2006 Sep 25; [Epub ahead of print]

[16] Wynia MK. Ethics and public health emergencies: rationing vaccines. Am J Bioeth. 2006 Nov-Dec;6(6):4-7.

[17] Eubank S, Guclu H, Kumar VS, Marathe MV, Srinivasan A, Toroczkai Z, Wang N. Modelling disease outbreaks in realistic urban social networks. Nature. 2004 May 13; 429(6988):180-4.

[18] Burke DS, Epstein JM, Cummings DA, Parker JI, Cline KC, Singa RM, Chakravarty S. Individual-based computational modeling of smallpox epidemic control strategies. Acad Emerg Med. 2006 Nov;13(11):1142-9.

Address for correspondence

Toomas Timpka, MD. PhD.
Department of Computer Science
Department of Social Medicine and Public Health
Linköping University
SE-581 83 Linköping
Sweden
Email: tti@ida.liu.se

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Effectiveness of a Chronic Disease Surveillance Systems for Blood Pressure Monitoring

Damian Borbolla[a], Diego Giunta[bc], Silvana Figar[bc], Mercedes Soriano[b], Adriana Dawidowski[b], Fernan Gonzalez Bernaldo de Quiros[ac]

[a] *Department of Medical Informatics, Hospital Italiano de Buenos Aires, Argentina*
[b] *Department of Epidemiology, Hospital Italiano de Buenos Aires, Argentina* [c]*Department of Medicine, Hospital Italiano de Buenos Aires, Argentina*

Abstract

Information Technology (IT) enables health care providers to manage patients with chronic conditions through identification, follow up and administration of specific interventions. In our setting, we developed a surveillance system for chronic diseases. The aim of this study was to show its efficacy on monitoring blood pressure throughout a cluster randomized controlled trial. Patients without blood pressure registries (condition 1) or with high blood pressure measurements (condition 2) were randomized to be detected by the surveillance system or to receive usual care. The proportion of patients with at least one blood pressure measurement within three months of follow up was 49.9% (207 patients) in the intervention group and 37% (195) in the control group ($p<0.001$) for condition 1. And 61% (224) vs. 50% (239) respectively ($p=0.002$) for condition 2. Patients under the surveillance system have higher proportion of blood pressure measurements, showing this study an improvement on the process of care with this IT tool.

Keywords:

hospital information systems; chronic disease; hypertension; disease management; medical records systems, computerized; population surveillance; randomized controlled trials

Introduction

Chronic conditions are the major cause of illness, disability and death in Argentina like in other countries [1, 2]. Attempts to accomplish an effective control of chronic prevalent diseases are nowadays mandatory. In Argentina, studies performed in different socio-economic levels, showed that hypertension control is heterogeneous and not always done [3, 4].

Information systems generally offer a number of benefits in health care. Chronic Disease Management Systems (CDMS) are an effective information technology tool to improve the management of chronic diseases. CDMS enable health care providers to manage patient with chronic conditions such as hypertension, diabetes or asthma [5]

From an epidemiological point of view a CDMS might help to perform epidemiological surveillance. This means the ongoing, systematic collection, analysis, interpretation, and dissemination of data, being the ultimate aim of all surveillance systems, support disease control [6].

Similar to the surveillance of infectious and other acute disease, surveillance for chronic diseases has been largely implemented [7, 8]. Until now chronic diseases surveillance systems did not allow the quick identification and intervention of detected uncontrolled cases, as has been achieved in acute diseases.

Our setting, a Health Maintenance Organization (HMO) (Plan de Salud del Hospital Italiano de Buenos Aires), is an University Hospital based prepaid health care system. The Hospital is a 650 beds tertiary center with high standards in quality of health care and information systems [9-12]. To achieve control of chronic conditions we started in August 2000 a Disease Management Program (DMP) which its efficacy for hypertension control has already been demonstrated [13]. To improve DMP performance we developed in the year 2005 a surveillance system to detect and rapidly intervene patients with chronic disease [14].

The aim of this study is to show the efficacy of our surveillance system as a strategy for improving the performance of the hypertension care process.

Materials and methods

Study design

We performed a cluster randomized pragmatic controlled trial

Setting

In our HMO around 140.000 middle class residents of Buenos Aires are attended. Most of the HMO health care system is based on primary care physicians (PCP). Primary care is capitated and delivered at the Hospital and/or at any

Selected for best paper award.

of the twenty four HMO's peripheral outpatient care centers.

The total number of PCPs working in our HMO is 182, taking care of 50.110 patients with chronic conditions (average capita 735 patients ranging from 80 to 1415 patients).

Randomization

PCPs with more than 100 capitated patients were included in randomization.

Patients without blood pressure registries (see rules below) seeking for an appointment with their PCP on the following 15 days were prospectively randomized.

In order to achieve a balanced randomization, physicians were stratified in three categories according to the number of appointments scheduled: >5 to <10, > 10 to < 15 and > 15.

Within each strata, physicians were categorized in a random way to be 0 (not included in the surveillance system=control group) or 1 (enrolled in the surveillance system=intervention group).

Finally, according to their physician allocation number (0 or 1) patients were (or were not) enrolled in the surveillance system.

Patients allocated to the intervention group were detected by 2 conditions, patients in condition 1 were those without blood pressure registries, and patients in condition 2 were those with high blood pressure measurements, and listed for the DMP by the system every time they ask for physician's appointments. Patients in the control group were not listed during the 3 months of the study period. (Figure 1)

Sample size

Required sample size was 2200 patients, calculated to detect a difference of 10% in the proportion of patients with their blood pressure measured, considering an alpha error of 1%, a power of 80%, and a desire precision of 2%. We multiply by a design effect of 1.3 to adjust for the additional variability of the cluster design. 88 clusters were needed.

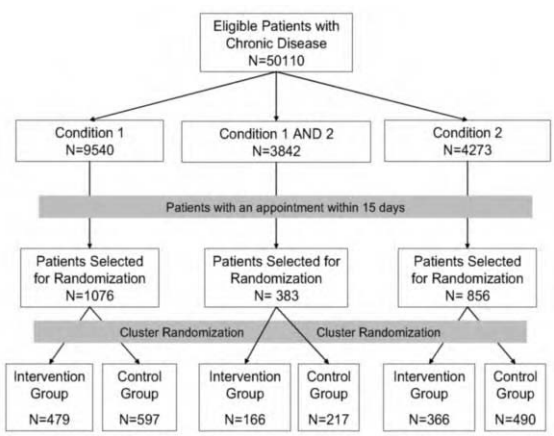


Figure 1- Flow chart of study subjects

Study outcomes

According to the standard outcome metrics for disease management programs published by the Disease Management Association of America [15], patients with chronic conditions must have their blood pressure measured at least once every 6 months, and uncontrolled hypertension patients must have this measurement done in every physician encounter.

The main outcome measure was the proportion of patients with at least one blood pressure measurement during the three months period, among those patients identified by the system according to the above criteria.

The main outcome was expressed separately for patients without registries during the last 6 months (condition 1) and for patients with uncontrolled hypertension (condition 2).

Secondary outcomes were mean systolic and diastolic blood pressures.

Statistical analysis

Means and proportions are shown with their respective standard deviations (SD) and 95% confidence intervals (95% CI) respectively. Comparability of baseline and pre-post intervention characteristics was ascertained by T test (either for single or paired samples) for continuous variables and Chi square test for proportions. The significance was considered at 1% level.

Intention-to-treat analysis was performed maintaining patients in their original groups. We used STATA 8 to perform the statistical analysis.

Intervention group

These patients were identified by the surveillance system and were followed by their PCP (usual care) and by the disease management program.

Surveillance System:

Patient Identification: the Master Patient Index (MPI) is the cornerstone of our health Information system [16]. Since 1999 our HMO has an Electronic Health Record (EHR) that is used by all physicians [12, 17]. Every entry in the record system, including pharmacological prescriptions and studies ordered (physician order entry), have to be attached to a medical problem. Each medical problem, diagnoses, risk factors and past history are automatically codified using terminological service based on SNOMED CT as referent vocabulary [18, 19].

The surveillance system uses information from two different sources:

1- EHR (medical problems, laboratory results, vital signs, pharmacological prescriptions)

2- Appointment Scheduling Software

Follow up: With data provided by this sources the surveillance system generates different surveillance lists(Figure 2), with the patients being classified as:

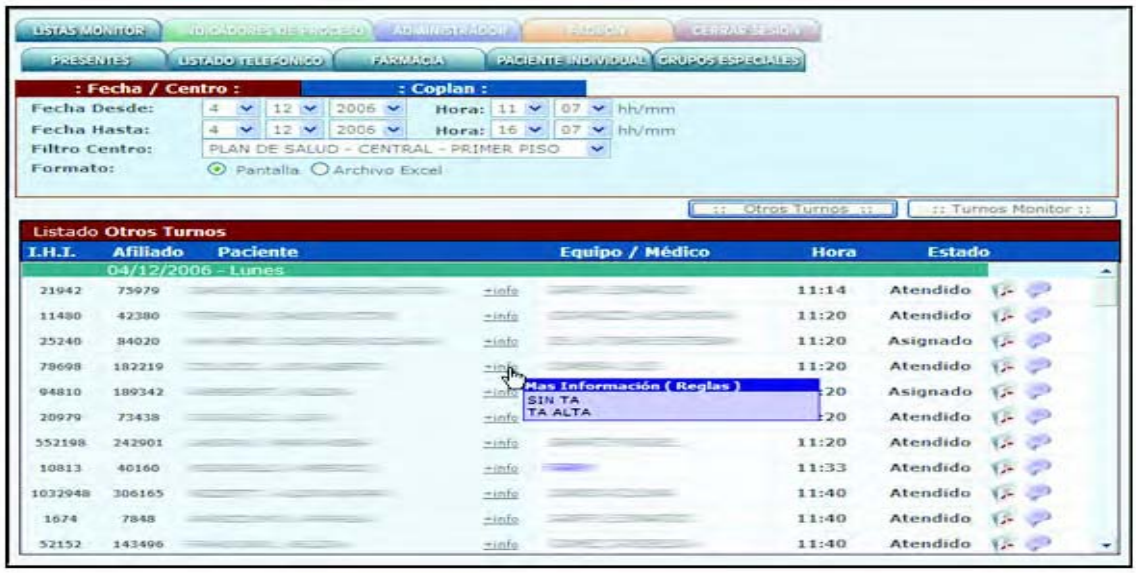


Figure 2 - Surveillance system lists

Not needing active intervention: hypertensive patient with blood pressure measurements recorded in EHR over the last six months with values under 140/90, or 130/80 in case of known diabetes.

- Patients to be enrolled: uncontrolled hypertensive patient (>140/90) or smokers or incident stroke not yet included.
- Patients already enrolled with interventions pending: hypertension patient without blood pressure measured in the last 6 months.

There are more than 10 rules running on our system to identify different medical conditions. In this study we tested the below two rules:

- Patients without blood pressure measurement in the last 6 months (condition 1)
- Patients with a high blood pressure measurement in the last appointment (condition 2)

Surveillance lists were daily matched with the Appointment Scheduling Software.

Appointments from the identified patients were marked generating a remainder for the receptionists so that they sent patients to program's office before physicians' visits.

Disease management program

Once in the program office, assistants measured and recorded risk factors, blood pressure, weight and height.

They reminded patients, verbally and with specially designed leaflets, to record their blood pressure weekly, to keep lifestyle changes and to comply with pharmacological treatment. They also invited patients to attend educational sessions that were previously demonstrated to be efficacious[20]

All the data was systematically recorded on the EHR before physician appointments so that the PCPs could make use of this information during consultations.

Control group

Patients in this group received usual care. This included health care administered by their PCP and by Disease Management Program only under PCPs request.

Results

A total of 2315 patients were included, 1011 were randomized to the intervention, and 1304 to usual care. 67 % attended to the scheduled appointment in both groups. Among them the median number of visits to primary care physicians was 1.93 (1.35) for the intervened group and 1.71 (SD 0.88) in the control group ($p < 0.01$) over the follow up period.

There were no statistical significant differences in basal characteristics between groups as is shown in table 1.

20% of patients control group were referred to the disease management program by their PCP.

Among those who fulfilled criteria for condition 1, there were 207 (49.9% with 95% IC from 45 to 55%) patients that had their blood pressure measured in the intervention group and 195 (37% with 95% IC from 33 to 41%) in the control group ($p<0.001$).

Among those who fulfilled criteria for condition 2, there were 224 (61%) patients that had their blood pressure measured in the intervention group and 239 (50%) in the control group ($p=0.002$).

Table 1 - Basal characteristics

Variables	Intervention Group (N=1011)	Control Group (N=1304)	P value
Mean age in years (SD)	73 (10.54)	73.77 (9.58)	0.73
Women, n (%)	661 (65.4)	846 (64.9)	0.801
Patients with hypertension, n (%)	919 (90.9)	1181 (90.6)	0.784
Patients with diabetes, n (%)	168 (16.6)	224 (17.02)	0.721
Patients with cardiovascular disease, n (%)	74 (7.3)	70 (5.4)	0.054
Patients with lipid disorders, n (%)	536 (53.0)	713 (54.7)	0.426
Patients with Stroke, n (%)	57 (5.6)	59 (4.5)	0.223
Smoker Patients, n (%)	214 (21.6)	260 (19.9)	0.468
Ex smoker Patients, n (%)	30 (3.0)	31 (2.4)	0.379
Mean Basal Systolic Blood Pressure, condition 2 (SD)	152.78 (14.06)	153.80 (13.32)	0.193
Mean Basal Diastolic Blood Presure, condition 2 (SD)	82.15 (11.15)	83.17 (10.58)	0.103

The groups did not differ significantly either initially (table 1) or at final assessment with regard to blood pressure (140/78 mm Hg vs. 138/78 mm Hg; p =0.162 and 0.914 respectively for systolic and diastolic). However both groups statistically reduced systolic blood pressure with a mean difference of 13 mm Hg ($p < 0.001$)

Discussion

The main outcome of our study is that patients in the intervention group had higher proportion of blood pressure measurements during follow-up, showing a statistically significant improvement with the intervention in the process of care of these patients.

Surveillance of chronic diseases has been based on population surveys or administrative data. Although this classic strategy is useful to plan future changes in clinical care process, it does not allow an active action on patients under the surveillance system.

Our program aims to monitor, detect and intervene as soon as possible those patients not fulfilling good quality of care outcomes. In this sense this program is more alike to active surveillance systems implemented for acute diseases than to those implemented for chronic diseases.

Although we were able to show improvement in process outcomes we could not find differences in clinical outcomes between groups. Several reasons could explain this discrepancy: this was a short term study (3 months), usually not enough to obtain clinical outcomes results. As PCPs` usual care in our Hospital includes sending patients to DMP, and there was a significant proportion of patients in the control group that actually received the intervention (20 % contamination). Also it has already been shown that there is more variability in outcome process than in clinical outcomes.

In previous reports, chronic disease management systems are shown as IT tools that could replace EHR [5], we think that this kind of surveillance systems must be integrated into a health information system, and the basic requirements that these IT applications should have are:.

- Identify, trace and continuously monitor chronic ill-patients to obtain a well cared population. Alert those patients reluctant to receive health care.
- Rank possible treatments, pharmacological and educational interventions, in terms or their greatest potential impact on outcomes.
- Keep adherence along the time in these chronic and asymptomatic conditions. Include IT solutions to schedule and remind preventive and needed interventions intelligently tailored to each patient profile

Further issues concerning the identification of barriers to comply with this chronic care model either from patients or physicians points of view should be studied and addressed.

References

[1] WHO. Informe sobre la salud en el mundo 2004 anexo estadístico 2004 [cited; Available from: http://www.who.int/whr/2004/es/

[2] PAHO. Base de Datos de Indicadores Básicos en Salud de la OPS. 2006 [cited; Available from: http://www.paho.org/Spanish/DD/AIS/cp_032.htm

[3] Janson J FS, González B, De Quiros F, Waisman G, Galarza C, et al. Prevalence of hypertension among ambulatory elderly patients in a private university hospital in argentina. J Hypertens 2002;20((Supl. 4): S85.).

[4] Przygoda P, Janson J, O'Flaherty M, Waisman G, Galarza CR, Alfie J, et al. Lack of effective blood pressure control among an elder hypertensive population in Buenos Aires. Am J Hypertens. 1998 Aug;11(8 Pt 1):1024-7.

[5] Jantos L HM. IT Tools for Chronic Disease Management: How Do They Measure Up: California HealthCare Foundation; 2006.
[6] CDC. Guidelines for evaluating surveillance systems. MMWR. 1988;37(S-5).
[7] Canada PHAo. Chronic Disease Surveillance in Canada: Health Surveillance Coordination Division,Centre for Surveillance Coordination Population and Public Health Branch, Health Canada; 2003.
[8] CDC. Indicators for Chronic Disease Surveillance. MMWR. 2004 September 10, 2004 53(RR11):6.
[9] Baum A, Figar S, Severino J, Assale D, Schachner B, Otero P, et al. Assessing the impact of change in the organization of a technical support system for an Health Information Systems (HIS). Medinfo; 2004; San Francisco, California, USA.; 2004. p. 1367-70.
[10] Gonzalez Bernaldo de Quiros F, Tavasci I, García Martí S, Paris M, Baum A, Martinez M, et al. Snomed CT and ICPC-2E cross mapping: Coding to facilitate the access to information. In: WONCA, editor. 17th World Conference of Family Doctors; 2004 October 2004; Orlando, Fl, USA: WONCA 2004 scientific comitee; 2004. p. 168.
[11] Luna D, Bernaldo de Quiros FG, Garfi L, Soriano E, O´Flaherty M. Reliability of secondary central coding of medical problems in primary care by non medical coders, using the International Classification of Primary Care (ICPC). Medinfo. 2001;10(Pt 2):300.
[12] Luna D, Otero P, Gomez A, Martinez M, García Martí S, Schpilberg M, et al. Implementación de una Historia Clínica Electrónica Ambulatoria: "Proyecto ITALICA". In: Ceitlin M-R, M., editor. 6to Simposio de Informática en Salud - 32 JAIIO; 2003 2-5 Septiembre 2003; Buenos Aires, Argentina: Sociedad Argentina de Informática e Investigación Operativa (SADIO); 2003.
[13] Figar S, Waisman G, De Quiros FG, Galarza C, Marchetti M, Loria GR, et al. Narrowing the gap in hypertension: effectiveness of a complex antihypertensive program in the elderly. Dis Manag. 2004 Fall;7(3):235-43.
[14] Gonzalez Bernaldo de Quiros F, Dawidowski A, Baum A, Soriano M, Bravo M, Figar S, et al. Medical Program Administrator: a Study and Functional Analysis ff Software for Follow-Up and Control of a Patient Population With Chronic Diseases. Technol Health Care. 2004 December 2004;12(5):369.
[15] Standard outcome metrics and evaluation methodology for disease management programs. American Healthways and Johns Hopkins Consensus Conference. Dis Manag. 2003 Fall;6(3):121-38.
[16] Garfi L, Navajas P, Gomez A, Luna D, Bernaldo de Quiros FG. Implementación de un sistema centralizado para la identificación de pacientes en un hospital de alta complejidad. In: Leguiza J-D, A., editor. 5to Simposio de Informática en Salud - 31 JAIIO; 2002 Septiembre 2002; Santa Fe, Argentina: Sociedad Argentina de Informática e Investigación Operativa (SADIO); 2002. p. 11-8.
[17] Gonzalez Bernaldo de Quiros F, Soriano E, Luna D, Gomez A, Martinez M, Schpilberg M, et al. Desarrollo e implementación de una Historia Clínica Electrónica de Internación en un Hospital de alta complejidad. In: Ceitlin M-R, M., editor. 6to Simposio de Informática en Salud - 32 JAIIO; 2003 2-5 Septiembre 2003; Buenos Aires, Argentina: Sociedad Argentina de Informática e Investigación Operativa (SADIO); 2003.
[18] Lopez Osornio A, Montenegro S, García Martí S, Toselli L, Otero C, Tavasci I, et al. Codificación múltiple de una lista de problemas utilizando la CIAP-2, CIE-10 y SNOMED CT. 3er Virtual Congress of Medical Informatics - Informedica; 2004; 2004.
[19] Lopez Osornio A, Gambarte ML, Otero C, Gomez A, Martinez M, Soriano E, et al. Desarrollo de un servidor de terminología clínico. 8mo Simposio de Informática en Salud - 34 JAIIO; 2005; Santa Fé, Argentina: Sociedad Argentina de Informática e Investigación Operativa (SADIO); 2005. p. 29-43.
[20] Figar S, Galarza C, Petrlik E, Hornstein L, Rodriguez Loria G, Waisman G, et al. Effect of education on blood pressure control in elderly persons: a randomized controlled trial. Am J Hypertens. 2006 Jul;19(7):737-43.

Acknowledgments

We would like to thank Carlos Musri for his help in the accomplishment of this paper

Address for correspondence

Dr. Damian Borbolla: damian.borbolla@hospitalitaliano.org.ar
Medical Informatics Residency. Department of Medical Informatics. Hospital Italiano de Buenos Aires. Gascón 450. Ciudad Autónoma de Buenos Aires. Argentina. (C1181ACH).
Tel/Fax:+54-11-4959-0507

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Nurses and Computers. An International Perspective on Nurses' Requirements

Carol S. Bond

Institute of Health and Community Studies, Bournemouth University, England

Abstract

This paper reports the findings from a Florence Nightingale Foundation Travel Scholarship undertaken by the author in the spring of 2006.

The aim of the visit was to explore nurses' attitudes towards, and experiences of, using computers in their practice, and the requirements that they have to encourage, promote and support them in using ICT.

Nurses were found to be using computers mainly for carrying out administrative tasks, such as updating records, rather than as information tools to support evidence based practice, or patient information needs. Nurses discussed the systems they used, the equipment provided, and their skills, or more often their lack of skills. The need for support was a frequent comment, most nurses feeling that it was essential that help was available at the point of need, and that it was provided by someone, preferably a nurse, who understood the work context.

Three groups of nurses were identified. Engagers; Worried Willing and Resisters. The report concludes that pre-registration education has a responsibility to seek to ensure that newly qualified nurses enter practice as engagers.

Keywords

Nursing Informatics, ICT, Attitudes towards computers. Education.

Introduction

In the UK the government made Information and Communication Technologies (ICT) a priority area when the National Health Service (NHS) published Information for Health in 1998[1] The strategy identified the need for computerisation of many support systems, including the introduction of nationally available electronic health records, networked services to support appointment systems, electronic handling of a wide range of diagnostic tests and results and a networked prescription service. Information for patients as well as professionals was a key element of the programme, the strategy stated:

> *'People need information about health and healthcare in many different circumstances. Patients want to know more about what is wrong with them and how they can best look after themselves. Carers or relatives or friends seek information on behalf of others'*

Providing information to patients was seen as 'an integral part of the local clinical care process', it was however recognised that a culture change was necessary to achieve this aim.

The need for these developments to be managed nationally was identified, and they are currently being implemented through the National Programme for IT (NPfIT) and the Connecting for Health programme. This ambitious programme will see all patient administration, patient records (including nursing care records) and communication between professionals being computerised.

Nurses need to be engaged with these developments. At the very least they need to have the skills and knowledge to use the systems efficiently and safely. They need more than just the skills to use systems provided by the NHS. They also need to have the skills to support patients and their family/carers in meeting their information needs, and to have an understanding of the systems that will enable them to transfer their skills to new systems as they are developed.

Nurses however generally have poor ICT skills [2, 3, 4,] and are resistant to the introduction of ICT [5]. Heather Tierney-Moore, the nursing clinical lead for Connecting for Health has stated that

> *'The problem is that IT is a big turn-off for most nurses…. If things are branded as IT it's unlikely nurses will bother to pick them up, let alone engage with them,'*

A National Audit Office report into the implementation of the National Programme [6] stressed that the lack of IT skills within the NHS were a risk to the timely implementation of the programme.

Just what skills are needed is not universally agreed. The NHS[7] has published competencies required of nurses, which includes a range of knowledge and skills, e.g. clinical informatics and information security, as well as basic IT skills. In spite of this the perception found in some studies[8, 9] is that all nurses need are basic IT skills.

One of the aims of Information for Health[1] was to establish, amongst other things, a culture to ensure that NHS clinicians would be able to access the information that they need to provide effective patient care. In the case of nurses the evidence suggests that this culture is not yet established.

Background to the study

With the support of a Florence Nightingale Foundation Travel Scholarship a five week visit to New Zealand was undertaken to talk to nurses about their requirements both in what a computer system should provide, and what support they considered essential.

The study was undertaken in New Zealand for a variety of reasons. The New Zealand health system and culture is similar to that in the UK. The challenges faced, identified in the Health Information Strategy for New Zealand [10] are very similar challenges to the UK:

- an ageing population, which will increase the pressure on our health sector
- rising incidences of chronic diseases such as diabetes and cardiovascular disease
- the re-emergence of some diseases
- the emergence of new infectious diseases such as Severe Acute Respiratory Syndrome (SARS)
- new technologies that are making more effective treatments available, often at higher prices.

The IT systems underpinning the health care systems are also under review. The New Zealand strategy contains many of the same aims and initiatives as the UK. There are however subtle but important differences in focus. Whilst the UK is currently focusing on NPfIT commissioning and purchasing national computer and records system, New Zealand considers the need for local customisation of the systems to be an important feature of their strategy.

New Zealand is also putting information for clinicians and patients higher in the priorities than the NPfIT is. The initial launch of the New Zealand strategy was the 2001 WAVE - Working to add value through e-information, project which emphasised the information needs of clinicians. The 2005 Health Information Strategy for New Zealand[10] also makes the need for Consumers and their advocates to have access to information to make informed choices about health and independence a top priority.

New Zealand has the benefit of a lower profile of developments than the UK has. This meant that there was not the high level of critical news stories to influence nurses' reactions to IT developments and therefore there was clearly the potential for UK nursing to benefit from the experience of New Zealand nurses.

The visit

Visits were made to specialist informatics groups, educational institutions and Health Board managed hospitals. These were on both New Zealands's North and South islands. In the time available for the visit it was not possible, nor was it the aim of the travel scholarship, to construct a rigorous structured research project. Several organisations hosted me for periods varying from a day to a week, and each visit reflected the local expertise and circumstances and therefore each was a unique experience. I am grateful to each for the programme that they put together for me. Because of the variety of locations visited the study tour was approached as 'conversations with nurses', rather than as a structured research project, although the underpinning ethical considerations were based on normal research ethics. The aim was to explore nurses' experiences and listen to their views and suggestions.

All participants were given a brief overview of the nature and purpose of the study tour, and permission was sought to take notes of conversations and in meetings. Anonymity was promised unless specific exceptions were agreed, therefore generally neither the location of the information gathered nor the name of participants has been identified.

The people who participated in the conversations fell into two main groups.

- Academics & specialists in health and nursing informatics
- General Nurses in practice.

Information from nurses in practice

I talked with approximately 50 qualified nurses about their computer use and wishes. Most qualified nurses were spoken to on a one to one basis, although small groups were used e.g. when speaking to staff in shared offices. Nurses came from a wide rage of disciplines including community care, intensive care, day care, mental health, children's, acute medical and surgical areas. Nurses spoken to split into 2 distinct groups. Senior & specialist nurses (clinical specialists, nurse educators and clinical nurse leaders) and general nurses. I also met several larger groups of nurses in the invited lectures that I gave.

Initial questions were similar for each group, and conversations started with my asking about the nurse's experience of using computers, and the support they had received or felt that they needed. The conversations then developed according to the experiences of the nurses. Nurses generally mentioned using a combination of patient record systems and care planning systems. Fewer staff mentioned using computers to access evidence based information, or information for their patients.

One group of staff spoken to were educators in practice. These were employed by hospitals and did not have formalised links with higher education. This group were quite anti-computer. One thought that they belonged only in high dependency areas, but not in general wards. Several had the view that computers took nurses away from patients. On behalf of the staff they supported they also made the point that many nurses had poor IT skills, found online learning difficult, and that they had neither the computer access, nor the protected time, to use it.

The comments from nurses covered three main areas.

Systems in use

Often nurses did not feel that the systems in use meet their needs. One very heavily criticised problem was that there were often duplicate systems in use (especially paper and computerised systems) as systems did not have the ability to talk to each other. This was felt to be especially true for systems in different organisations (e.g. hospital and com-

munity) which made transferring patients time consuming. The need to enter the same data into different systems, or more than once, was also criticised.

In one location nurses had the choice of using a computerised patient care planning system, or a hard copy one. Most nurses, especially those who liked the computerised system, complained that they never knew where to look for information as not all the computerised records were printed out and put in the patient's file.

Some nurses, especially those who were uncertain about using computers or who found them complicated, felt that a familiar 'look and feel' would help them, the example given was if the navigation was similar to that in Microsoft programmes, as the nurse was familiar with them, but found the interface of professional systems to be cumbersome and difficult to learn.

Email was sometimes seen as a poor way of communicating; some general nurses felt that senior nurses expected them to use it too much. Many nurses suffered from a common problem for users of email systems, that of receiving too many irrelevant emails which had been copied unnecessarily to large numbers of people.

Nurses wanted systems that would save them time, eg send orders to pharmacy when a prescription is written, or identify equipment necessary for a care pathway and place the order automatically. Some nurses, especially those who were uncertain about using computers or who found them complicated, felt that a familiar 'look and feel' would help them.

Having to remember several passwords, all with a different length and expiry date was seen as a big problem, and as leading to writing them down or sharing other peoples logins rather than people logging out after use.

Equipment

In order to use computerised systems nurses wanted greater availability of equipment. Senior nurses generally had access to their own computers, however ward staff had to share, and in some locations nurses did not consider that they could access computers when they needed to. Nurses did not generally want more desk based systems. Computers that could be used at the bedside, or wherever else patients were, were in demand.

Skills

Some nurses felt that they did not have the skills, or the confidence, to use systems. This did not generally apply to senior nurses, or to nurses who had completed their training in the past couple of years.

Most nurses felt that more support needed to be provided. Training was available, and most nurses who mentioned skills did appear to know how to access training or the helpdesk in their organisation. The support available however did not meet the nurses' requirements, as it was not available at the point of need, or delivered by other nurses.

Nurses who considered themselves to have poor skills did not generally feel comfortable with worksheets or other written guidance. Computer based training materials, especially if nurses had to use them in their own time, were also unpopular.

Most nurses who mentioned the need for support wanted something available that fitted with their work patterns (e.g. on the ward) and that addressed immediate problems when they arose.

Information from health/nursing informatics experts

Nursing Informatics experts agreed that within New Zealand, as with England, nurses had a wide range of IT skills and informatics knowledge, but that many nurses lacked the skills and knowledge to fully engage with the informatics agenda.

Unsurprisingly the majority of experts saw the use of computerised systems as being beneficial to nursing. They were more aware of the potential for quality improvement, for example systems being used to provide reports on tests that were ordered but not carried out so nurses can check if anything is outstanding and what action is needed (although this was also seen as being a new nursing task that took nursing time). There was a strong view that new systems should promote change (quality enhancement) not just computerise the current way of doing things.

This group could see also scope for a range of improvements to the available computer systems. Practicality was high on the list. Equipment was seen as needing to be small but lightweight to make it portable around a ward. Wireless networking was considered a good way of enabling portability. Speed of the equipment and programmes was also important, the systems needing to be able to work at the same speed as the nurse.

The need for systems to support communication across the health sector was considered essential for sharing information about patients and to ensure that notes etc. would always be available when and where needed.

Nursing informatics experts felt strongly that managers also needed to be knowledgeable about IT. Support from the top was considered to be particularly important for the work culture to be one where computer use is seen as being a part of care giving.

Decision support systems were wanted, but with the proviso that they must be linked to up-to-date evidence, and have the ability to be updated rapidly.

What nurses want

Generally nurses wanted computer systems that would make their lives easier. This included saving them time by automating tasks such as stock ordering. One essential element to meet this need was that systems need to talk to each other so that information from all points of care were included, and that data entry was minimised.

Offering good quality decision support and access to evidence based information was not as high a priority for many practice based nurses, although it was considered a higher priority by the nursing informatics experts. Those who did want it, wanted it to be available when and where they needed it, and for it to regularly updated.

A very important requirement was that computers needed to be available wherever the nurse and patient were. Wireless networks and portable computers were suggested as the best way of meeting this need. Computers at workstations were not popular for several reasons, including the pressures to find a free computer at the end of shift to update records. Point of care data entry not only meets nurses' requirements but will also contribute to improved patient care by allowing records to be updated contemporaneously when memories are fresh, a great improvement on handwritten notes being scribbled and kept until the nurse can get back to a free computer.

Passwords were a big problem for many nurses. Each individual system often requires its own login information, each renewable on a different cycle. Nurses were almost unanimous in wanted this simplified. The ideal solution was seen as each nurse having just one password that would give entry to all systems.

Support was a frequently mentioned need. Nurses wanted help available that fitted with their work patterns (e.g. on the ward) and that addressed immediate problems when they arose. Although training on systems was seen as important the need for ongoing support was also identified. Most nurses did not want this through manuals or computer aided learning programmes that they had to use in their own time. Nor was the ability to contact a helpdesk that would get back to them, often a couple of days after their enquiry, seen as meeting their needs. The most popular support method mentioned was for a specialist nurse to be available to come and give one to one help when and where problems were encountered. A nurse was requested rather than an IT person because there was a feeling that a nurse would understand the context of the situation, and what the nurse was trying to do and how they needed to do it. Nurses, especially those who lacked confidence, did not feel that IT specialists 'talked their language' or saw their problems in the same way that they did. This supports a study carried out in 2003 with rural New Zealand health professionals, including practice nurses, by Janes at al [11]. This study found that many respondents reported having poor IT skills. These professionals reported that learning methods that included some social interaction were more popular than computer based learning.

The four biggest barriers to the use of computers that were identified were:

- The co-existence of paper based systems, meaning that nurses didn't have to engage with the computerised systems. This was seen as leading to computerised systems being incomplete and therefore promoting the use of paper-based systems.
- Systems being slow and not user friendly so that using the computer took longer than doing the same task did (or had done) in a paper-based system. Linked with this was a distrust of computers with the fear that they would increase workload by making tasks that were previously done by administrative staff part of the nurse's workload.
- Lack of support when and where it was needed. Nurses did not see waiting for help as being acceptable when a problem was stopping them doing work that they needed to do.
- Computers not being available where and when they were needed. A culture of using computers not being seen as being as important as giving patient case was often mentioned. Nurses keen to engage would like to have computers available at the bedside (or consultation) so that they become part of care giving rather than part of a separate administrative workload.

Nurses tended to focus more on using systems than they did on accessing information to support care. Part of this may just be a lack of awareness of what is available, but part is also likely to be attributable to a culture where spending time on a computer is not seen as being as valuable an activity as spending time with a patient, irrespective of what is actually being done in either case.

A thematic analysis of comments about attitudes to computers produced three distinct groups:

The engagers

Nurses who used computers quite extensively. As well as using the systems that they were required to use they also mentioned using computers to access research and library resources to support evidenced based care. One nurse from the UK who fell into this group felt that New Zealand systems offered more flexibility than the systems experienced in the UK. Another commented that she wouldn't want to see nursing without good computer systems to support it. This group were more willing to tolerate imperfect systems and to see ways that they could be improved.

The worried willing.

Nurses who were willing, but felt that they lacked the skills to use systems confidently. One nurse, who was finding that patients and families were using the Internet for information felt that they expected her to be competent as well. She considered that education programmes should be including this so that newly qualified nurses had these skills from the outset, and that programmes should be available for qualified staff to catch up. This group struggled with imperfect systems and wanted access to help and support that met their needs.

The resisters

Nurses who did not want to use computers. The comments from nurses in this group included; that paperwork was easier before computers, and that with new computerised systems nurses were having to do work that ward clerks did previously. These nurses also tended to see computers as taking nurses away from patient care. This group felt that poor systems justified their not using them.

Before my visit to New Zealand a project was undertaken with an English hospital looking at nurses use of the hospital intranet. Nurses who participated in that study expressed very similar comments to the New Zealand nurses.

Conclusions

Many of the essential elements of systems identified by New Zealand nurses are already key elements of the UK National Programme for IT. These include the 'do once and share' concept, and having one integrated health record, available to all healthcare providers. The National Library for Health is also part of the National Programme, giving nurses access to reliable up to date information to support research and evidence based care.

There are however aspects of implementation that may not currently be present. Computing at the point of care was an important requirement. Achieving this may contribute to nurses seeing computers as being part of nursing care rather than being seen as a task that takes them away from patients.

Nurses who are comfortable working with computers appear to be much more tolerant of failures in the systems and most importantly want to develop systems that meet their needs. The 'willing worried' nurses who feel that they lack skills but are willing to engage want support that meets their needs. If this is met there is clearly the potential to shift these nurses into 'engagers'. There is also a risk that if their needs are not addressed they will join the resisters.

Whatever the resisters may wish, computers are not going to disappear from healthcare, rather their use is going to increase. Changing the views of these nurses is a challenge for staff development. Including informatics targets in appraisal will be important to achieve this. The aim however must be to stop nurses starting their careers as resisters and seek to ensure, through effective pre-registration education programmes, that newly qualified nurses enter practice as engagers.

Acknowledgments

This work was supported by a Florence Nightingale Foundation Travel Scholarship, sponsored by the English Department of Health.

References

[1] NHS Executive. (1998). Information for Health. An Information Strategy for the Modern NHS 1998 – 2005. London.

[2] Bond, C, S. 2003. Nursing Informatics for All. E-health for All, 8th Nursing Informatics Congress. Rio De Janeiro

[3] Bond, C, S. 2004. Surfing or Drowning: Students Nurses Internet Skills. Nurse Education Today. 24(3) pp169-173

[4] Griffiths, P. & Riddington, L. 2001. Nurses' use of computer databases. *Health Information and Libraries Journal.* 18, pp2-9

[5] Davidson, L. 2005. The Magic Glue. *e-health insider*. online at http://www.e-health-insider.com/comment_and_analysis/index.cfm?ID=93 [Accessed 1 Aug 2005]

[6] National Audit Office, 2006. Department of Health. The National Programme for IT in the NHS. Report by the Comptroller and Auditor General. HC1173 Session 2005-2006; 16 June 2006. HMSO.

[7] NHSIA. 2001. Health Informatics Competency Profiles for the NHS. Winchester: NHS Information Authority, Ways of Working With Information programme.

[8] Murphy, J. Stramer, K. Clamp, S. Davis, S. Grubb, P. & Gosland, J. 2002. Health Informatics Education for Healthcare Professionals. London: RHIED / Department of Health.

[9] Murphy, J. Stramer, K. Clamp, S. Grubb, P. Gosland, J. & Davis, S. 2004 Health informatics education for clinicians and managers — What's holding up progress? International Journal of Medical Informatics. 73, pp205-213.

[10] New Zealand Health Information Strategy Steering Committee. 2005. Health Information Strategy for New Zealand. Wellington: New Zealand Ministry of Health.

[11] Janes R, Arroll B, Buetow S, Coster G, McCormick R, Hague I. Rural New Zealand health professionals' perceived barriers to greater use of the internet for learning. Rural and Remote Health 5 (online), 2005: 436. Available from: http://www.rrh.org.au/articles/subviewnew.asp?ArticleID=436 [accessed 2/3/2007]

Address for correspondence

Dr Carol S Bond. IHCS, Bournemouth University. B103, Bournemouth House, Christchurch Road, Bournemouth, England. BH1 3LH. Telephone +44 (0) 1202 504356. Email cbond@bournemouth.

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

The Need for Global Certification in the Field of Health Informatics: Some Ethical Issues

Eike-Henner W. Kluge[a]

[a] *Department of Philosophy, University of Victoria, Canada*

Abstract

In the past, the training of health information professionals (HIPs) has focussed almost exclusively on technical matters, the concerns of software developers and purveyors have essentially centred on security and functionality, and health care providers have mainly worried about costs and efficiency. This paper outlines some ethical threats that are ignored by such a purely technical focus and argues that because of the increasing globalization of health care delivery through e-Health, and because of the international threats to confidentiality posed by legislation such as the US Patriot Act, the health informatics community should pursue a project of global certification for HIPs that includes information ethics as an integral component. It also argues that a corresponding certification process for health care institutions and software developers should be initiated.

Keywords:

certification, informatics, ethics

Introduction

Health care is in the process of evolving from a paper-based to an electronically-based technology. While the speed and penetration of this evolution varies from country to country [1] it has already begun to assume international dimensions with the advent of e-Health, and it is merely a matter of time before it becomes a global reality. As in all cases when a new technology is applied to an established modality, this change has not been without concerns and challenges. In the beginning, most concerns had been technical in nature and had focussed on electronic techniques of data collection, storage, manipulation and communication as well as on reliability, quality control, accessibility and usability [2]. More recently, the portability and integration of electronic health records (EHRs) have become of issues of major concern [3], and the standardization of their architectures and the interoperability of networks have become the focus of careful consideration [4]. Similarly, the standardization of nomenclatures has begun to receive considerable attention, and syntax and semantics have been the subjects of extensive investigations [5-7].

Nor has the debate focussed solely on technical issues. From the very beginning it was clear that, given the extremely personal and sensitive nature of patient health data, ethical and legal issues would figure prominently in these developments and that issues such as consent to data collection, control and secondary use of the collected data, as well as privacy, security etc. would have to be treated not merely as add-ons to technical development but as integral factors in the development and deployment of the technology itself. It was also realized that technological security and privacy measures are only as good as the conduct of the professionals who are involved in their application. Consequently there was a move to develop ethical guidelines for health information professionals, and in October of 2002, the General Assembly of the International Medical Informatics Association approved a Code of Ethics for Health Information Professionals [8] that was intended to function as a template for its member organizations.

Ethics and the professions

Health information professionals

It would therefore seem reasonable to expect that the education of a conscientious, well-trained and literate health information professional (HIP) should include not merely mention of, but also training in, the relevant ethical issues, and should instil a familiarity with the fundamental ethical principles that should govern the collection, communication and handling of EHRs. It is therefore surprising that with the exception of one statement by the Council of Ministers of the European Union [9], neither the professional literature nor the premier discussion of the subject - the Recommendations of WG1 of IMIA on Education in Health and Medical Informatics [10] - stresses that ethics education and certification should form a standard part of the training of HIPs. A literate person is someone who is well educated and familiar with the relevant topics and the literature [11]. One would therefore have thought that a properly educated and literate HIP could be expected to demonstrate competence in information ethics.

However, what is at issue is more than merely a matter of literacy. What is at issue is the very status of health informatics as a profession. Professions are characterized by the fact that their members usually have undergone an intense period of post-secondary education in a specialized subject, provide a socially important service and belong to an activity-related organization that has a code of ethics. Formal or statutory recognition of a profession usually confers

the right to set licensing and/or certification requirements, the right of self-government and the duty to enforce appropriate standards [12-14]. In practical terms, the formal or statutory recognition of a profession may be delayed but the profession may have *de facto* recognition as a profession simply because of the nature of what it does and because of the fiduciary or quasi-fiduciary nature of its relationship towards clients and towards society. Nursing is here a good example. Nursing was not formally recognized as a profession until fairly recently and yet, for all that, nurses were considered professionals because of the fiduciary nature of the nurse-patient relationship and the dependence of patients on nursing care. Health informatics meets the above criteria and therefore, *de facto,* is a profession [15]. Therefore HIPs may be held to professional standards, the lack of formal recognition of health informatics as a profession notwithstanding.

This is especially important at the present time because EHRs and e-Health present challenges of privacy, security, communication, handling, manipulation and inter-jurisdictional operability that did not exist with purely paper-based records. Moreover, given the increasingly global mobility of HIPs as well as the diversity of legal data-handling provisions and regulations that they encounter as they move from jurisdiction to jurisdiction, the only constant that defines appropriate professional behaviour remains the IMIA Code of Ethics.

In more specific terms, the appropriate functioning of contemporary health care systems depends not merely only smooth and efficient technological operation but also on appropriate behaviour on part of the professionals who make these systems technically possible. The successful deployment of e-Health, therefore, assumes not merely a secure and interoperable technical infrastructure but also an integrated and comparable set of professional standards for the informatics professionals who are responsible for its development and operation, where this set of standards is not confined to considerations of technological competence but also includes an ethical component that measures the competence of the professionals in identifying and dealing with ethical issues that arise in the handling and communication of EHRs. This, in turn, necessitates some means not merely for assessing and certifying these qualifications but also for enforcing the relevant standards and, if necessary, for administering disciplinary measures in a transparent and globally consistent manner.

As yet, only limited steps have been taken in this direction even in the domain of technology [9]. There is no international set of technological proficiency standards and no internationally recognized certification process for HIPs. Matters stand even worse with respect to ethics. Thus, not all national informatics associations have formally adopted the IMIA Code of Ethics as a template for their own statement of ethics. This holds true even of some organizations who are members of IMIA. However, unless and until globally integrated professional codes of ethics have been adopted and are included in a standardized education that leads to formal certification, the actual delivery of health care in the electronic age will remain threatened by this lacuna in the education of the professionals on whom the delivery of that care depends. The inability to identify and deal with ethical issues sets the stage for technologically competent but ethically inappropriate behaviour that easily leads to legal confrontation and to litigation on the national as well as on the global level. Explicit and standardized training and certification would minimize such a risk. In an ideal world, such standards and certification would ultimately be referable to an independent body that would function in association with the WHO, and that would rely on the expertise of professional organizations such as IMIA. It remains to be seen, whether the profession will adopt such a stratagem.

Health care professionals

As an addendum, it should be noted that this plea for the inclusion of ethics in training and certification extends beyond HIPs. e-Health and telemedicine cannot function without the participation of health care professionals (HCPs). Consequently it would seem not merely appropriate but also prudent to have informatic qualification and certification standards for HCPs who engage in e-Health and telemedicine, and to include an information-ethics component in their training. Given that currently there is no mechanism for ensuring internationally comparable standards in medical qualification, this may pose an even greater challenge than the medical certification of HIPs. However, absent international standards for health care professionals in informatics, e-Health will never be as good - or as secure - as it could be. IMIA and its affiliate national organizations may therefore wish to explore the issue with the corresponding organizations in the field of health care.

EHR system purveyors, institutions and the ethics of certification

The discussion so far has concentrated on professionals; however, the concept of certification includes more than proficiency: It also applies to the systems that are certified and to the protocols that are involved in their operation. Consequently, EHR system certification is also relevant for purveyors of EHR systems and for the institutions that operate them. In each case, there arise serious ethical issues, and in each case the relevant certification should reference these matters.

Purveyors

The ethical issues that arise for purveyors centre not merely in functional efficiency and reliability - issues that traditionally have been dealt with under the legal rubric of product liability - but also in security from unauthorised intrusions into the system. Most EHRs run on existing operating systems, and these have millions of lines of code. It is therefore a virtual certainty that they include backdoors that potentially allow a compromise of system security. However, purchasers such as health care institutions (as well as individual providers) who run such systems when using EHRs owe an obligation of confidentiality to the subjects of these EHRs. The health care providers, therefore, purchase the relevant operating systems not merely with the expectation that the systems will

function as efficiently as advertised but also that they will meet relevant and applicable privacy and security requirements. Consequently, from an ethical perspective, purveyors of operating systems have two choices: *either* develop entire operating systems from the ground up under strict quality control that allows them to certify that every element of their new systems is not only fit-for-the-purpose but also secure; *or* develop a standardized way of evaluating the vulnerability of their systems and establish an independent body that will be able to certify the vulnerability level of the existing systems. Unless at least one of these alternatives is adopted, purveyors of operating systems will continue to rely on the expertise of their lawyers to deal with the issues that arise, and will leave to the legal process what is essentially a matter of business ethics. With due alteration of detail, similar comments apply to developers of EHRs.

In this connection, one further issue needs mention. Some countries - notably the United States - have passed legislation that allows their security and intelligence agencies to demand that companies that are incorporated in their jurisdictions, and that the subsidiaries under the control of these corporations, provide the agencies with any and all items – the phrase used is "any tangible things (including books, records, papers, documents, and other items) ..." - if, in the estimation of the agencies, access to such items will assist them in evaluating and pursuing what the agencies believe to be are security threats [16].

The source codes for EHR software and for the operating systems on which they run are clearly included under the rubric of "other items." It is therefore ethically incumbent on any certification system that might be developed to include information as to whether the developers have complied with a request that has been made under such legislation.

However, this presents a serious ethical dilemma if the enabling security legislation follows the example of the *USA Patriot Act.* That *Act*, under §215 (2)(d), stipulates that "No person shall disclose to any other person (other than those persons necessary to produce the tangible things under this section) that the Federal Bureau of Investigation has sought or obtained tangible things under this section." The dilemma for developers and purveyors, then, is the following: If they do what is required by such a law, they will *de facto* mislead their purchasers about the security of their systems by withholding the fact that their systems are potentially penetrable by the security agencies of the country in which the purveyors or developers of the system are incorporated. On the other hand, if they do what is ethically appropriate and reveal that the codes have been made available to these security agencies, they will be breaking the law [17]. Under the circumstances, it may be appropriate for system developers and purveyors to consider whether ethical business practice allows them to be incorporated in a locality such as the United States, which could legally require them to falsify any such certification. Moreover, it may be appropriate for IMIA to examine whether it should develop an appropriate policy in this regard. Neither the IMIA Code of Ethics nor any other IMIA policies or declarations currently address this issue, whether that be in a direct or even an indirect fashion.

Institutions

In a world that is haunted by the spectres of efficiency and financial bottom lines, buying a service from the cheapest possible provider is considered good institutional practice. Information service outsourcing is therefore proceeding apace, even in the field of health care [18]. However, health care institutions owe an obligation of confidentiality to their patients, which means that they have a duty to ensure that their operating systems and the EHRs that are under their control or in their possession are secure. Consequently they have an ethical obligation to purchase only from purveyors who can provide appropriate security-certified software. More specifically, it means that they have an obligation not to purchase operating systems, EHR software or related programmes that are security-uncertifiable because the purveyors are subject to provisions such as those that are found in the *US Patriot Act*. In the event that they cannot purchase such software, they have an obligation to inform their patients that they cannot guarantee the confidentiality that is normally expected in the field of health care.

Moreover, institutions themselves are subject to a certification process that determines whether they meet appropriate institutional operating standards. This means that institutional certification processes should include a rubric that identifies whether an institution meets the informatics privacy and security standards that have been traditional in health care. It may be that such a certification process would have to examine whether the institutions should purchase their systems from purveyors who are subject to legislation such as the *USA Patriot Act,* and to rank the relevant institutions accordingly. As for those health care providers who are themselves subject to provisions such as he *USA Patriot Act* - and this would include institutions that are currently certified under JCAHO (Joint Commission on Accreditation of Healthcare Organizations) guidelines - these clearly have a legally mandated duty to provide EHRs when requested to do so by a duly empowered authority. However, that does not mean that they must abandon their traditional ethical duty of confidentiality towards their patients or that they must violate the duty of openness as stipulated in the IMIA Code of Ethics [8]. They can both fulfil their ethical duties and meet the requirements of the law by making it known to all incoming and/or prospective patients that they cannot guarantee confidentiality because the records in their possession will be subject to the provisions such as those contained in the *USA Patriot Act*.

Conclusion

The delivery of health care has become a multi-billion dollar global industry, and the rationalisation of health care delivery through the use of EHRs and e-Health systems is proceeding apace. It has generally been recognized that the quality of health care is best served through the certification of health care institutions. To date, this certification

has proceeded independently of informatic considerations. The time has come for this to change and for informatic parameters to be integral to any such certification. The time has also come for the concept of certification to be extended to include HIPs, since these are instrumental to the functioning of EHR-based health care systems and of e-Health. The preceding has identified some ethical and legal issues in this regard, and has argued that an independent and global certification process should be developed. It may be time for organizations such as IMIA to exert every effort in this direction, and to make strategic alliances with other bodies in order to achieve this end.

References

[1] Schoen C, Osborn R, Huynh PT, Doty M, Peugh J, Zapert K. On The Front Lines of Care: Primary Care Doctors' Office Systems, Experiences, and Views in Seven Countries. Health Affairs 2006:25:6: w555–w571.

[2] SEISMED Consortium. Data Security for Health Care vol. II, Technical Guidelines. Amsterdam: IOS Press, 1996.

[3] Marsh A, Grandinetti L, Kaurame T, eds. Advanced Infrastructures for Future Healthcare. Amsterdam: IOS Press, 2000.

[4] Blobel B, Pharow P eds. Advanced Health Telematics and Telemedicine: The Magdeburg Expert Summit Textbook. Amsterdam: IOS Press, 2003.

[5] Rooksby J and Kay S. Clinical Narrative and Clinical Organization: Properties of Radiology Reports, Proceedings of MedInfo 2001; 680-684.

[6] Peleg M, Boxwala A, Tu S, Greenes RA, Shortliffe EH, and Patel VL. Handling Expressiveness and Comprehensibility Requirements in GLIF3. In: Patel VL, Rogers R and Haux R, eds. Proceedings of the 10th World Congress on Medical Informatics. Amsterdam: IOS Press, 2001; pp.241-245.

[7] Kushniruk AW and Patel VL. Cognitive approaches to the evaluation of healthcare information systems. In: Anderson JG and Aydin C, eds. Evaluating the organizational impact of healthcare information systems, 2nd ed. New York: Springer-Verlag, 2005; pp. 144-173

[8] IMIA. Code of Ethics for Health Information Professionals, accessed November 20, 2006 at http://www.imia.org/code_of_ethics.html.

[9] Council of Europe Committee of Ministers. Recommendations No. R (90) 21 of the Committee of Ministers to Member States on Training Strategies for Health Information Systems*:* Recommendation on training strategies for health information systems. http://www.coe.int/t/e/social_cohesion/health/recommendations/Rec(1990)21.asp

[10] IMIA. Recommendations on Education in Health and Medical Informatics. Meth Inf Med 2000: 39: 267-277.

[11] The Shorter Oxford English Dictionary. Oxford: Clarendon Press, 1959.

[12] Bayles M. Professional Ethics. Englewood Cliffs, N.J.: Prentice Hall, 1985.

[13] Freidson E. Professional Powers: a Study of the Institutionalization of Formal Knowledge. Chicago: Chicago University Press, 1986.

[14] Parsons T. Essays in Sociological Theory. Glencoe, Ill.: The Free Press, 1954.

[15] Kluge E-HW. Professional ethics as a basis of legal control of health care information. Int J Biomed Comp 1996: 43**:** 33-37.

[16] Uniting and Strengthening America by Providing Appropriate Tools Required to Intercept and Obstruct Terrorism Act. Short title: USA Patriot Act. (HR 3162) §215.

[17] Kluge E-HW. Ehealth, the USA Patriot Act and other hurdles: the black lining on the silver cloud. Br J Healthcare Comput Info Manage 2006: 23(9): 16–23.

[18] Vijaya K. Teleradiology Solutions: Taking expertise to hospitals in US. Express Health Care Management, Issue dtd. 16th to 29th February 2004, accessed September 25, 2006 at http://www.expresshealthcaremgmt.com/20040229/innews07.shtml.

Address for correspondence

Eike-Henner W. Kluge, PhD
Professor, Dept. of Philosophy
University of Victoria
Victoria, BC, Canada, V8W 3P4
e-mail: ekluge@uvic.ca

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Using Fieldwork in Analyzing Ethical Issues Related to IT in Health Care

Ellen Balka[a], Christine Reidl, Ina Wagner[b]

[a] Simon Fraser University and Vancouver Coastal Health, Canada [b] Institute of Technology Assessment and Design, Vienna University of Technology, Austria

Abstract

This paper describes how an understanding of everyday conflicts that have ethical implications – what we call 'situated ethics'— can be explored through ethnographic field techniques in healthcare settings. Our approach to ethics is followed by findings from two ethnographic case studies focussing on issues arising as information technologies such as electronic patient records and automatic drug dispensing machines are introduced into varied health sector workplaces. By close and careful observation of these technologies in use and by incorporating narrative accounts from different perspectives the complexity and entangledness of real life occurrences are revealed. Our data suggest that several types of ethical issues (e.g., issues related to intellectual property, literacy, standardization, transparency, work ethics, and equitable allocation of resources) can be identified through fieldwork, and can have an impact on identification of everyday ethics in healthcare.

Keywords:

ethics; ethics, institutional; principle-based ethics, ethics, professional; narration; work; ethnography

Introduction

Our interest in this paper is in understanding the situated, everyday conflicts that have ethical implications – what we call 'situated ethics'. We developed an approach to studying 'situated ethics' as part of the Canadian Action for Health project which investigates the role of the internet, electronic patient records, computerized information systems and other forms of information technology in the health sector.

Our aim of studying 'situated ethics' is to bring everyday conflicts about IT in health care and their moral dimension to the fore. Ethical issues reveal themselves in the complex dynamics which unfold in everyday situations between actors, such as doctor and patient, nurse and management, clinical personnel and technical support/vendor. Their situated character lends itself to the narrative form – storytelling. The situated, everyday conflicts are connected to larger ethical issues – they are instantiations of such larger issues. They are often also connected to legal issues. We argue that understanding the context and the mundane details of situations in which ethical issues arise, and grounding them in people's everyday experiences may make it easier to address, and deal with ethical issues.

'Situated ethics' is about ethics as an integral part of everyday action. It treats ethical problems as researchable through social science methods. It claims neither to offer 'solutions' to the identified issues nor to be a substitute for a professional case deliberation but it can prepare the grounds for such a case deliberation. Hence one of our aims is also to explain the difference between the work of professional ethicists who analyze ethical issues, eventually helping people to come to a moral judgment of a situation and plan appropriate action; and the work of social scientists, who through qualitative research methods -observation and interviews -unravel ethically problematic aspects of everyday life/work situations.

Here we describe how to study the ethics of everyday practices, using (ethnographic) fieldwork. We make a case for 'narrative ethics'. We use vignettes based on fieldwork in two research projects on IT in health care to explain our approach to studying ethical issues and we reflect on how our approach may enrich the discourse about ethics and IT in health care.

How to study ethical issues

Researchers have for some time criticized the principled philosophical approach to ethics-in-practice. One major point of objection is that the principled approach assumes common moral intuitions, failing to recognize the existence of multiple cultural and religious traditions. For example, in her study of ethical decision making in cases of terminal diagnosis Turner [1] introduces the concept of 'local moral worlds' as meaningful for understanding different modes of moral reasoning. Another point of criticism is that clinical/medical ethics has been couched in general and abstract terms and focuses on dilemma-type issues. Guillemin and Gillam [2] argue that when it comes to looking at day-to-day ethical issues, these can rarely be phrased in the form of a dilemma. Using an example of a woman informant disclosing that her husband has been sexually abusing her daughter, [2] argue this can be interpreted as a classical ethical dilemma of whether to breach confidentiality or prevent harm. However, there are also more immediate ethical concerns, like whether or not to take up what the woman disclosed, in which words, what tone, whether to switch off the tape recorder, abandon the

interview, and offer help. These practical but highly relevant ethical concerns cannot be framed as dilemmas. Guillemin and Gillam [2] talk about 'ethically important moments', moments of response where wrong can be done. They focus on sensitizing people to ethical tensions, and through this enable ethical action, rather than prescribe specific types of responses.

In her study of different types of medical decision-making, Hall [3] proposed an approach, based on the work of Johnson, who sees the basis for moral sensitivity in the 'moral imagination'. Her notion of narrative ethics puts imagination and interpretation at the centre of ethical decision-making. Narrative ethics includes constructing and telling one's own story and comprehending the story of the other. This may include decisions such as: What is my story? What is important to relate? What is the best way to express what I consider to be the ethically most relevant material? Narratively crafting our understanding allows us to enter a decision-making process which includes contextual elements. The imaginative capacity is revealed in the images, metaphors, and symbolism the narrator uses, and in the small details s/he fills in. Hall [3] stresses that narrative ethics and arguing on the basis of rules and principles are not mutually exclusive but can complement each other.

The notion of 'narrative ethics' resonates well with methods of ethical case deliberation that see moral intuitions embedded in narratives. Participants are encouraged to bring forward/argue through stories and organizers of a deliberation process are required to provide suitable case descriptions. Throughout the literature we find the term 'vignette' for such descriptions. Vignettes are short, narrative accounts of a 'case' from the field that illustrates one or more ethical-legal issues. They are constructed on the basis of extensive practical experience in the field or of fieldwork material produced through research. They may be real cases or hypothetical cases based on experience. When simulated, a vignette can be constructed to look at the same situation from different perspectives or contain one manipulated variable (e.g. the gender or ethnicity of one of the people in the scenario). When based on field work, a vignette can describe a real-life dilemma as it occurred. Advantages include that all contextual information required to understand and analyze the issues are at hand; and that the situation described is sufficiently complex and enriched with detail.

Ideally, vignettes should be written so that readers understand the context (of people, tasks, IT support, organization, cultured practices, history, etc.) in which ethical issues arise. Vignettes should have a clear focus on describing the issues. Video material, cartoons and other techniques have been used for producing lively accounts. A 'good' vignette is defined as an account that has "great hermeneutic power in its capacity to enhance our understanding of behaviour" [4, p. 83]. One characteristic is ambiguity, which leaves space for participants to define the situation in their own terms. Another important feature is narrativity, which introduces specificity and detail, helping to understand the context in which an ethical issue arises. Both encourage and enable reflexivity and discourse. Our interest is in using vignettes as part of qualitative research and in vignette writing based in fieldwork. This ensures that vignettes genuinely represent phenomena that occur in real world settings and that they reflect the complexity of an area being studied. Moreover, "the ethical decision-making in such a situation has a narrative context, in which competing moral stories containing strongly rhetorical elements present themselves" [4, p. 84].

The cases

We have worked with ethnographic fieldwork material for constructing vignettes, and we elicited vignettes from other project partners. The material we collected – 34 vignettes from 16 different projects – differs vastly. Vignettes can be rather short and 'typifying', (putting forward a case that may be encountered in other places) or they can be very specific and rich in social detail.

The two cases we selected for this paper are very different. The Canadian case is about an automatic drug dispensing machine (ADS) that had just been introduced in a hospital; hence, it is about a system in use, the problems users encountered and how these were handled. The Austrian case is about the introduction of an electronic patient record (EPR) in five clinics and deals with the introduction process, the problems that came up and how these were addressed.

Case 1: New automatic drug dispensing machines

In 2003, several units of a large hospital moved into a new building, and several new technologies were introduced, including ADS. The vignette contained here emerged from extensive observationally based fieldwork (see [5, 6] for more detail). These machines consist of cabinets with a keyboard and screen. A nurse logs on by keyboard, enters information about her patient, and selects the medications she administers to her patient from a computerized list. Once the medications are selected the drawer containing the medication opens, and the bin containing the desired medications is automatically unlocked.

When the ADS was introduced, nursing staff had difficulty locating the name of drugs they had to administer in the machine's pre-programmed drug inventory. Nurses using the ADS shortly after its introduction reported that the names of drugs included in the ADS' inventory list differed from the naming conventions in use prior to the introduction of the ADS. We observed several instances of nurses searching through pre-programmed drug names in the cabinet before finding the correct name of the drug they wished to administer. Another problem appeared to be that the nomenclature programmed into the ADS (generic drug names) differed significantly from the nomenclature that was in use previously on the hospital units (both generic and trade names).

Although the ADS accommodates naming conventions different from those programmed into it, the planned integration of systems both within one hospital and across hospitals necessitated use of a common drug nomenclature across multiple computing platforms. As a result of this

standardization, the convention which had been in use to include common brand names after the generic drug name on medication orders ended, leaving nursing staff to translate between doctor's orders (that typically use brand names such as lasix) and generic names (such as furosemide) used by the central pharmacy, and now programmed into several computer systems. The location of drugs within the ADS were also standardized.

Another issue related to the ADS concerns the processes put in place for inventory control and ADS discrepancy reporting. One of the main justifications used for the implementation of ADS has been inventory control. Drug inventories are controlled through the ADS in part through count back procedures, where staff count back, or verify that the locked storage drawer compartments that house the drugs contain the quantity of drugs indicated on the ADS screen. Each time medication is removed from the ADS, the user is prompted to verify that the quantity of the listed drug is stored in the machine. In instances where the quantity of medication in a drawer varies from the quantity listed, staff are prompted to address the discrepancy, by either choosing a pre-defined reason from a menu of options (e.g., discrepancy already documented in report; error in previous count back quantity; error in previous remaining quantity; medication not removed; or unexplainable loss–see report), or entering a user supplied reason for the discrepancy. Dealing with a discrepancy in the manner dictated by the ADS required staff to count the quantity of medications on hand – a time consuming process when a frequently used, well stocked medication (e.g., 200 tablets) is involved.

Case 2: Introducing an EPR into 5 oncology clinics

The 5 clinics participating in the EPR project form part of the Community Hospital Organization (CHO), the association of hospitals, nursing homes and geriatric centers run by the city of Vienna. The CHO is one of the biggest health service institutions in Europe with an average of 400.000 admissions per year and a staff of 32.000. Our description is based on interviews with different stakeholders, an in-depth study of paper-based and computer-supported documentation practices in three clinics (ONC1, ONC2 and ONC4), and participation in several user group meetings.

The decision to introduce such a system was a central-political one, initiated by one of the five clinics (ONC4) which has been successfully running an EPR for 12 years and had hoped to convince the CHO that their system should be implemented in the other clinics. The CHO in turn was interested in the project since it was in line with their strategy to replace the old hospital information system (HIS) and to generally look for unified solutions/products, since these seem easier to maintain, and hence more economical. The project had the support of 4 of the 5 clinic heads, who delegated the work of formulating a requirements document to interested physicians in their clinic, with the experienced ONC4 in a dominant role. It was decided that ONC1 and ONC2 should start implementing the system. Several user group meetings were conducted at ONC2 who seemed ready to work with the system.

Confronted with the system which was demonstrated by the vendor, physicians at ONC2 started an intense discussion about a variety of issues, none of which seemed as easy to resolve as the CHO had imagined. A major concern was the interface to ordering lab exams and viewing the results, which is currently done within HIS. This was immediately declared as technically very difficult, hence too costly, by both the vendor and the CHO. The next point was the ordering of chemotherapies, for which each of the hospitals has a different system run by their pharmacy. The question of who should be responsible for the process and who should be responsible for the maintenance of the drug catalogue – clinicians or the pharmacy – arose. There was also the question of interfacing here. Without interfaces, a nurse or physician, after having documented in XX (the new system) what should happen with the patient, a blood test, an X-ray, a nuclear-medical exam, etc., would have to log into another program, identify the patient a second time, and again type in the order. There was also concern about how to integrate orders issued by external consultants. Questions arose about protocols, which in chemotherapy seem to change quite often – how many should be supported? Who should do the work of updating them?

There were heated discussions about the different templates physicians need, from the anamnesis sheet (a long and a short version) and the discharge letter to different kinds of overview sheets (patient status, overview of activities). How should they be designed? Which data can/should be automatically inserted from the system? How much standardization should be designed into a template? What kind of overviews should be enabled? etc. It became clear that at what seemed the start of implementation a time-consuming process had started, which uncovered contextual richness and highly relevant detail that would require much more attention than both the CHO and the vendor were prepared to give.

Discussion

The vignettes presented here in a shortened version are fieldwork accounts, rich with social detail. They do not just address one particular ethical issue but tell a story that needs to be analyzed to identify areas of ambiguity and complexity. Our analysis follows what Steinkamp and Gordijn [7] describe as 'clinical pragmatism'. This protocol is based on a structural analogy between clinical judgment and the structure of ethical reflection. It requires 'contextual factors' to be taken into account in "concentric circles. Implying first and foremost the perspectives of the other professional groups as well as their contributions to care giving" [7, p. 237]. Narrative elements, representing moral intuitions and stories are to be included. Institutional policies have to be examined. In this mode, we use the case descriptions to identify conflicts that can be related to ethical issues such as transparency, standardization, work ethics, privacy and confidentiality, intellectual property, equitable allocation of resources, literacy, etc. Different stakeholders' perspectives on issues are considered and we describe how organizational as well as inter-organizational relations mediate these issues.

Identifying the ethical issues

For identifying ethical issues we use a framework which draws on several sources. Biomedical ethics [8] introduces a four-principles approach – autonomy, justice, beneficence, and non-maleficence. Related to virtue ethics, the ethics of care originally emerged as a feminist critique to traditional theorizing (see the pioneering work of Gilligan [9]). It focuses especially on personal relationships and character traits that are valued in them. It has been reformulated as an ethics of responsibility, which with respect to science and modern technology stresses issues of accountability and liability. Furthermore, given the many opportunities information and communication technologies offer, it is vital to take account of the need for privacy, respecting people's right to maintain boundaries, but also to preserve privacy, autonomy, confidentiality, and solitude. Other ethical issues connected to ICT are transparency – awareness of and the ability to understand IT systems and their implications – and literacy. This widening catalogue of principles or issues reflects the plurality of perspectives and the complexity and specificity of areas such as modern technology and health [10]. Here we discuss only a few of the categories of ethical issues identified. In the Canadian case these are:

Issues regarding intellectual property – The use of ADS will result in the collection of more sophisticated data about drug use by patients during hospital stays than has previously been available. Drug companies have been trying for years to gain access to British Columbia health data, as BC has one of the most comprehensive data sets in North America. Who owns the data that the ADS collects and consolidates? Should patients have to consent to having anonymized data collected during their hospital stay used at some future point for health research? Which health researchers should have data access?

Issues of literacy – Gaps in computer literacy among nursing staff became evident with the introduction of the ADS. Also, issues of literacy emerged again in relation to the nomenclature issue -pharmacy staff felt everyone (nurses, doctors and pharmacy staff) should use the generic names for drugs, because this constitutes better practice. However, as staff struggled when the brand names disappeared, differences in the drug literacy of pharmacy staff and other staff become evident, along with physicians' non-compliance with a policy to use only generic names when prescribing.

Issues of standardization – The problem of drug naming conventions raises issues of standardization. Standardization was adopted for good reasons, but has had unintended consequences. The generic nomenclature is not in common use among floor staff who must now translate doctor's orders (which are often written in trade names) into generic names. Staff may spend more time looking up drug names than they did in the past. Standardizing the location of drugs within the ADS was intended to make it easier for staff who move between floors to use the ADS. However, staff on some floors were frustrated by having to retrieve commonly used drugs from the bottom drawer. In both cases it is not clear if the benefits gained from standardization out weighed the costs.

The issues in the Austrian case were somewhat different:

Issues of transparency – As local IT experts where not involved from the project's start, misunderstandings arose among physicians concerning system requirements and interfaces. For example, it was not clear that the new system would not support easy retrieval of lab results. Also, local system experts were not aware of hardware requirements of the new system, with the consequence that these were not included in local budgets. Nurses had been invited to attend the user meetings but it remained unclear how they might be affected by the implementation of the new system. At the same time the roll out of another IT project in support of care documentation was taking place and its status was far more advanced. How should nurses decide which system will meet their requirements best?

Issues of literacy – Vendor and project management asked physicians to provide specifications of working routines and work flow to be supported by the new system, although they lacked the experience and skills required for writing a useful requirements document. It turned out that the way physicians specified their requirements lacked relevant social detail so that something they had thought was a key requirement – easy retrieval of lab results – was formulated in a rather general and vague way, for example not defining when and in which form lab results should be communicated. This led to misunderstandings when translated into technological solutions by IT experts. The 'medium' of participation – a traditional requirements document – turned out to be inadequate.

Issues of work ethics – The system has been designed exclusively from the physician's perspective although nurses will presumably have to work with it too – how can their perspective be included? There is also the issues of time physicians don't have the time to participate in the set up of the project. For reasons of cost control physicians' possibility to work overtime has been limited to 20 hours a month. Whether physicians get compensatory time off for their collaboration in the implementation process is up to the head of their department. Also, the new system requires attention to issues, such as complicated login procedures or regular maintenance and updating of drug catalogues and protocols for chemotherapies inside and across clinics that have not been accounted for.

Issue dealing with equitable allocation of resources – As soon as it was decided that the oncology departments would get a new documentation system all further investment into the completion or updating of any old systems in place was stopped. This has led to a continuous slow down in data processing in some of the involved clinics.

Stakeholder perspectives

In the simplest sense, the stakeholders in the Canadian case are the central hospital pharmacy (pharmacy management, those who re-stock the ADS), the information management department of the hospital (responsible for the units' connection to the hospital information infrastructure), nurses on the units, their managers, and the patients. Other stakeholders are people who might use ADS data – one compelling reason for introduction of sys-

tems such as ADS is to improve data availability for research. The ADS implementation also provides insights about ethical issues related to assumed actors and assumed use contexts, both of which become stakeholders in system implementation processes. The ADS was designed with a U.S. market in mind, replete with all the record keeping required to bill for 100% of all services rendered or medication dispensed. This was reflected in the machine interface. Although in Canada the reasons for introducing the system did not include the ability to track costs associated with each patient, the financial need to do so within the US market influenced software design and subsequently work practice in a Canadian hospital.

The Austrian case reveals a complex network of actors, with the CHO as central hospital organization, the vendor, the five clinics with their own interests, and future users of the system – physicians and nurses. The perspective of ONC4, one of the main players, is of special interest here. In the 1990s they developed their own documentation system. When attempts to get funding for necessary improvements failed, the idea to promote purchasing of a standardized documentation system for all oncology departments arose. Responding to the call for tender was seen as a way to get a complete system update and at the same time standardize the systems of the other big oncology departments, thereby improving the possibilities to exchange data and promote research. But their system did not win – it was too expensive. Although it had been continuously improved over the years and adapted to changing work practices, these efforts and advantages were not included in the rating of the system. A lack of attention to the effort and expertise of one actor contributed to jeopardizing the whole project.

Mediating (inter) organizational relations

In a previous paper [11] we discussed the fragmented character of complex organizations and interorganizational relationships as a source of problems in systems development. Looking at the ethical issues in the oncology case we can identify fragmentation of (inter) organizational relationships as one of the reasons for the problems encountered. Systems development takes place in several loosely coupled social arenas between heterogeneous actors -CHO, vendor, the different clinics with differing interests, different occupational groups in the clinics, and within different hospitals, vendor, and varied occupational groups in the ADS case. We can also see action in other arenas shaping the process: the CHO using its power position, curtailing the needs for updating old systems by stopping financing.

Looking at the ethical issues in the ADS case, we see that a lack of attention to work design contributed to some of the problems, particularly related to the timely process of counting back commonly used drugs. Problems associated with which drug nomenclature were programmed into the ADS (generic or generic and trade names) were heightened by (inter) organizational relationships, which on the one hand did not adequately take into account current work practices (e.g., of doctors who use trade names in writing prescriptions), and on the other hand were not clearly visible to staff who had to cope with the consequences of the use of only generic names were programmed into the new machines – a decision that was taken because of inter-organizational relationships, but was not visible to staff. This lack of transparency, combined with a lack of transparency associated with how the communication of drug count discrepancies were reported to the pharmacy with the ADS undermined medication safety and inventory control.

Conclusions

The 'politics of systems design' is not a new agenda (see e.g. [12]). Our notion of 'situated ethics' proposes to use descriptions of (work) situations involving technology, to look into the moral dimension of technologies at work – how they are introduced, how they affect people and the quality of their work. The cases we have presented are based on extensive fieldwork. They are by no means simple; they reflect the complexity and entangledness of real life occurrences. As we have tried to show, the problems emerge in the social detail which can only be uncovered through close and careful observation of the technologies in use and/or the process of introducing the technologies in complex settings.

References

[1] Turner L. Bioethics in a Multicultural World: Medicine and Morality in Pluralistic Settings. Health Care Analysis 2003: 11(2): 99-117.

[2] Guillemin M, Gillam L. Ethics, Reflexivity, and 'Ethically Important Moments' in Research. Qualitative Inquiry 2004: 10(2): 261-280.

[3] Hall K. Medical decision-Making: An Argument for Narrative and Metaphor. *Theoretical Medicine* 2002; 23: 55-73.

[4] Wilks T. The Use of Vignettes in Qualitative Research. Qualitative Social Work 2004: 3(1): 78-87.

[5] Balka E,. Kahnamoui N. Technology trouble? Talk to us! Findings from an ethnographic field study. In: Proceedings of the eighth conference on Participatory design: Artful integration: interweaving media, materials and practices - Vol 1. New York: ACM Press, 2004; 224-234.

[6] Balka E, Kahnamoui N, Nutland K. (in press). Who's in Charge of Patient Safety? Work Practice, Work Processes and Utopian Views of Automatic Drug Dispensing Systems. International Journal of Medical Informatics. In Press: Epub ahead of print

[7] Steinkamp N, Gordijn B. Ethical Case Deliberation on the Ward. A Comparison of Four Methods. Medicine, Health Care and Philosophy 2003: 6: 235-246.

[8] Beauchamp TL, Childress JF. Principles of Biomedical Ethics. 4th ed., Oxford: Oxford University Press, 1994.

[9] Gilligan C. In a Different Voice. Psychological Theory and Women's Development. Cambridge: Harvard University Press, 1982.

[10] Wagner I. Informatique Médicale. Nouvelle encyclopédie de bioéthique. H. G. and M. J.-N. Brussels, De Boeck Université 2000: 529-535.

[11] Gärtner J, Wagner I. Mapping Actors and Agenda Political Frameworks of System Design & Participation. The Human-Computer-Interaction Journal 1996: 11(3): 187-214.

[12] Blomberg J, Suchman L, et al. Reflections on a Work-Oriented Design Project. Human-Computer-Interaction 1996: 11(3): 237-266.

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Handling Consent to Patient Data Access in a Hospital Setting

Gunnar René Øie[a], Herbjørn Andresen[b], Inger Anne Tøndel[c]

[a] *Department of Computer and Information Science, Norwegian University of Science and Technology, Norway*
[b] *Section for Information Technology and Administrative Systems, Faculty of Law, University of Oslo, Norway*
[c] *SINTEF ICT, Norway*

Abstract

The right to use patient data in treatment is based on the conditions of a need to know and patient consent. In electronic health records, these two conditions can be applied in various ways. We study the handling of consent in two Norwegian hospitals, with a view to how access control and consent handling can be integrated across electronic systems that process patient data. A workshop was held, where two consent handling scenarios were simulated, one in-hospital, the other external. Activities were identified and tied to roles and to the documents and systems used. Electronic systems were found to support the execution of the scenarios to some extent. The electronic functions used in-hospital were consent storage and logging of access; access control was not sufficient. When sharing information externally, the typical approach is a declaration signed by the patient or a referral; such external information sharing should be supported by specific functionality. A first step towards integrated access control is integrated consent handling.

Keywords:

confidentiality, informed consent,
computerized patient records

Introduction

Health care workers, such as physicians, nurses and managers have a need to access sensitive patient data in order to treat and manage patients. Patients have a right to be given good treatment as well as a right to privacy. Confidentiality of patient data is both a requirement for privacy, and an argument for openness between a health care worker and a patient. Confidentiality may thus enable the worker to get more important information of a potentially embarrassing nature from the patient; confidentiality may, however, also keep previously recorded data out of the worker's reach. Depending on the circumstance, confidentiality is thus both a hindrance and support in the work of healing illness. A health care worker's right to access patient information is typically based on two conditions: A need to know, and patient consent.

Need to know and patient consent

Norwegian health legislation contains a diversity of rules governing the legal basis for when and on what conditions a health care worker may or may not have access to patient information. The principle of "need to know" could be interpreted as an abstraction to those rules governing access, they include an obligation of secrecy, several conditions where consent may be needed, some degree of a right for the patient to participate in decisions concerning himself, and even a few situations where patient information may be revealed to third parties on the health care worker's discretion.

There is a complex body of legal acts influencing the rights to access patient information. For simplification, we mention the three most important acts regulating consent as a legal basis for access rights.

The Health Register Act is a Norwegian implementation of the EU Directive 95/46/EC. The term "consent" is explicitly defined in the Health Register Act; a valid consent must be informed, explicit, and given freely. The function of the consent in the Health Register Act is to serve as a legal basis on which responsible enterprises process the information.

The Health Personnel Act contains obligations and rules of conduct for the health care workers as professionals, regardless of whether they are self-employed or employed by a hospital, or have any other kind of assignment as a health care worker. The term "consent" is not explicitly defined in the Health Personnel Act, and for some important situations (e.g. sharing information with another health care worker involved in the same treatment episode) implicit patient consent will be sufficient. In these cases the patient consent is presumed, but the patient may object to the sharing of information.

The Patients' Rights Act also contains rules on patient consent. While the Health Register Act and the Health Personnel Act are mainly focused on the obligations to obtain the consent when it is needed, the Patients' Rights Act views consent as a means to ensure that the patient has a proper understanding of his own situation and can take part in important decisions concerning the treatment.

In the case of a patient not being able to make valid decisions on his own behalf, consent may be given by a duly empowered proxy.

Access control supporting consent management

Patient data are increasingly held and processed in electronic health record (EHR) systems, with potential for integration between systems. The access control mechanisms of existing EHR systems contain, to some degree, representations of the users' rights to access information. The access rights represented in the systems are in part authorisations based on organisational properties (i.e. roles and/or organizational units) and in part an interpretation of the legal basis for accessing information. Support for consent in access control imposes restrictions at a very fine level of granularity, in some cases as detailed as the relation between one specific health care worker and one specific patient.

Objectives

The objectives of this study are: Finding whether consent can be successfully managed with existing EHR systems; identifying challenges in consent management and access control for patient information inside a hospital and between health care providers; identifying implications, if any, that our findings have for integration of access control.

In the rest of this article we first present academic work that is related to consent management and security; we then present the methodology and results of our survey; discuss the results and conclude the paper.

Related work

A questionnaire survey in New Zealand [1] studied patients' attitudes towards sharing their health information. One of the major findings of this survey was that 90 % of the respondents had incomplete or no knowledge of how their information was shared, and that they would prefer to be consulted about the distribution of their information. Based on this study it is suggested that future systems should incorporate sophisticated and flexible access control policies that can be adapted to meet the preferences of individual patients.

Adams and Blandford highlight the importance of understanding the users' work practices and their relevance [2]. The paper claims that to effectively implement systems to protect organizational and personal security, one needs to understand the organization, and its social structure, norms and work practices.

Kluge [3] discusses some limitations to the suitability of consent as a means to protect the patients' right to privacy. For instance, denying consent may not harm the rights of other persons. The use of consent is also limited by the principle of impossibility: If the patient expects treatment he may not refuse necessary access to relevant data.

Coiera and Clarke [4] categorize electronic consent (e-Consent) along two axes: Form of consent that the patient can express (model), and function of consent in the system. The three functions of consent identified were as follows: Patient consent captured as a legal record only; the system requires the health care worker to verify that the patient's consent has been read and understood; the system automatically executes the consent and uses it as access control rules, i.e. as a "gate-keeper".

Methods

This survey was performed in one of the health regions of Norway in the fall of 2006. The aim of the survey was to get qualitative information on how patient consent is handled, and how it is related to access control. We chose to direct our survey towards the following types of participants:

- IT department representatives: These have knowledge of the IT systems in use, their possibilities and limitations. The representatives range from technical staff to project leaders and decision makers, including those responsible for information security.
- Representatives from decision makers in two of the hospitals of the health region. These have knowledge of the legal requirements, the overall decisions that have been made, and the overall problems experienced with today's solution.

Some important groups were left out in this initial qualitative study: Types of care givers involved in patient treatment, and administrative personnel like secretaries. We plan to include these groups in future more quantitative surveys.

Since our aim in this phase was to collect qualitative information, we chose to gather representatives from the above types of participant groups in a workshop. The workshop was organized as a process workshop [5]. This type of workshop had already been used successfully in other qualitative studies on access control in hospitals [6]. The focus of such a workshop is on finding out how important processes are handled today, identify those responsible and documentation and tools that are relevant. Based on the identified current processes one can discuss challenges and future solutions. We were interested in both how the processes are implemented in the organization, and how the IT systems support the processes.

The workshop was centred on the following scenarios:

- A patient A does not want one specific physician to have access to A's health information. What happens?
- A patient that earlier has given consent to share health information with another health care enterprise now wants to withdraw this consent. What happens?

Based on the scenarios, the participants identified activities, roles and documentation/tools, wrote them down on sticky notes and organized them on a process sheet. After doing this in small groups, a plenary session synthesised the results and associated each activity with roles and documentation/tools, as illustrated in Figure 1. We recorded audio and video to capture all details discussed.

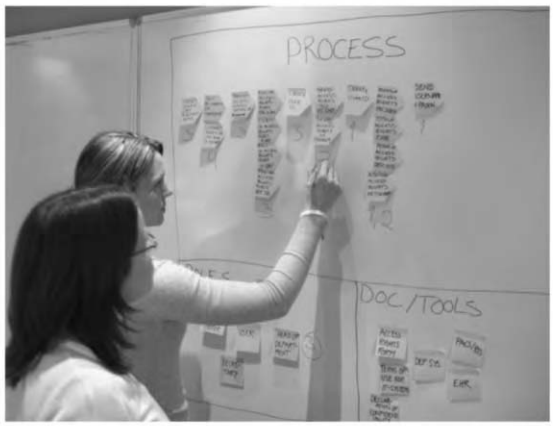


Figure 1 - Demonstration of process workshop plenary

Results

Scenario 1: denying access to a specific physician

The workshop participants considered the scenario to be realistic and to be something that occurs from time to time. In some instances of this scenario, the patient can not be adequately treated if the patient's request is to be carried out. This can occur if the physician and patient are at the same hospital department, but also if they are at different departments. In the former situation, even if there would be several other physicians present during the day, the physician in question would have to withdraw from certain parts of meetings. During the night and weekends, the physician in question might have been scheduled to be the only physician on call, meaning that the most available physician can not treat the patient. This can also happen in the latter situation, when the physician is at a different department, because the physician might be on call for several departments. In such cases, the patient would be better served by being treated at a second hospital; the first hospital might thus refuse planned non-emergency treatment on the grounds of its obligation to provide safe treatment.

The health care worker that speaks to the patient would first discuss the matter with the patient and colleagues to establish that the patient can be safely treated. Provided that the patient can be safely treated, the groups agreed that the patient's request would be carried out in the following manner.

The patient's request would be recorded in the EHR by the person responsible for the record or a secretary. Each medical record has a person responsible for it, as required by law. In some departments this person is registered in the record, while in other departments this role is taken by the department head or by the person in charge of medical systems.

The physician that should not access the patient's record would then need to be informed about the patient's request, warned about consequences and confirm that the warning has been understood. Subsequently, the patient would be informed about how his or her request was carried out.

After the patient has been discharged, the record responsible person must audit the access logs to confirm that the physician in question has not accessed the record of the patient in question. This would include access logs from both the dedicated medical record system and associated systems like medical image retrieval systems. If the patient wishes to confirm who has accessed the record, and requests the access logs, then the record responsible would also need to give these logs to the patient.

Even with these measures in place, the physician in question is able to access the records of the patient. The solution can still be considered adequate, if the patient is satisfied that the warning and subsequent log auditing reduces the risk of un-authorized access to an acceptable level. To completely remove access for the physician would, with today's access control solutions, involve keeping the record in a separate system, such as on paper, or transferring the physician to another hospital department for the duration of the patient's stay.

Scenario 2: withdrawal of consent to share information with other health care enterprise

When the workshop participants discussed this scenario, they first questioned the scenario. The common Norwegian definition of "health care enterprise" only includes those owned by the state, i.e. hospitals. General practitioners in private enterprises, as well as private clinics, are thus not included in this definition. There are more referrals and discharges between a hospital and a general practitioner than there are between hospitals, and the scenario might have been more relevant with a party called something else than "health care enterprise". The groups thus chose to expand the scope of the discussion to also include general practitioners, after first discussing inter-hospital information sharing.

Another point discussed was that in the hospitals represented, there is no mechanism for periodic handover of patient information, and as such few ongoing automatic handovers to break off.

There are mainly two types of handovers that are not based on an explicit consent given by the patient: Providing access to medical images in the region wide image system, and the routine sending of a discharge note to the referring health care worker when their patients are discharged from the hospital.

In other cases some form of instance by instance consent from the patient or other legal documentation is always needed. When giving information to the Norwegian patient compensation agency (Norsk pasientskadeerstatning), the patient consent is considered expressed by the patient filing a complaint with that agency, while it in other handovers, such as to the police or an insurance agency, is based on a signed declaration of consent from the patient.

In the plenary discussion, three different sub-scenarios of this scenario were identified and discussed. The first involved access to the imaging system, where it was stated

that once read access was given to a set of images the access could not be easily revoked.

The second and third sub-processes involved the referrals and discharge notes (the latter called "epicrises" in Norway). If the patient in these scenarios wants to withhold information from the referring health worker, the hospital staff needs to make a note of this in the referral and discharge module of the medical record saying that referrals and discharges should not be sent to the health care enterprise in question; this is only a note and would need to be discovered by the discharging personnel and be acted upon accordingly every time. There is thus no automatic method for stopping the sending of a discharge note to the referring health care enterprise.

Discussion

We discuss what the results tell us about the potential for consent management and for integration of policies and access control across systems.

Fulfilling denial of consent in present situation

As shown by the results, the current systems in use at the studied hospitals do not support the technical exclusion of one health care worker from accessing the records of one patient while still keeping access to the records of the other patients at the hospital department, or indeed at the same hospital. Even if the health care worker in question is temporarily re-assigned to another department within the hospital, emergency access would still be available.

Carrying out the procedure in Scenario 1 will only fail to deter the physician in question if this physician is willing to risk a near 100% chance of disciplinary action. For accessing the data after being told not to, and knowing that logs will be reviewed.

The procedure is suited to hospitals where default access is given to the records of patients at the same department as the health worker, but it would need to be adjusted for hospitals where the default rules are different. Some hospitals, for example, allow access to the records of discharged patients for several months after discharge, for the purpose of follow-up and avoiding using emergency or other non-standard access for answering questions from discharged patients and their referring health workers (e.g., General practitioner). In such cases, access logs audit may have to wait until this access expires, or to be repeated at intervals.

When auditing access logs, the record responsible person must also remember to audit access logs for departmental clinical systems.

Viewing consent in all hospital systems

In the hospitals studied, consent is documented in the primary medical record system, as required by Norwegian regulations. This system is simply called "the Electronic Patient Record (EPR)" within the hospital, although other systems such as laboratory and imaging systems also store patient data. These systems do not support the registering of consent, and some of them can be accessed without going through the primary record system. Thus, ensuring that a possible special consent or denial from the patient is acted upon would require that the health care worker checks the consent information in the record before accessing such systems; this would cause a deviation of the work flow.

One solution for better consent handling within the hospital would be accessing imaging, laboratory and department specific clinical systems through the primary EPR system; another is to add a common consent module to those systems. In this way, consent can function without substantially deviating work flow, and without risking that the health care worker forgets the consent (accidentally or deliberately). Some parts of the system could require the health worker to verify having read and understood patient consent and consent denials.

Fulfilling denial of consent with executable rules

Converting existing systems to using consent as access control "gatekeeper" [4] is a more involved process than that of displaying consent and consent denial in all systems with patient data. Integrating access control between systems inside a hospital and between hospitals would involve having a common definition model and language for representing the access rules, integrating role definitions and work descriptions and standardising information classes across systems. Today, the hospitals have established procedures for giving access to the individual systems, but this access is provided on the basis of each system being an "object" of information. Since patients may have much more detailed consent requests, e.g., concerning a diagnosis rather than concerning the type of system where the data exists, integration would need a much higher granularity of information objects.

Subject granularity

The issue of subject granularity has consequences for consent management. In access control, the subject is the person or system that is denied or granted access; subject granularity is thus a measure of how precisely we can define subjects. In other words, are we able to give access to individuals, or only to departments or entire hospitals? Work inside a hospital department is collaborative [7] and information flows through other channels. One physician may also have patients at several departments, or having tasks to do at other departments, necessitating some kind of access across departments [6].

It entails that denying access to a specific person in some cases means denying access to an entire department. This is because this specific person would be unable to function normally within such a highly collaborative department. As found in our study, by denying access to a single individual, it might also effectively make treatment at the hospital impossible without infringing on the patient's consent. Even weekday treatment of outpatients may become difficult; an outpatient might develop medical complications resulting in hospitalization, or the department might simply be a small one, meaning that the patient can not be treated without potentially infringing on denial of consent.

Withdrawing consent for other external actor

As shown by the results, the current systems in use at the studied hospitals support referrals and discharge notes that are sent as the patient is transferred. If the patient has a specific consent denial, denying that the referring health worker should get results on discharge, then this can be registered in the relevant part of the EHR system, but this is not a standard location of consent. This kind of consent denial may be improbable, but a consent function related to external exchange of information will become necessary.

As of today, information exchange not related to discharge and referral is based on a signed declaration by the patient. There will still be a need to consider what to do when such a signed declaration conflicts with a patient's prior expressed consent and finding out whether the consent is informed. With the arrival of new types of external information sharing, consent and consent denials should be shown at places in the system where a health care worker evaluates external requests for information.

Informing the patient

As in other cases of handling patient consent, the consent, or the denial, for access to data must be an informed one. As emphasized in the results, the patient must be aware of how patient data will be used, and that a denial of consent could lead to treatment becoming unavailable at that particular hospital.

If the patient is to directly control access with a computer interface [8], the patient also needs to be informed about consequences of such control. If the patient gives less or more access to patient data than what was intended, then the interface should make this evident to the patient.

Conclusion

Current EHRs that record consent claims can support the manual execution of patient consent denial, but with some caveats. If patient information is kept both in the application called “EHR” and in other systems like imaging systems and departmental systems, then the consent should also be used by both. Points in the system where information is transmitted to outside parties may need specific consent functionality that is stricter than that which applies to general access.

Although a wider integration of access control in the hospital and between health care providers is a goal, there is a beneficial step that can be taken first: Making sure that the consent is viewed by the health care workers at relevant points in the process of using the EHRs. Whether a consent viewer or gatekeeper is the best solution will depend on the setting and on the implementation of consent.

Acknowledgments

We would like to thank the workshop participants for their valuable contributions and their employers for letting us use some of their time. We also thank Per H. Meland and Lillian Røstad for assisting with the preparations for the workshop. We thank Øystein Nytrø for assistance and for supervising the project Integrated Access Control for Health Care Information Systems. The project is funded by the Research Council of Norway's IKTSOS programme.

References

[1] Whiddett R, Hunter I, Engelbrecht J, and Handy J. Patients’ attitudes towards sharing their health information. *Int J Med Inform*, 2006;75(7):530-541

[2] Adams A, and Blandford A. Bridging the gap between organizational and user perspectives of security in the clinical domain. *Int J Hum Comput Stud*, 2005;63(1-2): 175–202.

[3] Kluge EHW. Informed consent to the Secondary Use of EHRs: Informatic rights and their limitations. MEDINFO 2004;11(1):635-638.

[4] Coiera E and Clarke R. e-Consent: The Design and Implementation of Consumer Consent Mechanisms in an Electronic Environment. *J Am Med Inform Assoc*. 2004; 11:129-140

[5] Dingsøyr T, and Moe NB. The Process Workshop: A Tool to Define Electronic Process Guides in Small Software Companies. In: Proceedings of ASWEC 2004. Washington DC: IEEE Computer Society, 2004.

[6] Meland PH, Røstad L, and Tøndel IA. How to mediate between health information security and patient safety. International Conference on Probabilistic Safety Assessment and Management PSAM 8. 2005

[7] Bardram JE. The Trouble with Login – On Usability and Computer Security in Ubiquitous Computing. *Personal and Ubiquitous Computing*, 2005;9(6):357–367.

[8] Galpottage PAB and Norris AC. Patient consent principles and guidelines for e-consent: a New Zealand perspective. *Health Informatics J*, 2005;11(1):5-18

Address for correspondence

Gunnar René Øie, Department of Computer and Information Science, Norwegian University of Science and Technology, N-7491 TRONDHEIM, Norway
gunnarre at idi ntnu no

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

A Day in the Life of a Clinical Research Coordinator: Observations from Community Practice Settings

Sharib A. Khan MBBS, MA[1]; Rita Kukafka DrPH, MA[1]; Philip R.O. Payne PhD[1]; J. Thomas Bigger MD[2]; Stephen B. Johnson PhD[1]

[1]*Department of Biomedical Informatics, Columbia University, New York, NY*
[2]*Departments of Medicine and Pharmacology, Columbia University, New York, NY*

Abstract

One of the goals of the NIH Roadmap Initiative is to re-engineer the national clinical research enterprise, with an emphasis on information technology solutions. Understanding end-users' workflow is critical to developing technology systems that are grounded in the context of the users' environment and are designed to fulfill their needs. Community practices are becoming the prevailing setting for conducting clinical research. Few studies have assessed clinical research workflow in such settings. We have conducted a series of investigations to model the workflow and have previously reported on some basic aspects of it, like the lack of information systems to support the workflow. In this paper we describe finer details of the workflow, using results of observational studies. These findings highlight the needs and inefficiencies that suggest the kind of information system that must be developed to enhance collaboration, communication and improve efficiency. This preliminary investigation also opens ground for more extensive studies to further elucidate the workflow.

***Keywords*:**

clinical research, workflow, community practices, direct observations

Introduction

The National Institutes of Health (NIH) has embarked on the Roadmap Initiative, to achieve, among other goals, an increase in the clinical and translational research capacity of the United States [1]. This increase in capacity will require a substantial effort to develop and deploy information technology (IT) to improve efficiency, communication and collaboration among the diverse and distributed research networks and patient communities [1]. However, successful adoption of IT systems has always been a challenge, as evidenced by the slow adoption rates for Electronic Health Records (EHRs) and Computerized Physician Order Entry (CPOE) Systems in the United States and elsewhere [2]. Besides financial hurdles, numerous social and individual factors (such as understanding end user needs and workflow) have impacted the adoption of health care IT systems [2,3]. These factors must be considered even more carefully when considering a fundamental re-engineering of a national multi-stakeholder enterprise like clinical research. Such an onerous task requires one to address such fundamental questions as which stakeholder does the re-engineering target, what processes need to be addressed and how does one achieve the desired changes.

Columbia University's InterTrial project, which is funded as part of the Roadmap, has focused its re-engineering on community practices, which are becoming the prevailing setting for clinical research. Several factors, including lower administrative overhead and access to more representative subject populations, are motivating this shift [4]. To inform our re-engineering effort, we have undertaken a series of empirical studies (using qualitative and quantitative methods) to investigate stakeholder roles, organizational structures, workflow, information needs and communication patterns. Our multi-phase inquiry began with a focus group of clinical research coordinators (CRCs) from several member community practices of the Clinical Trials Network (CTN), Columbia University. This was followed by surveys, key informant interviews and preliminary field observations to develop an initial CRC workflow model [5]. This model (Figure 1) was used in a time-motion (TM) study to determine the workflow of a CRC in terms of what tasks (e.g., documentation, recruitment) and activities (e.g., scheduling a patient visit) were performed as part of research, what tools (e.g., phone, paper) were used to accomplish these activities and, ultimately, what outcomes were achieved (e.g., completed, incomplete).

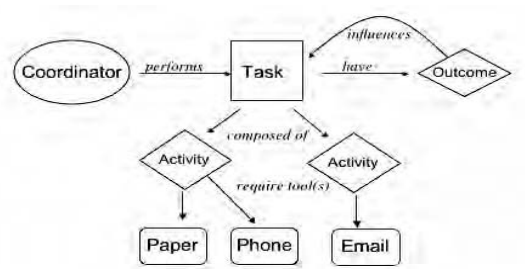


Figure 1 - Basic CRC Workflow Model

The TM study provided useful insights into the research workflow, highlighting such aspects as, the paper-driven

nature of work, the occurrence of several interruptions and the identification of often repeated and time-consuming activities [5]. However, it only provided snapshots of workflow and was insufficient to articulate the richness of the several interactions, conditions and circumstances influencing the workflow. For example, the TM study was unable to demonstrate who communicated with whom or to identify the reasons behind excess documentation. The observations described in this paper were done to elucidate this richness of the workflow.

Methods

In this study, we have used the method of direct observation to shadow CRCs in community practices and to record their work activities, interactions with other stakeholders, communication patterns, information needs and other aspects of workflow. Direct observation (and other qualitative methods) can provide a *deeper* understanding of a phenomenon, especially in the context of questions such as *why* and *how* [6]. This technique is also helpful in validating interview results and discovering unstated findings, since most respondents cannot precisely recall their work habits in an interview or survey [7]. Such studies are being increasingly undertaken in the biomedical domain to assess workflow [7].

We built on the findings from our initial time-motion study of three community practices by doing direct observations at three additional sites. For the purposes of this study, a community practice is defined as a physician office providing outpatient clinical care. Each of these sites also has a small research division with, typically, a single CRC managing some 8-10 trials. Those sites managing more clinical trials had a correspondingly larger team. Each of the three community practices was visited for three days, during which we conducted interviews with the CRC at the site, performed direct observations of the workflow and discussed certain work processes. The observations were recorded in field notes, which were later transcribed and analyzed for important workflow findings. In this paper we report on some aspects of workflow learned from these observations and supplement them with findings from the interviews, process descriptions and our earlier studies. What we describe is a *representative* day in the life of the CRC, though not all findings can be observed in the course of one particular day. We will not discuss workflow processes (e.g., scheduling a patient visit) in step-by-step detail (which were captured in our process discussions) but instead will highlight general crosscutting workflow themes. Example excerpts from our field notes are used to support the description of the findings. To make the paper readable and personable, we will allude to a fictitious CRC named Sarah, who epitomizes a composite of all the observed CRCs.

Findings

General aspects of sarah's workflow

Sarah is busy throughout the course of work. She is the *hub* of all work and information related to clinical research (if she is the main research staff) or all the trials under her supervision (if part of a research team). Sarah performs several tasks/activities in the course of her day, which can begin at 7.30 in the morning (particularly on days when patients are scheduled for a fasting blood glucose test) and go late in the evening. One of the characteristics of her workflow, also discovered in our earlier time-motion study, is the occurrence of regular interruptions (Figure 2).

> *As Sarah is completing a patient case report form (CRF), the Principal Investigator (PI) comes to ask about the protocol-related spirometry values. She points him to the protocol binder and he picks it and leaves. She returns to completing the form. After a minute the PI returns and says the binder doesn't have the value he is looking for. Sarah leaves the form and begins to search for another reference manual, which could have the value.*

Figure 2 - Excerpt showing an interruption

These interruptions usually happen because of incoming phone calls, or if some staff member (e.g., the Principal Investigator [PI] walks up to request some information). The occurrence of interruptions to workflow in clinical research parallel the interruptions noticed in other health care settings [8]. It is also common to see Sarah carrying out many activities in parallel, i.e., different activities being executed concurrently, but not necessarily requiring attention at the same time point. Sarah also engages in multi-tasking, i.e., actively doing two things at the same time (for example, speaking on the phone while flipping through the patient chart). Sarah usually does all types of activities related to research herself, i.e. at one point she could be taking calls, filling patient CRFs or examining a patient. There is some *division of labor*, i.e., different types of work being done by different people (e.g., phone calls being screened at the front desk or blood samples being drawn by a nurse other than the CRC) but this is usually limited to phone calls being received at the reception. During the course of her work, Sarah frequently moves in between shelves stacked with study binders, the examination room or the reception area equipped with the common fax and photo copy machines. This back and forth movement necessitates that she have at hand the information required to do a particular location specific activity. For example, having the right documents and the fax number before she walks to the fax machine. This movement may also require her to strategically economize her *travels* by grouping location specific activities (faxing and photocopy work). In a similar way, Sarah also makes simple strategies to group her work temporally, i.e., she may decide to do time-consuming paperwork in the afternoon when there are fewer interruptions.

Aspects of communication

Being the hub of information and work, Sarah must communicate with various stakeholders involved with research. These include the PI, other research and clinic staff, the sponsor, the monitor (some trials are managed by sponsor-appointed external monitors), patients and clinical laboratories. There are two noteworthy aspects about Sarah's communication channels. One is the apparent lack of access to other CRCs in nearby practices. This inability to reach out to other CRCs and create a peer-to-peer support network is a barrier to leveraging peer knowledge for self-help. Second is the authoritative *top-down* hierarchical directionality of the communication between Sarah and her communication partners (like the PI, sponsor and the

monitor), the nature of which usually results in Sarah's compliance with others' decisions.

Sarah communicates several times during the day, about wide-ranging topics including trial updates, obtaining documents, scheduling patient appointments or dispatching patient samples for blood work. Back-and forth communication (as can be discerned from the excerpt in figure 2, in which the PI returns twice to obtain the same information) is frequent and can stretch for hours or days, when, for example, a patient fails to show up for an appointment and repeated calls must be made to reschedule it. Communication occurs through various channels of which the most prevalent is verbal communication, followed by phone, fax and others. In fact, verbal communication happens in interesting ways (Figure 3), with Sarah at times, calling out to the entire office requesting a patient's medical chart ("chasing the chart" happens frequently) or to find out if a patient has called to cancel a visit.

Sarah comes out of the examination room and asks the office, "Does any one have this patient's chart?" Later that day, she realizes that a patient hasn't shown up yet. She asks, "Did John Doe come in today?" From her office space, both the receptionist and the medical record assistant are within earshot, affording her this small luxury.

Figure 3 - Excerpt of intra-office verbal communication

Multiple communications of the same information were observed in several instances. For example, during a patient visit, Sarah is required to dial an Interactive Voice Response System (IVRS) to obtain a code that confirms the occurrence of the visit. The code in some trials also serves to function as the code for the specific medication packet that must be dispensed to the patient. This information is communicated by the IVRS to Sarah during the call, is then later faxed and may also be sent by regular mail. Hence, three copies of the same information are communicated, in three different modalities, and each has to be maintained in the corresponding patient's study binder for later verification by the monitor. In contrast to multiplicity, there were instances of lost, delayed or unnecessary communication. Each of these sources of inefficiencies is particularly worrisome for Sarah. On occasions, she has completed and faxed documents only to be told (much later) that it was no longer required to fax that document. There have also been instances where the sponsor returned an entire batch of completed CRFs because an old version of the form was used. She once received two document packets for a study that ended months before. In most instances, the communication is dropped from the "other" side but Sarah has to face the consequences and has little influence to correct such mistakes (usually due to the top-down directionality of communication). She accepts it as "the nature of her job".

The burden of documentation

Documentation can indeed be a burden for Sarah. She spends considerable time doing it [5]. Most of the documentation deals with completing source documents (the main record of a patient visit, which she usually tries to complete during the patient visit) and then transcribing pertinent information into the corresponding CRFs for the study. The paper-based forms make the process of documentation tedious because they necessitate redundant data entry. For example, she needs to enter a patient's trial identifiers, trial name and other such information on every page of the CRF. Similarly, there is considerable overlap between reporting forms addressed to different stakeholders. A serious adverse event (SAE) report form used to notify the IRB maybe organized differently than that for the sponsor, even though both contain almost the same information, requiring duplicate data entry. Duplicate documentation effort is also necessitated when an entire batch of CRFs has to be redone because of version change (as explained earlier). Another particularly troublesome aspect of documentation is answering queries (a list of questions pertaining to the data in a CRF) sent by the trial monitor. To resolve the queries, Sarah has to look up old source documents or the patient's medical chart to revise the queried data element's value. This process can be laborious, especially when the queries refer to CRFs completed months or even years before. Besides completing CRFs, Sarah also has to write letters to the IRB or to a patient's primary care physician to apprise them of the patient's enrollment and progress in the trial. Most of the CRF and other forms are paper-based, though some trials provide software (some even provide exclusive trial laptops) for electronic data capture (EDC). Sarah finds EDC easier and more convenient because it offers automatically populated patient identifiers and immediate data integrity prompts. But such supporting systems have their own problems.

Information management and information needs

During her day's work, Sarah is actively managing several information exchanges in various formats and from diverse stakeholders. Her primary tools in managing the bulk of the information are the protocol and the individual patient binders. The protocol binders contain all the documents pertaining to the study. It is not uncommon to see protocol binders that are several inches thick and contain hundreds, if not thousands, of pages. Every notification of an amendment must be attached to the appropriate section in the binder and the monitors verify these during audits. Similarly, all patient-specific information (like new laboratory results, newly filled CRFs) must be filed in patient binders. The binders can quickly occupy much of the prime office space; half of Sarah's office is just shelves upon shelves of binders. It is necessary to store older binders until the sponsor permits destroying them or sending them offsite. Other information management tools commonly used by Sarah are a to-do list and a calendar. She uses a to-do list to remind her of the immediate or impending tasks. She usually maintains this list on a simple piece of paper and uses several highlighters to mark different activities in different colors (for example, highlighting blood work in red). She frequently uses her calendar to mark dates corresponding to her to-do list. She maintains her personal work calendar separately from the calendar that is maintained for the entire office (it can be in paper or electronic format), which she consults when scheduling visits.

Sarah has different information needs. She routinely seeks trial-related information (an important finding of the time-motion study [5]), which includes ascertaining the next visit for a study patient or determining a study's inclusion/exclusion criteria. Sarah also refers to her address-book or

the numerous artifacts (like small sticky notes littered all about her computer and office space containing phone numbers or other snippets of information) while making phone calls to sponsors, monitors, patients or the postal service for scheduling pickups. She typically needs to have several pieces of information to complete any activity. For example, while scheduling a radiology test for one of her patients, she needs to know, among other things, the trial inclusion/exclusion criteria, the patient's name, date of birth, contact information, preferred test dates, and the sponsor's name. Artifacts are routinely used by Sarah to take quick notes or post reminders around her. These notes provide easy reference to oft-needed information. The use of visual and spatial cues is a significant information retrieval strategy employed by Sarah. She has learned the spatial location of the protocol binders or other objects used in her work and can often find information at a glance. Similarly, the different color codes on her to-do list or of the objects in her surrounding, act as strong cues to narrow down her vast search space in a few milliseconds. As shown in the excerpt in Figure 2, she was quickly able to determine, by the spatial organization of her protocol binders, the location of the binder sought by the PI.

Use of information technology

Sarah does utilize a variety of computer-based tools to accomplish her work. She is adept at checking her email, accessing a website (she usually tracks updates in trials that maintain a website). She often uses word-processing programs and also knows how to use spreadsheets. Despite no formal computer training, Sarah can type quickly (25-30 words per minute), understand the concept of files or folders (which come in handy when managing her documents) and is aware of the security threats posed by viruses. She does EDC in two trials, one of which requires her to send data using a dial-up modem with which she has become familiar. As mentioned earlier, Sarah does note that computer systems improve efficiency but she highlights their quirks too. For example, a troubling aspect of using one of the EDC systems is the enforced password expiration. The software accepts her current password for only 60 days, after which she is required to change it. However, she uses the system *only once* in 90 days, to complete the CRF of the only patient in the trial who visits once in 3 months, by which time her password has expired. She then has to go through a laborious process of getting a new one from the sponsor's customer service hotline. The other vexing issue with computer software for Sarah is the lack of almost any control over the organization of the software or its features.

Discussion

The above findings illustrate interesting aspects (summarized in Table 1) of Sarah's workflow including some of the inefficiencies. These inefficiencies become more pronounced when examined in light of information management principles. First, the inefficiencies identified can be partially attributed to the *paper-based* workflow processes, which are limited by the nature of paper as a medium for information management. Paper forms are difficult to copy and disseminate, provide no error checking and are limited to one point in space or time (causing "chasing the chart" phenomenon).

Table 1 – Summary of workflow findings

Workflow themes	Important aspects
General	Frequent interruptions, multi-tasking, minimal division of labor
Communication	CRC is hub, several channels, multiple, dropped or lost communication
Documentation	Paper-based, redundant, some EDC
Information Mgmt./needs	Binders, to-do lists, calendars, artifacts, trial information sought, visual cues used
IT tools	Email, productivity software, EDC issues

Second, there is evidence to suggest considerable *redundant data entry*. For example, patient visits are first recorded in source documents and then transcribed to CRFs. Next, the CRFs require filling the same fields many times over or filling the same information for different purposes or audiences (the SAE example discussed earlier). Such a process can directly result in both transcription and omission errors. A consequence of redundant data entry is redundant update/delete efforts when data changes over time. Third, our examples indicate, that the processes have minimum provision for *reuse of data*. This is particularly troublesome, as in the case of managing medication logs. As mentioned earlier, Sarah obtains a medication code through an IVRS system while dispensing medications. She is required to log this information in a drug-dispensing log and also tally this with a drug inventory list, which she uses to order new supplies. Both the lists could be automatically generated through the sponsor's IVRS because the system has a record of all the drugs dispensed (and also of drug supplies received). In fact, the system could easily be programmed to send out new drug supplies automatically, based on the dispensing and previous order history. Going further along the lines of reuse of data, a system that allows the research team to share information on frequently sought information or repeated tasks, could drastically reduce time spent in repeatedly seeking the same information from other colleagues or slower mechanisms. As was mentioned earlier, trial inclusion/exclusion criteria, patient study visit schema, contact information and commonly used forms are frequently sought by all members of the research team and these could be obtained faster if provided in a collaborative, accessible and easily maintainable space like a FAQ web page. Fourth, there appears to be a lack of standards to specify the workflow processes pertaining to clinical trials. It was commonly noted that same tasks/activities (like *draw patient blood samples*) or other semantic concepts, were being represented using different syntactic phrases. Similarly, there are no standards for designing the various forms used in trials. Using standards would greatly allow reuse of data. Recently reported works on clinical research task vocabularies [9] and the initiatives led by the National Cancer Institute to create clinical research interoperability stan-

dards, would greatly aid information exchange among research information systems.

Several other shortcomings in the workflow merit discussion. The communication mechanisms frequently fail to provide up-to-date information or result in several duplications (that add up to cost in dollars and time). These can be corrected by adopting a paradigm of a central information repository that always stays up-to-date. For example, a trial-specific website (some sponsors provide this now) that hosts the latest versions of all required documents and forms, news updates and allows for the provision of such updates to be broadcast as emails or via really simple syndication (RSS) feeds, can greatly cut down the delays in communicating such updates via faxes or mail. In some cases, there is simply a lack of appropriate tools to carry out tasks. For example, some trials don't provide a handy one-page reference for the inclusion/exclusion criteria or the protocol visit schema (which are so frequently sought), leaving Sarah to devise her own reference sheets. Similarly, she needs to prepare her own checklists to remind her of the steps needed to complete a long sequence of activities correctly or create templates for letters. When tools exist, they may not have been designed keeping the workflow of the coordinator in mind. The example of the EDC system's password expiring every 60 days when Sarah only uses the system every 90 days is a good illustration of this disconnect. Similarly, getting a site's address printed in the "from" field of a shipping label (that is affixed on the boxes used to dispatch blood samples) can save Sarah some time. The IVRS system referred to above can also be better designed. It was not uncommon to see Sarah lose enthusiasm when navigating the depths of a complex IVRS system. The impersonal tone of the automated voice asking her to "press 1 if your entry is correct, press 2 if its wrong or 3 to go back to the previous menu" at every selection step, adds monotony and can be made more pleasant. Another important observation was that the information needed to complete a task or activity resides in multiple places including patient study binders, patient's medical charts, emails, protocol binders, or with the PI or some other person. The unavailability of any one of them can delay the completion of an activity.

It is important to emphasize that we don't intend to suggest that the clinical research workflow is *totally broken* or that the inefficiencies should be attributed to the CRCs. In fact, the resourcefulness (as demonstrated by Sarah in designing her own checklists), persistence and the *just get it done* attitude of the CRCs is very commendable. We briefly highlight a few strengths. One, an interruption in workflow is not always undesirable as it may support work through dissemination of unknown information through social exchanges. Two, audits by trial monitors do serve as useful checks to identify documentation errors. Third, the use of visual and spatial cues is a very rapid and effective mechanism for information retrieval as they narrow down the search space very acutely.

These observations have significantly informed our re-engineering efforts which are now focused on developing a web-based resource named the WorkWeb. The WorkWeb is intended to bring together, the currently distributed information resources (documents, phone numbers, etc), tools (a shared calendar and other applications designed to support specific tasks) and stakeholders (PIs, CRCs). It aims to enhance collaboration by supporting the creation of peer-to-peer support groups (via forums and blogs). The WorkWeb will also provide mechanisms for delivery of educational resources (like Good Clinical Practices training) through online multi-media delivery. As desired by the CRCs, almost the entire system will be under the control of the end-user who can add/modify content as they see fit and use features as needed, representing a paradigm shift of empowering our CRCs.

We acknowledge the limited nature of our observations because of which we may have overlooked (or incompletely understood) some aspects of the studied workflow. However, given the paucity of research in this field, we believe this is an important first step and contend, "One doesn't have to know everything in order to understand something [10]."

Conclusion

Clinical research is at the cusp of a fundamental re-engineering that involves substantial adoption of IT. Understanding workflow is critical to the design of the right IT solutions and their ultimate adoption. The InterTrial project is advancing efforts to elucidate the nature of workflow associated with clinical research in community practices in a bottom-up and empirical manner. These studies only represent preliminary work that needs to be further enriched by other (qualitative) inquiries before a detailed workflow model can be elicited. The initial results have informed our re-engineering efforts and we believe, illustrate the types of knowledge necessary to ensure the successful deployment of IT in community-based clinical research.

Acknowledgement

We thank our funding agency, the NHLBI and the community practices and CRCs for contributing their time to this study.

References

[1] Zerhouni E. The NIH roadmap. Science. 2003 Oct 3;302(5642):63-72.

[2] Ash, J, and Bates, D. Factors and forces affecting EHR system adoption: 2004 ACMI discussion. JAMIA. 2005;12:8-12

[3] Kukafka R, Johnson SB, Linfante A, Allegrante JP. Grounding a new IT implementation framework in behavioral science: a systematic analysis of the literature on IT use. J Biomed Inform. 2003;36 (3):218-27

[4] Sung NS, Crowley WF Jr, Genel M, et al. Central challenges facing the national clinical research enterprise. JAMA 2003;289:1278–87.

[5] Khan SA, Payne, PRO, Kukafka R, Bigger, JT, Johnson, SB. Modeling clinical trials workflow in community practice settings. Proc AMIA. 2006.

[6] Creswell JW. Research Design. Qualitative, Quantitative and Mixed Methods Approaches. Sage Publications. 2003.

[7] Unertl KM, Weinger MB, Johnson KB. Applying Direct Observation to Model Workflow and Assess Adoption. Proc AMIA. 2006

[8] Brixey, J. A Study of Interruption in a Level One Trauma Center. Proc AMIA. 2005.

[9] Dietzer J, Payne PR, Starren JB. Coverage of Clinical Trials Tasks in Existing Ontologies Proc AMIA, 2006.

[10] Wolcott HF. The art of fieldwork, Walnut creek, CA. Alta

[11] Mira Press, 1995

Address for correspondence

sharib.khan@dbmi.columbia.edu

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

A Survey of U.S.A. Acute Care Hospitals' Computer-based Provider Order Entry System Infusion Levels

Dean F. Sittig, PhD [1,2,3] , Ken Guappone, MD[2], Emily M Campbell, RN, MS[2], Richard H. Dykstra, MD[1, 2], Joan S. Ash, PhD[2]

[1]*Northwest Permanente, Portland, OR;*
[2] *Oregon Health & Science University, Portland, OR;*
[3]*Improve-IT Institute, Toronto, Ontario*

Abstract

We developed and fielded a survey to help clinical information system designers, developers, and implementers better understand the infusion level, or the extent and sophistication of CPOE feature availability and use by clinicians within acute care hospitals across the United States of America. In the 176 responding hospitals, we found that CPOE had been in place a median of 5 years and that the median percentage of orders entered electronically was 90.5%. Greater than 96% of the sites used CPOE to enter pharmacy, laboratory and imaging orders; 82% were able to access all aspects of the clinical information system with a single sign-on; 86% of the respondents had order sets, drug-drug interaction warnings, and pop-up alerts even though nearly all hospitals were community hospitals with commercial systems; and 90% had a CPOE committee with a clinician representative in place. While CPOE has not been widely adopted after over 30 years of experimentation, there is still much that can be learned from this relatively small number of highly infused (with CPOE and clinical decision support) organizations.

Keywords:

Computer-based Provider Order Entry,
health services research

Introduction

Several recent surveys of U.S. hospitals have estimated that less than 10 percent of all U.S. hospitals currently have Computer-based Provider Order Entry (CPOE) available, and of those that do, it is not known what percent of all orders are entered by clinicians [1, 2]. An earlier assessment of clinical information systems' capabilities (incl. CPOE) at five, early Davies Award winning institutions [3], was completed by Doolan et al. near the end of 2001 [4]. They found that the percentage of orders entered by clinicians at these leading institutions varied from less than 5% to well over 95% and that many different types of CDS were available. In 2004, Ash et al. found that CPOE "saturation", or "the percent of orders entered by physicians using a computer vs. other mechanisms" had a bimodal distribution (i.e., in hospitals with CPOE, 28.2% had less than 10% of their orders entered via CPOE, while 35.2% reported over 90% of their orders were entered using CPOE)[1]. In addition, of those hospitals that are using CPOE, no one knows the full extent to which real-time clinical decision support (CDS) has been incorporated into the clinicians' new CPOE-centered workflow [9].

The goal of this study was to further our understanding of CPOE and CDS infusion levels, both through improved definitions of CPOE and CDS infusion, as well as to actually measure these levels nationwide. Therefore, we developed and fielded a survey to help clinical information system designers, developers, and implementers better understand the infusion level, or the extent and sophistication of CPOE and CDS feature availability and use by clinicians within acute care hospitals across the United States of America.

Background

Infusion

Infusion of technology is described as the degree to which one uses an innovation in a more complete and sophisticated manner [5], and it occurs as information technology applications become more deeply embedded within the organizations' work systems [6].

Measuring the infusion level of any type of information system is difficult [7]. To compound the difficulty, separating the CPOE system functionality from the remainder of the clinical information system (CIS) can be even more difficult. We used the following definition for Computer-based Provider Order Entry (CPOE) - Entry of orders directly into the computer by the person responsible for making decisions about the patient's care. This person could be a physician, advanced practice nurse practitioner, or physicians' assistant with ordering privileges [8].

Methods

Initial survey development

We based our initial survey on that of Doolan et al. and added several additional questions about CPOE and deci-

sion support features after an extensive review of the literature [9,10,11,12] [see Appendix A for a copy of this initial 42-item survey]. Following IRB approval, we identified institutions who had successfully implemented CPOE by reviewing the literature, consulting with experts in the field, and talking with representatives from several of the major CPOE vendors. We included two sites (Wishard Memorial Hospital, Indianapolis, IN & Brigham & Womens' Hospital, Boston MA) from the original Doolan article for comparison. From each institution, we identified a key individual to answer our survey and contacted them directly via telephone or email. We sent a copy of the survey to each of our key informants and asked them to fill out as much of it as they could as of January 1, 2005. We followed up each returned survey via email or telephone to clarify and verify their responses.

Final survey development

We used the responses from these 16 organizations, discussions with several outside clinical informatics' experts, and our experience in the field of clinical information systems to create our final, shorter and more focused 6 question survey [see Table 1.] along with a weighted scoring system. Briefly, we created this 6-item survey by eliminating questions in which nearly all the organizations provided similar or highly correlated responses (e.g., 4 questions regarding order entry system attributes were replaced by one question that asked whether they had implemented CPOE for medications, clinical laboratory tests, or radiology examinations), those questions that upon further reflection did not add meaningful information to our survey (e.g., descriptive information about the hospital such as number of annual admissions or make of the CPOE system), those questions that were difficult for interviewees to understand and/or answer (e.g., the type of order communication interfaces in existence), nearly all of the detailed questions regarding specific types of clinical decision support (e.g., drug-drug interaction checking or preventive care reminders), questions about additional CPOE-related applications such as an electronic medication administration record or use of barcode medication administration, and we combined all of the organizationally-related questions (e.g., presence of a Chief Medical Informatics/Information Officer – CMIO, or type and extent of information technology support in clinical areas) into one question (i.e., Do you have a CPOE-related oversight committee with physicians on it?). All of the questions were scored on a scale from 0-1, except for "number of years the CPOE system has been active". For this question we gave organizations that had been using their CPOE system for greater than 10 years an extra point.

Development of final survey scoring categories

To allow us to combine the answers from the six survey questions, we created an ad hoc scoring system that ranged from 0 – 1 for each of the individual questions. We created the individual scoring ranges for each question based on an extensive review of the literature [13] and discussions with various experts in the field clinical information systems. Table 1 shows the actual scoring categories along with each question.

Table 1 - Questions from the final version of the survey.

Questions	Scoring categories
What percentage of all orders throughout the hospital is written by physicians using CPOE?	<50% = 0; ≥ 50% = 1
Which of the following types of orders do the physicians enter – medications, labs, radiology?	≤ 1 type=0; 2 types =0.5; 3 types =1
Can you sign in once to get to all aspects of your clinical information system?	no = 0; yes = 1
Do you have any decision support (e.g., order sets; drug-drug interactions; pop-up alerts) at the time of order entry?	≤ 1 type=0; 2 types =0.5; 3 types =1
Do you have a CPOE system-related committee (with physicians on it)?	no = 0; yes = 1
How many years has the CPOE system been active?	< 1yr = 0; <3 yrs = 0.25; <5 yrs = 0.5; <10 yrs = 1; ≥ 10 yrs = 2

Final survey site selection

We used the HIMSS Analytics database to identify 448 hospitals that identified themselves as "having implemented CPOE" (i.e., all sites reporting to have CPOE (the presence of a CPOE application that was "live and operational") from over 4500 hospitals. We then added all 113 Veterans' Affairs (VA) hospitals to this list. Using this combined list of 561 hospitals we systematically began to contact them. We sent email messages to all organizations for whom we could find email addresses. In addition, we telephoned all organizations for which we had a telephone number. For each of the remaining organizations we searched on the Internet in an attempt to find a contact number for each organization. In the end, we were able to find either a telephone number or an email address for all 561 organizations, although we were not able to verify their accuracy. We attempted to contact all of these organizations. Following completion of the survey, we carefully compared respondents to non-respondents using number of licensed beds and ownership as proxies for "other important organizational characteristics"[14], as well as geographic location and teaching status.

Results

Final survey administration

We were able to conduct telephone interviews with staff at 299 of the 561 acute care hospitals targeted, discovering that a large number (89) listed as having CPOE did not,

and finding that a number of hospitals (34) have policies against doing surveys. Consequently, using the Institute for Social and Economic Research (ISER) calculation for response rate [15], which takes into consideration the ineligibility of some sites (e.g., listed as having CPOE when they actually do not), our response rate (based on 176 valid interviews) was 47%.

Using logistic regression (forward, stepwise, likelihood ratio), we compared respondents and non-respondents. There was no difference in response rate with respect to number of licensed beds or teaching hospital status. Ownership type differed between respondents and non-respondents. The VA hospitals were most likely to respond, followed by other governmental hospitals, private institutions and non-profits (compared to VA, governmental LR = .63, private not for-profit LR = .38 and private for-profit LR = .35; $p<0.001$). There was also a small difference in response rate by geographic location with a site in the Northeast region most likely to be a responder, followed by the West, South and the Midwest, in that order (compared to South, Northeast LR = 2.11, West LR= 1.65 and Midwest LR = .93; $p=0.007$).

The length of time that CPOE had been in place ranged from 6 months to 25 years (median = 5 years). The percentage of orders entered electronically ranged from 1-100% (median = 90.5%). Greater than 96% of the sites used CPOE to enter pharmacy, laboratory and imaging orders; 82% were able to access all aspects of the clinical information system with a single sign-on; 86% of the respondents had at least 3 types of clinical decision support (order sets, drug-drug interaction warnings, and pop-up alerts) even though nearly all hospitals were community hospitals with commercial systems; and 90% had a CPOE committee in place.

Using our ad hoc scoring system (Table 1 Scoring Categories), we calculated a numeric infusion score (on a scale from 0 – 7) for each hospital (e.g., CPOE installed for 6 years receives score of 1). Fifty percent of the hospitals had an infusion score greater than or equal to 5.25 (7 max.). Figure 1 shows a graph of the cumulative percentage of all organizations surveyed in terms of total infusion score attained from the survey. From this one can identify the infusion level scores that place an organization in the 10^{th} (3.8) through the 90^{th} (5.9) percentiles.

Discussion

The most striking finding from our survey was the high degree of infusion at the vast majority of institutions we surveyed. We interpret this to mean that if organizations decide to, and are actually able to implement a CPOE system, they do it quite thoroughly as evidenced by a 50^{th} percentile survey score of 5.25 (7 max.). In addition, it was reassuring to find that most of the organizations surveyed (86%) had implemented at least some form of clinical decision support in three different areas. This is critical from our view since the vast

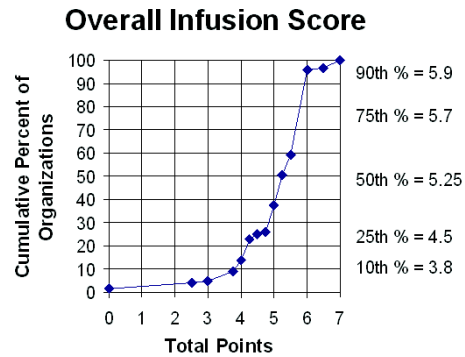


Figure 1 - Graph of cumulative percent of organizations vs. overall infusion score.

majority of the benefits accrued from CPOE result from the clinical decision support that can be simultaneously provided [16].

Also of interest was the fact that over half of the hospitals surveyed had been using CPOE for more than 5 years. This goes to reinforce our belief that complete implementation and incorporation of CPOE into the normal clinical workflow takes a significant (at least 5 years) period of time. In addition, the relatively high percentage of all orders entered via CPOE (median 90.5%) highlights the fact that once CPOE is in place, having both paper-based, manual order entry along with CPOE is not a good work practice and potentially increases the complexity of clinical workflow processes and can lead to additional unintended consequences [17].

Study limitations

The major limitation of this survey was the relatively low response rate of 47%. On the other hand, there were no differences between respondents and non-respondents with respect to hospital size or teaching status and only small differences with respect to ownership type and geographic location.

Conclusion

We have developed and tested a new CPOE infusion measure that demonstrates a very high level of infusion at those hospitals that have successfully implemented CPOE. Using this measure we have identified several key components of CPOE that nearly all successful sites have installed. While CPOE has not been widely adopted after over 30 years of experimentation [18], there is still much that can be learned from the sites that have successfully implemented these systems and incorporated various real-time clinical decision support system features into the clinicians' workflow.

Acknowledgements

This work was supported in part by grant LM06942-02 from the National Library of Medicine. We like to acknowledge the tremendous efforts on the parts of our telephone survey team including: Caroline Wetherton, Ashley Jones. We would like to

give a special thanks to those individuals who helped with our first survey: Abel Kho, Eric Poon, Jason A. Lyman, Karen L. Hughart, Brian R. Jacobs, Terri Andrews.

Appendix A. Initial Survey sent to 16 organizations with known successful Computer-based Provider Order Entry implementations.

Hospital characteristics
Number licensed acute care beds
Annual admissions
Annual Emergency Department visits
Hospital Type
Clinical service profile
CPOE Availability [Measured or (est.)]
Maker of CPOE system
Years since first in-patient go-live
Hospital Locations with CPOE (% of units)
ICU's
General Nursing units
Emergency Department

Order Entry System Attributes
Medications/Therapeutics i.e., IV medications, Blood products, Chemotherapy, Parental nutrition
Diagnostic tests(i.e., Pathology laboratory, Radiology)
Clinical consults / referrals (i.e., Specialty in-pat. consultations, out-patient, community nursing)
Patient Care or Ancillary Therapies (i.e., Nursing, Dietary, PT, OT, RT)
Order Communication Interfaces (e.g., Paper, fax, electronic 1-way, electronic 2-way, etc.)
Pharmacy
Clinical Pathology laboratory
Radiology

Clinical Decision Support Types Available Medication ordering
Subsequent or Corollary orders (e.g., order ACE inhibitors prompts for serum creatinine level [19]
Context-sensitive information retrieval (e.g., Info based on patient or drug-specific information [20]
Patient-specific relevant data displays (e.g., potassium levels when ordering digoxin)
Dose adjustments based on Renal impairment [21], Age (either pediatrics or elderly), Weight
Automatic Formulary checking
Order by indication (e.g., enter hypertension, get suggested medications)
Drug-drug interaction checking [22]
Drug-allergy interaction checking [23] by specific drug or by drug family?
Display of medication cost
Alternative/substitute Medication suggestions
Order sets [24]
Hospital-wide
Departmental
Personal
Diagnostic test ordering
Duplicate order checking
Display past results before ordering new test [25]
Charge display [26]
Preventive care reminders [27]

CPOE-related applications
Medication Administration Record [28]
Barcode Medication Administration [29]

CPOE-related Organizational support
Chief Medical Informatics/Information Officer
Multi-disciplinary CPOE design review/oversight committee
Multi-disciplinary Clinical Decision Support review/ oversight committee
By-laws modified to account for CPOE
IT support in clinical areas, how available
Multi-disciplinary Clinical Decision Support review/ oversight committee
By-laws modified to account for CPOE
IT support in clinical areas, how available

References

[1] Ash JS, Gorman PN, Seshadri V, Hersh WR. Computerized physician order entry in U.S. hospitals: results of a 2002 survey. J Am Med Inform Assoc. 2004 Mar-Apr;11(2):95-9.

[2] Cutler DM, Feldman NE, Horwitz JR. U.S. adoption of computerized physician order entry systems. Health Aff (Millwood). 2005 Nov-Dec;24(6):1654-63.

[3] Davies Award winners: http://www.himss.org/ASP/ davies_organizational.asp

[4] Doolan DF, Bates DW, James BC. The use of computers for clinical care: a case series of advanced U.S. sites. J Am Med Inform Assoc. 2003 Jan-Feb;10(1):94-107

[5] Saga V, Zmud RW. The Nature and Determinants of IT Acceptance, Routinization, and Infusion. Proceedings of the IFIP TC8 Working Conference on Diffusion, Transfer and Implementation of Information Technology 1993; 67 - 86.

[6] Orlikowski, WJ, Robey, D. Information technology and the structuring of organizations. *Information Systems Research*, 2(2), 143-169;1991.

[7] Zmud RW, Apple LE. Measuring technology incorporation / Infusion. J Prod. Innov. Manag. 1992; 9:148-155.

[8] Sittig DF, Thomas SM, Campbell E, et al. Consensus Recommendations for Basic Monitoring and Evaluation of In-Patient Computer-based Provider Order Entry Systems. Information Technology and Communications in Health (ITCH), Victoria, BC Feb. 2007. (in press).

[9] Osheroff JA, Pifer EA, Teich JM, Sittig DF, Jenders RA. *Improving Outcomes with Clinical Decision Support: An*

Implementer's Guide. Health Information Management and Systems Society, 2005.
[10] Kuperman GJ, Gibson RF. Computer physician order entry: benefits, costs, and issues. Ann Intern Med. 2003 Jul 1;139(1):31-9. (used to find CPOE features)
[11] Haug PJ, Gardner RM, Tate KE, et al. Decision support in medicine: examples from the HELP system. Comput Biomed Res. 1994 Oct;27(5):396-418.
[12] Kuperman GJ, Teich JM, Gandhi TK, Bates DW. Patient safety and computerized medication ordering at Brigham and Women's Hospital. Jt Comm J Qual Improv. 2001 Oct;27(10):509-21.
[13] www.ohsu.edu/academic/dmice/research/cpoe/biblio.shtml
[14] Longo DR, Hewett JE, Ge B, Schubert S. The long road to patient safety: A status report on patient safety systems. JAMA. 294;2858-2865, 2005.
[15] Lynch P, Beerten R, Laiho J, et al. Recommended standards for final outcome categories and standard definitions of response rate for social surveys. Institute for Social and Economic Research, University of Essex ISER Working Papers No. 2001-21, www.iser.essex.ac.uk/pubs/workpaps/pdf/2001-23.pdf
[16] Kaushal R, Jha AK, Franz C, et al. Brigham and Women's Hospital CPOE Working Group. Return on investment for a computerized physician order entry system. J Am Med Inform Assoc. 2006 May-Jun;13(3):261-6.
[17] Campbell EM, Sittig DF, Ash JS, et al. Types of unintended consequences related to computerized provider order entry. J Am Med Inform Assoc. 2006 Sep-Oct;13(5):547-56.
[18] Sittig DF, Stead WW. Computer-based physician order entry: the state of the art. J Am Med Inform Assoc. 1994 Mar-Apr;1(2):108-23.
[19] Overhage JM, Tierney WM, Zhou XH, et al. A randomized trial of "corollary orders" to prevent errors of omission. J Am Med Inform Assoc. 1997 Sep-Oct;4(5):364-75.
[20] Cimino JJ, Elhanan G, Zeng Q. Supporting infobuttons with terminological knowledge. Proc AMIA Annu Fall Symp. 1997;:528-32.
[21] Chertow GM, Lee J, Kuperman GJ, et al. Guided medication dosing for inpatients with renal insufficiency. JAMA. 2001 Dec 12;286(22):2839-44.
[22] Ishikura C, Ishizuka H. Evaluation of a computerized drug interaction checking system. Int J Biomed Comput. 1983 Jul;14(4):311-9.
[23] Kuperman GJ, Gandhi TK, Bates DW. Effective drug-allergy checking: methodological and operational issues. J Biomed Inform. 2003 Feb-Apr;36(1-2):70-9.
[24] Payne TH, Hoey PJ, Nichol P, Lovis C. Preparation and use of preconstructed orders, order sets, and order menus in a computerized provider order entry system. J Am Med Inform Assoc. 2003 Jul-Aug;10(4):322-9.
[25] Tierney WM, McDonald CJ, Martin DK, Rogers MP. Computerized display of past test results. Effect on outpatient testing. Ann Intern Med. 1987 Oct;107(4): 569-74.
[26] Bates DW, Kuperman GJ, Jha A, Teich JM, et al. Does the computerized display of charges affect inpatient ancillary test utilization? Arch Intern Med. 1997 Nov 24;157(21):2501-8.
[27] Dexter PR, Perkins S, Overhage JM, et al. A computerized reminder system to increase the use of preventive care for hospitalized patients. N Engl J Med. 2001 Sep 27;345(13):965-70.
[28] Mekhjian HS, Kumar RR, Kuehn L, et al. Immediate benefits realized following implementation of physician order entry at an academic medical center. J Am Med Inform Assoc. 2002 Sep-Oct;9(5):529-39. (eMAR)
[29] Patterson ES, Rogers ML, Render ML. Fifteen best practice recommendations for bar-code medication administration in the Veterans Health Administration. Jt Comm J Qual Saf. 2004 Jul;30(7):355-65.

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

E-Healthcare in India: Critical Success Factors for Sustainable Health Systems

Udita Taneja[a], Sushil[b]

[a] *University School of Management Studies, GGS Indraprastha University, Delhi, India*
[b] *Department of Management Studies, IIT Delhi, India*

Abstract

As healthcare enterprises seek to move towards an integrated, sustainable healthcare delivery model an IT-enabled or e-Healthcare strategy is being increasingly adopted. In this study we identified the critical success factors influencing the effectiveness of an e-Healthcare strategy in India. The performance assessment criteria used to measure effectiveness were increasing reach and reducing cost of healthcare delivery. A survey of healthcare providers was conducted. Analytic Hierarchy Process (AHP) and Interpretive Structural Modeling (ISM) were the analytical tools used to determine the relative importance of the critical success factors in influencing effectiveness of e-Healthcare and their interplay with each other. To succeed in e-Healthcare initiatives the critical success factors that need to be in place are appropriate government policies, literacy levels, and telecommunications and power infrastructure in the country. The focus should not be on the IT tools and biomedical engineering technologies as is most often the case. Instead the non-technology factors such as healthcare provider and consumer mindsets should be addressed to increase acceptance of, and enhance the effectiveness of, sustainable e-Healthcare services.

Keywords:

e-Healthcare, critical success factors

Introduction

Proponents of adopting increasingly high levels of technology in the healthcare field justify this by stating that technology helps to improve the quality of healthcare provided while reducing overall costs. Qualitatively it is easy to see the level of healthcare being improved but as to how costs are cut is not so obvious. In this increasingly cost-conscious era, however, rigorous studies must be undertaken to measure the effectiveness of adopting newer ways of delivering healthcare. In the developing world e-Healthcare is often proposed as a solution to certain healthcare problems – accessing the rural population and trying to bridge the huge gap between the demand and supply of healthcare services [1, 2]. As resources are limited, especially in a developing country, it is important to ensure success of new and sustainable ways of healthcare delivery [3, 4]. Based on a literature survey we identified ten critical success factors that influence effectiveness of e-Healthcare in terms of their impact on increasing reach and reducing cost of healthcare delivery. We then attempted to determine the relative importance of these factors as well as understand the interplay among them in the Indian context.

Critical success factors for e-healthcare

Data warehousing and data mining

Appropriate data warehousing and data mining techniques are important as online patient record storage and retrieval of relevant data for medical decision making is an indispensable component of any e-Healthcare paradigm [5, 6].

Expert systems

Decision support systems for diagnoses and for demographic analysis for public health programs are extensively applied in the delivery of e-Healthcare [7].

Data access control

The healthcare sector in the West has stringent requirements for data security and controlling access to confidential patient data. In the United States the Health Insurance Portability and Accountability Act (HIPAA) was passed for improved patient privacy and data security [8]. In India, however, laws and standards for protecting a patient's electronic healthcare record are still in their infancy.

Biomedical engineering technology

Appropriate biomedical engineering technology is necessary to support telemedicine applications [9].

Telecommunications and power infrastructure

Reliable networks will be of prime importance as e-Healthcare activities increase [10]. Nowhere is this more important than in India where an uninterrupted power supply is yet to be realized in most parts of the country. This will need to be coupled with a high-bandwidth, zero-downtime telecommunications network to support e-Healthcare delivery.

Government policies

Government policies concerning healthcare, education, infrastructure, technology, insurance, and legal issues all have a bearing on the success of e-Healthcare in India. A radical change in the healthcare delivery process is necessary to achieve success in the e-Healthcare arena [11]. It is,

therefore, important for the government to be actively involved in this paradigm shift in healthcare delivery.

Healthcare insurance

Health insurance is another sector where the entry of third party administrators and private players is set to change the way healthcare coverage is provided in India. Providing insurance cover for illness as well as maintaining wellness is another factor to be taken into account.

Literacy levels

A literate population with a minimum level of computer awareness is essential for the success of e-Healthcare. Technology is useless when faced with ignorance and an inability to use it appropriately and effectively.

Consumer mindset

A literature review on the effect of telemedicine on doctor-patient communication showed that telemedicine was favored in approximately 80% of the studies [12]. In India, however, there exists a face-to-face culture that will have to be overcome and the use of technology actively encouraged to ensure the success of e-Healthcare.

Healthcare provider mindset

To be able to fully utilize the potential of e-Healthcare, healthcare providers will have to be responsive and committed to telemedicine [13]. Effective change management will be necessary to overcome resistance to adapting to new ways of delivering healthcare.

In light of the above, we assessed the critical success factors that influence effectiveness of e-Healthcare in terms of their impact on increasing reach and reducing cost of healthcare delivery.

Methods

Analytic Hierarchy Process (AHP) [14] was used to determine the relative importance of the success factors that influence the effectiveness of e-Healthcare. Interpretive Structural Modeling (ISM) [15] was used to determine the interplay between the success factors that influence effectiveness of e-Healthcare and establish the strategic drivers necessary for success. The independent variables correspond to the ten success factors influencing the effectiveness of e-Healthcare. The dependent variable is effectiveness of e-Healthcare. Increasing reach of healthcare services and reducing cost associated with healthcare delivery drives the effectiveness of e-Healthcare. These are used as the performance assessment criteria for measuring effectiveness of e-Healthcare. In light of the above, reach and cost were treated as the mediating variables connecting the independent variables and the dependent variable. This is shown in Table 1.

Table 1 – Analytical tools and associated variables

Analytical Tool	Variables
AHP	Objective: Select critical success factors influencing the effectiveness of e-Healthcare initiatives Criteria: Increasing reach, Reducing costs Alternatives: Data warehousing and data mining, Expert systems, Data access control, Biomedical engineering technology, Telecom and power infrastructure, Government policies, Healthcare insurance, Literacy levels, Consumer mindset, Healthcare provider mindset
ISM	Issues: Data warehousing and data mining, Expert systems, Data access control, Biomedical engineering technology, Telecom and power infrastructure, Government policies, Healthcare insurance, Literacy levels, Consumer mindset, Healthcare provider mindset Contextual relationship: Has an impact on

AHP

Based on the above objective, two criteria and ten alternatives, a vector ranking the ten alternatives was obtained from a group AHP performed on the inputs provided by individual healthcare providers. A Microsoft Excel spreadsheet was programmed to compute the group AHP rank vector. The ranking in this vector denotes the perceived importance of the alternatives.

ISM

The ten critical success factors were treated as issues to be assessed within the framework of a contextual relationship. The contextual relationship refers to each issue having an impact on another issue. These success factors were used to develop a group ISM based on the inputs provided by the same individual respondents. To develop a group ISM, the majority answer was taken for each question from the individual answers provided by the respondents. GMU ISM software, a Microsoft DOS based software, was used to perform the ISM calculations. The ISM shows, in a graphical form, the contextual relationship between the issues in terms of whether an issue has an impact on another issue. This helped determine the strategic driver issues that have an impact on the other issues.

Data collection

An exploratory study was carried out with data gathered using a survey instrument tailored for AHP and ISM and consisting of two sets of questionnaires. The questionnaires were personally given and explained to thirty-one healthcare providers out of which eighteen responded. Eleven of these respondents were doctors from a large

government teaching hospital, four were doctors in private practice, one was an e-healthcare manager for a large private chain of hospitals, and two responses were obtained from academics specializing in information technology and healthcare management.

Results

The results obtained from the analytical tools used on the questionnaire responses from the healthcare providers surveyed are given below.

AHP results

A group AHP computation was performed on the collective set of inputs obtained from all the respondents. The group AHP results, in terms of the ranking of the criteria and alternatives with their relative weights, are as given in Table 2. Healthcare providers, as a group, perceive increasing reach to be more important than reducing cost in increasing the effectiveness of e-Healthcare. Keeping in mind the large population of India that is currently underserved in terms of basic healthcare services, any initiative that attempts to provide healthcare services should increase the reach of services.

Healthcare providers rank literacy levels as the most important factor influencing reach. Consumer and healthcare provider mindsets along with telecom and power infrastructure follow in close succession as to their importance in increasing reach. Without a literate population it is not possible to fully utilize e-Healthcare services. The mindsets of people as well as the basic infrastructure issues are important for reaching a larger number of people.

With respect to reducing cost, literacy levels are once again considered the most important factor, with telecom and power infrastructure a close second, and government policies in third place. For e-Healthcare to service the country a literate population is desirable. When dealing with an illiterate population greater costs are incurred and this is reflected in ranking literacy levels as the number one factor affecting the cost of e-Healthcare. The infrastructure costs also have a bearing on e-Healthcare costs and hence they are in second place. Government policies, in third place, also have an impact on the cost of e-Healthcare.

Table 2 – Group AHP ranks

Criteria

Increasing reach	0.5778
Reducing cost	0.4222

Alternatives with respect to increasing reach

Literacy levels	0.1951
Consumer mindset	0.1298
Healthcare provider mindset	0.1274
Telecom / power infrastructure	0.1262
Government policies	0.1042
Expert systems	0.0703
Data access control	0.0665
Healthcare insurance	0.0644
Data warehousing / data mining	0.0632
Biomedical engineering technology	0.0531

Alternatives with respect to reducing cost

Literacy levels	0.1644
Telecom / power infrastructure	0.1537
Government policies	0.1337
Consumer mindset	0.1070
Healthcare provider mindset	0.1055
Healthcare insurance	0.0887
Data warehousing / data mining	0.0877
Expert systems	0.0679
Biomedical engineering technology	0.0536
Data access control	0.0378

Overall ranking

Literacy levels	0.1821
Telecom / power infrastructure	0.1378
Consumer mindset	0.1201
Healthcare provider mindset	0.1181
Government policies	0.1167
Healthcare insurance	0.0747
Data warehousing / data mining	0.0736
Expert systems	0.0693
Data access control	0.0544
Biomedical engineering technology	0.0533

In the overall ranking, literacy levels are considered the most important factor influencing the effectiveness of e-Healthcare. The telecom and power infrastructure is ranked second with consumer and healthcare provider mindsets a close third and fourth. As mentioned earlier a literate population is a prerequisite for effective e-Healthcare delivery both in terms of increasing reach and reducing cost. Without telecom and power any e-Healthcare initiative will not function properly. The mindsets of both the healthcare providers and consumers are also important for e-Healthcare to gain acceptance as a means of giving and receiving healthcare services. The technology aspects of the business such as appropriate IT and biomedical engineering technologies are ranked the lowest in their impact on e-Healthcare.

ISM results

A group ISM was developed from the complete set of inputs obtained from all the respondents and the digraph is shown in Figure 1. Government policies are the most important strategic driver having an impact on the other factors. Relevant government policies need to be in place to accelerate the pace of infrastructure development in the country without which e-Healthcare cannot hope to achieve any measure of success. Government policies also have a direct impact on the literacy levels in the country that will, in turn, drive changing consumer mindsets.

The telecom and power infrastructure in the country is the second most important strategic driver for e-Healthcare initiatives. With an adequate infrastructure in place the healthcare provider mindset will be influenced positively in favor of e-Healthcare as a successfully delivery mechanism. The information and biomedical engineering technologies are not as critical in e-Healthcare delivery as can be see from the digraph.

Discussion

Healthcare providers, as a group, perceived increasing reach to be more important than reducing cost in increasing the effectiveness of e-Healthcare. In the overall AHP ranking, literacy levels were considered the most important factor influencing the effectiveness of e-Healthcare. The telecom and power infrastructure was ranked second, with consumer and healthcare provider mindsets a close third and fourth. A literate population is a prerequisite for effective e-Healthcare delivery and without telecom and power e-Healthcare will not work. The mindsets of both the healthcare providers and consumers are also important for e-Healthcare to gain acceptance. The technology aspects of the business such as appropriate IT and biomedical engineering technologies were ranked the lowest in their impact on e-Healthcare. The ISM digraph identified government policies as the most important strategic driver. Government policies need to be in place to accelerate the pace of telecom and power infrastructure development in the country. Government policies also have an impact on the literacy levels in the country, also an important driver for e-Healthcare effectiveness. The information and biomedical engineering technologies are not as critical in e-Healthcare delivery as can be see from the digraph. The results from both tools show that non-technology issues such as government policies, telecom and power infrastructure, and literacy levels in the country are more important than technology issues.

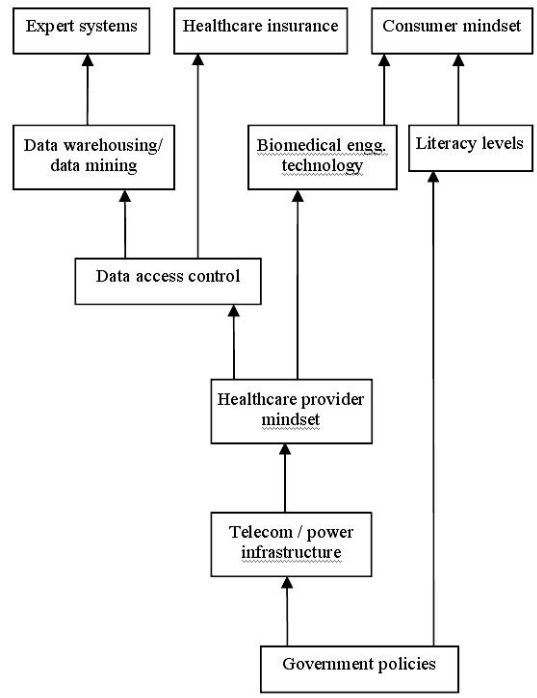


Figure 1 – Group ISM digraph

Conclusion

We have sought to evaluate the ten critical success factors that influence the effectiveness of e-Healthcare in India. We have attempted to determine the relative importance of these factors as well as understand the interplay between them. Increasing reach and reducing cost were the performance assessment criteria that were used. AHP was used to rank the factors. ISM was used to establish a causal relationship diagram between the critical success factors and identify the strategic drivers. To succeed in e-Healthcare initiatives the critical success factors that need to be in place are appropriate government policies, literacy levels, and telecommunications and power infrastructure in the country. The focus should not be on the IT tools and biomedical engineering technologies as is most often the case. Instead the non-technology factors such as healthcare provider and consumer mindsets should be addressed to increase acceptance of, and enhance the effectiveness of, sustainable e-Healthcare services.

References

[1] Kaul B. Health boosters. Business World 2002: 18 March: 26-29.
[2] Singh G. A healthy initiative. Business World 2001: 24 December: 26-27.

[3] Wootton R. Telemedicine and developing countries - successful implementation will require a shared approach. J Telemed Telecare 2001: 7(5) (Suppl 1): 1-6.
[4] Wootton R and Hebert MA. What constitutes success in telehealth? J Telemed Telecare 2001: 7(6) (Suppl 2): 3-7.
[5] Müller ML, Ganslandt T, Eich HP, Lang K, Ohmann C, and Prokosch HU. Towards integration of clinical decision support in commercial hospital information systems using distributed, reusable software and knowledge components. Int J Med Inform 2001: 64(2-3): 369-377.
[6] Staccini P, Joubert M, Quaranta JF, Fieschi D, and Fieschi M. Modeling health care processes for eliciting user requirements: a way to link a quality paradigm and clinical information system design. Int J Med Inform 2001: 64(2-3): 129-142.
[7] Gell G. Side effects and responsibility of medical informatics. Int J Med Inform 2001: 64(2-3): 69-81.
[8] Duncan M, Kleinberg K, Gabler J, Rishel W, Gunasekaran S, and Klein J. Healthcare predictions for management: 2002 and beyond. Gartner Research and Advisory Services, 2002.
[9] Engelmann U, Schröter A, Schwab M, and Meinzer HP. Reality and perspectives in teleradiology: a personal view based on personal experiences. Int J Med Inform 2001: 64(2-3): 449-59.
[10] Aas IHM. A qualitative study of the organizational consequences of telemedicine. J Telemed Telecare 2001: 7(1), 18-26.
[11] Jardine I, Griffiths J, Kelly J, and Navein J. Telemedicine options appraisal for a health district - a strategic clinical and business case. J Telemed Telecare 2001: 7(5) (Suppl 1): 16-19.
[12] Miller, EA. Telemedicine and doctor-patient communication: an analytical survey of the literature. J Telemed Telecare 2001: 7(1): 1-17.
[13] Sjögren LH, Törnqvist H, Schwieler A, and Karlsson L. The potential of telemedicine: barriers, incentives and possibilities in the implementation phase. J Telemed Telecare 2001: 7(5) (Suppl 1): 12-13.
[14] Saaty TL. How to make a decision: The analytic hierarchy process. Eur J Oper Res 1990: 48(1): 9-26.
[15] Sage AP. Methodology for large-scale systems. New York: McGraw-Hill, 1977.

Address for correspondence

Dr. Udita Taneja
Email: utaneja@vsnl.com Phone: +91 98 185 52305

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Health Information Systems Adoption: Findings from a Systematic Review

Maryati Mohd. Yusof[a], Lampros Stergioulas[b], Jasmina Zugic[b]

[a] *Dept. of System Mgmt. & Science, Faculty of Info. Tech. & Science, National University of Malaysia, Malaysia*
[b] *School of Information Systems, Computing & Mathematics, Brunel University, UK*

Abstract

Earlier evaluation studies on Health Information Systems (HIS) adoption have highlighted a large number of adoption problems that were attributed to the lack of fit between technology, human and organisation factors. Lessons can be learned from these evaluation studies by identifying the most important factors of HIS adoption. In order to study the adoption issue, a qualitative systematic review has been performed using a recently introduced framework, known as HOT-fit (Human, Organisation and Technology fit). The paper identifies and highlights the following critical adoption factors: technology (ease of use, system usefulness, system flexibility, time efficiency, information accessibility and relevancy); human (user training, user perception, user roles, user skills, clarity of system purpose, user involvement); organisation (leadership and support, clinical process, user involvement, internal communication, inter organisational system, as well as the fit between them. The findings can be used to guide future system development and inform relevant decision making.

Keywords:

Medical informatics, informatics, information systems, evaluation studies, framework.

Introduction

Although we are currently witnessing a rapid growth in the development and use of Health Information Systems (HIS), HIS stakeholders (purchasers, patients and physicians) experience familiar levels of disappointment associated with more general system development. Such disappointing systems are mainly caused by the inability of HIS to match the work patterns and settings of healthcare, which results in barriers to the use of HIS. In general IS development, past experience shows that overemphasizing technical issues while overlooking social issues can result in system failure [1]. For HIS, the lack of fit among the main organisational elements such as organisational strategy, structure, management process, the skills of people and their roles, and technology contributes to a large number of system failures in public health [2]. The most significant socio-technical issues therefore need to be identified in order to enable HIS to improve, drawing upon the past experiences, and to assist the health care process. The aim of this paper is to present the main findings of a systematic review of selected case studies on HIS adoption in clinical practices.

Theoretical framework

The proposed Human, Organisation and Technology-fit (HOT-fit) evaluation framework was developed after a critical appraisal of previous HIS and IS evaluation studies [3]. It makes use of the IS Success Model [4,5] in categorizing its evaluation factors, dimensions and measures. The IS Success Model is adopted on the basis of its comprehensive, specific evaluation categories, extensive validation and its applicability to HIS evaluation. In addition, the IT-Organisation Fit Model [6] is used to complement the IS Success Model by integrating its featured organisation factors and the concept of fit between the human, organisation and technology factors (See Figure 1).

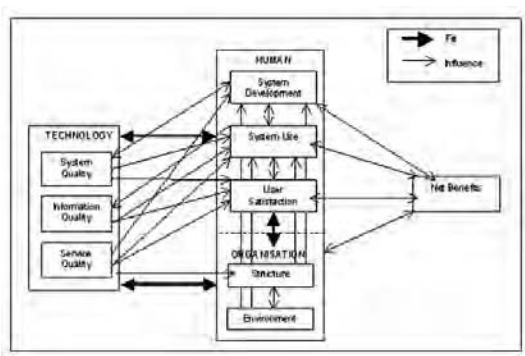


Figure 1 – Human-Organsation-Technology Fit (HOT-fit) framework

Methods

The research approach employs a subjectivist, systematic review strategy with a combination of both qualitative and quantitative methods. A subjectivist approach was chosen in order to gain an extensive understanding of the healthcare context surrounding the HIS through detailed, insightful explanation of the study [7]. A systematic review was adopted to synthesise both qualitative and quantitative research in a comprehensive manner.

This systematic review consists of a number of discrete activities [8,9,10]: 1) identification of the need for a review, 2) formulation of a research question, 3) development of a review protocol 4) location of studies, 5) selection of primary studies, 6) quality assessment of included studies, 7) data extraction, 8) data analysis, 9) interpretation of results, 10) report writing.

The search for articles that addressed issues pertinent to HIS evaluation was conducted using a number of databases ranging from medical, informatics and engineering field. The search terms were individualized for each database and included terms for: (health or medical) informatics, system, computer based, evaluation and assessment. In this systematic review, evaluation study is defined as a methodical, empirical assessment of health information system in clinical settings. The type of study design selected is *case study* as it provides understanding and insight into the research question. Thus, other study designs, such as experiment and survey are not included in this systematic review. HIS stands for any computer based information systems that involves human interaction used in healthcare settings. All computer or knowledge based training and education systems for healthcare professional are excluded in the review, as they are not directly related to clinical care (or clinical use of technology). Searches were limited to human subjects and articles published in English, between 1985 and 2005, and were completed from January to February 2006. The studies were searched both electronically and manually in order to cover a comprehensive range of literature. The selection of studies involves a multi-stage process [8] (See Figure 2).

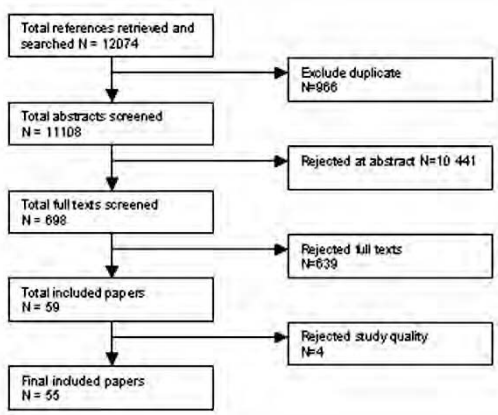


Figure 2 – Methods of study selection

Results

Study description

Fifty five studies from various resources listed in Appendix A and at www.ftsm.ukm/mmy/sr1985-2005list were selected to be analysed. The participants in the selected studies range from top managers to clerical staff, with doctors and nurses being the main informants. Their numbers vary from 3 to 1500. Most studies employ a combination of these data collection methods: documentation, archival records, interview, questionnaire, observations and physical artifacts. About half of the studies used a mix of qualitative and quantitative methods. The review has found that, nearly 40% of the studies focused on a combination of technology, human and organisation factors of the HIS adoption; the rest of the studies focused on either single or combination of two of those factors. In terms of the evaluation time, 44 out of 55 (80%) studies took a summative approach, while only 9 studies chose both formative and summative approaches and 2 studies used a formative approach. The type of HIS evaluated and their occurrences are as follows: Electronic Medical Record (EMR) (13); Telemedicine (10); Computerised Provider Order Entry System (CPOE), Clinical IS (CIS) (8), Picture Archiving and Communication System (PACS) (7); Clinical Reminder System (CRS) (4); Patient Care IS, Radiology IS (RIS), Hospital IS (2) and Laboratory IS (1).

We used the following qualitative research appraisal criteria for assessing all empirical papers to be included in our review by judging seriously flawed papers [11]: clarity of 1) aims and objectives, 2) research design and process by which their findings were produced; appropriateness of analysis method; results credibility and importance; and conclusions drawn. All 55 studies were assessed to be of sound quality. Qualitative quality assessment showed that nearly half of the studies have excellent study quality and the other half have acceptable study quality.

Principal findings

Adoption factors were identified according to technology, human and organisation category outlined in the HOT-fit framework, as well as new evaluation dimension (system development) and measures identified from the themes emerging from the extracted data. Table 1 shows the occurrences of critical adoption factors according to their category and dimension.

Overall, data analysis shows that successful HIS adoptions tend to have many occurrences of human and organisation factors relevant to strong human support and collaboration namely user involvement, relationship with IT people, teamwork, leadership, top management support, system champion and medical sponsorship. Highly positive cases of HIS adoption also had a clear purpose and benefits of system and employed effective installation strategy. On the contrary, unsuccessful HIS adoptions showed a high number of conflict among stakeholders, negative user perception and user resistance. Interestingly, the occurrences of technology factors are almost equally the same within the positive and negative adoption. In the system quality dimension, the following measures are cited more in negative adoption compared to positive ones: system flexibility, system sophistication and turnaround time. Meanwhile, most studies of positive HIS adoption reported a high number of information quality measures in system net benefits, particularly information accessibility, relevancy, legibility and format. In the negative adoption studies, apart from information relevancy, other information quality measures were not cited.

Table 1 – Dominant factors of HIS adoption

Technology	Human	Organisation
System Quality	**System Development**	**Organisational Structure**
Ease of use N=28 Usefulness N=24 Time (response & turnaround) N=23 Flexibility N=18 Reliability N=13 Technical support N=13	User involvement N=13 System purpose N=12 Installation strategy N=11	Leadership & support N=34 Clinical process N=20 Internal communication N=11 Resources N=10
Information Quality	**System Use**	**Organisational Environment**
Accessibility N=26 Relevancy N=20 Completeness N=13 Legibility N=12 Format N=10	User training N=24 User perception N=18 User skills N=17 User attitude N=10 User roles N=14	Inter organisational systems N= 10

Technology

Ease of use was identified as one of the highest adoption factors cited in this review. In many studies, positive HIS adoption has been linked to the ease of system use in a number of ways including friendly user interface, 'point-and-click' features and free-text entry [12,13]. Ease of use is a major incentive to system use. According to a physician, "… it needs to be convivial, easy to use, and functional. If you don't have that, physicians get fed up and don't want to use it anymore" [14]. Unfavourable comments on ease of use include 'cumbersome', 'unintuitive' and 'restrictive' [15,16]. Poor usability results in many problems, such as error in the use of patient record as the record was not visible enough and excessive complexity of screens [17,18].

Within the system quality dimension, system usefulness was the second most quoted factor. Various systems functions including query, messaging, data storage and access, reporting, image manipulation, and educational tool were described as beneficial to users. Examples of useful system features include prescription writer, templates, document imaging, minimal free text format, graphical display and order sets.

Time efficiency in terms of response time and turnaround time of HIS was highly regarded by user, as reported in many studies. Users viewed this as the most important aspect of CPOE to be considered; perceptions of time vary according to role [19,20] An organisation where a CIS was implemented had to make trade-offs between adding new functionality and keeping the system response speed to a level that was acceptable to clinicians on a constant basis [21]. Meanwhile, many studies reported long turnaround time that affected healthcare delivery. Examples include accumulated time for entering orders for a whole group of patients at a time that can add up to hours [20]. Data entry and capture was time consuming, seldom error-free and distanced the user and patient, thus resulting in patient dissatisfaction [22].

System flexibility has been frequently described as one of the most important factors of HIS adoption. Flexibility of a system was viewed as the degree to which it personalized departmental or individual requirement and the ability to integrate to other HIS. For example, the Southeast Health Center staff felt that the EMR was easy to adopt and flexible enough to accommodate different provider requirements. Moreover, the EMR was found to promote increased efficiency and effectiveness in the billing process, attributed largely to the system ability to automatically link clinical diagnoses with the disease and health related problem codes [12].

Positive effects of HIS adoption were commonly associated with information accessibility and relevancy. Information accessibility benefits users and organisation in many ways. Increased access and availability of EMRs patient information to multiple users has resulted in increased efficiency (e.g., less staff time searching for charts; time saver; speedy and improvement of physicians decision making) [12,23,24]. Information relevancy was highly regarded by clinicians. For instance, an Imaging System was found to provide real patient experience; it helped students associated a visual image to a patient history and eased the linked of verbal reports to visual symptoms [24]. In a telemedicine project, visual information provided an essential dimension for healthcare provision (e.g. they enable detailed assessment of patients, including for example emotional states and skills) [25].

Human

Most successful HIS implementations were attributed to the provision of extensive, continuous user training and support [19,20]. Some users had some basic computing skills before participating in system training. In a practice, for example, staff who were pleased with their training had the chance to familiarize themselves with computers and play computer games, months before the system operation [23]. A common problem identified with training was that time needed for training constantly conflicted with clinicians' busy schedule [18,26,27]. User perception was highly cited as a contributing adoption factor of HIS. Low adoption studies reported the use of HIS that was perceived as "add-on" rather than integral task; cumbersome and clerical functions [28,29].

Many studies cited the influence of user roles and skills on system use. Effective system use in many organisations was impeded by the hierarchical structure of the organisation, which affected the way users perceived the system and unable to respond to its deployment [30]. Residents were reluctant to enter patient data, regarding it to be a traditional task for support staff [28]. A RIS project team acknowledged the importance of having the right combination of personalities and talents to achieve the outcome [31]. In contrast, the dramatic pattern of decline in the frequency of telehealth consultations was attributed to the

lack of physician familiarity with equipments, particularly in emergency situations [25].

Clarity of purpose and expected benefits has been very influential in both highly positive and negative HIS adoption. During the early stage of a successful RIS adoption, the project team set an objective that was very visible to the staff; their completion helped provide a jump-start as well future progress to the whole project [31]. A problematic telehealth installation was partially caused by improper needs assessment [25]. User involvement was also identified as an important HIS factors. At the service level of a telemedicine project, "the optimal approach appears to be one that is highly inclusive in the full involvement of stakeholders…" [25]. Over time, continuous improvements suggested by users yielded to a simpler, user friendly and comprehensive version of the system [22].

Organisation

More than half of the studies cited stalwart individuals in both leadership and support positions as one of the most important factors for implementing HIS. Ash et al., [19] categorized essential people to the successful implementation of CPOE. These people were also very influential on the adoption of other types of HIS. Their roles are classified into three major levels: the leadership level, including the Chief Executive Officer (CEO); the clinician level, including champions, opinion leaders, curmudgeons, and the clinical advisory committees and the bridger level, which includes those who do training and support and interact directly with users (bridgers translate user needs to the higher level and vice versa). In addition, top management support can also be included in this special people category. These special people have been very instrumental and influential in gaining users buy-in through strong support, commitment, persistence and assistance in overcoming fear and even providing technical support.

The adoption of HIS was also commonly linked with the ability of system to fit with the clinical practices. Other relevant issues of clinical process are heavy workloads that leave clinicians with limited time availability to undergo training and use HIS [18,26,27].

Discussion of the findings

Many human factors were identified as barriers and enablers to HIS adoption. User training and user perception were particularly common, followed by user skills/ knowledge, user roles, user involvement, system purpose and installation strategy. The aforementioned critical technology factors such as ease of use, system flexibility, and information usefulness are related to its ability to align with human and organisational needs. Organisation also plays a major role in supporting the uptake of HIS through leadership, resources and moral. Many instances of fit were identified in the included studies. In particular, the fit between system and clinical practices was quite common. Cases with misfit between the system and clinical practices were more common compared to the cases with good fit. In one US hospital, the design of the CPOE was clinician-oriented and the physicians' current workflow was preserved to a great extent [26]. An exemplar of internal fit was identified in a large scale CPOE implemention where both workflow and system requirements were accordingly changed to fit with each other [32]. In contrast, an EMR was disintegrated into current clinical practices and its documentation was interspersed among other more important physician's tasks [27]. Lehoux et al. [30] concluded that problems in the EMR implementation indicated a design that interfered with the clinical practices it was supposed to support.

Conclusions

Many adoption factors of HIS have been identified from this systematic review. Different types of systems have certain unique adoption factors, but they nevertheless share a lot of common adoption factors. The variables that seem to have a major determinant effect on adoption of Health Information Systems are ease of use, system usefulness, system flexibility, technical support, response time/ turnaround time, information accessibility, information relevancy, clarity of system purpose, user involvement, user training, user perception, user skills/ knowledge, user roles, clinical process, champion/ medical sponsorship, leadership, internal communication, and resources. All three technology, human, and organisation factors are equally important, in addition to the fit between them. Findings from this systematic review can be used to inform decision making in the investment and development of HIS and IS in general.

The presented review of HIS adoption factors has also demonstrated the capability of the HOT-fit evaluation framework to address the essential components of IS, namely human, organisation and technology, and the fit between them. This evaluation framework is potentially useful to researchers and practitioners for conducting rigorous evaluation studies of other IS or IT applications in healthcare settings or any other domain. The systematic review can be extended further by including hand searched, high quality conference papers. From this study's findings, recommendations for improving HIS adoption can be identified to guide future development of HIS.

Acknowledgments

We gratefully acknowledge the funding received from both the Public Serv. Dept. of Malaysia and from Univ. Kebangsaan Malaysia (National Univ. of Malaysia) that helped sponsor this study. We would also like to thank Prof. E. Ammenwerth and Prof. J. Wyatt for their valuable suggestions on this review.

References

[1] Anderson JG. Clearing the way for physicians' use of clinical information systems'. Comm ACM 1997;40(8): 83-90.

[2] Southon FCG, Sauer, C., Dampney, K. Lessons from a failed information systems initiative: issues for complex organisations. Int J Med Inform 1999;55(1):33-46.

[3] Mohd. Yusof, M, Paul, RJ, and Stergioulas, LK, Towards a Framework for Health Information *Systems Evaluation.*

Pro 39th Hawaii Int Conf Syst Sc 4-7 January 2006. Kauai, Hawaii, USA.
[4] DeLone WH, McLean ER. Information Systems Success: The Quest for the Dependent Variable. Inform Syst Res 1992;3(1):60-95.
[5] DeLone WH, McLean ER. Measuring e-commerce success: applying the DeLone & McLean Information Systems Success Model. Int J Electron Comm 2004;9(1):31-47.
[6] Scott Morton MS. The Corporation of the 1990s. New York: Oxford University Press; 1991.
[7] Friedman CP, Wyatt JC. Evaluation Methods in Medical Informatics. New York: Springer-Verlag; 1997.
[8] Centre for Reviews and Dissemination. Undertaking Systematic Reviews of Research on Effectiveness: CRD's Guidance for those Carrying Out or Commissioning Reviews. (CRD Report 4 2nd Edition). York: University of York; 2001 Report No.: 4.
[9] Berman NC, Parker, R.A. Meta-analysis: neither quick nor easy. BMC Medical Research Methodology 2002;2.
[10] Pai M, McCulloch, M, Gorman, JD, Pai, N, Enanoria, W, Kennedy, G, Tharyan, P, Colford, JM, Jr. Systematic reviews and meta-analyses: An Illustrated, step-by-step guide. Natl Med J India 2004;17(2):86-95.
[11] Greenhalgh, T & Taylor, R. How to read a paper: Papers that go beyond numbers (qualitative research) Brit Med J 1997;315:740-743.
[12] Townes Jr. PG, Benson DS, Johnston P, Vaughn C. Making EMRs really work: The Southeast Health Center experience. J Ambulatory Care Manage 2000;23(2):43-52.
[13] Wang X, Gardner RM, Seager PR. Integrating computerized anesthesia charting into a hospital information system. Int J Clin Monit Comput 1995 May;12(2):61-70.
[14] Gagnon MP, Lamothe L, Fortin JP, Cloutier A, Godin G, Gagne C, et al. Telehealth adoption in hospitals: an organisational perspective. J Health Organ Manag 2005;19(1):32-56.
[15] Jones MR. "Computers can land people on Mars, why can't they get them to work in a hospital?" Implementation of an Electronic Patient Record System in a UK Hospital. Methods Inf Med 2003;42(4):410-415.
[16] Aarts J, Doorewaard H, Berg M. Understanding Implementation: The Case of a Computerized Physician Order Entry System in a Large Dutch University Medical Center. J Am Med Inform Assoc 2004 May 01;11(3): 207-216.
[17] Ash JS, Gorman PN, Hersh WR, Lavelle M, Poulsen SB. Perceptions of House Officers Who Use Physician Order Entry. Proc AMIA Symp; 1999; 1999. p. 471-5.
[18] Southon FCG, Sauer, C., Grant, C.G. Information Technology in Complex Health Services: Organisational Impediments to Successful Technology Transfer and Diffusion. J Am Med Inform Assoc 1997;4(2):112-124.
[19] Ash JS, Stavri PZ, Dykstra R, Fournier L. Implementing computerized physician order entry: the importance of special people. Int J Med Inform 2003a; 69(2-3):235-250.
[20] Ash JS, Gorman PN, Lavelle M, Payne TH, Massaro TA, Frantz GL, et al. A Cross-site Qualitative Study of Physician Order Entry. J Am Med Inform Assoc 2003c;10:188-200.
[21] Doolan DF, Bates DW, James BC. The use of computers for clinical care: a case series of advanced U.S. sites. J Am Med Inform Assoc 2003 Jan-Feb;10(1):94-107.
[22] Pliskin N, Glezerman M, Modai I, Weiler D. Spreadsheet evaluation of computerized medical records: The impact on quality, time, and money. J Med Syst 1996 APR;20(2):85-100.
[23] Wager K, Lee F, White AW, Ward DM, Ornstein SM. Impact of an Electronic Medical Record System on Community-Based Primary Care Practices. J Am Board of Fam Pract 2000 September/October;13(5):338-348.
[24] Kaplan B, Lundsgaarde HP. Toward an evaluation of an integrated clinical imaging system: identifying clinical benefits. Methods Inf Med 1996 Sep;35(3):221-229.
[25] Moehr JR, Schaafsma J, Anglin C, Pantazi SV, Grimm NA, Anglin S. Success factors for telehealth-A case study. Int J Med Inform 2005 51(4):294-300.
[26] Kuperman GJ, Teich JM, Gandhi TK, Bates DW. Patient safety and computerized medication ordering at Brigham and Women's Hospital. Jt Comm J Qual Improv 2001;27(10):509-521.
[27] Gosbee JW, Clay M. Human factors problem analysis of a voice-recognition computer-based medical record. Pro 6th Annual IEEE Symp Comp-Based Med Syst; 1993; p. 235-240.
[28] Zheng K, Padman R, Johnson MP, Diamond HS. Understanding technology adoption in clinical care: clinician adoption behavior of a point-of-care reminder system. Int J Med Inform 2005 Aug;74(7-8):535-543
[29] Massaro TA. Introducing physician order entry at a major academic medical center: I. Impact on organisational culture and behavior. Acad Med 1993;68(1):20-25.
[30] Lehoux P, Sicotte C, Denis J. Assessment of a computerized medical record system: disclosing scripts of use. Eval Program Plann 1999 11;22(4):439-453.
[31] Tobey ME. Paperless medical records: moving from plan to reality. Radiol Manage 2004;26(4):36-42.
[32] Ash JS, Gorman PN, Lavelle M, Stavri PZ, Lyman J, Fournier L, et al. Perceptions of Physicians Order Entry: Results of A Cross-Site Qualitative Study. Methods Inf Med 2003b;42:313-323.

Appendix A: Sources of Selected Studies

Acad Med (1) Am Heart J (1) Am J Roentgenol (2) Annals of Fam Med (1) Annals Long-Term Care (1) Brit Med J (2) Crit Care Med (1) Eval Program Plan (1) Health Serv Manage Res (1) Home Health Care Serv Q (1) IEEE Trans on Eng Manage (1),Int. Congr Ser (1) Int J Circumpolar Health (1) Int J Med Inform (7) J Adv Nurs (1) J Ambulatory Care Manage (1) J Digit Imaging (1) J Health Org Mgmt. (2) JHIM (1), J.Manage Med (1) J Med Syst (2) J Telemed Telecare (2) J Am Board of Med Pract (1) J Am Med Inform Assoc (5) MD Computing (1) Methods Inf Med (3) Pediatrics (1) HICSS'04 (1) IEE Symp (1) Pro AMIA Symp (1) Proc Int ACM SIGGROUP (1) Radiol Manage (1) Sociology Health Illness (1) Telemed J E-health (2) Jt Comm J Qual Improv (1), Topics Health Info Manage(1).

Address for correspondence

Maryati Mohd. Yusof, email: mmy@ftsm.ukm.my
maryati226@yahoo.com

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Telemedicine Portal and Standardized Clinical Documents

Ivan V. Emelin[a], Radik A. Eltchiyan[a]

[a] *General Research Computer Center of the Russian President's Management Department, Russian Federation*

Abstract

The paper describes the experience of creating a flexible telemedicine portal intended to improve a business process of information exchanges needed to manage real-time telemedicine events. The portal structure allows use of digitally-signed standardized clinical documents based on the HL7 Clinical Document Architecture.

Keywords:

telemedicine, telehealth, web portal, CDA,
digital signature

Introduction

Telemedicine is spreading over the Russian Federation mostly by efforts of the enthusiastic healthcare authorities, physicians and engineers using different kinds of telecommunication channels, communication and terminal equipment for real-time teleconsultations and telelearning [1]. Usually these real-time telemedicine events are rendered by the well-equipped Russian medical centers specialized in cardiovascular surgery, neurosurgery, oncology and other medical disciplines. Their clients are mostly the tertiary hospitals located in many Russian regions. There are also the mobile telemedicine units intended for field use – the telemedicine trucks and even the telemedicine trains. There is no common administration for all the participants of this telehealth network.

There is a lot of preliminary information exchanges needed to manage real-time telemedicine events. Typically these exchanges are made using phones, electronic mail and file transfer protocols. It is a time- and labor consuming error-prone task. Nevertheless it has one substantial advantage – it is quite affordable to the medical staff having limited knowledge of information and communication technologies.

This paper describes the telemedicine portal that may considerably lighten the burden of this preparatory work.

Use case

The real-time telemedicine consultation is to be negotiated by many people including a patient, an attending physician, the administrators of the participating telemedicine units (TU), the consultants and the accountants. The attending physician sends a referral to a local TU administrator. The administrator looks for the appropriate consulting medical center and forwards the referral to the TU administrator of this center (host TU administrator). This administrator checks validity and completeness of the referral, asks for additional patient data if needed, finds the consultants that may fulfill the referral, schedules the videoconference, sends reminders to the participants of the teleconsultation and does a lot of other related duty. This process may iterate before the videoconference.

The consultants write the referral notes after the videoconference and sign them. The host TU administrator sends notes and a bill to his vis-à-vis who in turn forwards the notes to the attending physician and the bill to his accountant.

The similar use cases may be written also for telelecture and for other types of telemedicine events.

Telemedicine portal solution

There are many possible solutions that may help to optimize the business process described above. We had set the following base principles helping to choose the better solution:

1. The solution shall be flexible: the TU administrators shall be able to customize it using non-custom tools.
2. It shall be secure: only the direct participants of the telemedicine event may have access to information related to this event. Clinical documents and their attachments are to be signed electronically. The crypto providers used shall conform to the national requirements.
3. It shall be affordable: the end users may get access to information having no more than office applications and web browser on their computers. They are to be able to use this solution having only base knowledge of these applications.
4. It shall be scalable: it shall support thousands telemedicine events simultaneously.
5. It shall supply the possibility to enter and use the standardized clinical documents based on HL7 CDA Release 2 standard.

The web portal may offer the collaboration environment conforming to these principles. There are a lot of telemedicine portals worldwide intended for similar use, for example TempoBy in Germany [2], Apollo's telemedicine portal in India [3], U.S. Military Health System portal [4].

But they are not flexible enough and not as affordable as we need.

Telemedicine portal structure

Each TU has its own web site on our telemedicine portal. A TU administrator has privileges to customize this site using non-custom authoring tools.

Each real-time telemedicine event is moderated by one TU (host). An administrator of this TU creates a separate site (workspace) for the telemedicine event and gives full access to this site to the administrators of other TUs (clients) participating in this event.

The telemedicine portal contains a central gallery of site templates. There is a separate template for each type of telemedicine event and also for TU site. Each site may have its own gallery of site templates. A TU administrator may construct his own gallery of site templates and may use them instead of templates stored in the central gallery.

This structure is shown on figure 1 as a UML class diagram. It ensures *portal flexibility.*

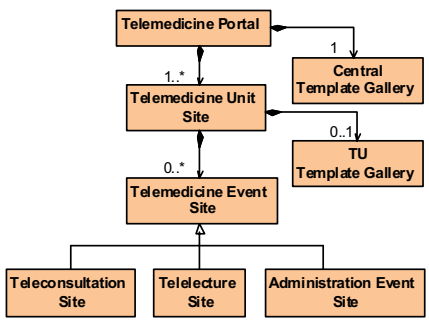


Figure 1- Telemedicine portal structure

A TU administrator grants the proper access rights to a telemedicine event site only to direct participants of this event. All the clinical documents needed for the event will be stored only on event site. There is no central electronic medical records (EMR) database. The telemedicine event site is accessed using the Secure Socket Layer (SSL) and the national crypto providers. All this ensures *security* and *privacy* of the telemedicine portal.

The sample of TU home page is shown on figure 2. It contains general information needed to communicate with this TU, a form library of the current referrals, a list of the scheduled telemedicine events, a list of the consultants and a list of the consulting TUs.

The TU administrator may reject or adopt a referral. If the referral is adopted the TU administrator creates a new entry in the list of the telemedicine events. A new workspace is created for this entry using a teleconsultation site template. The sample of a teleconsultation workspace home page is shown on figure 3.

Figure 2- Sample of a TU home page

Figure 3- Sample of a teleconsultation home page

This page contains the form libraries for the referrals, the discharge notes and the referral notes, a document library for the attachments to the referral (for example medical images) and other lists. It is intended to improve collaboration of TU administrators, the referring physician and the consultants and may be customized by TU administrator to optimize this collaboration.

Standardized clinical documents

There is a special form template for each form library. This template contains XML-schema, allows to create XML-document conforming to this schema and to store the created document in the library. As many documents as needed may be stored in each form library.

The first our idea was to construct the form templates using XML-schemas defined in CDA Release 2 [5] because CDA has good chances to be adopted as ISO standard. But we faced two main problems:

- CDA schemas per se are to complicated for template designer used
- They do not allow digital signatures

To resolve these problems we have constructed an “envelope” containing the following namespaces:

- Slightly modified CDA namespace
- Local namespace
- Digital signature namespace

The local namespace is used to construct XML-forms intended for entering referrals, discharge notes and referral notes. The local namespace is mapped to the CDA namespace immediately before submitting the form to a form library so this form will contain a CDA-compliant part. Digital signatures may secure the CDA namespace, the local namespace or both.

TU administrators may change the appearance of XML-forms using a standard office application. They may add, remove or change form fields and so on. But they cannot change internal XML-schemas using simple tools so the changed forms will still produce standardized XML-documents.

It is also possible to upload XML-documents immediately from the EMR system. In this case the local namespace may be not used at all and the uploaded documents may be viewed using web browser and XSL-transformation.

Conclusions

Our experience of deploying and using the portal described above shows that telemedicine portals may improve the business process of information exchanges needed before and after real-time telemedicine events. The structure of these portals will depend on the business process used. The hospital chain may use the portal together with centralized EMR. Our solution assumes that the participants of the telemedicine portal consider the portal as customizable collaboration environment. Different business processes may require another solution.

The experience of using CDA with digital signatures shows that it may be necessary to standardize the structure of an envelope containing CDA-compliant part and stored on the telemedicine portal. In principle CDA-documents may be sent as the binary enclosures in HL7 V3 messages. But the telemedicine portal stores documents, not messages and there still remains the problem of using XML-digital signatures.

IHE proposes to send XML-digital signatures as separate documents linked to the clinical documents to be signed [6]. This approach is not very convenient if we want to store the signed clinical documents on a portal, view them and modify document appearance using simple tools affordable to many medical users. So our proposal to standardize the envelope structure may help to involve more users in exchange of standardized clinical documents.

Currently our telemedicine portal is used by TU of the Russian President's Management Department. We promote it to several other Russian TUs. The process of portal adoption is rather slowly mostly due to the requirements to buy and install the national crypto providers. But electronic mail as a popular transport for the referrals and other clinical documents becomes less convenient because spam filters used by the biggest e-mail providers reject the messages from time to time and may blacklist other providers. We suppose that this unpleasant situation will fasten the portal adoption.

The method described above of using a web portal for real-time telemedicine events is patent pending in the Russian Federation.

References

[1] Eltchiyan R., Emelin I., Fedorov V., Mironov S., Stoliar V. Telemedicine in Russia. In: M.Fieschi et al. (Eds). Proceedings of the 11th World Congress on Medical Informatics. Amsterdam: IOS Press, 2004, pp. 953-955.

[2] The new telemedicine portal at the University of Munich Hospital. http://www.medical.siemens.com/siemens/en_INT/gg_gs_FBAs/files/references/telemedizin-portal_e.pdf .

[3] Videoconferencing Clinical Programmes. http://www.apollohospitals.com/hcTeleMed_video.htm.

[4] MHS enterprise eHealth portal. http://www.tricareonline.com.

[5] HL7 Clinical Document Architecture Release 2. http://www.hl7.org.

[6] IHE IT Infrastructure Technical Framework. Supplement 2005-2006. Document Digital Signature Integration Profile. June 15th, 2005. http://www.ihe.net/Technical_Framework/upload/IHE_ITI_TF-Supplement_Digital_Signature-PC_2005-06-15.pdf

Address for correspondence

Emelin Ivan, General Research Computer Center, 6, Vozdvisgenka ul., Moscow, 125009, Russian Federation.
Phone +7(495)202-2633,
e-mail emelin@pmc.ru.

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

The Development of an Information System and Installation of an Internet Web Database for the Purposes of the Occupational Health and Safety Management System

I. Mavrikakis[a,1], J. Mantas[a,1] , M. Diomidous[a,1]

[a]*Department of Public Health, Faculty of Nursing, University of Athens, Greece*

Abstract

This paper is based on the research on the possible structure of an information system for the purposes of occupational health and safety management. We initiated a questionnaire in order to find the possible interest on the part of potential users in the subject of occupational health and safety. The depiction of the potential interest is vital both for the software analysis cycle and development according to previous models. The evaluation of the results tends to create pilot applications among different enterprises. Documentation and process improvements ascertained quality of services, operational support, occupational health and safety advice are the basics of the above applications. Communication and codified information among intersted parts is the other target of the survey regarding health issues. Computer networks can offer such services. The network will consist of certain nodes responsible to inform executives on Occupational Health and Safety. A web database has been installed for inserting and searching documents. The submission of files to a server and the answers to questionnaires through the web help the experts to perform their activities. Based on the requirements of enterprises we have constructed a web file server. We submit files so that users can retrieve the files which they need. The access is limited to authorized users. Digital watermarks authenticate and protect digital objects.

Keywords:

occupational health and safety, workflow management, computer networks, World Wide Web, file submission, questonnaries

Introduction

The management of occupational health and safety constitutes a field which concentrates the interest of enterprises. Particularly the enterprises with a modern concept of management, not only care to comply with the legal obligations as far as it concerns occupational health and safety, but also implement a health and safety management system, in order to improve the work conditions. Quality management systems are developed and implemented into enterprises following standards. For health and safety management, ISO 18001 is widely applied. For the development of such systems the workflow and information management between experts and enterprises is very important and interesting. The use of computer networks and the whole infrastructure is significant for this purpose. Such a system can be administered via the WWW. An external expert can handle the information in a way to correct and to intervene to improve the quality of occupational health and safety services. We have developed such a system in order to allow authorized users of each enterprise, mainly those responsible for occupational health and safety, to communicate and submit files. On the other hand, we can use comments for each submission to explain the files in a more executive way and retrieve them. For the purpose of managing and detecting some very important workflows we use questionnaires; the processing of which are very useful for management and decision making in this field. The depiction of interest is important because we adopt new technologies and methods in order to accommodate the procedures for the occupational health and management system.

The interest of the potential users responsible for occupational health and safety is included on the analysis cycle of the software. At the second phase we are searching for the requirements of the enterprises analyzing the answers of the questionnaires. The evaluation of the results by the experts will lead us to create pilot based applications for communication between enterprises and selected branches.

At the second phase of the implementation work we are developing a system which will be responsible to inform the interested parties about the subjects of occupational health and safety management system. On the other hand this system is going to store information about educational

1 Health Informatics Laboratory, Faculty of Nursing, University of Athens, 123 Papadiamantopoulou street, 11527 Athens, Greece, email imavrik@nurs.uoa.gr, jmantas @cc.uoa.gr

issues. A relational database is a scientific tool for the purposes of storing and retrieving useful information.

We insert new technologies to depict the daily situations on the working conditions and in order to support the work of the administration and the work of the employees. We initialize a measuring system based on ISO 18001 in order to evaluate and shape new working conditions. The instability at the working conditions forces us to initiate some rules to protect and promote health and safety of the employees. We are trying to develop a system which will be useful to the users. Such a system after the frequent use must be improved in order to keep the practical features and be more operational. The improvements of such a system is not an easy task. We must find the principle components which support our system and follow some equations to intervene to improve the state of the system. We should pose some limits for the states of the system. In case we exceed them we must enable some actions to correct and to prevent the unsafe conditions. We should restore to the initial regular state the system in case we find that we have exceeded the initial parameters.

In order to achieve the above ideas we should install some means of communication between enterprises. The Internet supports this communication. The protocol on which the Internet is based is TCP/IP [9]. The IP address is a 32-bit number fundamental to Internet addressing and routing. It is also used for addressing the servers and clients. TCP/IP provides a realiable transport service which is used by most Internet Applications. Examples of applications are electronic mail (e-mail), File Transfer Protocol (FTP) and Web Page access through the World Wide Web via HTTP. The evolution of the e-mail was the incorporation into the web.

Another issue that we are going to analyze is videoconferencing. Based on some protocols we are going to insert some categories of research subjects. The compression based on H.263 protocol. The difference in quality of video between the protocols H.263 and H.261. We are searching for the limitations of teleteaching environment which are included on web enviroments. We depict a system of applications and services.

Some issues on computer networks

Interconnection networks play a major role in the performance of modern enterprise management. There are a lot of factors which can differentiate the design of each network.

- Performance requirements (on time and reliable decisions from the informations we collect): All operations are usually performed by explicit message passing or by accessing shared variables. To reduce the message latency, we reduce the idle time of processes and memory access time to remote memory locations [6]. The delay of information trasmission between two points is a very important aspect in case we take some useful decisions. The validity of information is also important, if we can collect information from all the enterprises we have reliable data, otherwise the latency of communication between the enterprises and the experts may cause an unreliable source of information.
- Scalability: As we add more enterprises in the network, we should proportionaly increase the network bandwidth, the I/O bandwidth and the memory bandwidth. If we are not able to use a scale factor for all the above requirements, at the network, it may become a bottleneck for the rest of the system decreasing the overall efficiency accordingly [6].
- Incremental expandability (example of ISPs): Customers are unlikely to purchase a computer network with a full set of computers and other electronic devices. More enterprises may be involved in the network until a system's maximum configuration is reached. We should find a way that the new nodes we add, do not decrease the performance. For example if we have 10 enterprises and we want to install an internet service provider we may not use a provider exactly for 10 enterprises, we should predict that after a long time we may provide our services to 15 enterprises, so we must use a provider with incremental expandability in order to reach the new requirements.
- Partition ability: It depends on the work we want to do. We devide the work in several partitions and then we dedicate certain tasks to computers of network [7]. This partition is very efficient for the performance of the network. We can use it also for the design of the network, we configure a network for the requirements of our work and we do not configure the work to the requirements of the network. For example we can use one node of the network as file server, another as database server, another as network server etc. Partition ability may also be required for security reasons [6].
- Simplicity: It is useful for the customers who understand the design and can easily exploit their performance. Otherwise the network may not be so efficient [6].
- Distance span, Locality: There are appropriate mechanisms which can reduce the noise during the transmission of the data, but there are a lot of constraints mainly on the distance between nodes on this domain. The use of optical links solve these problems equalizing the bandwidth of short and long links up to a much greater distance than when copper wire is used [6]. The locality is a very importat parameter, we must know where the enterprises are placed in a map in order to design a network.
- Physical Constraints: The operating temperature control, the wiring length limitation and the space limitation are constraints for the network design. We must be careful in case we put together a lot of wires, the overheating may cause damages to the wires [6].
- Reliability and Repairability: It should be able to transmit information reliably. In addition interconnection networks should have a modular design allowing upgrades and repairs [6].

- Expected Workloads: The network design should be robust. The performance should be efficient independently of the wide range of traffic conditions [6].

Cost Constraints: We should find an optimal solution between the cost of the implementation and the performance of the network. A solution which is probably expensive it is not necessery and efficient [6].

Interconnection networks

At the design of a network we should predict its connectivity with other networks, such as wide area networks [18]. We may need to connect our network with a network of an enterprise and to communicate with an establishment of the enterprise, in order to find useful informations for our work. We can connect our network with leased lines of the telephone network, in order to use services such as voice over ip, teleconference, video conference etc. Interconnection networks exist between parallel and distributed systems [4], [7]. In general purpose the usage of the above systems is the efficient solution of a problem. The network design helps us to divide our work into small tasks and try each computer to solve these small tasks. Then we combine the solution of each node to the final [7]. For those systems the network should reduce the message latency in order to be efficient. A decomposition technique based on interconnection network should care for the load balancing of the work between computers. A central unit can recursively divide the work, sending partition of work to nodes and control the overall work.

Meanwhile an application can not use general rules, the services we want the network offer to their users and the quality of services determine the parameters of a network. There are a lot of issues in network design [6], [27], [28]. First issue is how many routers we are going to use. Having a topological map of the whole network where shall we put them? The connections are going to be wireless (data links), which routing algorithm we are going to use, Dijkstra or smth else. There are also some challenging issues such as the network bandwidth and the time delay of packets. Computer networks individual for enterprises are crucial for their speeds in transfering data and for the internal communication.

Programming and configuring an internet web database

For the purpose of installing a web database we used the Apache Server Program and a Database Management System. The apache server enables some files of the host computer to be accessible from the Internet via the HTTP. For the communication of the database we use some script programs written in PHP language. The accessible files via HTTP are HTML forms which communicate with PHP programs [29]. For safety reasons we used some global variables (sessions) on php scripts in order only experts and executive persons from enterprises could administer the internet web database. The information on the Internet Site is accessible to everyone. On the site we can find which enterprises are interested on the domain of Occupational Health and Safety. We can find the executive persons of each enterprise which are responsible to communicate with us. We install a communication between the database and one program in java [8] in order to compute some interesting statistical values from the dataset. The internet web site which host a part of our work is located at the address http://healthandsafety.nurs.uoa.gr/sdyaergasia.php .

In order to construct an Internet Web Site, we find the system requirements and the users requirements. Then we design the web site in order every user to find easily whatever he wants on the domain of occupational health and safety. Then we configure the web site in order to achieve a communication between the site and the database. The updates for news on research in occupational health and safety are appropriate. The updates on the software are also important. The file management we present [29], uses the classic way of FTP. We insert documents, files, data and multimedia [21] in a directory which is structured in such a way for displaying into the www. Using some additional attributes to the above attributes we can cluster the data into groups [20]. The development of an Internet Web Site for efficient storing and indexing documents is a pioneering research work on the domain of Occupational Health and Safety. These attributes help the users to retrieve the data easily. They are inserted by the system administrator or by the composer of digital data. The metadata are all the appropriate elements which describe the uploading documents. Some operations that a system administrator can do are the uploading of a document and the updating of metadata using a simple web interface.

We develop this internet web site in order the executive person of an enterprise be informed about the issues of Occupational Health and Safety.

We are heading the Semantic Web [17]. Through the Semantic Web the Internet can support workflow models. Workflow models are abstractions revealing the most important properties of the entities participating in a workflow management system. The workflows are of three types: (1) Human-oriented (a person is responsible to execute a workflow), (2) System-oriented (we use databases and decision support systems [16, 22] in order to find answers to our queries) and (3) Transactional workflows (consisting of a mix of tasks, some performed by humans, others by computers and support selective use of the transactional properties for individual activities or for the entire workflow). The workflow models consist of entities that interact with others in an Interconnection Network, and they participate in a management system [1, 17].

Administration by authorized users

The administration of an Internet web site could be performed via the web. The web interface is quite familiar to everyone who wants to use it. We use HTML forms to upload documents. We insert metadata and documents in a database using php scripts. Experts on the domain of Occupationl Health and Safety are responsible to update the site with news and with the latest scientific research. The web interface which supports the administration can be seen in Figure 1.

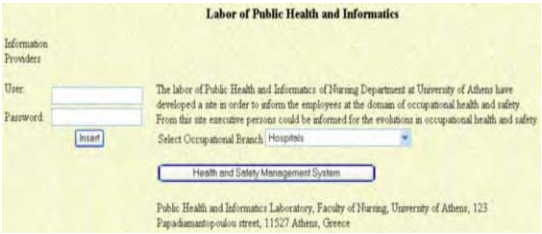


Figure 1 - Web interface for administrating and presenting information on occupational health and safety

There are some constraints at the insertion of data. We can insert only text documents, Microsoft Word Documents, and documents which can be read by the Adobe Acrobat Reader. Meanwhile we may need to insert files for educational reasons, such as power point files, video and images. On multimedia files we can insert a digital watermark in order to authenticate the constructor of the object [15]. We can insert a list of notes for each document in order to have a dialogue about certain ideas [27, 28]. This dialogue is important for improvements in work.

The data we insert at the internet site are peer reviewed in order to be reliable. We also protect the authenticity of the digital objects inserting to them digital watermarks resistant to attacks such as compression, smoothing and geometric transforms [15] (mainly for multimedia data).

In the web site http://healthandsafety.nurs.uoa.gr/sdyaergasia.php, figure 1, developed in the Department of Public Health in the Nursing Department, University of Athens, we have configured the server to be able to find all the relevant information useful to this task. We can search the documents [20] placing keywords. There is also the possibility of browsing in all data lists.

Services provided by the system and applications on distant learning

The transmission of video in real time is critical for designing this system. We make a lot of experiments broadcasting the enviroment and scenes and we are going to transmit through the Internet some lectures. The main program we used for the experiments is a product of Microsoft which is named Windows Media Encoder. We are encoding audio and video using certain sampling rates and analyzing the video on certain number of frames per second. All these experiments made in order to adapt the rate of transmission data with the channel capacity avoiding delays. We also search the number of clients a video server can serve with certain criteria of quality of services. The video on demand has as target to inform enterprises about the subjects of common interest. The subjects will be registered electronically among all the speakers and finally the revised lectures will be inserted in the information system in a corrected form. Meanwhile there are some difficulties on the transmission of multimedia data: firstly, the subject about the quality of service; second, the decompression of the data; and, finally, the removal of the jitter.

Another issue is that the system can provide training packages for the protection and promotion of health at the workplace and the implementation of safety procedures [26]. These packages are disseminated through the network or applied by e-learning techniques [2]. The function and the effectiveness of those packages will be evaluated at the ability of installing integrated solutions and applications to support experts on Occupational Health and Safety.

Conclusion

In this paper we have described a web application, which is administrated by authorized persons only. We intend to inform executive persons of each enterprise for the evolutions in Occupational Health and Safety. This site will inform and allow communication among enterprises. Documents and data based on ISO 18001 will be submitted, processed and presented in appropriate format in order to be useful to the interested executive persons. The corrections and the control of the whole infrastructure is significant to our work. Our target is to improve occupational health and safety, to find new policies, to design new systems, to control new procedures and, finally, to review the whole structure in order to be comprehensible to the users. An application has already been installed at hospital. The application will be evaluated continuously in order to achieve the standards of ISO 18001.

Acknowledgement

The present study was funded through the Operational Programme for Education and Initial Vocational Training (O.P. "Education") in the framework of the project "PENED 2003 – Support of University Research Groups" with 75% from European Social Funds and 25% from National Funds.

References

[1] Aalst W, Hee K. Workflow Management, Models, Methods, and Systems, Cooperative Information Systems Series.

[2] Apostolakis ., Balsamos P., "Distant Learning Education of Health Officials", Health Survey November - December 2005.

[3] Balaouras P., Stavrakakis I., Merakos L., Potential and limitations of teletaching environment based on H.323 audio-visual communication systems. Computer Networks 34 (2000).

[4] Date, C. J, Introduction in the Database Systems, vol , Computer Books Publications Kleidarithmos, 6th American Publication (2003).

[5] Diomidous M, Verginis I, Mantas J. The construction of a simulation-based system for the development of powerful and realistic models and practicals for healthcare professionals.. IEEE Trans Inf Technol Biomed 1998;2(3):174-82.

[6] Duato J., Yalamanchili S., Ni L., "Interconnection Networks, An Engineering Approach", Morgan Kaufmann Publishers.

[7] Elmasri R., Navathe S., Basic Principles of Database Systems, Translation – Editor Michalis Chatzopoulos, Diavlos Publications

[8] Jaworski J., JAVA 2 Platform, Unleashed The Comprehensive Solution, 1999

[9] The Internet Protocol Journal, June 2000 Volume 3, Number 2

[10] Karvounarakis G, Kapidakis S. Submission and repository management of digital libraries, using WWW. Computer Networks 2000; 34:861-72.

[11] Liaskos J, Mantas J. Documenting nursing practice by using ICNP on the Web. Stud Health Technol Inform. 2003;95:806-11.
[12] Liaskos J., Mantas J. Measuring the user acceptance of a Web based nursing documentation system. Methods Inf Med. 2006;45(1):116-20.
[13] Liaskos J, Mantas J. Nursing Information System. Stud Health Technol Inform. 2002;65:258-65.
[14] Malamateniou F, Vassilacopoulos G, Mantas J. A search engine for virtual patient records Int J Med Inform. 1999 Aug;55(2):103-15.
[15] Mavrikakis I., Digital Watermarking applied for Digital Image Protection, Master of Science Thesis, Computer Science Department, Heraklion 2002.
[16] Mavrikakis I., Pattern Analysis with use of Support Vector Machines Applied to Image Processing. Degree Examination Paper, University of Crete, Computer Science Department, Heraklion 2000.
[17] Marinescu D., Internet - Based Workflow Management Toward a Semantic Web, Wiley Series on Parallel and Distributed Computing, Albert Y. Zomaya, Series Editor.
[18] Nieuwpoort R., Maassen J., Bal H., Kielmann T., Velderna R., Wide Area Parallel Computing in Java, Department of Computer Science, Vrije Universiteit Amsterdam, The Netherlands.
[19] Panagiotou E., Panagiotou G., Hospital Laboratory Information System (LIS), Epithewrhsh Hygeias November – December 2005.
[20] Park L., Ramamohanarao K., Palaniswami M., "A Novel Document Retrieval Method Using the Discrete Wavelet Transform", ACM Transactions on Information Systems, Vol. 23, No. 3 July 2005 Pages 267-298.
[21] Shirmohammadi S., Saddik A., Georganas N., Steinmetz R., Web-Based Multimedia Tools for Sharing Educational Resources. ACM Journal of Educational Resources in Computing Vol. 1, No. 1, Spring 2001.
[22] Suykens J., Gestel T., Brabanter J., Moor B. and Vandewalle J., Least Squares Support Vector Machines. World Scientific
[23] Terpos A., Research of Information Systems on Occupational Health and Safety section EL.IN.Y.A.E., Athens 2000.
[24] Tziritas G., Labit C., Motion Analysis for Image Sequence Coding. Elsevier, 1994.
[25] Velonakis E., Tsalikoglou F., Occupational Health and Safety Management System in Hospital. Scientific Publications, Parisianou. Athens
[26] Velonakis E., Mantas J., Mavrikakis I., A site of communication among enterprises for supporting Occupational Health and Safety Management System, 20th International Congress of the European Federetion for Medical Informatics, Maastricht Netherlands, 2006
[27] Walrand J., (Translation M. Anagnostou), Communication Networks, Papaswthriou Publications (1997)
[28] Walrand J., Communication Networks A First Course, Second Edition, WCB McGraw-Hill.
[29] Welling L., Thomson L., PHP and MySQL Web Development, M. Giourdas Publishers.

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Reliability Assessment of Home Health Care Services

Stergiani Spyrou, Panagiotis Bamidis, Vassilis Kilintzis, Irini Lekka, Nicos Maglaveras, Costas Pappas

Lab of Medical Informatics, The Medical School, University of Thessaloniki, Central Macedonia, Greece

Abstract

In this paper, a model of reliability assessment of services in Home Health Care Delivery is presented. Reliability is an important quality dimension for services and is included in non-functional requirements of a system. A stochastic Markov model for reliability assessment is applied to patient communication services, in the field of home health care delivery. The methodology includes the specification of scenarios, the definition of failures in scenarios as well as the application of the analytical model. The results of the methodology reveal the critical states of the Home Health Care System and recommendations for improvement of the services are proposed. The model gives valuable results in predicting service reliability and, independently of the error types, it can be applied to all fields of Regional Health Network (RHN).

Keywords:

home health care delivery, reliability, quality in health

Introduction

Patient communication services in the health delivery domain are supported by information and communication technology (ICT) systems [1],[2],[3]. Contact centers in home health care delivery are a modern way to provide services to patients efficiently and are mostly implemented as call centers or portals on the Web. The purpose of this work is to present the application of an innovative model to assess quality indicators of the provided services in home health care delivery domain. The innovation is in the direction of reliability assessment on the analysis, the specification and the verification phase of developing such systems, with the ultimate scope to lead to a concrete system at the implementation phase. The analysis and the results are based on scenarios of the services of a contact center for chronic disease patients.

Reliability analysis for information systems supporting health care delivery services or even hospital information systems is not met on the literature. This work has been based on the technical report [1], where a model based on stochastic reliability model (Markov model) is applied to health care delivery services.

The analysis has been applied to a project under the title 'Quality Healthcare Management and Well-being through INTERLIFE services (INTERLIFE[1])' [4] implemented in the Lab of Medical Informatics[2], The Medical School, Aristotelean University of Thessaloniki. INTERLIFE is a generic and modular contact center platform for the communication, management, processing and assessment of multimedia medical information, and the provision of high-quality pervasive tele-health services to the citizens. It aims at reaching a high number of patients in a simple way and with low-cost budgets through telephone and WEB communication. The aim is to perform a clinical, technical and economic evaluation of the INTERLIFE system when applied in real environments for the management of two chronic conditions: chronic heart failure (CHF) and obesity/diabetes. The INTERLIFE platform at AUTH is based on the CHS system developed in the context of the CHS program [5], IST-1999-13352 that had been evaluated mainly for its technical features.

The model can be applied to any field of systems supporting health care delivery and it gives valuable results in the area of quality assessment of services and, specifically, reliability assessment in health care.

Materials and methods

INTERLIFE Contact Center

The methodology has been applied to INTERLIFE Contact Center [4]. It is described below with emphasis on the scenarios used in our model. The INTERLIFE core system offers the infrastructure that allows home-monitored patients to easily communicate with the INTERLIFE Contact Center by exchanging their data. During the LAB pilots, patients are offered three interfaces to choose from: simple telephone, mobile phones and WEB. The cheapest technology with the widest penetration is the use of regular telephone with call-center automation. At the Contact Center of INTERLIFE, patient data are stored so they can be securely accessed by the clinicians for reviewing either through a client/server application or through a WEB application. Technical infrastructure and support is provided by the INTERLIFE technical team at the LAB and the Department of Cardiology of Aristotelean University

1 INTERLIFE: stands for the project 'Quality Healthcare Management and Well-being through INTERLIFE services (INTERLIFE)'

2 LAB: stands for the Lab of Medical Informatics, The Medical School, Aristotelean University of Thessaloniki

Selected for best paper award.

of Thessaloniki hosted at the AHEPA Hospital of Thessaloniki, Greece which participated in the study. Patients on this trial are chronic disease patients.

The main contact center services are:

- Transmission of patient data
- Educational sessions in the form of voice messages
- Patient and physician communication via voicemail for problem reporting, advice, etc.

The session describing the patient contact with the contact center is shown in Figure 1. The two scenarios implemented and included in the trial setting are described in Tables 1 & 2.

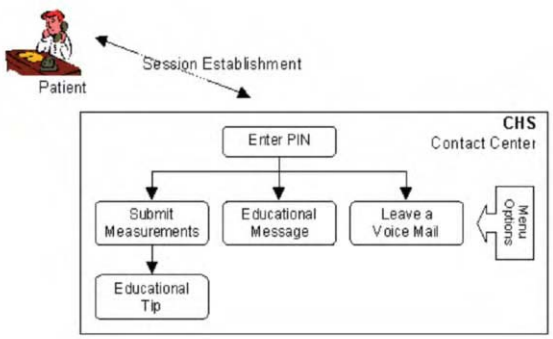


Figure 1 - The patient session using the automatic call center

First, patient authorization takes place. The user can dial 1 on the phone to enter the measurement session (menu choice 1 of contact center) or 2 to enter the educational session (menu choice 2 of contact center) at the main menu of the contact center. The voice mail menu option was excluded for the trial.

Each patient session is recorded in a (text) log file with the interaction details. Log files have been used for the assessment of technical failures, performance and system usage. Moreover, analysis of those log files gave the transition probabilities and other information, such as error reports, that constitute the necessary parameters for our model. A patient calls in the center to report the measurements three times per week (three sessions per patient per week). In this trial, **2.887** patient sessions are included.

The model's focus is to assess how reliable patient contact is with the center, based on the following dimensions:

1. The patient data input errors regarding the usage of the system
2. The technical infrastructure problems that obstruct the successful termination of a scenario.

Applied to our model, those factors can determine the reliability of the system in view of assessment of technical failures and patient data input errors.

Table 1 - Scenario for measurement session

Purpose of Measurement Session: Patients transmit parameter patient data.
1. Patient calls the call center to report the measurements frequently.
2. Patient enters their PIN by dialing it into the phone to be authorized by the center.
3. The patient dials measurements such as body weight, temperature, heart rate, blood pressure and answers (yes or no) to five simple questions; for example: 'Were you breathless during the night?', 'Are your legs swollen?' etc. The selected questions target two objective symptoms (paroxysmal nocturnal dyspnea and peripheral edema), two subjective symptoms (fatigue and dyspnea) and a gentle reminder to take their medicine.
4. Voice verification of each measurement value is provided immediately after it is keyed in. There is the option to correct the keyed-in value after entry and verification.
5. In case of no entry, the patient is repeatedly prompted to enter a value.
6. The patient is informed when the measurement values have been saved.
7. Goodbye tip of the day contains advice according on patient's situation.
*[The following actions are not recorded in log file:]*Daily review of patient data will be performed by a clinician at the contact center; the cardiologist will be in contact with the patient for advice in special cases during the week.

Table 2 - Scenario for educational session

Purpose of Educational Session: Patients Listen to Educational Voice Messages
1. Steps 1 & 2 are the same as the measurement session.
2. Patient dials the 2-digit code of the category of educational message they are interested in
or
3. Patient can interrupt educational session by exit (0) key.
4. The repetition (*) and the interruption keys (#) can be initiated at any point even while listening to a message.
5. The system always keeps track of what the user has heard.
6. A goodbye tip ends the session.

Scenarios and failure behavior

Reliability is the probability that a system will operate without failure for a specified number of natural units or a specified time – known as the mission time [8]. Reliability, R(t), is related to failure probability, F(t), by $R(t) = 1 F(t)$ where, $F(t)$, is the probability that the time of failure is less than or equal to t. To model software reliability, one must mostly consider the fault introduction that depends mainly on the characteristics of the software system and the development process [1]. The steps to describe fault introduction to systems are mainly (1) describe the scenarios that the system implements and (2) identify the failures of the scenarios' activities.

In this work, we are going to predict and assess reliability for processes based on a scenarios analysis of contact center of home care delivery. The idea is based on [1]. The methodology is based on scenarios as the main focus is to predict the quality dimension of reliability for processes at the early stage of building a home care system.

The methodology as presented in [1] and customized for the needs of the presented work is:

1. Describe behaviour of the contact center using scenarios.
2. Draw state transition diagrams based on the analysis on previous step and annotate them with transition probabilities to also describe the failure behaviour of the system.
3. Use Markov modelling to calculate the reliability of the system described by the scenarios.
4. Experiment with transition probabilities of states. or reliability factors of states, to find the most critical phases of the system which are the outcome of the Markov modelling.

A state transition diagram, as needed in step 2 of the methodology for INTERLIFE Contact Center, is presented in the results part of this paper.

The transfer of control of such a diagram has a Markov property and [1] we model software execution behavior with an absorbing Discrete Time Markov Chain (DTMC) with a transition probability matrix $P = [p_{ij}]$, where p_{ij} is the probability to transfer the control from module (state) i to module (state) j [6]. We assign multiple absorbing states (F0, F1, F2, F3 etc) according to the failure behavior denoting failures, and state T denoting correct output and successful termination. Then the transition probability matrix can be calculated based also on the reliabilities of each state R_{si} which denotes the Reliability of component-state of State S_i, $P_{Si\text{-}Sj}$ denotes the probability to transfer control from state S_i to state S_j. Based on a transition probability matrix and the DTMC model with absorbing states, we can calculate [1],[8] table $A= [a_{ik}]$, where a_{ik} is the probability of reaching an absorbing state k starting from a transient state I and the probability to reach a terminating state is calculated.

Results

This section presents the results of applying the model to INTERLIFE project. As described previously in the methodology, the following scenario specification is the identification of failures of the scenarios' activities. The failures of the system include the patient data input errors and technical infrastructure failures. However, it is important to note that the methodology does not depend on the type of errors that are specific to each project according to its functionality. The INTERLIFE system has to check for possible failures due to:

- Patient errors (mistyping, bad measurement technique)
- Instrument errors
- Transmission errors.

The types of patient data input to the system are: (1) parameter data (e.g. weight, blood glucose), (2) signal data (e.g. heart rhythm, ECG) and (3) image data (e.g. wound image from digital camera).

Figure 2 shows state transition diagram for INTERLIFE scenarios that includes the failure behaviour as described above. Abbreviated names for the states are included in parenthesis in capital letters following each state in the diagram. Abbreviations are used in the state transition tables and the rest of the figures.

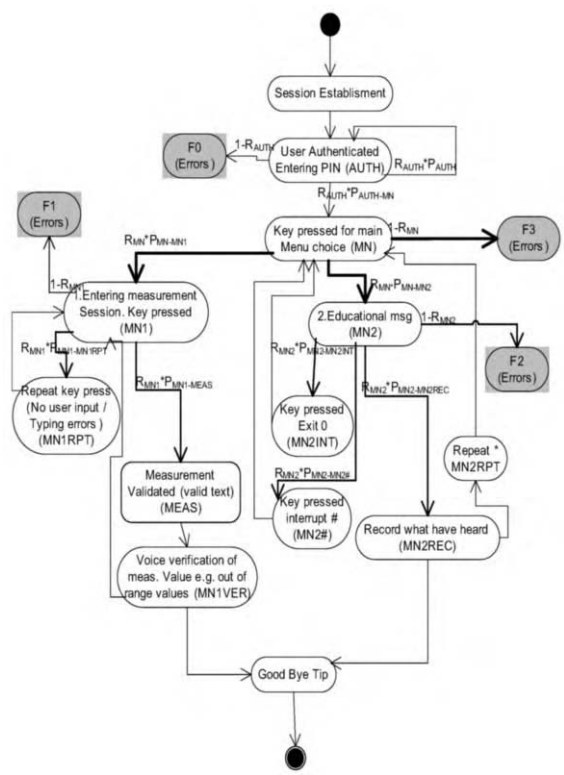


Figure 2 - State transition diagram for INTERLIFE scenarios

The diagram has been annotated with transition probabilities (P_i) and reliability factors of states (R_i). For example,

Transition probability from state MN to MN2, considering that reliability of state MN is R_{MN}, is given by $R_{MN}*P_{MN\text{-}MN2}$ in the state transition diagram. When no probability is assigned it is assumed to be equal to 1.

Analysis of the log files of the trial the probabilities for the case of **2.887** total patient sessions are shown in Table 2.

Table 2 - Probability reaching a terminating state starting from any state in the scenario

Probability	*Value*	*Probability*	*Value*
P_{AUTH}	0.051	P_{F1}	0.077
$P_{AUTH\text{-}MN}$	0.754	$P_{MN1VER\text{-}MN1}$	0.241
P_{F0}	0.195	$P_{MN1VER\text{-}T}$	0.759
$P_{MN\text{-}MN1}$	0.923	$P_{MN2\text{-}MN2INT}$	0
$P_{MN\text{-}MN2}$	0.053	$P_{MN2\text{-}MN2\#}$	0
P_{F3}	0.024	$P_{MN2\text{-}MN2REC}$	0.837
$P_{MN1\text{-}MEAS}$	0.668	P_{F2}	0.163
$P_{MN1\text{-}RPT}$	0.255	$P_{MN2REC\text{-}T}$	1.0

The following results are obtained from analysis of the log files: of a total of 552 errors in the measurement session, there are **165** patient errors and **387** instrument and transmission errors.

Applying the model, the transition probability matrix is created and then matrix A, describing the probability to reach a terminating state is shown in Table 3. Title T in second column denotes successful termination of the scenarios. Title F1, F2, F3 in the following columns denotes failure at measurement session, at educational session and at initiating menu choice respectively.

Table 3 - Probability reaching a terminating state starting from any state in the state transition diagram

Table A	**T**	**F0**	**F1**	**F2**	**F3**
AUTH	0.6719	0.2058	0.0966	0.0000	0.0256
MN	0.8460	0.0000	0.1217	0.0000	0.0323
MN1	0.8682	0.0000	0.1318	0.0000	0.0000
MN2	0.8374	0.0000	0.0000	0.0000	0.1626
MEAS	0.9682	0.0000	0.0318	0.0000	0.0000
MN1VER/PRMT	0.9682	0.0000	0.0318	0.0000	0.0000
MN1RPT	0.8682	0.0000	0.1318	0.0000	0.0000
MN2INT	0.8460	0.0000	0.1217	0.0000	0.0323
MN2#	0.8460	0.0000	0.1217	0.0000	0.0323
MN2RPT	0.8374	0.0000	0.0000	0.0000	0.1626
MN2REC	1.0000	0.0000	0.0000	0.0000	0.0000

From the matrix the following results can be excluded: implementing the scenarios and starting form initial state, AUTH the probability to have:

(1) Successful termination (T) is equal to 0.6719

(2) Termination with error F0 is equal to 0.2058

(3) Termination with error F1 is equal to 0.0966

(4) Termination with error F2 is equal to 0.0

(5) Termination with error F3 is equal to 0.0256.

It is important to note that, excluding patient errors from measurement session (165 errors), the total number of errors at this session would be 387 (instrument and transmission errors) and the total reliability of the system becomes **0.5512**. If patients could perfectly use the system (at measurement session only), the reliability of the system is significantly improved.

Conversely, if there are only patient errors in the measurement errors (excluding instrument and transmission errors), the reliability of the system would be **0.6719**.

Experimenting with transition probabilities, such as the probability of reaching the absorbing state F1, the total reliability of the system increases dramatically (shown in Figure 3). In addition, by varying the reliability of states, one can see how the total reliability of the system is affected. More complicated and descriptive behavior of the system can be the outcome of applying the model, as shown in Figures 3 and 4.

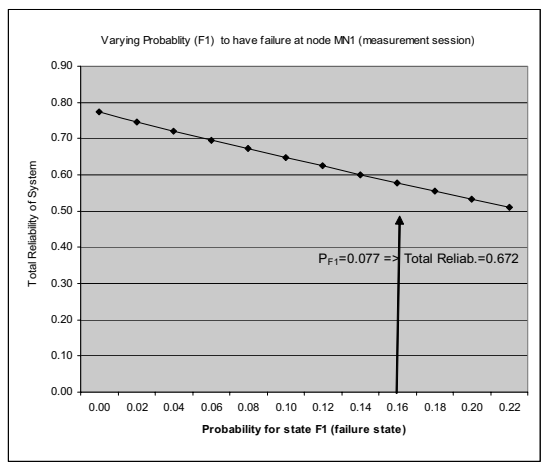


Figure 3 - Varying probability to have failure at measurement session and reach failure state F1

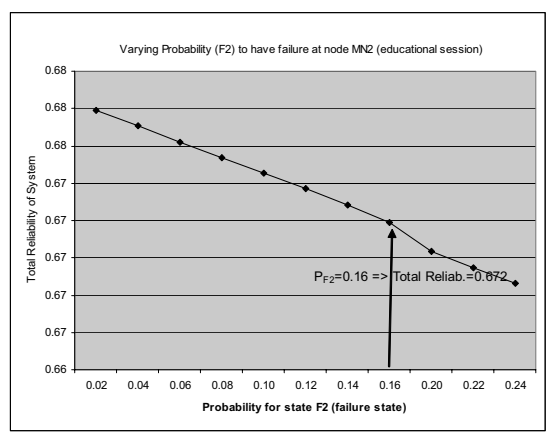


Figure 4 - Varying probability of failures at educational session and reach failure state F2

Discussion and conclusion

Analysis of quality issues of home health care systems and reliability modelling techniques provides an innovative way to predict performance measure, even at the design phase of the system. The model presented can be applied at the design, verification and implementation phases of systems supporting health information systems and give valuable results predicting the reliability of services.

The approach focuses on defining the sources of errors and failures, and the impact those errors can have at services. It does not depend on the type of errors. The more elaborated the failure analysis is, the more accurate the results of the model can be. It becomes feasible then to identify the critical states of the system in the scenarios [1] and propose interventions for improving the provided services.

Even when prototyping or implementation data are not provided, this model provides a way to estimate or predict reliability. However, the more data that are available, the more accurate the results the model will produce. The depth of analysis to describe the states depends on the analyst, where elaborated states would reveal analysis in detail. It is valuable to analyze failures and errors in the design phase of the system to have records of the failures.

The methodology's limitations are linked with the need to have an elaborate requirement analysis of the system to define the state diagrams in detail. In the methodology presented, no severity analysis was included.

Future focus will be directed towards enhancing this analytical calculation, with the combination of software tools for simulation and evaluation.

The presented analysis could be the preliminary analysis to calculate factors of the systems like Mean Time To Failure (MTTF) or Mean Time To Repair (MTTR) [7], [9], [10], [11], which define crucial factors for the service-level agreement contracts for the organizations.

The model and the analysis can be applied to any type of health information system. The results in the area of home health care systems are important for improving the provided services to patients.

References

[1] Spyrou S. et al. "A Methodology for Reliability Analysis in Health Networks", IEEE Trans Inf Technol Biomed (under review)

[2] Voss H, Heimly V, Sjögren LH. The Baltic Health Network – Taking Secure, Internet-based Healthcare Networks to the Next Level, Proc of MIE2005 - The XIXth International Congress of the European Federation for Medical Informatics, Studies in health technology and informatics, 2005;116: 421-426.

[3] Vartzopoulos D, Spyrou S, Minaoglou E, Karolidou V, Bamidis PD. Establishing a Regional Contact & Service Centre for Public Health Care: The case in Central Macedonia Greece, Proc of MIE2005 - The XIXth International Congress of the European Federation for Medical Informatics, Studies in health technology and informatics, 2005;116:917 - 922.

[4] Prentza A, Maglavera S, Maglaveras N. Quality Healthcare Management and Well-Being through INTERLIFE Services: New Processes and Business Models, 1st D2H2 Conference, Arlington, VA, pp. 109-112, 2006.

[5] Maglaveras N. et al. The Citizen Health System (CHS): a modular medical contact center providing quality telemedicine services. IEEE Trans Inf Technol Biomed. 2005 Sep;9(3):353-62.

[6] Bolch G, Greiner S, de Meer H, Trivedi KS. Queueing Networks and Markov Chains: Modeling and Performance Evaluation with Computer Science Applications (The White Book), John Wiley, NY, 1998. ISBN number: 0-471-19366-6.

[7] Tsopelas P, Platis A. Performability indicators for the traffic analysis of wide area networks, Reliability Engineering and System Safety, 2003; 82:1–9.

[8] Musa J. Software Reliability Engineering. McGraw-Hill, ISBN 0-07-913271-5

[9] Platis A. A Generalization of the Performability indicator. Special Issue on The Second Euro-Japanese Workshop on Stochastic Risk Modelling for Finance, Insurance, Production and Reliability, Computers & Mathematics with Applications, An International Journal, 2006; 51:239-246.

[10] Limnios N, Ouhbi B, Platis A, Sapountzoglou G. Non-parametric Estimation of Performance and Performability for Semi-Markov Processes. International Journal of Performability Engineering, 2006; l.2(1).

[11] Koutras VP, Platis AN. Resource Availability Optimization for Priority Classes in a Website. In Proc 12th IEEE International Symposium Pacific Rim Dependable Computing PRDC'06, 18-20 December 2006 - Riverside, California, IEEE Press, accepted.

Acknowledgment

This work was supported in part by the IST eTEN INTERLIFE project (contract number C517340).

Address for correspondence

Stergiani Spyrou, Phd Student, Lab of Medical Informatics, The Medical School, Aristotelean University of Thessaloniki, Tel:+30 2310 368822, Fax: +30 2310 368808 e-mail: spirou@med.auth.gr

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

From the Description of Activities to the Identification of Risks for Clinical Management: a Proposal of Building, Merging and Sharing Knowledge Representations of Care Processes

Pascal Staccini[a,c], Michel Joubert[b], Rémy Collomp[c], Jean-François Quaranta[c], Marius Fieschi[b]

[a]*Laboratoire STIC Santé, Faculté de Médecine, Université Nice-Sophia Antipolis, Nice, France*
[b]*LERTIM, Faculté de Médecine, Université de la Méditerranée, Marseille, France*
[c]*Pôle Information – Risques – Evaluation, Centre Hospitalier Universitaire de Nice, France*

Abstract

Management of clinical processes and hospital activities takes advantage of business process reengineering methodology. It is now recognized that care process modeling must integrate the definition of goals and the assessment of risk. Two kinds of issues have been outlined: 1) the lack of an integrated model to identify and describe processes and their components according to a functional point of view; and 2) an increasing amount of documents that hospital staff members have to create, collect, index and maintain. As initial models focused only on a structural view of activities, we reviewed different sources of standards and norms to extract and classify a set of metadata aimed at describing any activity and its outcomes. The model includes links to structured terminologies to name attributes or value them. An object-oriented information model has been created and implemented to test the relevance and the feasibility of the modeling approach. Conceptually speaking, this model gives opportunity to bridge tacit and explicit knowledge. Practically speaking, limits to generalization remain partly due to the lack of a template processes database.

Keywords:

process assessment (health care), risk management, systems analysis, systems integration, hospital information systems.

Introduction

During the last twenty years or so, French hospital management has been impacted by successive policies which led on to: 1) the end of "limitless" budgets and the setup of a national and external process of evaluation, certification and regulation; 2) the emerging of integrated information systems, shifting individual centered practices to collaborative working methods; 3) the increasing desire for transparency, whether coming from the state or the patient; and 4) the end of the dumb patient, captive and submissive recipient, becoming an enlightened client aware of his(her) power and duty. Therefore, hospital staff, from caregivers to managers, have had to join and collaborate for a new way of hospital management, consisted of *saying*, *doing* and *justifying*. If the "doctor manager" is going to be the emergent logic of hospital management, a big issue is to put everyone on a collaborative track, to create synergy and be able to justify the survival or the continued existence of the establishment.

Context

From now on, to comply with the national and mandatory process of certification, it has been recognized that business process reengineering (BPR) methodology suited fairly well to represent hospital activities [1]. Its application to health care facilities is related: 1) to the optimization of patient flows [2] and of care deliveries [3], especially in emergency and acute care wards [4,5]; 2) to the evaluation of the performance of clinical pathways [6] or clinical training scenarios [7]; and 3) to the design of information systems for identifying and sorting out activities to simulate the potential performance of target systems, and, further, to organize and synchronize users views [8,9]. Whatever the objectives are, the initial work of staff members always starts with the description of activities performed towards the patient. Requirements for the quality and the security of care delivery make them to pay attention to several normative specifications such as the national guide for certification published on behalf of the French Health Higher Authority [10] as well as quality management standards framing the production of services (ISO 9000:2000) [11], the environmental requirements (ISO 14001:2004) [11] and the occupational health and safety (OHSAS 18001:1999) [12].

Documenting activities and identifying related risks, as precisely and as relevant as possible, necessitate usually to involve a lot of people in order to produce lots of documents. From a hospital level point of view, such a publishing action needs to be structured, planned and monitored. Relevant indexation and accuracy maintenance of the content is crucial, in order to make material be rapidly retrieved and easily used by professionals. Solutions can be found out with implementing project and digital libraries management tools. From a clinical staff member point of view, the initial work of describing care activities, writing procedures, and pointing out their potential risks can be suitably framed using "continuous quality improvement" tools [13]. But, critical issues remain, such as: 1) making authoring staff member be aware of other similar activities described previously and/or elsewhere in the hospital; providing him(her) with the ability: 2) to "re-use" pieces of data (and so preserve the overall coherence of the system), 3) to interconnect new pieces of data with existing ones (in order to enrich or expand processes or sub-processes); and 4) to be informed of updates according to his(her) profile.

As an example of issues, regarding the methodology suggested at a national level [14] to create and structure a hospital-based medication process, steps can only be described from a clinical point of view (ordering, dispensing and administrating). A multi-level analysis is not clearly identified. Thus, trade actors (pharmacist, nurse, and physician) are solely motivated to build their own view of the system without embracing the whole process, starting with the referencing of medication, following with the monitoring of patient and ending with the report of adverse events. For each process component, only one level of specification is available, making difficult the identification of risk characteristics. Such a methodology gives an opportunity to make staff be aware of the necessity of activity modeling. But is lacks methodological and collaborative tools to share the generated material between actors or processes. Applying such a guide results in a poor quality process model, potential misunderstandings and mistakes, as well as difficulties to review, organize and browse documentation.

Basically, solutions addressing such issues deal with knowledge representation and ontology, merging health related concepts and enterprise modeling concepts, such as: 1) structured sources of explicit knowledge (e.g. diagnoses and procedures) ; 2) emergent enterprise manufacturing research approaches [15] (attempting to create a globally shared conceptualization of organizational domains as reservoir of concepts and re-usable components for organizational modeling); 3) non standardized representations (types and meta-description of processes, activities, roles and risks).

Objectives

Our proposal is centered on the "clinical content management" part of the overall hospital information system. It aims at identifying and describing any care process, its components and its outcomes, according to a single view merging three aspects of clinical management: structure/resources, strategy and safety. To achieve this integrated view, three sets of metadata have to be defined: 1) the activity/process metadata set; 2) the indicator metadata set; 3) the risk metadata set. This paper aims at proposing a pragmatic approach in order to identify metadata sets, to describe a process component, and to structure a minimal information model to be implemented in a cross-referencing documentation system. Discussion will highlight the benefits and the limits of this bottom-up approach, as well as the derivate potential usage of the tool.

Materials and methods

A process is a structured collection of activities, a set of partially ordered steps or scenarios according to which a user interacts with its environment (other users, patients, context of practice), has a set of responsibilities and reaches goals [16]. The basic process model identifies: 1) agent: an actor (human or machine) who performs a process element; 2) role: a coherent set of process elements to be assigned to an agent as a unit of functional responsibility; 3) artifact: a product created or modified by the enactment of a process element. This "macro" definition does not offer enough meaningful details to frame a generic set of attributes. The "process element", the activity, must be categorized and refined.

Primary work has been published, reporting a methodology for applying enterprise modeling tools to explicit health care practices, especially in the field of transfusion medicine [17]. Concepts and views of care process modeling have been extracted from a structured design and analysis technique, called IDEF0/SADT [18]. The atomic element of this modeling technique is called "ICOM box" (I for Input, O for Output, C for Control, M for Mechanism). It is intended to express the notion of any form of activity or task. Semantics of this box is: "under Control an Input is transformed into Output by Mechanism". Using this representation, a clinical agent can easily tell what is going to be done, who is going to do it, and how it will be done. But he cannot tell why it will be done and who is dependent on its being done. These questions are related to the purpose, the strategy and the regulation controls which initiate or not an activity. Due to the lack of "functional" attributes to qualify components, goals and outcomes of any activity, sub-process or process, we analyzed normative reference frames as listed above.

First, action verbs were listed by merging analysis of clinical documents and verbs describing learning objectives [19]. Then, description of an activity was refined with a grammar extracted and rearranged from the conceptual domain of Enterprise Ontology [20]. Goals of clinical activities were described according to a grid aimed at creating guidelines ontology [21]. Terms to qualify an indicator came from the FD X 50-171 norm [22] and from categories of healthcare quality measures defined by the Agency for Healthcare Research and Quality [23]. Regarding the metadata set for risk identification, we studied the Failure Mode, Effects and Criticality Analysis (FMECA). This method is used in industry to prevent process or product failures. It has been applied in health care to investigate risk issues [24]. We picked up attributes to define actions related to risk control. Finally, concepts of the "Risk paradigm" were taken from the environment metadata record, promoted and maintained by the Environmental Protection Agency [25]. Terms used to name or qualify information elements, as well as terms to value attributes, were put into a global dictionary with their respective source. The resulting information model was built according to UML notation. Material produced in our hospital was used to test the relevance of the model and the feasibility of this approach. It consists of clinical processes descriptions (transfusion medicine, organ transplant, clinical pathways in cancerology, medication delivery process, etc.), procedures and other quality documents published by our hospital staff members.

Results

Various kinds of processes can be identified within an enterprise: individual human work processes, organizationally-defined processes, operational processes, production processes, management and change processes, etc. [16]. In the context of clinical activities, the main processes encountered are individual human-based processes (e.g. care delivery processes with diagnosis screening, therapeutic ordering processes, etc.), organizationally-defined processes (e.g. patient registration, patient discharge, medication delivery, appointment records, etc.), and operational processes (e.g. surgical act performing, monitoring of medication administration, etc.).

Table 1 - Terminology extracted from organizational, management and modeling sources

Core terms	Refined terms	Components / Attributes terms [Values]	Source
Activity		Action verb (see table 2)	ISO/Ha
Input			ISO/IDEF0
	Object	Data; Material; Product; Service; Men	ISO/IDEF0/EO
	Environment	(see below)	ISO
Control			IDEF0
	Methods	Scenarios; Strategy (see below)	ISO/EO/Ha
	Environment	Context; Triggers; Conditions	ISO/Ha
	Constraints	Temporal [periodicity, delay]; Scheduling [priority, ranking]; Cost; Requirement	OC
Mechanism			IDEF0
	Men	Person; Role [doer; owner; partner]; Function; Skill; Capability;	ISO/EO
	Machines		ISO
	Material		ISO
Output			ISO/IDEF0
	Object	Data; Material; Product; Service; Men	ISO/IDEF0/EO
	Risk	(see below)	EO
	Indicator	Category [resource, process, outcome, satisfaction, conformity; safety; patient centeredness; behavior; access; effectiveness; timeless; environmental condition; environmental performance]; Field; Parameters; Computation; Periodicity	ISO/AHRQ
Strategy			EO
	Goal	Type [action, knowledge]; Action verb (see table 2); Object (see above)	EO/Ha
	Objective	Criterion	EO
	Assumption	[critical; non-critical]	EO
	Influence Factor	[critical; non-critical]	EO
	Success factor	[critical; non-critical]	EO
Risk			EO/FMECA
	Exposure		EPA
	Cause	Methods; Environment; Mechanisms *(see above)*	FMECA/ISO
	Effect		FMECA/EPA
	Priority	Severity, Probability, Detection	FMECA
	Corrective action		ISO/FMECA
	Preventive action		ISO/FMECA

AHRQ: Agency for Healthcare Research and Quality [23]; EPA: Environmental Protection Agency [25];
EO: Enterprise Ontology [20]; FMECA: Failure Modes and Effects Analysis [24]; IDEF0/SADT: [18]; ISO: International Organization for Standardization [11]; Ha: Hashmia et al. [21]

The resulting terminology for naming activity, sub-process and process components is listed in Table 1. Sources are mentioned for each term. The core terms refer to the starting schema based on IDEF0/SADT notation. For each action, a strategy details goal(s) and objective(s). Strategy is also part of the methods used to achieve an action because it mentions influence factor(s). Indicator has been considered as an output of action. Risk information model combines description items (exposure, effect) and assessment items (priority). Risk (or failure) has been considered as an output of the activity. Some components are used to describe cause of failure such as: methods, men, machines, materials and medium (environment), which are the "5M" categories of the Root Cause Analysis methodology [13]. Priority, corrective and preventive actions are three elements extracted from the FMECA methodology. They are part of the "risk paradigm" state by EPA: exposure, effects, assessment (identify, prioritize), and management (correct,

prevent). Action verbs are classified according to four domains as listed in table 2. They are used to name the action underlying an activity or a goal. An object-oriented information model has been built (figure 1), based on this metadata repository.

Table 2 - Action verbs to name actions and goals (from [19])

Creative behaviors	Problem solving behaviors
ask, create, document, group, inform, modify, predict, prescribe, question, review, revise, search, simplify, synthesize, validate, write	analyze, appraise, assess, classify, combine, compare, decide, deduce, diagnose, evaluate, explain, formulate, generate, induce, infer, interpret, interpolate, plan, rank, rate, schedule, substitute, translate
General discerning behaviors	**Practical behaviors**
associate, choose, collect, define, describe, detect, differentiate, estimate, identify, indicate, list, locate, match, omit, order, pick, place, point, recognize, select, separate	apply, calibrate, compute, calculate, convert, demonstrate, diagram, insert, keep, lengthen, limit, manipulate, operate, practice, prepare, remove, replace, report, reset, score, set, specify, straighten, time, transfer, transport, use, weigh

This information model has been implemented with an object-oriented framework called Zope. This web service is based on libraries containing elementary information objects which can be associated to describe activities and compose sub-processes and processes. Each element is described by means of an identity form and belongs to a category. In order to populate libraries, users start from structured documents (templates are available to guide them to write procedures and draw schemas). They sort these electronic documents according to their purpose and content. They extract relevant information to create data objects in the libraries. Doing this, they easily index each document according to the business process it belongs. As a final result, the browsing of the database makes user able to cross-navigate from activities to documents and backwards.

Discussion

Documentation of care processes is from now on a founding part of care management in order to comply with the quality and certification rules. One must pay attention to it. Creating, collecting, indexing, and publishing electronic documents are becoming big issues for hospital staff members. Scientific and evidence-based information is not enough to answer daily questions raising throughout care delivery processes. Beyond the narrative of diseases, diagnosis procedures, therapeutics algorithms and recommendations, staff member must be aware of internal policies and external legal constraints, as well as evaluation methods to review the relevance of care and the risk occurrence throughout clinical pathways.

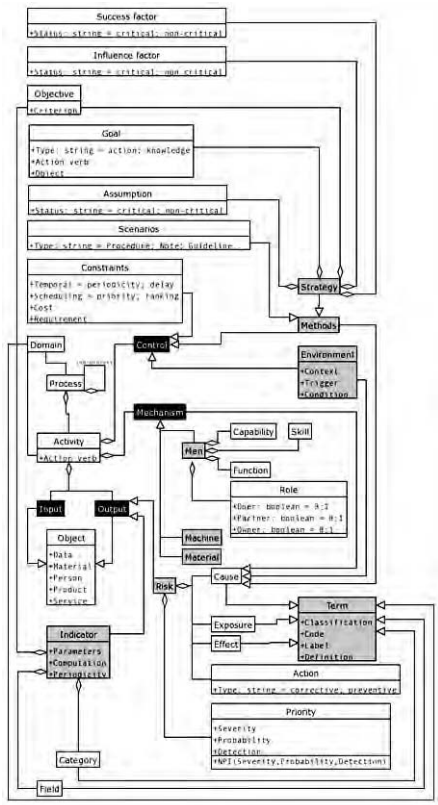


Figure 1 - Object information model for describing an activity (IDEF0/SADT core elements are black colored; new elements deduced from quality and safety tools are grey colored; terms to name indicator and risk elements belong to classifications.)

Basically, the information model we built is activity-based and follows common design conventions. But, it also integrates the definitions of strategy and risk/failure as elements of a functional view of processes [26]. Combined into a unified metadata repository, this set provides staff members with an integrated framework to create documents while indexing them at the same time. Links to standardized terminologies enrich the model providing users with material to query external knowledge databases. The resulting "business process" content management system can be browsed according to a given role (a physician, a nurse…), a category (a procedure, a technical note, a form, a legal rule…), a type of activity (prescribe, validate, synthesize…), and a goal or an output (all tasks with potential adverse events). A hierarchy of information elements is automatically generated and provides users with an overall map of the processes and components managed in the hospital.

In terms of knowledge, the elements of care delivery express the knowledge of collaborative practice (who, when, where), the knowledge of the content of each care process (what, how to) and the knowledge of the indication and the outcomes of the care process (why). In terms of information systems, providing clinical actors with sharable knowledge is a key feature to create collaborative

practice. Our model helps users to value data elements with medical concepts: domain of a process/activity, field of an indicator, exposure, effect and cause of a risk/failure. Doing this, a link is set up between knowledge of practice and explicit knowledge (Figure 2).

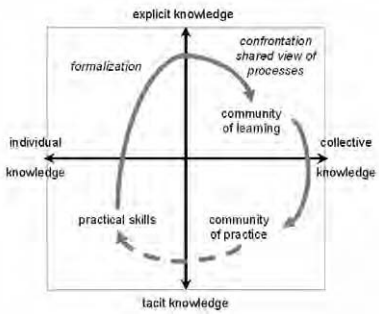


Figure 2 - Linking explicit and tacit knowledge by means of structuring and sharing views of processes

Testing the relevance and the feasibility of the model has been carried on in different domains of care delivery while focusing on risk assessment (transfusion medicine and medication delivery process). Nevertheless, applying limits raised, more likely related to the level of granularity choosen to define activities and to generate documents than related to the understanding of the information model itself. The lack of documents at the beginning of the project makes people uncomfortable, but the review of already published materials is also a burden that cannot be easily avoided. As a conclusion, the purpose of our work was to demonstrate the feasibility and the utility of a rich information set to describe care processes and their outcomes, focusing on risk management. Issues remain related to the compliance of staff members. They can be solved with the building of a "startup" knowledge database describing a "virtual" hospital enterprise.

References

[1] Plsek PE. Systematic design of healthcare processes. Qual Health Care 1997;6:40-8

[2] Hall R, Belson D, Murali P, Dessouky M. Modeling patient flows through the healthcare system. In: Hall R, editor. Patient flow: reducing delay in healthcare delivery. New York, NY: Springer; 2006. pp. 1-44

[3] Morrison BP, Bird BC. Healthcare process analysis: a methodology for modeling front office and patient care processes in ambulatory health care. Proceedings of the 35th conference on Winter simulation 2003:1882-6.

[4] Ceglowski A, Churilov L, Wassertheil J. Knowledge discovery through mining emergency department data. Proceedings of the 38th Annual Hawaii International Conference on System Sciences (HICSS '05) 2006;5:142-3.

[5] Wong C, Geiger G, Derman YD, Busby CR, Carter MW. Healthcare process analysis: redesigning the medication ordering, dispensing, and administration process in an acute care academic health sciences centre. Proceedings of the 35th conference on Winter simulation 2003:1894-902.

[6] Bowers J, Lyons D, Mould G, Symonds T. Modelling outpatient capacity for a Diagnosis and Treatment Centre. Healthcare Management Science 2005;8(3):205-21.

[7] Lusignan S, Pritchard K, Chan T. A Knowledge-management Model for Clinical Practice. Journal of Postgraduate Medicine 2002;48(4):297-303.

[8] Amberg M, Gräber S. Specifying hospital information systems using process modeling, In: Brender J, Christensen JP, Sherrer JR, McNair P editors. Medical Informatics Europe '96. Amsterdam: IOS Press; 1996. pp. 1037-41.

[9] Ehlers F, Ammenwerth E, Haux R, Pohl U, Resch F. Analysis and modeling of the multi-professional treatment process: preliminary results. Medinfo 2001;10(Pt 1):324-8.

[10] HAS. Certification des établissements : les indispensables à télécharger. http://www.has-sante.fr/portail/display.jsp?id=c_435008 (last visit: 12/004/2006).

[11] ISO. ISO 9000 and ISO 14000 – in brief. http://www.iso.org/iso/en/iso9000-14000/understand/inbrief.html (last visit: 12/04/2006).

[12] The Occupational Health & Safety Group. The Health and safety & OHSAS Guide. http://www.ohsas-18001-occupational-health-and-safety.com/index.htm (last visit: 12/04/2006).

[13] Agency for Healthcare Research and Quality. Quality Tools. http://www.qualitytools.ahrq.gov/ (last visit: 12/03/2006).

[14] HAS. Organisation du circuit du médicament en établissement de santé. Fiche thématique. 2005.

[15] Gehrmann A, Ishizu S. A framework for modeling, designing and simulation of management systems. A generic standard approach. AI, Simulation & Planning In High Autonomy Systems. March 6 -8, 2000, Tucson.

[16] Kueng P, Kawalek P. Goal-Based Business Process Models: Creation and Evaluation. Business Process Management Journal 1997;3:17-38.

[17] Staccini P, Joubert M, Quaranta JF, Fieschi M. Mapping care processes within a hospital: from theory to a web-based proposal merging enterprise modelling and ISO normative principles. Int J Med Inform 2005 Mar;74(2-4):335-44.

[18] Knowledge Based Systems, Inc, IDEF Family of Methods, http://www.idef.com/idef0.html (last visit: 12/04/2006).

[19] Anonymous. Writing Course Objectives and Program Objectives. http://www.lco.edu/facstaff/curric/writing_course_objectives.htm (last visit: 12/04/2006).

[20] Ushold M, King M, Moralee S, Zorgios Y. The Enterprise Ontology. The Enterprise Ontology. The Knowledge Engineering Review 1998;13:special issue.

[21] Hashmia N, Boxwala AA, Zaccagnini D, Fox J. Formal Representation of Medical Goals for Medical Guidelines. In: Fieschi M, Coiera E, Yu-Chan JL, editors. Medinfo 2004. Amsterdam: IOS Press, 2004; p. 1633.

[22] AFNOR. Système de management de la qualité: indicateurs et tableaux de bord. Paris: AFNOR, 2000

[23] Agency for Healthcare Research and Quality. National Healthcare Disparities Report: Background on the Measures Development Process. http://www.ahrq.gov/qual/nhdr02/nhdrprelim.htm (last visit: 12/04/2006).

[24] Anonymous. Failure mode and effects analysis. A hands-on guide for healthcare facilities. Health Devices 2004 Jul;33(7):233-43.

[25] US Environmental Protection Agency. Environmental Information Management System. http://www.epa.gov/eims/ (last visit: 12/04/2006).

[26] Neiger D, Churilov L, zur Muehlen M, Rosemann M. Integrating risks in business process models with value focused process engineering. Proceedings of the 14th European Conference on Information Systems (ECIS 2006). 2006. Göteborg, Sweden: AIS.

Address for correspondence

Pascal Staccini – Laboratoire STIC Santé UFR Médecine, 28 av. de Valombrose, 06107 Nice cedex 2, France
e-mail: pascal.staccini@unice.fr
url: http://medecine.unice.fr – url: http://www.essquad.fr

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Near-Miss and Hazard Reporting: Promoting Mindfulness in Patient Safety Education

Leanne M. Currie[a,b,c], Karen S. Desjardins[a], Patricia W. Stone[a], Tsai-ya Lai[a], Eric Schwartz[a], Rebecca Schnall[a], Suzanne Bakken[a,c]

[a] *Columbia University School of Nursing, New York, USA*
[b] *New York Presbyterian Hospital, New York, USA*
[c] *Department of Biomedical Informatics, Columbia University, New York, USA*

Abstract

Patient safety efforts advocate for transforming healthcare from a culture of blame to a non-punitive "culture of safety." One of the most challenging hurdles is to encourage healthcare practitioners to be mindful about their activities. To promote mindfulness during clinical education for entry-level nurses, we developed a web-based dangerous situation and near miss reporting system for a cohort of baccalaureate nursing students (N=156). For this curricular innovation project, we provided wireless handheld devices to students who were required to submit one report for each of their clinical days in the medical-surgical, pediatric, psychiatric, obstetrics and community settings. During a ten-week period, students submitted 1487 reports. Of these, 63% were dangerous situations and 37% were near misses. The most frequently occurring dangerous situations were poor infection control practices. The most commonly reported near misses were medication errors. Free text comments from students identified inadequate patient identification and poor documentation as commonly occurring dangerous situations observed.

Keywords:

Medical Errors, voluntary reporting, patient safety education

Introduction

Despite efforts to improve safety of healthcare environments [1, 2], many patient safety concepts are not fully integrated into educational curricula and are instead acquired when clinicians begin clinical orientation for their job [3]. Although there is a growing interest in integration of these concepts into healthcare education [4], formal patient safety training is minimally represented in medical, nursing and pharmacy schools. Internationally, academic programs for patient safety are beginning to be formalized. In the United States the Institute for Healthcare Improvement is partnering with medical, nursing and pharmacy schools to formalize patient safety education [5], however, patient safety curricula have yet to reach the main stream.

Background

Safety culture dimensions

Promotion of a safe culture in the healthcare environment is increasingly being seen as a critical factor in reducing medical errors [6]. A safe culture is one in which identification of errors is seen as an opportunity to improve the system. Historically, an error was viewed as *an incident*, which was committed by an individual who was then reprimanded. Unfortunately, this stimulated individuals to cover up errors for fear of being punished. Moving from a culture of blame and shame to a culture of safety is an actionable goal in healthcare. However, the perception that error reporting will lead to reprimand persists as a barrier to voluntary reporting [7].

Towards the goal of enhancing patient safety, several dimensions of safety culture have been identified [8]. These dimensions include, i) leadership commitment, ii) professional salience, iii) presence of a non-punitive environment, iv) systems for error reporting, and v) communication. Leadership commitment is defined as the degree to which a leader perpetuates a culture of safety (e.g., fixes problems identified by staff; promoting a blame free environment). Professional salience is the degree to which healthcare professionals perceives that they can perform their work safely. A non-punitive environment is one in which error reporting does not result in punishment and in which systems for error reporting are in place to capture all errors (including near misses). Lastly, communication methods must be in place and communication must be encouraged to promote a culture of safety [8].

Two elements that, if identified early, can ostensibly lead to system improvement include staff identification of dangerous situations (or hazards) and near misses [9]. Dangerous situations include those situations that make the environment or care process hazardous or unsafe. Near misses are those situations in which an event nearly happens, but in which an *interception* prevents the event from reaching the patient. Two types of interceptions have been identified in relation to near misses: 1) a planned interception which occurs when a system is in place to prevent an event from happening (e.g., two nurses checking transfusion); and 2) an unplanned interception which occurs when an event is prevented serendipitously. Double-checking blood products is a systematic barrier to prevent error that is widely used to promote patient safety. Encouraging vol-

untary reporting of dangerous situations and near misses is a method to identify risks so that institutions might be able to develop other barriers to prevent errors from occurring as well as promote the clinicians' mindfulness regarding patient safety.

Electronic error reporting

Electronic error reporting is increasingly common [10]. Nurses perform approximately 50% of voluntary reporting in electronic systems [10, 11], likely because incident reporting was historically considered a nursing activity. Milch and colleagues examined more than 92,000 reports and found that 13% of the reports were near misses and 14% were environmental problems. However they also found that reporting rates were variable between institutions and that physicians reported less than two percent of the events [11]. Low physician reporting rates were also noted by Tuttle [10]. The implications of who performs reporting is unclear.

The purpose of this report is to describe a curricular innovation project in which mindfulness of patient safety was promoted by providing baccalaureate nursing students with a Web-based reporting system to document dangerous situations and near misses observed during their clinical experiences in multiple sites in a large metropolitan area.

Methods

Website development

We developed a Web Portal, which was modeled after the electronic Medical Event Reporting System for Hospitals (MERS-TH), a system in use at Columbia University Medical Center (CUMC) and elsewhere [12]. Table 1 shows the event type categories used in this project. The categories *infection* and *food/nutrition* were not in the MERS-TH system, but were added because they were considered important based on the clinical expertise of the team.

The website was accessed by students from the student page at the school of nursing. Figure 1 shows the interface from which the student would login. Once logged into the website, the student was asked two questions:

> **Q1:** On your shift today, were there any "dangerous situations" that could cause a future event?
>
> **Q2:** On your shift today, were there any near misses (i.e., events that almost happened)?

For each of these questions there were 13 and 26 associated questions, respectively. The student was able to select as many elements as he or she wanted to document. In addition to the questions and sub-categories, a free text comment section was provided at the end of each question.

Student training and reporting

Use of the Hazard and Near Miss reporting system was part of an overall Patient Safety education series, which was integrated into the students' curriculum. The students received training to use the web-based system via a demonstration and one-on-one support if necessary. Academic leadership decided that students would be required to submit one report for each of their two clinical days every week during their clinical time.

Data Analysis

Data were analyzed for frequencies and proportions of each reported question and sub-category. In addition, comments were analyzed for themes using content analysis techniques.

Table 1 - Dangerous situation and near miss event-type categories with examples

Event Categories	Examples
Accident (non-fall)	Needle stick, electrical shock, burn, poisoning
Environmental Hazard/Safety	Body fluid exposure, chemical exposure, chemotherapy spill, hazardous material spill
Equipment/ Device	Equipment malfunction, poor maintenance, inappropriate use, non-availability
Fall	Factors related to the individual or the environment
Food/Nutrition*	Diet and NPO orders
Infection*	Sterile precautions, hand washing
Laboratory	Laboratory orders or results
Medication	Prescribing, ordering/ documenting, administering, monitoring
Patient Disappearance	Increase risk of patient disappearance
Procedure/ Treatment	Consents, delays, wrong procedure/treatment, failure to perform
Restraint	Improper bedrail use and other types of restraint use
Transfusion	Sample collection or product administration
Other	Another type of risk

* Items added to extend MERS-TH categories

Results

Student reporting

Data for this project were collected over two five-week clinical rotations. During the ten-week period, 156 students submitted a total of 3086 reports with responses to the two questions (dangerous situations and near misses), for a total of 6172 responses. Of these, 1487 (24%) responded 'Yes' and 4685 (76%) responded 'No'. Of the 'Yes' responses, 933 (63%) reported Dangerous Situations and 554 (37%) reported Near Misses. Tables 2 and 3 display the rates and percentages for each of the subcategories for each report. There were almost twice as many reported dangerous situations as near misses.

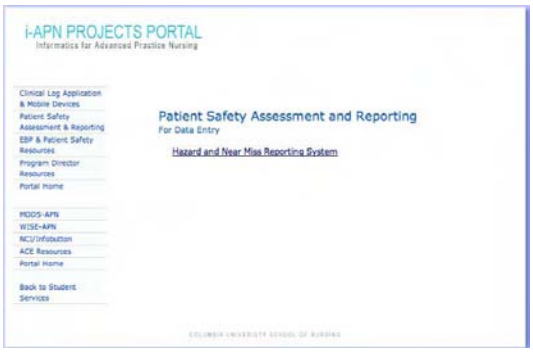


Figure 1. Hazard and Near Miss Reporting Website

Near miss reports

Of the near misses, 208 (38%) were noted to have a planned interception in place and 346 (62%) were intercepted by an unplanned act. The most frequently reported near miss was medication-related and was rescued by an unplanned interception. Transfusion-related near misses were relatively infrequent as were laboratory-related near misses (see Table 2).

Table 2 - Reported Near Misses (N = 554)

Category	Interception Type	n (% of all near misses)
Accident (non-fall)	Planned	23 (4.2)
	Unplanned	25 (4.5)
Environmental Hazard / Safety	Planned	16 (2.9)
	Unplanned	17 (3.1)
Equipment / Device	Planned	25 (4.5)
	Unplanned	39 (7)
Fall	Planned	13 (2.3)
	Unplanned	35 (6.3)
Food/ Nutrition	Planned	10 (5.5)
	Unplanned	22 (3.8)
Infection	Planned	33 (6.0)
	Unplanned	29 (5.2)
Laboratory	Planned	2 (3.7)
	Unplanned	11 (1.9)
Medication	Planned	43 (7.8)
	Unplanned	69 (12.5)
Patient Disappearance	Planned	15 (2.7)
	Unplanned	11 (1.9)
Procedure/ Treatment	Planned	15 (2.7)
	Unplanned	29 (5.2)
Restraint	Planned	2 (3.7)
	Unplanned	8 (1.4)
Transfusion	Planned	2 (3.7)
	Unplanned	3 (0.5)
Other	Planned	9 (1.6)
	Unplanned	48 (8.7)

Note: Percentages do not total 100% due to rounding

Dangerous situation reports

The most commonly occurring dangerous situations were environmental hazards, equipment or device failures, infection control processes, and medication-related situations. Only five transfusion-related situations were reported and only seventeen laboratory-related situations (see Table 3).

Table 3 - Reported Dangerous Situations (N = 933)

Category	n (% dangerous situations)
Accident (non-fall)	49 (5.25)
Environmental Hazard/ Safety	134 (14.4)
Equipment/Device	135 (14.5)
Fall	51 (5.5)
Food/Nutrition	74 (7.9)
Infection	138 (14.8)
Laboratory	17 (1.8)
Medication	128 (13.7)
Patient Disappearance	33 (3.5)
Procedure/Treatment	64 (6.9)
Restraint	18 (1.9)
Transfusion	5 (0.5)
Other	87 (9.3)

Note: Percentages do not total 100% due to rounding

Free text comments

Seven hundred and forty-four free text comments were submitted; 495 were associated with the dangerous situation questions and 249 were associated with the near miss questions. Table 4 provides examples of comments for reported dangerous situations.

Table 4 - Examples of comments for dangerous situations

Response Categories	Example
Accident (non-fall)	Needle left on patient's bed
Environmental Hazard/ Safety	Wet floor
Equipment/Device	Broken BP cuff
Fall	No rails in bathroom
Food/Nutrition	Milk left on counter

Response Categories	Example
Infection	No alcohol hand cleanser
Laboratory	Urine specimens without labels in bathroom
Medication	Medication cart drawer left open
Patient Disappearance	Doors open on psychiatric unit
Procedure/Treatment	Delay with post op patient getting medication
Restraint	Patient restrained using pins and gauze post operatively
Transfusion	No comments reported for transfusion events
Other	Illegible documentation

The major themes that were noted were as follows: 1) poor infection precautions related to poor hand washing, lack of soap or alcohol-based hand cleansing solution; 2) environmental hazards such as construction, wet floors; and 3) medication-related issues such as a medication cart being left open and expired medication on a code blue cart. Of particular interest were the comments related to the 'Other' category for both questions. Of 87 reported dangerous situations in the 'Other' category, 79 had comments. The common themes across the dangerous situation comments for the 'Other' category included: 1) privacy issues with patient information being communicated in public spaces, patient charts being left in public location or a computer screen with patient information being on display; 2) patient identification bands either not being on patients' wrists (or ankles); 3) patient data concerns such as the wrong labels in a patient's chart, documenting on an incorrect patient and finding the wrong note in a patient's chart; and 4) three comments noted physician and nurse handwriting as being illegible.

Of the 57 'Other' reports for the near miss question, 45 comments were submitted. Two comments related to near misses with planned interception reported a missing wristband and documentation on the wrong chart. Of the 48 near miss reports with unplanned interceptions, 43 had comments. These comments addressed a wide variety of near miss events including inaccurate allergy documentation, nearly placing baby in the wrong bassinette, and near patient disappearance in the psychiatric setting. Tables 4 and 5 provide examples of comments in each of the categories.

Table 5 - Examples of comments for each reported near miss

Question 2: On your shift today, were there any events that almost occurred ("near misses" or "good catches")?		
A planned interception occurs when a system is in place to prevent an event from happening (e.g., two nurses checking transfusion). Unplanned interception occurs when an event is prevented serendipitously.		
Response Categories	**Interception Type**	**Examples**
Accident (non-fall)	Planned	Needle on counter, placed in sharps container
	Unplanned	Baby bathed in baby warmer, which got very hot, baby could have been burned
Environmental Hazard/Safety	Planned	Construction taking place. Negative air flow vent working
	Unplanned	Noticed that soap was out in all the bathrooms had maintenance come and refill them
Equipment / Device	Planned	Electronic thermometer broken, new thermometer obtained
	Unplanned	Glucose monitor reading incorrectly, second device used
Fall	Planned	Nurses' aide found patient trying to walk by self
	Unplanned	Patient's roommate prevented patient from getting out of bed
Food/ Nutrition	Planned	Diabetic diet not given to patient; nurse checked when handing out tray
	Unplanned	Wrong food tray given to patient, patient's family notified staff
Infection	Planned	Family members not gowning up before visiting patient with contact precautions
	Unplanned	Staff not informed about TB positive patient, Unit Clerk found out when talking to Lab
Laboratory	Planned	Labels hastily placed on samples, but double checked with nurse
	Unplanned	A patient's birthday was recorded incorrectly on a lab form, and was noticed by the administrator of the clinic, and fixed

Medication	Planned	Wrong medication drawn into syringe, nurse noticed when double checked vial
	Unplanned	Pharmacist refilled medication that had been discontinued. RN caught mistake
Patient Disappearance	Planned	Patient with psych issues had tried to escape from her room several times and had wandered out of the unit during the night before
	Unplanned	Patient left floor to smoke, found when nurses' aide went for lunch
Procedure/ Treatment	Planned	NPO patient who's surgery was cancelled waited very long time for food
	Unplanned	Isolation patient unable to reach call bell
Restraint	Planned	Patient had all 4 bed rails raised. Two rails lowered because 4 side rails is a restraint
	Unplanned	Restraints inappropriately applied by previous shift, reapplied by nurse
Transfusion	Planned	Two units to be given, but five units sent by blood bank, caught when double checking
	Unplanned	Nurse entered patient information into wrong chart then realized her mistake
Other	Planned	Illegible charting, second nurse asked for clarification
	Unplanned	Wrong chart was put with the wrong patient, quickly corrected once doctor looked at it

Discussion

We developed a Web-based patient safety education curricular innovation project in which we asked baccalaureate nursing students to report dangerous situations and near misses during their clinical experiences. We provided students with web-enabled handheld devices to enter these reports. Dangerous situations accounted for more than twice as many reports as near misses, which may indicate that dangerous situations are more visible and therefore easier to report for entry-level students. Alternatively, near misses may occur less frequently because systems are typically in place to catch these types of events, or perhaps, the difference resulted from the fact that students at this stage of learning might not notice the nuances of near miss type events.

Infection control practices were one of the most common reports from this student group. Infections are not typically reported in a patient safety reporting system, instead, they re collected by the infection control nurse in a separate system. However, the ability to examine near miss and hazardous situations related to infections may improve infection control practices. Three reports indicated that patients on isolation precautions were not frequently monitored by staff. In one instance, the student reported that the patient had not been seen 'all night'. This may help to elucidate the report by Stelfox, Bates and Redlemeier who identified higher rates of falls in patients on isolation precautions [13]. We intentionally did not track actual incident reports in this project, because the students were taught to report actual events via the appropriate methods at their sites. Therefore, such outcomes cannot be identified.

The second and third most commonly reported dangerous situations were environmental and equipment problems. The list of problems included ongoing institutional construction and water on floors with a large time delay before clean up. These hazards could be easily remedied by improving the safety system. Medication mishaps were the most commonly occurring near miss with unplanned interceptions being more common than planned interceptions. Although the proportion was small, it would be important to identify the root causes of these problems. A very small number of reports identified problems with blood product processes. The two comments related to blood products indicated that the near misses were caught by processes in place in the organization. Double checking blood immediately prior to blood administration is an example of a barrier that is in place very *near* to the patient. Although not a specific category in our reporting system, lack of patient identification bracelets, documentation in the wrong chart and other patient identification methods was a common report. This item could have been documented in the procedure/treatment section, but this was not intuitive to students, thus creation of a communication problems item might be suggested. Some comments were not categorized correctly which is likely due to the novice status of students in this phase of nursing education, however, most comments were relevant and represented mindful thinking.

Conclusion

Encouraging students to participate in reporting during their studies may help to improved voluntary reporting in general. More importantly, it may help to improve students' mindfulness regarding the proximity of barriers in relation to the patient. The wide variety of comments demonstrated that the students were indeed mindful of the clinical environment despite being novices. It is hopeful that integrating patient safety concepts and voluntary reporting into the curriculum will instill safety culture concepts sufficiently early to ensure safe practice throughout one's career.

Acknowledgments

The Wireless Informatics for Safe and Evidence-based Advanced Practice Nurse Care (WISE-APN) project is sponsored by an educational grant from the Health Research Services Administration (HRSA). The authors thank the members of the Mobile Decision Support for Advanced Practice Nursing (MODS-APN) team for their advice and assistance.

References

[1] Committee on Data Standards for Patient Safety. Patient Safety: Achieving a New Standard for Care. Washington, DC: Institute of Medicine; 2003.

[2] Committee on Quality of Health Care in America. To Err Is Human: Building a Safer Health System. Washington, DC: National Academies of Press; 1999.

[3] Bakken S, Cook SS, Curtis L, Desjardins K, Hyun S, Jenkins M, John R, Klein WT, Paguntalan J, Roberts WD, Soupios M. Promoting patient safety through informatics-based nursing education. International Journal of Medical Informatics. 2004;73(7-8):581-9.

[4] Effken JA, Carty B. The Era of Patient Safety: Implications for Nursing Informatics Curricula. J Am Med Inform Assoc. 2002 November 1, 2002;9(90061):S120-3.

[5] Institute for Healthcare Improvement. IHI Health Professions Collaborative Projects. 2006 [cited, November 2006]; Available from: http://www.ihi.org/IHI/Topics / HealthProfessionsEducation /EducationGeneral/

[6] Longo DR, Hewett JE, Ge B, Schubert S. The Long Road to Patient Safety: A Status Report on Patient Safety Systems. JAMA. 2005 December 14, 2005;294(22):2858-65.

[7] Potylycki MJ, Kimmel SR, Ritter M, Capuano T, Gross L, Riegel-Gross K, Panik A. Nonpunitive Medication Error Reporting: 3-Year Findings From One Hospital's Primum Non Nocere Initiative. Journal of Nursing Administration. 2006;36(7-8):370-6.

[8] Agency for Healthcare Research and Quality. Hospital Survey on Patient Safety Culture. . Update 2006 [cited 2006 December 3]; Available from: http://www.ahrq.gov/qual/hospculture/

[9] Kaplan HS. Getting the right blood to the right patient: the contribution of near-miss event reporting and barrier analysis. Transfus Clin Biol. 2005 Nov;12(5):380-4.

[10] Tuttle D, Holloway R, Baird T, Sheehan B, Skelton WK. Electronic reporting to improve patient safety. Qual Saf Health Care. 2004 August 1, 2004;13(4):281-6.

[11] Milch CE, Salem DN, Pauker SG, Lundquist TG, Kumar S, Chen J. Voluntary Electronic Reporting of Medical Errors and Adverse Events. An Analysis of 92,547 Reports from 26 Acute Care Hospitals. Journal of General Internal Medicine. 2006;21(2):165-70.

[12] Kaplan HS, Callum JL, Rabin Fastman B, Merkley LL. The Medical Event Reporting System for Transfusion Medicine: will it help get the right blood to the right patient? Transfus Med Rev. 2002 Apr;16(2):86-102.

[13] Stelfox HT, Bates DW, Redelmeier DA. Safety of patients isolated for infection control. Journal of the American Medical Association. 2003 2003 Oct 8;290(14):1899-905.

Address for correspondence:

Leanne M. Currie, RN, DNSc
Columbia University
Mail Code 6, 630 West 168th Street,
New York, NY, USA 10032
Telephone: 1-212-342-3919
Fax: 1-212-305-6937
e-mail: lmc2007@columbia.edu

Chapter 2.

Data Repositories and Information Retrieval

This page intentionally left blank

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

A Territory-Wide Electronic Health Record – from Concept to Practicality: the Hong Kong Experience

Antonio CH Sek[a], NT Cheung[a], KM Choy[b], WN Wong[a], Anna YH Tong[a], Vicky H Fung[a], Michael Fung[c], Eric Ho[c]

[a] *Department of Health Informatics, Hospital Authority, Hong Kong SAR, China*
[b] *Chief Manager, Service Transformation, Hospital Authority, Hong Kong SAR, China*
[c] *Information Technology Division, Hospital Authority, Hong Kong SAR, China*

Abstract

In Hong Kong, a pilot project is being undertaken to implement a web-based electronic patient record system to allow integrated, real time patient based information to be shared in clinics, private and public hospitals. Such sharing aims to ensure that complete and accurate healthcare information is available to citizens' multiple points of care through a stable IT system. A challenge is to share this electronic information whilst ensuring privacy and security. Hong Kong has achieved its initial goals and pioneered in building a territory-wide electronic health record (EHR). This paper will outline the tasks involved, approach, method used and initial review of the pilot project. Barriers to implementation are discussed and critical success factors are identified.

Keywords:

Medical records systems, computerized; implementation; evaluation; Hong Kong; electronic patient records; public-private interface.

Introduction

The Institute of Medicine's 2000 report *To Err Is Human* generated a flurry of publicity on the issue of medical errors [1]. The report concluded that as many as 98,000 deaths occur annually in U.S. hospitals as a direct result of medical errors. In his review of patient safety issues in the Canadian healthcare system, Dr. Matthew Morgan stated [2] that "coordinated national EHR initiatives will cost less, save lives and prevent harm when compared to the status quo." An interoperable EHR is seen as the one solution that solves most of the problems. According to one study in the US, universal adoption of a complete EHR will save US$77.8 Billion annually [3]. In Canada, the federally funded Infoway has estimated savings of CAD $3.4 billion [4].

The race is now on for electronic health record adoption. Denmark is leading the pack. As such, Denmark has a public healthcare system and developed a national system to electronically communicate messages between healthcare stakeholders since 1994.

In recent years many other governments have also set up e-Health programs to implement EHR. However, success in Denmark has not been replicated in many other OECD countries and barriers to adoption have been identified, such as interoperability, security, and financial considerations.

It is estimated that only about five percent of all U.S. primary care providers currently use EHR systems [5], and other studies describe rates at under 10% in all inpatient settings [6]. EHR adoption is hindered by a more fundamental failure of the HIT market in general in the United States [7]. The concept of an interoperable EHR is easy to embrace, but difficult to implement. This paper will outline the method used in Hong Kong to build a territory wide EHR.

Background

Hong Kong has a population of over 7 million people. There are 44 public hospitals and 12 private hospitals in Hong Kong. All 44 public hospitals are governed and managed by the Hospital Authority (HA). In the past, private hospital treatment was often the choice for those who could afford them. However, improvement in public hospital standard since the establishment of the HA has substantially narrowed the quality gap between the public and private sectors. Currently, private providers accounted for around 5% of all inpatient bed days in Hong Kong [8]. Nowadays, patients often seek care in both public and private sectors, and medical manpower is about equal in the two sectors

HA adopted a centralized approach in developing its IT system for clinical care, the clinical management system (CMS). This has reduced IT cost per hospital, whilst ensuring compatibility. High mobility of both patients and doctors characterizes practices in Hong Kong. Therefore, inter-operability between different hospitals was one of the key aims in developing the clinical management system. A familiar user interface is important in enhancing the efficiency in a busy clinic setting. The HA CMS has very high user acceptance and is in use in everyday care delivery [9].

From 2000 onwards, a web-based system of patient lifetime record in public hospital has been developed - the

electronic patient record (ePR). This system provides a comprehensive and easy to use view of all patient encounters within the Hospital Authority in the lifetime of the patient. Tremendous responses have been received from both the patient and doctor perspectives. This can reduce cost by minimizing repetition of investigations and administrative cost and delay in obtaining patient's information for continuation of care from another hospital. It can reduce risk by having a complete, up-to-date medical history at the point of care. Now the benefits of this longitudinal electronic record are being piloted to the private sector and wider community.

Methods

The Public Private Interface – Electronic Patient Record (PPI-ePR) project commenced development in mid-2005 with the following objectives:

- to enhance collaboration between the public and private sectors;
- to allow continuity of care for the patients;
- to facilitate a free flow of patients between the two sectors; and
- to allow timely access of information.

A one-year pilot is currently being conducted to test the concept, the technical feasibility, including the protection of patient data privacy in the internet environment. During this phase, selected clinical records in the ePR are being shared (with the consent of patients) with participating practitioners through the internet, as depicted in Figure 1.

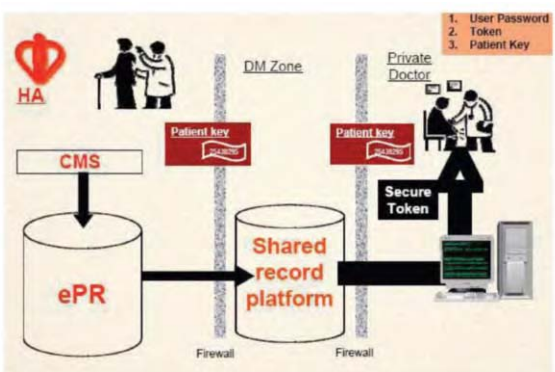


Figure 1 - Schematic diagram of Hong Kong ePR sharing mechanism

On the one hand, we want the free flow of information between the public and private sectors. On the other hand, we want the tightest security to protect patients' privacy and confidentiality.

When a patient consents to sharing his record from HA to the private sector, his record in HA will be extracted from the ePR in HA's internal network to a Public-private interface (PPI) database, where important data fields such as name, Hong Kong identity card number, date of birth and address are encrypted using triple DES encryption algorithm. The platform is protected by firewalls and intrusion detection systems. Patient will be given a PIN (personal identification number) when his record is put on the database.

Private doctors who wish to view patient's records need to register with the HA and agree to comply with the rules and regulations of the project. He will then be given a PIN and a security token. A 6-digit number will be refreshed once per minute on the token. These form the two-factor authentication for doctor's sign in.

Figure 2 - Two-factor authentication of doctors login via web browser

An enrolled doctor has to logon to the PPI-ePR system via web browser with his password and security code generated by the security token (figure 2). To access the patient's ePR record, the patient needs to provide his Hong Kong identity card number and his PIN for input at the web browser. Upon successfully authenticating the enquiry request, PPI-ePR will display the patient's record on the web browser, which is protected all the way through in an encrypted tunnel through the public internet. The contents shared will include: diagnosis, procedures done, discharge summaries, laboratory reports, radiological appointments and reports, medications, allergies and recent appointment schedules. This content covers much of ASTM Continuity of Care Record [10]. On the end-user workstation, we have been careful not to cache pages. As an additional security measure, an SMS message will be sent to the patient's mobile phone, alerting him of such access (figure 3).

Figure 3 - An SMS message showing ePR access

Results

From the start of the pilot in April 2006, enthusiastic responses have been received from many healthcare providers. They range from solo doctor clinics to private hospitals with more than 600 beds. Doctor-to-patient ratio ranges from 1:1 to 1:200 in elderly care homes. This creates different challenges to implementation of the project, but also trials implementations in different healthcare settings.

Doctor registration

We enlisted the help of the three largest doctors' association groups to recruit doctors, namely the Hong Kong Medical Association, Hong Kong Doctors Union and the Association of Licentiates of Medical Council of Hong Kong. A total of 117 doctors were registered this way.

Three group practices have been invited to join and signed a Memorandum of Understanding before their member doctors and nurses could join the project. 30 healthcare providers were registered via their groups. We also worked with five private hospitals in recruiting patients. In total, 131 healthcare providers (doctors and nurses) from private hospitals have joined the program.

We reached the quota of 500 doctors in one month and the waiting list has about 200 doctors, demonstrating the support and enthusiasm of the private sector for the program.

Patient registration

It has taken longer for patient registration to catch up with doctor registration. Since not all HK citizens will attend both private and public clinics, we hope to target our pilot patient population to those standing to benefit most from our pilot. Two target patient groups were those in elderly care homes, or in shared care program. Patients will be given drugs and special tests for chronic disease in the public whilst other problems are dealt with by their family doctors. The PPI-ePR acts as a bridge between the two sectors to provide real time, accurate information to enhance quality of care. As of March, 7800 patients have registered in the program and their private doctors are reviewing their medical record via the secure platform.

Future work will focus on expanding the program to cover more doctors and patients, and to make use of the Hong Kong SmartID card for registration and authentication. However, during this pilot, a few critical success factors and barriers to adoption have been identified, as discussed below.

Discussion

There are several critical success factors identified for community wide EHR. Protection of individual privacy is of paramount importance because it will take some time for citizens to have confidence in the system security and reliability. Having a pilot is a good way to pave for wider acceptance. Any information leak via the system will be a severe blow to the progress towards a Community wide EHR. Therefore, every precaution must be taken in the design of the system, while not affecting the ease of use.

Security and confidentiality

The PPI-ePR system houses a strong enterprise-wide encryption, two factor authentication, centralized database which allows audit trails, additional SMS and protected by surrounding firewalls and intrusion detection system. Since the start of the pilot, there is no report of security nor privacy issues.

An additional independent audit has been done in Aug 2006 to ensure that adequate security and control measures have been put in place to protect the privacy of patient information in accordance with the relevant requirements of the law.

Broad participation

Another important factor is broad participation, particularly from the private health sector, and building of community and stakeholder trust and confidence. In the pilot, over 500 doctors joined in the program and started to use the system. This accounts for about 10% of the private doctors in the market. A survey by the Hong Kong Medical Association showed that 87% of the private doctors are interested in using the PPI-ePR system. This supports our observation that response for doctor registration is very strong.

Nonetheless, there are certainly some barriers to adoption. The most obvious one is the lack of standards for community wide interoperability. In the US, electronic patient information resides on many isolated islands that have been very difficult to bridge. Hong Kong has taken a centralized approach in developing the ePR and thus solved most of the interoperability issues during its development, using the same data standard and vocabulary. The next phase will be to explore sharing of the HA system with the private sectors. However, a further development of the community-wide ePR is to encourage two-way sharing, so that private health records are also shared with the public sector, which leads to the need to delineate roles, responsibilities and ownership across sectors.

Another challenge is to identify who will make the initial investment in extending this pilot program to a community wide EHR. HA has taken the first steps in investing the manpower and infrastructure for the development of a pilot system. The subsequent maintenance and enhancement costs will need to be shared between the government and different stakeholders to make it financially sustainable.

Conclusion

Hong Kong has taken a centralized approach to develop the community wide EHR and sharing the contents with the private sectors. Ninety-five percent of inpatient records are electronic and shared to the remaining 5% of the healthcare sector. The implementation of EHR in the Hong Kong will ensure that complete and accurate healthcare information is available to Hong Kong citizens at the point of care. The primary goal of EHR is and will be to store and share this electronic information while guaranteeing privacy and security. We believe Hong Kong has demon-

strated a viable platform and pathway to a community wide EHR.

Acknowledgments

We would like to thank Dr Fung Hong, Ms Nancy Tse, Dr WL Cheung and Mr Andre Greyling and members of the PPI Working Group for their support and leadership. We appreciate the HA management team in cultivating this sharing culture and We would like to thank many of the stakeholders in making this pilot a success, thus paving our road to a community wide EHR.

References

[1] Kohn LT, Corrigan JM, and Donaldson MS. To Err is Human: building a safer health systerm. 1st ed. Institute of Medicine: National Academy Press, 2000

[2] Morgan MW. In Pursuit of a Safe Canadian Healthcare System. Safe Healthcare Vol. 5 No. 3 2004

[3] Walker J. The Value of Health Care Information Exchange and Interoperability. Health Affairs 24 (2005): 10-19.

[4] Romanow RJ. The Future of Health Care in Canada. Commission on the Future of Health Care in Canada. National Library of Canada, 2002.

[5] Bates, DW. A Proposal for Electronic Medical Records in U.S. Primary Care. Journal of the American Medical Informatics Association 2003:(10): 1-10.

[6] Ash JS. and Bates DW. Factors and Forces Impacting EHR System Adoption: Report of a 2004 ACMI Discussion. Journal of the American Medical Informatics Association 2005;12(1):8-12.

[7] Middleton B. Accelerating US EHR Adoption: How To Get There From Here, Recommendations Based on the 2004 ACMI Retreat. Journal of the American Medical Informatics Association 2005:(1): 13-19.

[8] Health and Medical Development Advisory Committee – Discussion Papers 2005 Landscape on Health Care Services in Hong Kong. http://www.hwfb.gov.hk/hmdac/ english/ dis_papers/dis_papers_lhcshk.html

[9] Cheung NT, Fung KW, Wong KC, Cheung A, Cheung J, Ho W, Cheung C, Shung E, Fung V, Fung H. Medical informatics--the state of the art in the Hospital Authority. Int J Med Inf 2001 Jul;62(2-3):113-9

[10] ASTM E2369-05 Standard Specification for Continuity of Care Record (CCR). 2006

Address for correspondence

Dr Antonio CH Sek, Department of Health Informatics, Hospital Authority Building, 147B Argyle Street, Kowloon, Hong Kong
Email: sekch@ha.org.hk

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Northern Territory Health*Connect*: Shared Electronic Health Record Service Implementation Experiences and Benefits Realised in Indigenous Health

Stephen Moo, John Fletcher

Northern Territory Department of Health & Community Services

Abstract

This presentation summarises the learnings from the HealthConnect Northern Territory (HCNT) Shared Electronic Health Record Service (SEHR) from Trial to Implementation and the emerging benefits realized as the project is implemented across the Northern Territory of Australia.

The presentation:

- *explores the challenges and experiences of implementing a SEHR service in urban and in some of the remotest regions on the Australian continent.*
- *demonstrates the emerging health benefits e-Health can provide in enabling the sharing of medical information between public and private health service providers in particular the service delivery and benefits provided to a highly mobile Indigenous population who currently experience the greatest health problems and experience difficulties accessing continuum of care created by factors which include remoteness, mobility and communication.*
- *explores the evolvement of the "opt in" consumer consent model adopted by the Territory for the implementation of the HCNT SEHR.*
- *advises of plans for future development, which inform other implementations, and NeHTA standards development for the implementation of the National SEHR Service.*
- *Informs project plans to incorporate NeHTA standards as they are developed and transition the HCNT SEHR to the National SEHR Service when implemented recognising the importance of developing strong partnerships with key stakeholders, in particular consumers, health care providers and system vendors who inform project development and implementation.*

Keywords:

Health*Connect,* e-Health, Health*Connect* NT Shared Electronic Health Record Service.

Introduction

Health*Connect* NT trial

Sponsored by the Australian and Northern Territory Governments, a Health*Connect* research and development trial commenced in Katherine in the Northern Territory in June 2002.

The Trial was established to help determine whether Health*Connect* could improve continuity of care in remote and rural areas of Australia by enabling health care providers timely access to vital health information.

The Trial centered on the development and utilisation of a secure Health*Connect* repository.

With a registered consumer's consent, health service providers participating in the Trial were able to send event summaries of health events including medical summaries and hospital discharge summaries via a secure network to the repository. Once in secure storage, other authorised health care providers involved in the Trial were able to access consumer medical summaries with consumer consent.

Consumers and providers embraced the concept of Health*Connect* in the Territory and 1,800 consumers, mostly from remote areas were registered to participate in the Trial by March 2005.

There were 49 registered Providers involved in the Trial. Users of Health*Connect* NT included Hospital medical officers and district medical officers, some accident & emergency nursing staff, remote clinic doctors, remote clinic nurses and Aboriginal Health Workers.

Consumer consent

During the early stages of the Trial Providers were required to obtain consumer consent at each occasion of care to retrieve and view consumer health information on Health*Connect* and to send a summary of that occasion of service to Health*Connect.* This created some issues in early 2004, where Providers became concerned that the model in their view could be medically dangerous and were also concerned that they could be legally exposed if they failed to obtain consent or forward information to Health*Connect.* Consumers also complained that it was humbug to be asked to consent each time they visited a health care provider. Many stated that the reason they had joined Health*Connect* was to enable providers to access their health information and expected their information to be sent to Health*Connect* after a consultation as a matter of course.

Provider and consumer concerns regarding the requirement to consent at each occasion of care culminated in a major consumer and provider workshop being held in Katherine. The outcome of this workshop resulted in a new consent model being introduced in the latter stages of the Trial with full consumer and provider support. The new consent model complied with national privacy legislation and required a Health*Connect* registered consumer to consent on registration to their health information being sent, retrieved and viewed from the repository on each occasion of service, unless at a particular occasion of service the consumer instructed the provider that he or she did not want the information viewed or sent. This model was adopted by the Territory and proposed by the Territory as the preferred model for national implementation.

Evaluation

The phase 2 evaluation conducted during 2004 found that the Health*Connect* concept was unanimously supported by providers, health care managers and consumers involved in the Trial and that providers and consumers strongly supported the implementation of Health*Connect* across the Katherine Region and the Northern Territory. There was a strong belief among providers and consumers that Health*Connect* would provide major benefits in the delivery of coordinated health care across the Territory. There were also a number of lessons learnt from the evaluation, which are included in the "Lessons Learned from the Medi*Connect* and Health*Connect* Field Test and Trials" Report.

Promotion

During the Trial and as a major promotional tool supporting implementation, promotional material was developed using MARVIN. The development of MARVIN is a result of a collaborative partnership between Industry, Community and Government in the NT and contributes part of its existence to the Australian Flexible Learning Framework's (Framework) LearnScope professional development project. MARVIN allows community members to develop the learning and training resources themselves, typing in their own messages and recording their own voices, in their own languages. What is then seen onscreen is walking, talking computer generated characters, most of them modeled on elders within the community or upon the learners themselves. Promotional material developed for the Health*Connect* NT Trial and more recently for implementation has received wide acclaim.

Health*Connect* NT shared electronic health records service - implementation

Following the success of the Trial and the enthusiastic support of consumers, providers and other major stakeholders, the Territory decided to implement the Health*Connect* NT (HCNT) Shared Electronic Health Record (SEHR) Service Territory wide and to investigate the possibility of expanding the SEHR to cross border regions in South Australia and Western Australia.

Planning for the phased implementation of the Health*Connect* NT SEHR commenced in the latter part of 2004/2005 and initial implementation commenced across the Katherine Region on July 1 2005.

Health*Connect* NT implementation aims to break down over time the Territory's distance barriers to health services delivery, particularly to highly mobile remote indigenous populations by enabling:

- participating providers to create and view shared electronic health records for consumers in participating remote communities; and
- participating hospitals to generate discharge summaries to participating providers automatically.

This will contribute to enhanced clinical services in participating remote communities, including greater ability to manage chronic disease and enhanced care for children in remote communities through greater capability to deliver primary care, childhood development, immunisation and nutrition services.

Consumer registration

Vast distances complicate the registration of consumers to participate in the Health*Connect* NT SEHR, remoteness of communities, climate (in particular wet season access) and cultural diversity. This provides significant challenges in terms of engaging with, and gaining the informed consent of, consumers particularly with Territorians for whom English is a second language and where there is a range of cultural considerations.

- New marketing and promotional materials were developed, tailored to communicate with, and educate consumers from diverse cultural backgrounds and language groups.
- A partnership with Medicare Australia was negotiated and new registration forms for implementation were developed to register consumers to participate in the SEHR service, to confirm Medicare registration details and facilitate new Medicare registrations. A detailed registration process was established with data compatibility checks undertaken to a strict criteria – data compatibility at time of registration is checked to ensure consumer demographic details are a complete match with home health centre CIS, Medicare and CMI (Hospital) data prior to information being entered into the SEHR.
- New pamphlets informing consumers of consent, privacy, provider access to records etc were prepared and issued to consumers at time of registration.
- Consumer Advisory Group (CAG) established. The Chair of the CAG and an Indigenous cultural representative of CAG are represented on the HCNT Implementation Steering Committee.
- A dedicated Consumer Services team was established to focus on consumer registration and provide secretarial support and information to the CAG.
- Consumer registration involved community and stakeholder promotion and the appointment of casual Local Project Officers to assist with registration and consumer identification. This created employment and skills

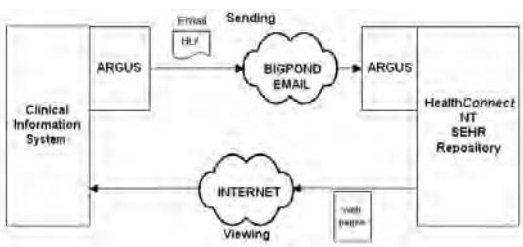
development opportunities in areas of high unemployment and community ownership in the program.

- Consumers across the region embraced the concept of Health*Connect* and readily committed to register to participate. A 90% uptake was realized in remote communities across the Katherine Region.
- Over 9,000 consumers registered to participate in the HCNT SEHR service by 30 November 2006, most from remote communities spread over a very large geographic area.

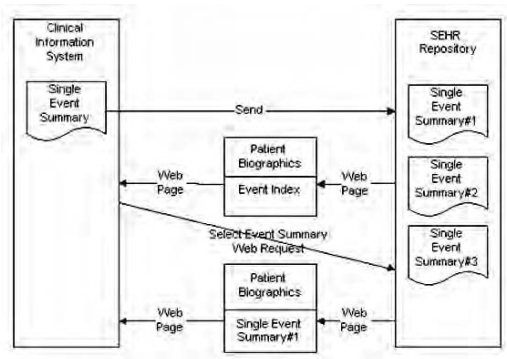
System capability

The HCNT SEHR system is designed around a central server sending and receiving medical information from numerous clinical information systems. The HCNT SEHR system receives secured medical summary messages (called event summaries) from feeder systems and stores them in the HCNT SEHR Repository database. Medical providers via the Internet can view these event summaries.

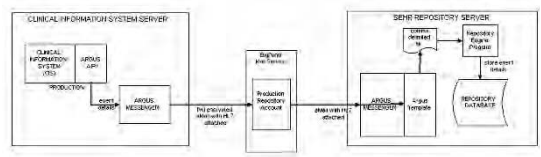
The HCNT SEHR system uses an Argus messaging tool to generate HL7 messages, containing the event summary, and attaches them to HeSA PKI encrypted message sent to the Repository Database via a Big Pond email account. Medical providers using their clinical information system can securely view, using SSL, via the Internet the medical information stored in the SEHR Repository for a specified consumer. The diagram below illustrates these concepts.



The event summaries generated by the feeder systems are usually single consultation events from health centres, or hospital emergency/inpatient discharge summaries, but may also be an initial health profile summary or a pathology test result. All these summaries are stored as discrete pieces and these events can be viewed specifically via the Internet directly or via a web window within a clinical information system. Accessing the medical data over Internet requires medical providers to enter a user id and password. An index listing all the event summaries for a patient is shown. Once selecting the appropriate event summary a web view of that event summary is sent to the viewer. The following diagram shows the basic overview of these concepts.

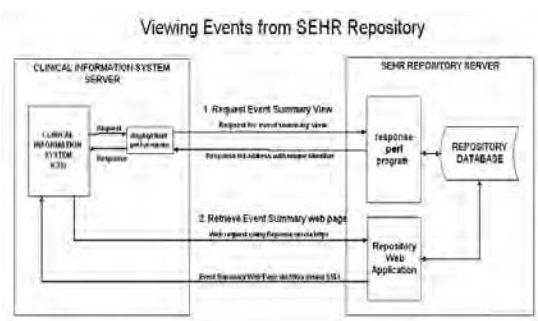


Sending and Viewing Technical Design in detail



The above diagram shows in detail the sending processes. The key to sending is the use of the Argus email tool, which has two key components; the first is the Argus API that allows the Feeder System to directly call functions that interact with Argus Messenger. The second Argus component is Argus Messenger, which can generate an email, attach files to them and encrypt the emails. These emails are sent to a single bigpond email account where it awaits retrieval.

At the HCNT SEHR Repository Argus Messenger is also running, automatically downloading emails from the bigpond email account about every 10 minutes. Argus Messenger has a set of predefined rules that require HL7 attachments on emails to be interpreted by Argus' Template tool, which convert the HL7 into a more readable comma delimited file. In turn a perl program called the Repository Engine reads the comma delimited file and stores the event summary against the appropriate patient in the Repository Database.



The diagram above describes the viewing processes required to allow a medical provider using their Clinical Information System to read the SEHR Repository event summaries.

The first step is a request to view made by the medical provider via their clinical information system, in which the provider must enter their HC NT SEHR user id and password. The HCNT SEHR Repository server through a response program returns a url address with a parameter of the selected patient's unique identifier, only if the provider's user id and password are accepted.

In the second step the clinical information system on receiving the returned url puts this url into its built-in web browser, which makes a web request for that url address. This url address uses https, which uses SSL encryption. The url address location is the SEHR Repository's web server, which returns the requested first web page containing the patient's demographics and an index of event summaries. The event summaries can be reached via links on the presented web pages.

System features via seamless integration into participating feeder systems include:

- automatic flagging notifying of consumer registration status.
- notification (prompt) to send event summaries including inpatient and emergency discharge summaries in hospital systems.
- viewing features of event summaries.
- sending of initial health profiles (IHPS).
- sending of pathology results.

Participating providers are able to create and view initial health profiles, medical event summaries and pathology results for consumers who have registered to participate.

Participating hospitals generate inpatient discharge and emergency event summaries to participating providers automatically for consumers who have registered to participate.

Daily notification to providers of inpatient and emergency discharge summaries for their clients is provided.

Medical summaries generated by the HCNT SEHR Service include:

Medical Summary of a consultation/encounter with Individual - includes Diagnosis, Medications, Immunisations, Observations, Risk Factors, and Allergies.

Hospital Discharge Summaries (ED and Inpatient) - includes Diagnosis, Procedures, Medications, Observations, Pathology Results.

Pathology Results.

Initial Health Profile - 2-year history of patient from their home Health Centre. Includes Diagnosis, Medications, Immunisations, Observations, Pathology Results, Risk Factors, and Allergies.

Current Health Profile – Will replace Initial Health Profile in 2007 and be generated at each occasion of service.

Provider engagement

- Providers enthusiastically supported the HCNT SEHR service, believing it would assist greatly in delivering health services to their clients across the continuum of care.
- 258 providers registered and participating in the HCNT SEHR service by 30 November 2006. These include Hospital medical officers and district medical officers, some accident & emergency nursing staff, remote clinic doctors, remote clinic nurses and Aboriginal Health Workers.
- Clinical Advisory Committee (CAC) established and meets monthly. Representation on the CAC includes Drs, Nurses, Aboriginal Health Workers and Allied Health Professionals. The CAC informs HCNT management and the Implementation Steering Committee on issues including: protocols, privacy, change management / provider engagement and training, QA and system requirements, capability, user friendliness and has a major influence in the system enhancement development program.
- Provider protocols were implemented and provider agreements at organizational level were agreed to. All providers apply to register, are trained in Health-*Connect* and are provided with user ID and Password access.
- HCNT hotlines were introduced to enable providers to obtain HCNT HCIDs for consumers visiting their health service from another location, to address user ID password / access issues and to report system faults.
- Quality Assurance procedures in conjunction with providers were established to monitor data quality and audit provider access.
- Considerable effort was undertaken in engaging providers and influencing change management to enable providers to incorporate utilization of the HCNT SEHR within normal work practice when utilizing feeder
system clinical information systems. A joint Communicare / HCNT SEHR user manual was developed by HCNT with input from participating providers and "how to" sheets were prepared for users of all participating CIS's.
- 17 sites actively participating in the SEHR service at November 2006 as well as the Royal Darwin, Katherine and Tenant Creek Hospitals.
- Provider participation increased markedly during 2005/2006 as a result of the increase in the number of registered consumers, efficient provider engagement practices, system reliability / major decrease in system failure and faults, system enhancements, data quality and the SEHR began delivering real service outcomes.

Governance

New Governance arrangements were implemented in 2006 to reflect Territory e-Health implementation and the HCNT Implementation Steering Committee (HCNTISC) was established.

The HCNTISC is responsible for providing advice on policy and business issues associated with the implementation of

HealthConnect NT in accordance with project objectives and the revised HealthConnect Implementation Strategy.

The following sub committees inform the HCNTISC

- Implementation Management Group
- Consumer Advisory Group
- Clinical Advisory Committee
- Information Technology Advisory Committee
- Indigenous e-Health Advisory Committee

Progressive implementation – major program outcomes:

The Northern Territory is progressively implementing the HCNT SEHR service across the whole of the Northern Territory targeting first all remote communities and urban-based Aboriginal Medical Services.

Implementation of the HCNT SEHR builds upon and complements the implementation of the HCNT Point to Point (P2P) Service and major e-Health initiatives being implemented by the Northern Territory Department of Health and Community Services in the Northern Territory public hospitals and Remote Health Services Program which include:

- Implementation of Advanced Medication Management and Point of Care Clinical Workstation in NT Hospitals.
- Implementation of Remote Communications Infrastructure and WAN connections into all departmental operated Remote Health Clinics.
- Progressive implementation of a standardised integrated Patient Recall and Chronic Disease Management System (Primary Care Information System) into all public Remote Health Clinics.

Results

The implementation of the HCNT SEHR service in the Northern Territory will achieve the following major outcomes:

- *Passing the Patient Safely Back to their Community* – the generation of electronic discharge summaries (including electronically generated discharge medications for patients of NT Hospitals) for inpatient and Emergency Department attendances by consumers registered for the expanded HCNT SEHR Service.
- *Caring for the Chronically Ill and Children* – critical health information about consumers registered for the HCNT SEHR service, including records of immunisations, medications, accessed by participating health providers across the NT, enabling those professionals to better manage chronic disease and high risk child health issues such as "failure to thrive".
- *Breaking Down the Distance Barriers by Improving the Co-ordination and Delivery of Health Services to Indigenous Populations* – engagement and maximisation of registration of residents in the Northern Territory receiving services from health providers participating in the expanded HCNT SEHR service, targeting first, remote indigenous communities, Aboriginal Medical Services operating in remote and major urban centres and small regional centres and towns across the Territory;
- *Greater involvement of consumers in their health care* – establish two (2) consumer resource centres in remote health centres to pilot strategies for consumer access to their SEHR, telehealth services and video-conferencing with relatives in hospital;
- *An informed community ready for participation in integrated health care* – the Northern Territory will make continued progress in the development of the communities' understanding of and willingness to participate in Health*Connect* through localHealth*Connect* marketing and promotions aligned to national e-Health Strategies.

Conclusion

The Northern Territory will build on the proven successful development and implementation of its HCNT SEHR Service. The Northern Territory's HCNT SEHR implementation is the most advanced Health*Connect* service implementation of its kind in Australia in terms of range of services, service coverage, provider and consumer participation. The HCNT SEHR is being expanded to target the most disadvantaged Indigenous populations with the greatest health needs, residing in rural and remote communities and major urban centres. There is clear evidence that Indigenous populations have embraced the concept of HealthConnect with 90 % voluntary consumer participation rate and 100% provider participation achieved. There is also clear evidence of benefits realisation for a highly mobile Indigenous population through improved coordination and quality of health service delivery spanning vast distances, in some of the remotest areas in Australia. These initiatives will provide significant learnings and opportunities to inform the development of emerging national and cross-jurisdictional e-Health implementation strategies and standards/infrastructure developments, that is, the NEHTA work program.

Acknowledgements

The Australian Government Department of Health & Ageing

References

[1] Health Connect Implementation Strategy, version 2.01. June 2005 (revised).
[2] Lessons Learned from the MediConnect Field Test and Health Connect Trials. April 2005.

Address for correspondence

Stephen Moo
Director Health*Connect* Northern Territory
Chief Information Officer
Department of Health & Community Services –
Northern Territory Government of Australia

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

A Simulation-Based Performance Analysis of a National Electronic Health Record System

Leonidas Orfanidis[a,b,c], Panagiotis D Bamidis[d], Barry Eaglestone[b]

[a] *South East European Research Centre, Greece.*
[b] *Department of Information Studies, University of Sheffield, United Kingdom.*
[c] *Department of Computer Science, Greece.*
[d] *Medical Informatics Laboratory, School of Medicine, Aristotle University of Thessaloniki, Greece.*

Abstract

This paper addresses through simulation experiments a number of technical issues which are raised during the development and operation of a National Electronic Health Record System (NEHRS). The simulation experiments represent the NEHRS performance for a variety of technological infrastructures, within the context of a realistic scenario. The scenario includes the estimation of the delays created in queues during the exchange of Electronic Patient Records (EPR) between different health service points. It is essential to clarify the delays derive from LAN and Internet technologies, the EPR encryption/ decryption, the HL7 message generation/parsing, and the databases. The results of this study identify how a number of technical aspects influence the NEHRS development and operation.

Keywords:

National Electronic Health Record System, simulations, performance

Introduction

Creation of national services for maintenance of and access to Electronic Health Records (EHR) has been a prominent issue in health information management and technology strategies in developed countries over the recent decade. In fact, EHR have been assigned a catalytic role within many national information management and technology strategies in healthcare. The UK, 'Information for Health: An Information Strategy for the modern NHS 1998-2005' [1] for example has presented a strategy based upon a vision of patient-centered data flowing freely across all areas of healthcare, to be accessed and utilised wherever needed. Similarly Canada [2], Australia [3], New Zealand [4], Denmark [5,6], Israel [7] have presented comparable plans for their countries. In the same manner, the European Union expects from all its member states to define, promote and achieve progress on e-Health by the end of 2009 [8].

However, National Electronic Health Record System (NEHRS) programmes, as for example the English, expand the NEHRS delivery date since the means to realize this vision remains problematic [9] [10]. A number of problematic aspects are the security and confidentiality issues that occur during data storage and transmission, the lack of standardisation (coding of medical conditions, vocabulary of medicine) and the need to support and reconcile different views on the structure and the content of the EHR due to the different users and their different needs (physicians, patients, nurses, administrative staff) [11]. Another critical issue which has been identified by the English National Health Service (NHS) is the physicians' unwillingness to use the EHR Systems of their Health Service Points (HSP) which in future will be connected building up a NEHRS. This problem is caused mainly by the lack of motivation and computer literacy of the physicians [12]. Particularly, 60% of physicians in England do not want to use available EHR systems in their HSPs, since they feel it overloads the amount of their work without adding any significant advantage [13]. On this problematic issue are also added the EHR Systems harms which are presented because of technical problems on these systems causing further delays on the work of the health care personnel. A representative example is the English case, where the NEHRS which is partly in operation broke down 110 times in four months during 2006 [14]. Our study focuses on these technical problems simulating the Electronic Patient Record (EPR) exchange process in an NEHRS environment and studying the factors which could cause these system harms. An EPR is a record of a healthcare episode, such as a course of treatment within a hospital, and therefore forms a component of an electronic health record (EHR), which is a birth to death healthcare history of an individual.

Although there are multiple studies on the factors which influence the NEHRS operation and development, only a few address the impact of technical design options. A similar study to the current investigation (by Huang and Liou [15]) has simulated the EPR exchange of text data measuring the number of users that can be served per hour in one of the health service points of Taiwan. However, their estimations were covering only two Internet link technologies and one EPR size. We extended the simulation experiments examining the NEHRS performance for 13 Internet link technologies, 3 sizes of text EPR, and adapting more precise estimation methods for the time spent for the

records' encryption/decryptions [16] and transfer over the Internet [17]. The expansion on the simulation experiments gave us the opportunity to elaborate extensively the results and retrieve better insights.

Methods

The general strategy employed in this study was to analyse the performance of a range of technical NEHRS configurations by simulating a realistic scenario. Performance was analysed in terms of response times and information capacity. The realistic scenario included the estimation of the service times for the generation/parsing of HL7 messages, the EPR query and EPR transfer over the Internet, the EPR encryption/decryption, the public key search, the EPR search in the database and the LAN transactions. The scenario also includes the use of smart cards which hold the unique identification keys for each EPR of the patient which owns the card.

Generation and parsing of text messages

Estimation of transmission speed is complicated by the overheads of employing HL7. The HL7 (2.3.1) standard presupposes the generation and parsing of RQC/RCI-I05 messages. In the simulations the estimated time for the generation and parsing of HL7 RQC/RCI – I05 messages takes the value calculated in a study by Huang and Liou [15]. Their estimation is based on averaging the time taken by 1000 actual HL7 encoding/decoding transactions using workstations comparable with those assumed in the current simulations. Accordingly, generation of RQC/RCI-I05 is estimated as 12 milliseconds (ms) and the parsing of RQC/RCI-I05 is 50ms. Note that these times are constant, since the HL7 overheads are independent of message size.

Encryption/decryption of text

Calculation of the delays due to text data encryption/decryption is more complex since it depends on three different aspects; (i) the size of the encrypted/decrypted file; (ii) the public key size and algorithm; and (iii) the specification of the computer performing the encryption/decryption. In the current simulations, the calculation is that of the Snyder and Weaver method [16]. Their estimations is based on averaging the performance measurement for a powerful PC (P4, 3GHz) performing encryption/decryption for a range of public key algorithms and key lengths. Specifically, their estimate is based on 400 different tests for each algorithm and key length in order to measure the average encryption/decryption times. We selected for our study the AES public key algorithm [16] with a 128 bits key length since this is a sufficient technology for EPR encryption/decryption, in order to maintain a sufficient level of security [18]. In the study by Snyder and Weaver, the results show that using the specific algorithm, the PC (P4, 3GHz) can encrypt 6.63 MB/seconds and decrypt 7.19 MB/seconds. Since we decided on the EPR sizes it was easy to calculate the encryption/decryption time (e.g. the encryption time for the 10KB EPR was: (10KBx1second)/6630KB = 0.0015 seconds).

EPR sizes

As in [9] [15] [16], record size is a parameter in our simulations. This is in order to investigate the relationships between information capacity and performance. In particular, this allows us to explore where certain thresholds, such as that of acceptability, are reached for given compromised ratios of information capacity and infrastructure performance. Accordingly, the record sizes that needed encryption/decryption in our study were varied as follows:

1. The RQC-I05 message is set to 5KB, a typical size characterising this HL7 part [15]. This size is determined by the HL7 protocol irrespective of the EHR size, and it therefore a constant within the simulation.
2. The RCI-I05 message, which includes the EPR, is set to 10KB, 100KB and 500KB, so as to range from minimal textual EPR content to the more comprehensive patient histories envisaged for NEHRS.

EPR and public key retrieval

In both use cases, EHRs are retrieved from HIS databases, the location of which is stored on the smart card. Estimation of the time associated with searching an HIS database to retrieve a requested piece of information is also based on the method developed by Huang and Liou [15]. Accordingly, it is assumed that the database is well indexed and the indexes are assumed to be memory resident. Consequently, query execution times can be estimated as that of a binary search, which has a service time of the order, $O(\log_2)$. The service time depends on the number of records that compose the database. Huang and Liou [15] calculated the query service time using the function $\log_2(N/760000)$, where 760000 is the number of records that can be searched in 1 second of query processing, and N is the total number of records in a database. Although the records' number in the database is a dynamic value, in order to carry out the simulations we defined to one million the records (in the database) since a static value is necessary for their accomplishment. Therefore, for 1 million records the calculation of database search time is as follows:

$$\log_2 (N/760000) = \log_2 (1000000/760000) = 0.396 \text{ seconds} \quad (1)$$

As regards the time associated with searching a public keys database to retrieve the public keys necessary for the encryption/decryption, we used the estimation by Huang and Liou [15]. Their estimation is considered to be appropriate for our study because it is based on real empirical data from a Taiwanese governmental organisation for the management of public keys (National Public Key Infrastructure Organisation (NPKIO)) and therefore the time demands for public key retrievals is considered to 0.4 seconds.

EPR and public key transfers

Another point to stress with respect to our simulation experiments is that the measurement of the network (Internet and LAN) service time has been estimated using the method of Menasce and Almeida [17]. Consequently, estimation of performance takes account of the next net-

work attributes: LAN Bandwidth - *(LAN Bandwidth in megabits per second)*; Max. LAN PDU - *(maximum Protocol Data Unit (PDU) size for the LAN's network layer protocol in bytes)*; LAN Frame Overhead - *(frame overhead of the LAN's link layer protocol, in bytes)*; Router Latency - *(router latency in microseconds per packet)*; Internet Link Bandwidth - *(bandwidth, in megabits per second, of the connection to the ISP)*; Internet Round Trip Time - *(Internet average Round Trip Time, in milliseconds)*; Internet Data Transfer Rate - *(Internet Data Transfer Rate, in Kbps)*; Browser Rate - *(rate, in HTTP operations/sec)*; Average Size of HTTP requests - *(Average size of the HTTP request sent by the browser to the server, in bytes)*; Maximum Segment Size - *(Maximum Transmission Control Protocol (TCP) segment size, in bytes)*; Record Size - *(EPR/Query size that is transferred in Kilobytes)*. We used the large combination of equations from [17] which involve the above network attributes and we estimated the EPR and public key transfer over the network.

The *Internet Link Bandwidths* and the *Record Size* had dynamic values that were changing during the simulations, so as to model the differences in infrastructure performance. In particular, the Internet Link Bandwidths varied from 56Kbps (PSTN) to 52Mbps (VDSL) and the EPR Size varied from 10KB to 500 KB. The rest of the simulated attribute values remained static [17].

The above static and dynamic values are the attributes which assisted in the estimation of the time demand (service time), M, for record and public key transfers, using a number of complex equations [17]. In particular, M or else the total service demand, was a summation of five different service demands which are shown in equation 2.

$$M = D_{LAN} + D_{router} + D_{OutL} + D_{Int} + D_{InL} \quad (2)$$

To understand the 5 different service (time) demands of equation 2 better, we will briefly explain all the steps of the EPR order process. The doctor sends the EPR order from her PC and this is received from the LAN station of her HSP (D_{LAN}). Then the HL7 standard is adapted on the order and the latter is encrypted. After that, the router receives the order (D_{router}) and forwards it though its outgoing link (D_{OutL}). Afterwards, the EPR order is transferred through the Internet (D_{Int}) and it is received from the LAN station (D_{LAN}) of the HSP which holds the record, having passed first through the incoming link (D_{InL}) and the router (D_{router}). Then the order is decrypted, the HL7 message is parsed and the search in the database is taking place in order to locate the record. Later on, the record is becoming an HL7 message, it is encrypted, and sent back though the outgoing link (D_{OutL}). The record is transferred through the Internet (D_{Int}) and it is received from the LAN station (D_{LAN}) which sent the EPR order originally, passing first through the incoming link (D_{InL}) and the router (D_{router}). The record is decrypted and parsed making the EPR available to the doctor.

Accordingly, we simulated NEHRS performance for different numbers of (user) requests. The number of simulated requests were 100, 500, 1000, 1500, 2000, 3000, 4000, 5000 per hour. The Queues which are presented during the EPR exchange process in the real order are the following: 1) Queue of the network service time for sending the query for public key, 2) Queue of the NPKIO, 3) Queue of the network service time for receiving the public key, 4) Queue of the RQC-I05 message generation, 5) Queue of the RQC-I05 message encryption, 6) Queue of the network service time for sending the RQC-I05, 7) Queue of the RQC-I05 message decryption, 8) Queue of the RQC-I05 parsing, 9) Queue of the search in the database, 10) Queue of the network service time for sending the query for public key, 11) Queue of the NPKIO, 12) Queue of the network service time for receiving the public key, 13) Queue of the RCI-I05 message generation, 14) Queue of the RCI-I05 message encryption, 15) Queue of the network service time for sending the RCI -I05, 16) Queue of the RCI-I05 message decryption, and 17) Queue of the RCI-I05 message parsing.

The equation which was used for the estimation of the response time of each queue is [17]:

$$R_q = 1/(\mu(q)-\lambda(q)) \quad (3)$$

where $\mu(q)=1/T$ (T is the time demand per request) and $\lambda(q) = n/t(q)$ (n is the number of requests and t(q), is the time taken to service n requests queued on q). The Internet technologies which were used for the simulation experiments: PSTN (56 Kbps), ISDN, IDSL (128 Kbps), ADSL (384 Kbps), Cable (1 Mbps), HDSL, ISDN PRI, T1 (1,544 Mbps), SDSL (2 Mbps), ADSL2 (12 Mbps), ADSL2+ (24 Mbps), T3 (45 Mbps), VDSL (52 Mbps).

Results

Applying the above calculation methods we estimated the NEHRS performance during the EPR exchange. Specifically, we experimented with the simulation variables to investigate the relationships between infrastructure performance and information capacity. Figure 1 presents the service times for the EPR exchange of 10KB which vary according to the different number of user requests and the kind of Internet technology. As it was expected the PSTN technology provides very slow responses even for 100 requests/hour (9.01 seconds). For 1000 requests the service time does not only overcome the 80 seconds but also the system is expected to stop the service. As it is mentioned in multiple studies [19] [20] the service time of a web system as in our case should not exceed 3 to 5 seconds. The simulation experiments also shown that even ISDN, IDSL, and ADSL are also slow providing service times larger than 5 seconds and the service stops on 2000 requests in an hour. The Cable, HDSL, T1, ISDN PRI and SDSL technologies offer faster service time lower than 5 seconds until the 2000 requests/per hour and stop the service on 4000 requests/hour. Finally, the ADSL2, ADSL2+, T3 and VDSL offer service times around 3 and 5 seconds until the 3000requests/hour and they can continue the service until the 5000 requests.

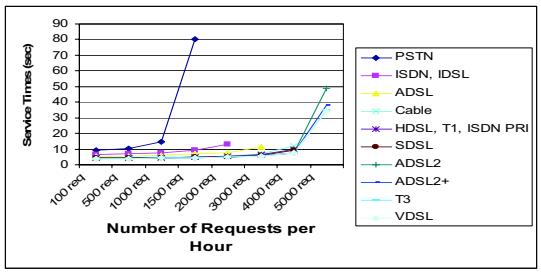


Figure 1 - Service Times for a variety of user requests and Internet technologies (EPR size 10KB was used)

Figures 2 and 3 intend to show which elements (queues) of the EPR exchange process cause delays. The results show that the lines (services times) for the EPR exchange of our scenario which includes everything and the EPR exchange without the HL7 standard are very close (almost tangents) which means that the HL7 overloads the above process with a very small amount of time which is approximately 0.125 seconds (Figure 2). On the other hand, the comparison between the lines (service times) of our scenario and the case of not having EPR encryption/decryption (PKI) shows that encryption /decryption process slows dramatically the process and it makes almost double the service time (Figure 3). Figures 2 and 3 also show how the service times are decreased while the Internet technologies are changed to faster ones.

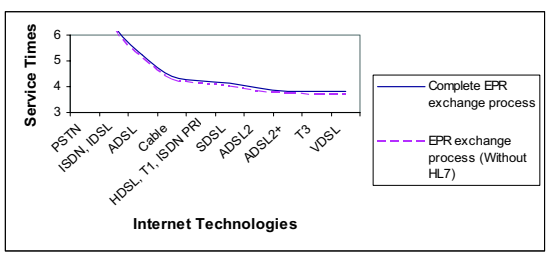


Figure 2 - Services Times for EPR exchange with and without HL7(10 KB EPR size was used, 100 requests/hour were taken into account)

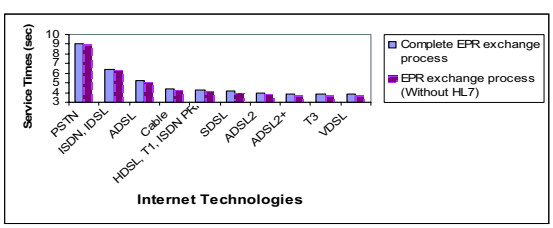


Figure 3 - Services Times for EPR exchange with and without PKI (10 KB EPR size was used, 100 requests/hour were taken into account)

The simulation experiments were accomplished by keeping the size of the database constant. It was interesting to be examined if and how much the database size can affect the EPR exchange process. We further expanded the simulation experiments and we estimated the service times for different database sizes (1, 10, 20, 50 and 100 millions records). We estimated the difference of service times (t) for 100KB EPR (t1) and 10KB EPR (t2) for 100 requests/hour while the database size was varied. The same estimation was repeated for 500 requests/hour. Figure 4 demonstrated that as the database size is increased, the service times are dramatically increased. In specific, service times exceed the 10 seconds for a 10 million record EPR database even for 100 requests.

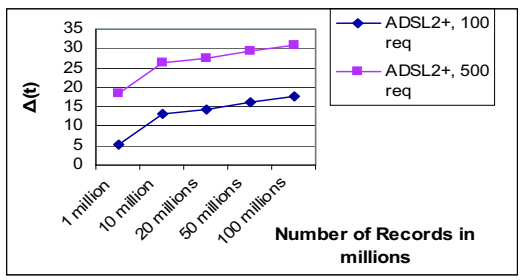


Figure 4 - Service Times increases with relation to database size, EPR size and number of requests. Top Curve: Difference of service times between 100KB and 10KB EPR size when a 500 requests/hour are taken into account. Bottom Curve: The same for 100 requests/hour

Discussion

The usefulness of the simulation experiments on the NEHRS research area is important. They assist the identification of the technical parameters considered to play a key role on the NEHRS operation and development. For example, in case of systems break downs as they exist in England, the simulations could indicate how to break the user requests (workload) by installing new workstations with autonomous Internet links. Moreover, even if the identification of the critical role of the kind of Internet link technologies was being expected, the database size seems that it has not been seriously considered until now. In addition, the simulation experiments demonstrate that the overhead of implementing HL7 is small, particularly when compared with the service time for PKI, which in effect doubles response time. It is envisaged that further elaboration of the simulation experiments will investigate the role of other factors such as different data structures and NEHRS architectures that can mitigate the slowdowns on the EPR exchange process.

Conclusions

The simulation experiments reached a number of interesting conclusions. Simulation of an NEHRS is a difficult process since there are parameters which have dynamic values and the absolute results' precision is not feasible. However, the simulations' precision was enough for examining the roles of the Internet link technologies, EPR sizes and user requests (per LAN station) on the NEHRS performance.

Moreover the experiments helped us to understand that the Internet link technologies cause long delays during EPR exchange and almost half of the NEHRS response time comes from the encryption/decryption process which is necessary for the data security. On the other hand, the HL7 standard, necessary for having data interoperability, loads negligible time on the total NEHRS response time. Finally, the database size increases dramatically the NEHRS service time and therefore solutions should be found to resolve this.

References

[1] National Health Services, Information for Health: An information strategy for the modern NHS 1998-2005, 1998, http://www.dh.gov.uk/assetRoot/04/06/24/28/04062428.pdf [Accessed 16 November 2006]

[2] Saranummi N, and Piggott D, Katehakis D, Tsiknakis M, Bernstein K. Regional health economies and ICT service: The PICNIC experience. Amsterdam: IOS Press, 2005.

[3] Commonwealth of Australia. HealthConnect implementation strategy: Version 2.0, 2005. http://healthconnect.gov.au [Accessed 15 July 2005].

[4] Health Information Strategy Steering Committee. Health information strategy for New Zealand, Ministry of Health, New Zealand, 2005 http://www.moh.govt.nz [Accessed 20 June 2006].

[5] Cannaby S, Wanscher C, Pedersen C, Voss H. The cost benefit of electronic patient referrals in Denmark full report. ACCA and Medcom. 2005. http://www.accaglobal.com [Accessed 10 April 2006].

[6] The Ministry of the Interior and Health, National IT Strategy for the Danish health care service 2003-2007. Sundhedsstyrelsen National Board of Health. 2003. http://www.sst.dk/publ/Publ2004/National_IT_strategy.pdf [Accessed 24 May 2006].

[7] Saiag E. The Israeli Virtual National Health Record: A robust national health information infrastructure based on a firm foundation of trust. In: Engelbrecht et al. (eds.). Connecting Medical Informatics and Bio-Informatics. 2005.

[8] Commission of the European Communities. e-Health – making healthcare better for European citizens: The European economic and social committee and the committee of the regions. Brussels. 2003. http://ec.europa.eu/information_society/doc/qualif/health/COM_2004_0356_F_EN_ACTE.pdf [Accessed 10 August 2006]

[9] National Health Services Information Authority, ERDIP Evaluation N6 - Product P1: Technical Options for the Implementation of Electronic Health Record Nationally, 2002. http://www.nhsia.nhs.uk/erdip/pages/docs_egif/evaluation/technical/n6p1tech-options-report.pdf [Accessed 20 February 2005].

[10] BBS News. Call for review of NHS IT upgrade. BBC Health. 2006. http://news.bbc.co.uk/go/pr/fr/-/2/hi/health/4896198.stm [Accessed 10 July 2006].

[11] Noris A. Current trends and challenges in health informatics, In: iSHIMR 2002. Proceedings of the 7th International Symposium on Health Information Management Research, Sheffield, England. 2-15, 2002.

[12] Orfanidis L, Bamidis PD, Eaglestone B. Data Quality Issues in Electronic Health Records: An adaptation framework for the Greek Health System. Health Informatics Journal 10:1 2004, pp.23-36.

[13] BBC News (2006). GPs dissatisfied with IT system. BBC Health. http://news.bbc.co.uk/go/pr/fr/-/2/hi/health/5028762.stm [Accessed 30 May 2005].

[14] Carvel J. NHS computer system breaks down 110 times in four months, Guardian Unlimited. 2006. http://www.guardian.co.uk/uk_news/story/0,,1875454,00.html [Accessed 25 October 2006].

[15] Huang E, Liou D. Performance Analysis of a Medical Record Exchanges Model, IEEE Transactions on Information Technology in Biomedicine 2006, (In Press).

[16] Snyder AM, Weaver AC, Dwyer S, Gay S, Bassignani M, Moynihan S. Impact of image encryption on radiology department workflow, Journal of Digital Imaging, 2006, (In Press) http://www.cs.virginia.edu/~acw/security/doc/Publications/Impact%20of%20Image%20Encryption%20on%20Radiology%20Department%20Workflow~1.pdf [Accessed 20 June 2006].

[17] Menasce D, Almeida V. Capacity Planning for Web Services. Prentice Hall, New Jersey, 2002.

[18] Computer Security Resource Center, Advanced Encryption Standard (AES), 2001. http://csrc.nist.gov/publications/fips/fips197/fips-197.pdf [Accessed 15 March 2006].

[19] Nielsen J. Designing Web Usability: The practice of simplicity. New Riders Publishing, 2000.

[20] Galbraith S, Davidson N. Quantifying the relationship between website download time and abandonment by users, Technical Paper, http://www.red-gate.com/products/ants_load/technical_papers/understanding_testing_results.htm [Accessed 20 April 2006].

Address for correspondence

Panagiotis Bamidis, Medical Informatics Laboratory, School of Medicine, Aristotle University of Thessaloniki, PO Box 323, 54124, Thessaloniki, Greece.

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Principles-Based Medical Informatics for Success – How Hong Kong Built One of the World's Largest Integrated Longitudinal Electronic Patient Records

Ngai-Tseung Cheung[a], Vicky Fung[a], Wing Nam Wong[a], Anna Tong[a], Antonio Sek[a], Andre Greyling[b], Nancy Tse[b], Hong Fung[c]

[a]*Health Informatics Section, Hospital Authority, Hong Kong SAR*
[b]*Information Technology Division, Hospital Authority, Hong Kong SAR*
[c]*New Territories East Cluster, Hospital Authority, Hong Kong SAR*

Abstract

Since 1994, the Hospital Authority has been developing and deploying clinical applications at its constituent 41 hospitals and 121 clinics. The Clinical Management System (CMS) is now used by over 4000 doctors and 20000 other clinicians on a daily basis to order, document and review care. The territory-wide Electronic Patient Record (ePR) has given clinicians an integrated, longitudinal, lifelong view of a patient's record.

Today the CMS and ePR form an essential clinical and management tool to the Hospital Authority. The CMS handles two million clinical transactions per day, and the ePR has over 6TB of data covering 57 million episodes for 7.9 million patients.

This paper describes how the HA has taken a principles-based approach to Medical Informatics to achieve its success in the enterprise-wide deployment and deep utilization of a comprehensive clinical information system.

Keywords:

electronic patient record, medical informatics

Introduction

The Hong Kong Hospital Authority (HA) was formed in 1990 to manage all public hospitals in Hong Kong. Today the Authority manages a Head Office, 41 public hospitals/ institutions, 47 specialist clinics and 74 general (primary care) outpatient clinics. Each year in the HA there are over one million inpatient episodes, 2 million emergency attendances and 13 million outpatient visits.

Clinical systems in the HA

In 1990 the HA was essentially a "greenfield" site, with very little computerization of any sort, let alone clinical computing. From 1991 onwards HA developed its basic information technology infrastructure, including financial, HR, patient administrative and departmental systems. In 1994 the HA began developing the Clinical Management System (CMS), an integrated clinical workstation giving clinicians access to all available electronic clinical information as well as providing direct entry of orders and care or patient documentation [1]. The Electronic Patient Record (ePR) was developed in 2000 to provide both a standardized repository of all the clinical data collected throughout the HA, as well as a clinician-friendly view into a comprehensive longitudinal lifelong record. The CMS and ePR are integrated across all settings – inpatient, outpatient and emergency, and have also been the platform for development of all subsequent clinical modules, including modules for different clinical specialties, the allied health disciplines and nursing.

By the turn of the century, the CMS had already been rolled out to all HA sites, and today the CMS and ePR are an essential clinical and management tool, handling over two million clinical transactions per day. The ePR contains the records from 57 million episodes for 7.9 million patients in a repository of over 6TB (not including radiology images) and receives 300,000 views per day. The total expenditure on system development and implementation of the entire clinical informatics portfolio has been under US$200million.

The wave that never breaks

Such a wide and deep penetration of computerized patient records is far from the norm. Despite promises of the "paperless" or "electronic" hospital for over 30 years, comprehensive adoption of such systems has not materialized in most countries – electronic medical records are "the wave that never breaks" [2]. The seminal Institute of Medicine report on the Computer-based Patient Record [3] released in 1991 shaped the thinking of a generation of systems, but implementation remained low. A decade later another pair of IOM reports on medical errors [4] and quality systems in medicine [5] has stimulated a fresh round of interest in the electronic medical record as a key enabler for quality improvement and error reduction. Other potential drivers include consumerism and the rise of personal health records [6] (PHRs), and estimates of positive returns from investments in electronic records, both at the institutional [7] and national levels [8].

Perhaps the age of ubiquitous clinical computing truly is imminent. However there are many indicators that a long

and hard road remains ahead. The Annual HIMSS Leadership Survey has shown an striking incongruence between intent and achievement in installing electronic medical record systems. In 2003 the survey showed that 19% of healthcare organizations in the US had fully operational electronic medical record (EMR) systems, and that 32% had begun installation. [9] In 2006 the same survey found that 24% now had fully operation EMR systems, and 42% had begun installation. [10] That is to say, in the last three years only a 5% increase in fully operational systems was seen.

Outside the US, some very well funded high profile national initiatives (such as the UK's £6billion Connecting For Health [11] and Canada's C$1.2billion Health Infoway [12]) are making headway, but are still years away from a nationwide EMR implementation at both hospital and outpatient settings.

So why is that in Hong Kong, with its comparatively modest expenditure on health IT, the wave has apparently broken? Success in clinical information systems is multifactorial [13], [14] and without doubt the nature of the healthcare system in Hong Kong and the leadership of the HA have played a major role. However this article will focus on one aspect that has been constant throughout the history of clinical systems development in Hong Kong – the integration of medical informatics into the leadership and execution of the clinical systems programme, and a consistent adherence to certain key principles and processes throughout this period.

Seven principles for medical informatics success

The customer is always right (well, almost)

This old adage from the retail industry is of course not to be taken literally, but points to the reality that meeting clinician needs is paramount [15] and that clinician resistance is often the critical factor leading to the failure of clinical system [16]. Focusing on the needs of the customer has always been of prime importance in clinical systems development in the HA. But to meet the needs of our customers, we must know who they are.

The clinicians using the system for care delivery are obviously the most direct customer. However, clinical and executive management, policy makers, payers, researchers and ultimately the patients are all customers of the clinical systems. To fully engage this large body of stakeholders in a manageable fashion, the HA has instituted formal Clinical IT Governance structures and processes. The key governance committee is the Clinical Informatics Program Steering Group (CIPSG), comprising nominated representatives from different hospitals, specialties, corporate areas and other constituencies. Each representative chairs a working group that looks at the detailed needs of that constituency, and reports back to CIPSG. The CIPSG in turns reports up to the board-level IT Governance Committee (ITGC).

The Health Informatics Team in close collaboration with the IT division runs the Clinical Informatics Program Office (CIPO), which is the executive arm of the CIPSG, and all developments, enhancements, collaborations and other activities are filtered through CIPO.

This constitutes the complete "top down / bottom up" approach to customer engagement in the HA, from detailed consideration of needs at the working groups and CIPO, to implementation decisions at CIPSG, to investment and strategic decisions at ITGC – with constant coordination and communication between the levels of governance (figure 1).

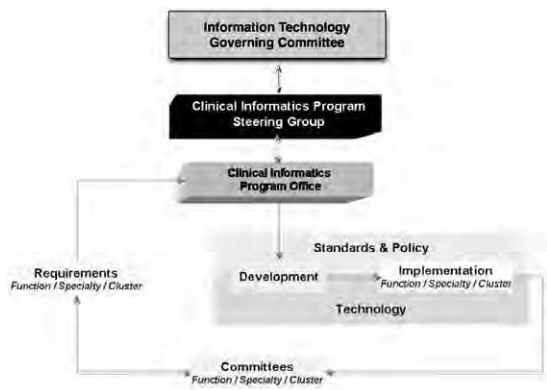


Figure 1 – Clinical Systems Governance Process

Win - win (and win - win)

It is often easy to create a IT-based program that provides benefits to patients or the organization, but at a cost to the clinicians asked to implement the program - often in the form of extra work, changes in process or simply loss of flexibility. Avoiding this problem by ensuring systems benefit, or at least are cost-neutral, to all stakeholders is vital to ongoing success. The four wins in this section heading refer to the four major stakeholder groups – 1) clinicians, 2) patients, 3) management and other secondary data users and 4) health informatics and ITD who need to ensure integrity of data and architecture.

Key considerations in enabling common benefits are: can the needs of individual clinicians be met with a standardized system? Can structured data be captured by clinicians at the point of care without using more time and impinging on clinical autonomy? Can the system adjust for different practices and workflows? Can the data gathered by the system flow onto management, quality assurance and other such uses?

The next two principles have been the key to answering these questions in the affirmative.

One step at a time (success breeds success)

In the complex domain of medical informatics it is extremely difficult, if not impossible to envisage a comprehensive solution to any problem de novo, and it is safe to assume that unexpected issues will arise [17]. In the HA "big bang" implementations are avoided wherever possi-

ble in favor of a rapid prototype design cycle, followed by one or two pilot sites with evaluation and redevelopment where necessary, followed by phased rollout to the rest of the organization.

Success stories are vital to the long-term health of clinical IT, and taking small steps allows continuous demonstration of success, developing a cadre of clinicians who have seen the benefits and understand the issues. A stepwise approach has also allowed us to make our mistakes earlier and on a smaller scale, providing continuous learning whilst minimizing the risk of catastrophic failure.

Medicine is an art and a science (and so is medical informatics)

Clinical systems sit at the uneasy interface between the vagaries of human biology/disease and the scientific requirements of quality management and evidence based medicine. There is a constant tension between the need for standardized, codified data and the clinical requirements for flexibility and expressivity. It has been shown that acceptance of structured entry is not uniform [18] and in the HA the optimum balance between the two is a key decision for any new module.

Although the degree of structure is negotiable, the way the data is codified is not – all systems must conform to the standards laid down in the HA's Information Architecture [19]. However the design of systems must always try to allow for flexibility whilst preserving structure.

Use it or lose it (data use begets data quality)

In the HA environment there are tens of thousands of users of varying levels of IT sophistication entering data from a variety of modules in different clinical settings. How do we ensure the quality of this data? The main strategy taken in the HA is to reuse the data wherever possible – "write once, use many".

One obvious reuse of data is secondary uses, such as quality management, planning and analysis. Where clinicians run their own quality or analysis programs they will soon see the need for good data, and we have developed sophisticated, clinician friendly data mining tools to allow clinicians to run their own ad hoc queries on the massive data warehouse that has accrued over the years.

However opportunities also abound for reuse of data in direct clinical care. The medications list on a discharge summary can be copied from the discharge prescription. The reason for doing a test can be drawn from the problem list. Key lab results can be automatically copied into a diabetic documentation template. Data reuse of this sort is a form of semantic interoperability – albeit *within* but not *between* complex clinical systems environments – and the Information Architecture provides the framework and standards to support this reuse.

Prioritize ruthlessly (first things first)

In the early days of clinical systems in the HA, overcoming clinician resistance and finding areas of benefit were of prime importance. However as a critical mass of "informatics aware" clinicians developed, there was fairly sudden switch from an informatics-driven "push" of clinical systems to a clinician-led "pull" of additional requirements. Today there are far more requirements than can be met by available resources, in terms of functionality, hardware or support for specific clinical initiatives.

Demand management is now crucial, and all requests must be evaluated, developed and either discarded or moved up the governance chain until the appropriate level is reached where a yea or nay decision can be made. At each level of governance evaluation criteria and processes have been put in place to enable an equitable approach which takes into account the needs of patients, clinicians and the organization as whole.

Embrace your Informaticians (and feed them well)

Implementing sophisticated clinical systems is a difficult endeavour, and there is no shortage of literature examining stumbling blocks [20], problems [21] and case studies of failure [22].

The principles listed in this paper all require constant attention, and a team that is able to navigate between clinical and organizational needs, technical limitations and possibilities, standards and architectures, policies and guidelines, change and demand management, and to do this over long time periods. In the complex HA environment, and in any large care delivery setting, this is not a task that can be done properly without dedicated, trained informaticians.

Conclusion

Healthcare operates in a sea of change – social, environmental, scientific and technological change. Electronic medical records are an increasingly important tool to enable healthcare delivery organizations stay afloat and perhaps prosper, but they can only do so with the necessary commitment, skill sets and a adherence to disciplines such as have been presented in this paper.

Acknowledgements

We would like to thank the members of Clinical Informatics Program Steering Group and all many others at the Hospital Authority who have contributed their time and efforts to shaping the CMS and ePR.

References

[1] Cheung NT, Fung KW, Wong KC, Cheung A, Cheung J, Ho W, Cheung C, Shung E, Fung V, Fung H. Medical informatics--the state of the art in the Hospital Authority. Int J Med Inf. 2001 Jul;62(2-3):113-9

[2] Berner ES, Detmer DE, Simborg D. Will the wave finally break? A brief view of the adoption of electronic medical records in the United States. J Am Med Inform Assoc. 2005 Jan-Feb;12(1):3-7

[3] Dick RS, Steen EB (1991). The Computer-based Patient Record: An Essential Technology for Health Care, Washington, D.C., National Academy Press

[4] Committee on Quality Health Care in America. To err is human: Building a safer health system. Washington, DC: Institute of Medicine; 1999

[5] Committee on Quality Health Care in America. Using information technology. Crossing the quality chasm:

A new health system for the 21st century. Washington, DC: Institute of Medicine; 2001
[6] Detmer D, Steen E. Learning from Abroad: Lessons and Questions on Personal Health Records for National Policy. Available at American Association of Retired Persons at: http://www.aarp.org/research/health/healthliteracy/2006_10_phr_abroad.html. Accessed 19 Apr 2006.
[7] Garrido T, Raymond B, Jamieson L, Liang L, Wiesenthal A. Making the business case for hospital information systems--a Kaiser Permanente investment decision. J Health Care Finance. 2004 Winter;31(2):16-25.
[8] Walker J, Pan E, Johnston D, Adler-Milstein J, Bates DW, Middleton B. The Value Of Health Care Information Exchange And Interoperability. Health Aff (Millwood). 2005 Jan-Jun;Suppl Web Exclusives:W5-10-W5-18
[9] 14th Annual HIMSS Leadership Survey, retrieved Nov 10 2006 from: http:/www.himss.org/2003survey/ASP/healthcarecio_final.asp
[10] 2006 Leadership Survey, retrieved Nov 10, 2006 from: http://www.himss.org/2006survey/healthcareCIO_final.asp
[11] NHS Connecting for Health. History of Our Organisation. Available at http://www.connectingforhealth.nhs.uk/about/history. Accessed 10 Nov 2006.
[12] Catz M, Bayne J. Canada Health Infoway – A pan-Canadian Approach. AMIA Annu Symp Proc. 2003; 2003: 807.
[13] Scott JT, Rundall TG, Vogt TM, Hsu J. Kaiser Permanente's experience of implementing an electronic medical record: a qualitative study. BMJ. 2005 Dec 3; 331(7528): 1313-1316.
[14] Ash JS, Fournier L, Stavri PZ, Dykstra R. Principles for a successful computerized physician order entry implementation. AMIA Annu Symp Proc. 2003;:36-40.
[15] Hier DB, Rothschild A, LeMaistre A, Keeler J. Differing faculty and housestaff acceptance of an electronic health record. Int J Med Inform. 2005 Aug;74(7-8):657-62
[16] Lapointe L, Rivard S. Getting physicians to accept new information technology: insights from case studies. CMAJ, 2006 May 23; 174(11): 1573–1578
[17] Campbell EM, Sittig DF, Ash JS, Guappone KP, Dykstra RH. Types of unintended consequences related to computerized provider order entry. J Am Med Inform Assoc. 2006 Sep-Oct;13(5):547-56
[18] O'Connell RT, Cho C, Shah N, Brown K, Shiffman RN. Take note(s): differential EHR satisfaction with two implementations under one roof. J Am Med Inform Assoc. 2004 Jan-Feb;11(1):43-9
[19] Cheung NT, Fung V, Kong JH. The Hong Kong Hospital Authority's information architecture. Medinfo. 2004;11 (Pt 2):1183-6.
[20] Sittig DF, Krall M, Kaalaas-Sittig J, Ash JS. Emotional aspects of computer-based provider order entry: a qualitative study. J Am Med Inform Assoc. 2005 Sep-Oct;12(5):561-7.
[21] Hendy J, Reeves BC, Fulop N, Hutchings A, Masseria C. Challenges to implementing the national programme for information technology (NPfIT): a qualitative study. BMJ. 2005 Aug 6;331(7512):331-6.
[22] Ornstein C. Hospital Heeds Doctors, Suspends Use of Software. LA Times, Jan 22, 2003.

Address for correspondence

Dr N.T. Cheung
Rm 121N, Hospital Authority Building
147B Argyle St
Kowloon
Hong Kong
email: cheungnt@ha.org.hk

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Knowledge-Level Querying of Temporal Patterns in Clinical Research Systems

Martin J. O'Connor[a], Ravi D. Shankar[a], David B. Parrish[b], Amar K. Das[a]

[a]*Stanford Medical Informatics, Stanford University, Stanford, CA, USA*
[b]*The Immune Tolerance Network, Pittsburgh, PA, USA*

Abstract

Managing time-stamped data is essential to clinical research activities and often requires the use of considerable domain knowledge. Adequately representing this domain knowledge is difficult in relational database systems. As a result, there is a need for principled methods to overcome the disconnect between the database representation of time-oriented research data and corresponding knowledge of domain-relevant concepts. In this paper, we present a set of methodologies for undertaking knowledge-level querying of temporal patterns, and discuss its application to the verification of temporal constraints in clinical-trial applications. Our approach allows knowledge generated from query results to be tied to the data and, if necessary, used for further inference. We show how the Semantic Web ontology and rule languages, OWL and SWRL, respectively, can support the temporal knowledge model needed to integrate low-level representations of relational data with high-level domain concepts used in research data management. We present a scalable bridge-based software architecture that uses this knowledge model to enable dynamic querying of time-oriented research data.

Keywords:

clinical trials, temporal querying, knowledge-based systems, Semantic Web, ontology

Time in clinical research databases

Relational databases have become an essential part of biomedical research projects needing to maintain, integrate and share data. In biomedical research projects ranging from clinical studies to genomics research, relational databases are typically used to store data and custom queries are written to extract subsets of the data into specialized tools to support study management and focused analyses. A serious shortcoming of this approach is that the data-processing steps are often customized to a particular analysis and database and thus do not generalize to other research projects. By its nature, however, the standard relational model does not adequately support important biomedical domain concepts, such as hierarchies and time; thus, the link between domain knowledge and data representation used in database querying is often implicit. The lack of support for temporal information on research study design (such as longitudinal patient observations or time-course experiments) at the database level can limit the investigation of causal phenomena that are central to biomedical research.

As a result, there is a critical need to provide querying methods that can operate at the *domain knowledge* level rather than the *database schema* level. To address this problem, we have developed end-to-end methodologies and a software architecture that permit design-time encoding and execution of temporal patterns needed for clinical research management. Our approach consists of three knowledge-based components: a temporal ontology, a temporal pattern specification language, and a database mapping tool. The design of these components is driven by the needs of the Immune Tolerance Network [1], a collaborative clinical research organization focused on developing new therapeutics in immune-mediated disorders. Our methodology bridges the gap between clinical-trial specification and clinical-trial implementation, which enhances compliance monitoring and data analysis within this research environment.

Knowledge and database disconnect

Many clinical research systems have significant requirements for the querying and management of temporal data. Trial design and compliance monitoring tasks, for example, typically revolve around evaluating temporal patterns among data. Example patterns (found as free text in a study design document) include: "Visit 3 for a participant must occur with three weeks of visit 2," "clinical assessments are required twice a week until day 28 or discharge from hospital," and "test is scheduled on weeks 4, 6, and 8 during treatment." When encoding such patterns, developers may face two types of disconnect between the initial specification of these patterns and their execution.

First, there can be a knowledge specification disconnect. Constraints are typically expressed as unstructured free text throughout a study design document. Their interpretation is heavily dependent on the context of the research protocol being encoded. Even core terms—such as, for example, a definition of participant visit—can be poorly specified. Is a visit a single encounter between a participant and a provider, or can it span several encounters? If it can span several encounters, what is the exact definition of the visit end? Producing a precise definition for a constraint can thus be difficult. In addition, the unstructured

Selected for best paper award.

constraint specification process can also result in gaps in the final specifications.

Second, there can be a database specification disconnect. Constraints are usually encoded in terms of data that is collected during a study's execution. These data are often stored in relational databases. The schema design of the databases often reflects the operational requirements of the study managers whose activities were not defined within the research protocol document. Constraints can thus be encoded at a level that is at least one step removed from their initial specification, with a consequent loss of precision. This difficulty is compounded by the fact that developers may not have direct access to the protocol authors. A related problem is that constraints may only be checked after data has been entered into a database and not when the data is collected, allowing non-compliant data to enter the system. As a result of these disconnects, the encoding and implementation of protocol-specified constraints as temporal patterns may not reflect the intentions of the designers. The quality of the trial data can thus become seriously compromised, which may not be noticed until the final stage of analysis.

Methods

To overcome these types of disconnect, we apply existing knowledge representation and temporal relational methods.

Knowledge representation language

Our approach relies on using the standard knowledge specification methods for the Semantic Web. The Semantic Web is a shared research plan that aims to provide explicit semantic meaning to data and knowledge on the World Wide Web [2]. The Web Ontology Language (OWL) has been designed as the language of the Semantic Web. OWL can be used to build ontologies that provide high-level descriptions of Web content. These ontologies are created by building hierarchies of *classes* describing concepts in a domain and relating the classes to each other using *properties*. OWL can also represent data as instances of OWL classes—referred to as *individuals*—and it provides mechanisms for reasoning with the data and manipulating it.

OWL provides limited deductive reasoning capabilities, however, so recent work has concentrated on adding rules to it. The Semantic Web Rule Language (SWRL) allows users to write Horn-like rules that can be expressed in terms of OWL concepts and that can reason about OWL individuals. SWRL provides deductive reasoning capabilities that can infer new knowledge from an existing OWL ontology. We recently developed the first SWRL editor [4]. It was written as an extension to Protégé-OWL [3], an open source framework that provides a suite of tools for constructing OWL domain models and knowledge-based applications. Our editor permits interactive creation, editing, reading, and writing of SWRL rules. We have also developed one of the first systems supporting inference with SWRL rules using the Jess rule engine.

Temporal relational model

Most modern clinical research systems store data within relational databases, which provide a well-defined data model and query language. However the relational model provides poor support for storing complex temporal information. For example, if a database row contains some temporal information, there is no indication as to the relationship between the timestamp and the non-temporal data in the row. Does the timestamp refer to the point at which the information was recorded, or to the point at which it was known? Other shortcomings include no standard way to indicate a timestamp's granularity, no support for automatic coalescing or merging of temporally overlapping data, and no standard means of writing queries with relative times or that refer to the current time [5].

Several proposed extensions to the relational model address these shortcomings. Most have focused on *valid-time* model [5-7, 9]. In this model, all facts have temporal extent and are associated with instants or intervals denoting the times that they are held to be true. When this model is used in a relational database temporal information is attached to all rows in a temporal table. This approach effectively adds a third dimension to two-dimensional relational tables. In these tables, every tuple holds temporal information denoting the information's valid-time. Conceptually, this representation means that every tuple is held to be true or valid during the time or times associated with this tuple. This consistent representation of temporal information allows standard temporal operators top be applied consistently and greatly simplifies the temporal reasoning task. Several SQL extensions have been developed that support these operators [6, 7].

Results

We used OWL and SWRL to develop a valid-time temporal model to support methods for temporal constraint specification in a protocol tracking system. We also developed a set of mapping tools to allow the use of this model with existing relational data. A knowledge-driven architecture was then implemented to support the efficient deployment and execution of system components.

Temporal ontology

We have encoded a temporal model in OWL [8] based on the valid-time temporal model. The core class modeling this association in the OWL ontology is called `Extended-Proposition`. This class models information that extends over time. It has a property called `hasValidTimes` that holds the time(s) during which the associated information is held to be true. This property is modeled by an abstract class called `ValidTime`, which has subclasses `ValidInstant` and `ValidPeriod`. `ValidInstant` has the property `hasTime`, and `ValidPeriod`, has the properties `hasBeginning` and `hasFinish`. These classes represent instants and intervals, respectively. Valid times also have granularities associated with them. Named points in time—often called *anchor points*—can be modeled as subclasses of the `ValidInstant` class. Temporal durations are

modeled using a `ValidDuration` class that holds a count and a granularity.

There are two types of extended propositions in the model: (1) *extended primitive propositions*, which represent data derived directly from secondary storage; and (2) *extended abstract propositions*, which are abstracted from other propositions. They are represented by `ExtendedPrimitiveProposition` and `ExtendedAbstractProposition`, respectively, in the temporal ontology. The extended primitive and abstract proposition classes can also hold a value in addition to its valid times. This value is denoted by the `hasValue` property. The value is any XML Schema data types, such as strings or integers.

These extended propositions can be used to consistently represent temporal information in ontologies. For example, a set of visits in a protocol tracking application can be represented by defining a class called `Visit` that subclasses the extended proposition class. It inherits the `hasValidTime` property from that class, which holds its visit times. Similarly, an extended primitive proposition can be used to represent a drug regimen, with a value of type string to hold the drug name and a set of periods in the valid time property to hold drug delivery times. These extended propositions can then be associated with a class using OWL properties.

Once all temporal information is represented consistently using the temporal ontology, SWRL rules can be written in terms of this ontology. However, the core SWRL language has limited temporal reasoning capabilities. A few temporal predicates called *built-ins* are included in the set of standard predicates, but they have limited expressive power. Fortunately, SWRL provides an extension mechanism to add user-defined predicates. We used this mechanism to define a set of temporal predicates to operate on temporal values. These predicates support the standard Allen temporal operators [10].

Using these built-in operators in conjunction with the temporal ontology permits expression of complex temporal rules. For example, in modeling visits in a protocol as extended propositions and the start of treatment of a participant as an anchor point, a new SWRL rule can indicate that a second visit in a particular protocol must occur within two weeks of the start of treatment anchor, as follows:

```
Participant(?p) ^ hasVisit(?p, ?v) ^
V2(?v) ^ temporal:hasStart(?v, ?startV2) ^
hasAnchor(?p, ?a) ^ StartOfTreatment(?a) ^
temporal:hasTime(?a, ?sot) ^
durationGreaterThan(?sot, ?startV2, 2, weeks)
-> NonConformingPatient(?p)
```

Temporal pattern specification

Our efforts to model clinical trials is driven by the needs of the Immune Tolerance Network (ITN; [1]), which develops new therapeutics for immune-mediated disorders. In collaboration with ITN, we have created a knowledge-based architecture (called Epoch [11]) to support the management of multi-site clinical trial protocols and the discovery of common tolerance mechanisms across multiple trials. We have focused our efforts on developing participant and sample tracking models [11], both of which must specify complex temporal constraints at the knowledge level. To meet this need, we used the temporal ontology to model the temporal dimension of core components in the model and then analyzed a range of ITN's protocols to determine the types of temporal constraints required by protocols and if our model could represent them.

In principle, SWRL rules could be used to express all constraints within the protocol tracking application. However, while relatively concise, SWRL rules are not suitable for non-specialists. As a result, we decided to define a high-level user-friendly constraint language to allow ontology developers to encode constraints at the domain level. These constraints are then mapped automatically to SWRL rules at run time.

The constraint language allows times to be specified as absolute or relative times. For example, an indication that something must start within two weeks of a start of treatment anchor is `SOT + WEEKS(2)`, where SOT is the name of the start of treatment anchor, to which an offset of two weeks is added. Offsets can be positive or negative and can be combined at different granularities. Offsets correspond to durations in the temporal model may also be referred to as such. For example, an offset of one month and two days can be specified as `MONTH(1) + DAYS(2)`.

In addition to named anchors, the constraint specification language can work directly with temporal propositions. So, for example, if visit number two is modeled as an extended proposition subclass `V2`, a constraint can refer to its start time as `V2.hasBeginning`. This syntax could, for example, express the expect start time of a third visit as a two month offset from the end of visit two as `V2.finish + months(2)`.

Temporally constrained entities in the protocol model are modeled using a *plan* that can hold temporal constraints specified in the constraint language. This class is modeled using an OWL `Plan` class and has a number of properties that can specify the expected temporal behavior of the associated protocol entities:

expectedStart The time the protocol entity is expected to start. This is specified using the constraint language.

expectedStartVariance The temporal uncertainty of the start time. This is expressed using constraint language offset clauses, *e.g.*, `WEEKS(2)`.

expectedFinish, expectedFinishVariance End time specifications for an entity.

expectedDuration How long this entity is expected to last. Specified using duration clauses, *e.g.*, `DAYS(2)`.

expectedCycles Used for cyclical specifications to indicate the number of times the entity is expected to repeat and the intervals between repetitions.

A plan class also has two properties that hold run-time time values for the protocol entity. These properties—called actualStart and actualFinish—can be compared against expected values when validating an entity for compliance. Using plan specifications in conjunction with the constraint language allows us to express a large range of constraints for protocol entities, which can be mapped to SWRL rules for execution.

Database mapping

In principle, developers could take biomedical data in an existing relational database, develop an ontology to describe those data, and then convert the data into a knowledge-based form for all future processing. Apart from the significant development effort involved, this solution does not scale well. Current ontology-specification tools, such as Jena or Protégé-OWL, do not support high data throughput. For large data sets, an alternate mapping solution is needed. We have developed a customized mapping tool, called Synchronus [12, 13], that supports both a direct relational-to-OWL mapping and also a lightweight mapping mechanism for large data sets. This tool maps relation data described in terms of the temporal ontology to OWL individuals. Essentially, it creates extended propositions from time-stamped relational data. It also supports the reverse mapping of extended propositions to relational data.

Two OWL ontologies are used to drive Synchronus: (1) a *schema ontology*, which is a knowledge-level description of a relational schema; (2) a *mapping ontology*, which describes how relational data are mapped to extended propositions. The schema ontology describes the structure of one or more databases that will be mapped. It contains descriptions of the tables in the database, such as the names of types of columns in those tables. The mapping ontology uses this schema ontology to describe the relational or temporal-relational tables to be mapped. Every extended proposition in the temporal model has an optional input and output storage descriptor. The descriptor uses the schema ontology to point to data that is stored in a database. Synchronus uses this descriptor to perform run-time transformations of the data between rows in a relational database and OWL individuals.

The direct relational-to-OWL data-mapping method has two main modes of operation: batch mode, where an OWL knowledge base is fully populated with relevant data from a database, or a database is populated from an OWL knowledge base; and dynamic mode, where propositions are mapped on demand. The latter mapping mechanism reads and writes objects represented by extended propositions without creating OWL individuals.

Bridge architecture

We have developed a bridge architecture to support the integration of relational databases and reasoning methods into a knowledge-driven system. The Figure shows a schematic of the architecture and its five main components: (1) a knowledge base; (2) a relational database; (3) Synchronus; (4) a method; and (4) the bridge itself. A bridge is a customized method to provides a specific computational task through the integration of one or more existing knowledge sources (such as an OWL knowledge base); data sources (such as a relational database); and data-processing mechanisms (such as a rule engine). The bridge resolves low-level differences of how these software components interact with each other through the communication of data and knowledge. A deployed bridge may work with several databases, methods, and, potentially, knowledge bases.

Each bridge is driven from its associated knowledge base. The knowledge base contains a number of ontologies that are used in deploying the bridge: (1) a *method ontology*, which describes at a high level the analytic method or methods being used; (2) a *mapping ontology*, which is used by Synchronus to map relational data; and (3) a *domain ontology* that describes the underlying application domain.

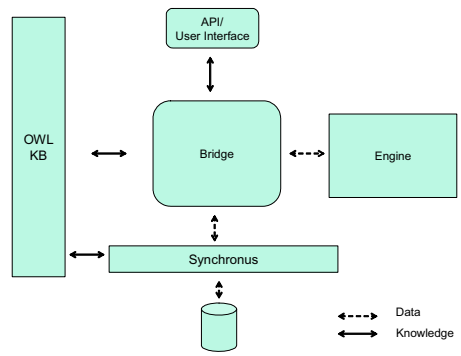


Figure 1 - Bridge architecture. Schematic showing how a bridge is deployed to isolate an existing analytic method (rule engine) from details of both an OWL knowledge base and an existing relational database, accessed through Synchronus

To support the validation of temporal constraints for clinical-trial management at ITN, we have developed a bridge architecture that provides the infrastructure necessary to incorporate rule engines into Protégé-OWL and execute SWRL rules specified in the knowledge base. The bridge provides a mapping layer that generates as input into the rule engine representations of all rules and relevant OWL classes, individuals and properties. A target rule engine implementation takes in these representations and implements them in the rule engine's native format.

Providing efficient data and knowledge access techniques is a central goal of our bridge design and implementation. For example, when translating OWL knowledge into an intermediate form, instead of transferring all knowledge in a knowledge base, only potentially relevant knowledge is represented. The bridge examines each SWRL rule and only represents OWL classes, properties and individuals that are referenced by those rules. Such references can be indirect, so the bridge must traverse the interrelationships between all OWL concepts mentioned in SWRL rules to ensure completeness. This step significantly reduces the amount of knowledge that needs to be reasoned with by a

rule engine and can ensure significant performance benefits.

Another optimization technique relates to data access. Extended propositions used in rules may be held in databases and accessed through Synchronus. SWRL rules can operate on these propositions using temporal built-ins. There is a fairly direct mapping from SWRL rules with temporal propositions to valid-time queries. This parallel structure can be exploited by the bridge to optimize its data access. The bridge examines each SWRL rule with temporal operators and looks for operators that temporally restrict the range of propositions. For example, if a temporal operator restricts the range of a proposition to dates after a particular time point, only data after than time will be requested from Synchronus. Because SWRL rules do not have disjunctions, this optimization process is not elaborate.

In our clinical trial management architecture for ITN, we currently use an implementation that invokes the Jess rule engine [3]. Our Jess implementation for a rule engine bridge employs the temporal built-in library for SWRL that we have presented and automatically undertakes the mapping of data and knowledge for knowledge-level querying of specified temporal patterns in a relational database.

Discussion

The gap between the specification of a study protocol and the management of resulting data can often be quite significant in clinical research systems, such as clinical trial management applications. To help close this gap, we have developed a set of end-to-end general methodologies for specifying and executing temporal patterns at the knowledge level rather than the database level. Our approach demonstrates that proposed Semantic Web standards for ontology and rule representation, OWL and SWRL, respectively, can support the knowledge model needed to integrate temporal representations of relational data with the domain-specific semantics needed to reason with them for biomedical and healthcare applications.

In contrast to previous work on constraint specification in clinical trials [14-15], our set of methodologies addresses the knowledge and database disconnect that exist in clinical research systems. Our approach requires that all relevant temporal knowledge on a study protocol and its corresponding data representation be encoded within an OWL ontology, which allows the uniform specification of temporal patterns in knowledge-level querying. Our bridge architecture supports robust optimization techniques to ensure that encoded constraints are automatically translated into an executable form at run time and are efficiently validated against study data held in an existing relational database.

Acknowledgments

This research was supported in part by the Immune Tolerance Network, funded by Grant NO1-AI-15416 from the National Institutes of Health (USA), and by a Pharmaceutical Research and Manufacturers Association Foundation Research Starter Grant in Informatics. The authors thank Valerie Natale for her editorial comments.

References

[1] Rotrosen D, Matthews JB, and Bluestone JA. The immune tolerance network: a new paradigm for developing tolerance-inducing therapies. J Allergy Clin Immunol 2002: 110 (1): 17-23.

[2] Berners-Lee T, Hendler J, and Lassila O. The Semantic Web. Scientific American 2001: 35 (May): 43-52.

[3] Knublauch H, Fergerson RW, Noy NF and Musen M.A. The Protégé OWL Plugin: an open development environment for semantic web applications. In: Third ISWC (ISWC 2004), Hiroshima, Japan, 2004; pp. 229-243.

[4] O'Connor MJ, Knublauch H, Tu SW, Grossof B, Dean M, Grosso WE, and Musen MA. Supporting rule system interoperability on the Semantic Web with SWRL. Fourth International Semantic Web Conference (ISWC 2005), Galway, Ireland, 2005; pp. 974-986.

[5] Snodgrass RT. On the semantics of 'now' in databases. ACM Trans Database Systems 1997: 22 (2): 171-214.

[6] Snodgrass RT (Ed). The TSQL2 temporal query language. Kluwer Academic Publishers: Boston, 1995.

[7] O'Connor MJ, Tu SW, and Musen MA. The Chronus II temporal database mediator. In: AMIA Annual Symposium, San Antonio, TX, 2002; pp. 567-571.

[8] Shoham Y. Temporal logics in AI: semantical and ontological considerations. Artif Intell 1987: 33 (1): 89-104.

[9] O'Connor MJ, Shankar RD, and Das AK. An ontology-driven mediator for querying time-oriented biomedical data. In: 19th IEEE Symposium on Computer-Based Medical Systems (CBMS2006), Salt Lake City, Utah, 2006; pp. 264-269.

[10] Allen JF. Maintaining knowledge about temporal intervals. Comm ACM 1993: 26 (11): 832-843.

[11] Shankar R, Martins SB, O'Connor MJ, Parrish D, and Das AK. Towards semantic interoperability in a clinical trials management system. In: Fifth International Semantic Web Conference (ISWC 2006), Athens, Georgia, 2006: pp. 901-912.

[12] Das AK and Musen MA. Synchronus: a reusable software module for temporal integration. In: AMIA Annual Symposium, San Antonio, Texas, 2002: pp. 195-199.

[13] Narayanan PS, O'Connor MJ, and Das AK. Ontology-driven mapping of temporal data in biomedical databases. In: AMIA Annual Symposium, Washington, DC, 2006: p. 1045.

[14] Weng C, Kahn M, and Gennari J. Temporal knowledge representation for scheduling tasks in clinical trial protocols. In: AMIA Annual Symposium, San Antonio, Texas, November, 2002; pp. 879-83.

[15] Deshpande AM, Brandt C, and Nadkarni PM. Temporal query of attribute-value patient data: utilizing the constraints of clinical studies. Int J Med Inform 2003: 70 (1): 59-77.

Address for correspondence

Martin J. O'Connor
{martin.oconnor, ravi.shankar,amar.das}@stanford.edu
dparrish@immunetolerance.org
Stanford Medical Informatics
Stanford University
251 Campus Drive, MSOB X275
Stanford, CA 94305 USA

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

A Discussion about the Importance of Laws and Policies for Data Sharing for Public Health in the People's Republic of China

Xiue Fan[a], Ping Yu[b]

[a]*Chinese Center for Disease Control and Prevention*
Department of Health, Beijing 100050, P.R. China
[b]*School of Information Systems and Technology, Faculty of Informatics, University of Wollongong, Australia*

Abstract

This paper introduces the current status of data sharing in the People's Republic of China. It discusses barriers to data sharing and proposes three key solutions to overcome these barriers in China. The establishment of national laws and policies for data sharing is considered the key prerequisite to ensuring the successful implementation of resource sharing activities in public health. Driven by established laws and policies, the relevant operational models should be developed. It is also important to have strategies in place to ensure the established laws and policies are implemented by various organizations in different jurisdictions. These discussions are supported by relevant local and international evidence.

Keywords:

data, data sharing, public health, law, policy, implementation

Introduction

Scientific data is the original measurement or characteristic of a person or thing that is collected from various scientific activities, or extracted from databases or data warehouses by various data mining techniques. It may be 'a clinical measurement, a laboratory record, a medication dosage, or even a listing of treatment'.[1] Only after scientific treatment to comply with certain data standards of the relevant regulation bodies can data be used effectively for various public health purposes. 'Data are an expensive and valuable resource. Data have market value; they represent a revenue stream for many academic and governmental agencies'.[2]

The essence of sharing public health data involves the sharing and publication of information. Data sharing among organizations is important for (1) developing an essential information infrastructure that supports public health services and functions; (2) reducing duplication of effort in data collection and increasing the usability of the data that has already been collected; (3) providing information effectively and efficiently to inform public health policy and management decisions; (4) improving communication and standardization of core procedures and assessment to facilitate best practice in health care[3] and (5) realizing the scientific, economic and social value of scientific data. Therefore, to benefit society, the data stored in various databases owned by various departments and jurisdictions should be open to the public.

This paper will focus the current status, challenges and prospects of data sharing for public health. First, it will introduce the current practice of data sharing in China. It will then discuss the challenges to this data sharing in China. Finally, through a comparison of the practice of several countries, it proposes three essential strategies to overcome the barriers encountered.

The current practice of data sharing in the People's Republic of China

A significant amount of data has been collected in China. According to statistics, there are several billion bytes of scientific data collected by various levels of government departments and scientific research organizations sponsored by the government. The three major sources of these data include: (1) the administrative data collected by the health administrative organizations; (2) the scientific data generated from the scientific activities undertaken by staff in research, business and management organizations funded by government; and (3) the scientific data generated from scientific or productive activities undertaken by privately owned scientific, business, education or health organizations. These data are scattered in the databases of various organizations located in different jurisdictions. Although there is a critical need to compare these data to support public health research and policy making, the differences in data elements, the method of collection, the coding system and the structure of database tables that store these data etc. make comparison of these data impossible. As these heterogeneous, incompatible data are not readily usable for statistical analysis and reporting, they have rarely been published, despite the significant effort and resources spent in collecting them.

In recent years, the importance of data sharing has been emphasized by the Central Government in China. After a two-year, continuous, persistent effort, the Department of Health has gained considerable experience in the development and establishment of data-sharing policies. The Department of Science in China has implemented a scien-

tific data-sharing project to establish a platform to facilitate the sharing of scientific data. To date, more than 600 databases of varying size and quality have been established. Meanwhile, resources have been invested in digitizing the original paper-based data and establishing various databases to store these data. Some organizations that own these databases have initiated data-sharing services, which have resulted in some social benefits. However, the level and scope of scientific data sharing in China is still very limited.

There are many barriers to effective data sharing and the following section will discuss these barriers. Then, three essential strategies are proposed to overcome these barriers in China.

The barriers to sharing scientific data in China

Scientific data sharing in China is yet to break away from the custody of departments in various jurisdictions. The appropriate technology and model is yet to be developed for large-scope data sharing, the integration of heterogeneous data, and the distribution of collaborations to researchers and policy makers. To date, there is no network-based platform that is able to provide data services for clients in different regions that is underpinned by technology services, effective support mechanism and quality assurance.[3] The identified challenges fall into one of the following four categories: (1) a lack of laws and policies to guide the practice of data sharing; (2) a lack of coordination of effort, which leads to numerous isolated data sources that cannot communicate with each other; (3) a lack of an enabling structure and operational mechanism to sustain data-sharing services; and (4) a lack of a culture of sharing data. The following sections will provide a detailed explanation of each of these four categories.

A lack of laws and policies to guide the practice of data sharing

'Laws provide the mission, functions, and powers of public health agencies, set standards for their actions, and safeguard individual rights … They are an essential tool for improving public health infrastructure and outcomes.'[5] Therefore, the establishment of laws and policies are prerequisites for achieving data sharing for public health. However, the relevant data-sharing policies, the standards for data exchange, quality assurance and information services, and the relevant terminology and classification system have yet to be developed in China. Without the above regulatory protections, the owners of the relevant databases will not give end users access to databases to download data for the purposes of scientific research, disease prevention and control. If they are lucky, these end users can acquire a very limited amount of data, such as abstract of a publication. Conversely, the utilization rate is poor for databases established through public funding, which is a waste of public investment.

A lack of coordination of effort, which leads to numerous isolated data sources that cannot communicate with each other

Different levels of government and their departments have their own needs and targets for data collection and storage. Currently, there is no mechanism to achieve collaboration on data sharing among local, provincial and central governments. The reasons for this are similar to those identified in Canada: 'unclear constitutional roles and responsibilities for public health and the potential for disputes to arise over funding and data sharing'.[6] As in Australia,[7] there is a lack of national vision or a mechanism to implement data sharing. The consequences are duplicated effort and overlap in data collection, and the establishment of numerous heterogeneous databases at central, provincial, city and county levels of government departments and the private sector. Therefore, interoperability and data sharing among information systems is not possible.

A lack of an enabling structure and operational mechanism to sustain data sharing services

Data sharing is a process of social system engineering that needs the coordination of government departments in various jurisdictions. The limitations of the current government organizational structure in China mean there is a serious lack of communication between organizations in various industry sectors. This lack of coordination in information management has led to an inability to identify the availability, place of storage, ownership, accessibility, scale and scope of data stored in a database. Although a national identification system has been established for public servants, currently the system does not include the increasing number of employees in the private sector and farmers. In addition, there is no common data exchange standard that clearly defines how the organizations should share data with each other. [8] Therefore, it is difficult to link two information systems together. Even when two information systems are connected, it is challenge to integrate their data as these data are captured in different formats and semantics.

A lack of a culture of sharing data

As in Australia,[7] there is a lack of incentives for data sharing among the different levels of government without direct reporting relationships. To share data among organizations, an inter-organizational information network should be established to link data sources within organizations. This novel model of information networking challenges the traditional organizational structure and ownership of data in China. Therefore, its implementation will only be possible after significant organizational cultural change has been achieved.

In the following section, we will propose three essential strategies for overcoming the barriers discussed and facilitating scientific data sharing in China.

Our proposed solutions to overcome the barriers for scientific data sharing in China

We believe the following five key factors have to be satisfied before sharing of scientific data can be realized: data sources, data organization and management, regulations about data sharing, the development of technology to enable data sharing, and the interaction between these four factors. Therefore, to overcome the barriers for data sharing and optimize the benefits of these scientific data to the public, three key strategies are proposed: (1) the establishment of laws and policies for sharing data for the sake of public health (this action should take into consideration the culture and political structure in China); (2) the development of a comprehensive data-sharing mechanism that is guided by established laws and policies; and (3) the establishment of a mechanism to safeguard the implementation of the laws and policies. The following sections will discuss these three strategies in detail.

The establishment of laws and policies for data sharing in the context of China's culture and political structure

For the purpose of establishing an effective approach, the experiences of foreign countries should be examined if these countries have already taken action to achieve data sharing and interoperability in health care. For example, the Turning Point Model State Public Health Act in the USA has established clear definitions about the acquisition, use and disclosure of identifiable health information, security safeguards and fair information practices.[5] In Britain, the use and sharing of personal information in the public sector is governed by a number of laws including (1) the law that governs the actions of public bodies (administrative law); (2) the *Human Rights Act 1998* and the European Convention on Human Rights; (3) the common law tort of breach of confidence; (4) the *Data Protection Act 1998* and (5) the European Union law.[9] There are laws that define information management practice in either the public or private sectors in Australia. Canadians have also established their code of practice.[10, 11] However, as cautioned by Hodge *et al.,*[5] 'reforming public health laws is a delicate process that involves good timing, political will, and willingness to compromise'. Therefore, the Chinese political, social, cultural and organizational context has to be carefully considered, and the relevant stakeholders have to be thoroughly consulted when establishing laws and policies.

Data sharing can only be realized through negotiation between the organizations or stakeholders who own the data. Therefore, an adequate security protection mechanism has to be in place to ensure the scientific data owned by various entities will only be made available to the general public after the explicit consent of these entities. The legislation should clearly define the intellectual property of the owner(s) of the data. The interests of various stakeholders (such as individuals, organizations and society), should be appropriately balanced; whereas the information related to national security and commercial confidentiality should be adequately protected.

Before the establishment of such national laws and policies, different jurisdictions or industries could establish their own short-term policies or regulations for public health data sharing. Only through such practice could data sharing be achieved and scientific evidence acquired through analyzing these data. Such evidence will, again, support scientific research or management decision-making.

The development of a comprehensive public health data-sharing model that is guided by laws and policies

The fundamental requirement for public health data management is the proper configuration of these data to maximize their potential in serving the society. The federal governments in Australia, Britain and Canada have taken the lead in their efforts to establish a national health information network for sharing health care information across the continuum of care.[12, 13, 14] It is proposed that the Chinese Central Government should take the leadership and coordination role and provide incentives to encourage various organizations, such as various levels of health departments and public and private health care organizations, to collect and share data.[15]

An incremental update model[6] that uses a five-step implementation may be useful in achieving data sharing in the short term. Baker[6] defined the five steps as (1) the data authority is identified; (2) data are consolidated/redistributed: making sure the authority has the data; (3) the data authority allocates a persistent identifier for each feature within each theme; (4) data are redistributed: the data authority redistributes data with identifiers attached; and (5) incremental updating of the whole process. As mentioned previously, a national identification system has been established for public servants. This is in the process of being extended to the general public, including the rural population.

The best scenario is the complete publication of scientific data after acknowledging and protecting the intellectual property of its owners.[16] Of course, appropriate compensation should be given to the owners of these data to provide incentives for them to share data. The relevant regulations should also be established to define the rights and responsibilities of the providers of public health data, as well as the end users of these data. A possible definition might be: if a customer is an end user of the data, they have no right to redistribute the data; whereas the data provider should share authorship with the end user in the publications that result from the data sharing. Whether the data leads to a product for the end user or not, the data provider should always be given financial compensation to maintain the data-sharing system.

The establishment of structure and process to enforce the implementation of laws and policies for data sharing

Necessary mechanisms should be in place to enforce the implementation of laws and policies for data sharing. In fact, China has established a comprehensive primary health care network from village to town to county to province.[17] The system was established in the early 1950s based on the former Soviet Union model of epidemic pre-

vention and education, with focus on the hygiene and prevention.[17] This network has played an important role in controlling and preventing infectious diseases for the past 50 years. It will also be an effective structure to enforce the implementation of laws and policies for data sharing once these laws and policies are in place. The China CDC has established a four-tiered network of disease control (centrality, province, city and county) that is managed at different levels. In addition, the China CDC has close working partnerships with non-government organizations such as the Chinese Medical Association, the Chinese Preventive Medicine Association and the Chinese Epidemiological Association.[17] Therefore, the structure for public health data-sharing management and services already exists.

The experience of other countries in this area can also be valuable. For example, to enhance public servants' understanding of the country's complex legal framework, the Department for Constitutional Affairs (Justice, Rights and Democracy) published *Public sector data sharing: guidance on the law* in 2003 as a map to guide the practice of data sharing in Britain.[13] In the USA, the implementation of environmental justice at the federal level was matched by state programs.[18]

Conclusion

In summary, any complex social problem can only be solved by following certain laws and policies. In addition, any creative theory and strategy can only be implemented after following certain procedures. The established laws, policies and procedures can only be effectively implemented after an adequate assessment, reporting and feedback mechanism is in place. Therefore, the establishment of sound laws and policies is the first step towards achieving public health data sharing. Afterwards, the implementation of the clearly defined laws and policies and a strong monitoring mechanism for this implementation will lead to improving data sharing for public health in China.

References

[1] Chapman R. New means for increasing data accessibility. In O'Carroll PW, Yasnoff WA, Ward ME, Ripp LH and Martin EL (eds) Public health informatics and information systems. New York, Springer-Verlag 2003**;** pp**.** 408-430.

[2] Lumpkin JR. History and significance of information systems and public health. In O'Carroll PW, Yasnoff WA, Ward ME, Ripp LH and Martin EL (eds) Public health informatics and information systems. New York, Springer-Verlag 2003**;** pp**.** 16-38.

[3] Hill K, Smith R *et al.* Falls clinics in Australia: a survey of current practice and recommendations for future development. Australian Health Review 2001; **24**(4): 163-74.

[4] Huang D, Guo C. A discussion about how to manage the sharing of scientific data. Beijing: Chinese Science and Technology Publishing Company 2002; (2) 40-60.

[5] Hodge JG, Gostin LO *et al.* Transforming public health law: the turning point model state public health act. Journal of Law, Medicine & Ethics 2006; **34**(1): 77-83.

[6] Baker T. A conceptual model: data sharing, incremental updating, and future topographic map production. The Globe 2002; **53**: 23-31.

[7] Wilson K A Canadian agency for public health: Could it work? Canadian Medical Association Journal 2004; **170**(2): 222-223.

[8] Xiao XM. A discussion about the safeguarding mechanism for sharing literature resources in China. Journal of Library 1997.

[9] DCA (Department for Constitutional Affairs, Justice, Rights and Democracy). Public sector data sharing: Guidance on the Law 2003. http://www.foi.gov.uk/sharing/toolkit/lawguide.htm

[10] Statistics Canada. Data-sharing agreements. http://www.statcan.ca/english/survey/business/sharing.htm

[11] Genome Canada. Data release and resource sharing policy 2005. http://www.genomecanada.ca/xcorporate/policies/DataReleasePolicy.pdf

[12] HealthConnect. http://www.health.gov.au/internet/hconnect/publishing.nsf/Content/home

[13] Department of Health. The National Programme for IT in the NHS, key elements of the procurement approach. http://www.dh.gov.uk/PublicationsAndStatistics/Publications/PublicationsProcurement/PublicationsProcurementArticle/fs/en?CONTENT_ID=4010183&chk=/H75W1

[14] Canada Health Infoway. http://www.infoway-inforoute.ca/en/home/home.aspx

[15] Xue G. Establishing scientific data sharing to increase the competency of science and technology in China. Fundamental Science in China 2003; 1: 5-9.

[16] Wang H. The strategic importance of scientific data. Xinghua Digest 2004; 11: 114.

[17] Lee L. The current state of public health in China. Annual Review of Public Health 2004; 25:327-339.

[18] Parris, Thomas M. Searching for environmental justice. (BYTES OF NOTE). *Environment* 2005; 47(6): 3.

Address for correspondence

Ping Yu
School of Information Systems and Technology
University of Wollongong
Wollongong, NSW 2522
Australia
E-mail address: ping@uow.edu.au
Tel: 0061 2 4221 5412
Fax: 0061 2 4221 4045

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Confidentiality Preserving Audits of Electronic Medical Record Access

Bradley Malin[a], Edoardo Airoldi[b]

[a]*Department of Biomedical Informatics, School of Medicine, Vanderbilt University, United States*
[b]*Lewis-Sigler Institute for Integrative Genomics & Department of Computer Science, Princeton University, United States*

Abstract

Failure to supply a care provider with timely access to a patient's medical record can lead to patient harm or death. As such, healthcare organizations often endow care providers with broad access privileges to electronic medical record (EMR) systems. In doing so, however, care providers may access a patient's record without legitimate purpose and violate patient privacy. Healthcare privacy officials use EMR access logs to investigate potential violations. The typical log is limited in its information, so that it is often necessary to merge access logs with other information systems. The problem with this practice is that sensitive information about patients and care providers may be disclosed in the process.

In this paper, we present a privacy preserving technique that enables linkage of disparate health information systems without revealing sensitive information. The technique permits any number of vested parties to contribute to audit investigations without learning information about those being investigated. We motivate the protocol in a real world medical center and then generalize the protocol for implementation in existing healthcare environments.

Keywords:

privacy, confidentiality, computer security,
electronic medical records systems, medical record linkage

Introduction

In the mid-1990's the National Research Council of the United States concluded that electronic medical record (EMR) systems increase the potential for the inappropriate disclosure of a patient's health information to parties both inside and outside of healthcare organizations [1]. The commission recommended that healthcare organizations design and adopt policies, as well as technologies, to prevent and address intrusions. Following recommendations from the NRC and other studies, such as [2], state and federal regulations were enacted in the United States to regulate the transmission, privacy, and security, of electronic personal health information [3, 4]. Regulations safeguarding medical information have been enacted by many other countries as well, including member states of the European Union [5], Japan [6], and Australia [7].

Technological protections for personal medical information have lagged behind policy ratification, which is due in part to the complexity of the healthcare environment. Many off-the-shelf EMR systems are equipped with role-based access control (RBAC) [8], a common policy requirement. However, RBAC is rarely enforced at point-of-care because a lack of medical record availability can cause patient harm or death. Rather, hospitals tend to use a "break the glass" model: they endow care providers with broad access privileges, but stress that harsh punishments will result for system misuse [9]. Nonetheless, unauthorized accesses occur. For instance, since 2002, the University of California, Davis has fired at least six employees, demoted one, suspended one without pay, and retrained eighty for inappropriate accesses [10].

System misuse is discovered through audits of medical records access logs. A recent survey found that 28 of 28 EMR systems incorporate audit capability [8]. Yet, only 10 of the systems alert healthcare administrators of potential violations. A principle challenge is that EMR systems do not always contain the necessary information to characterize violations. Oftentimes, it is necessary to gather information from other information systems. However, these systems can be controlled by different facets of the organization with diverse privileges and proprietary knowledge.

An EMR audit paradox

Consider the following example. The Vanderbilt University Medical Center (VUMC) is a large healthcare organization with a centralized EMR system. The EMR access log documents each medical record viewed, including a timestamp, the login of the care provider, and the medical record number (MRN) accessed. It is the role of the VUMC Privacy Office to audit the EMR access logs for inappropriate accesses.

A particular type of access privacy officials look for is pre-existing relationships between care providers and patients; e.g., "Are the care provider and patient coworkers at the university?" Such information is not in the EMR, but is in the university's human resources (HR) knowledgebase. Yet, no primary key exists between the EMR and HR databases. Thus, the common attributes are personal identifiers, such as personal names, demographics, or Social Security Number must be used to link the systems. However, the execution of a database join on patient's

identifiers, as shown in Figure 1, would reveal patient-specific information to human resources administrators, including information on patients that are not university employees. This investigation, though performed for surveillance purposes to detect wrongs committed against patients can violate patient privacy.

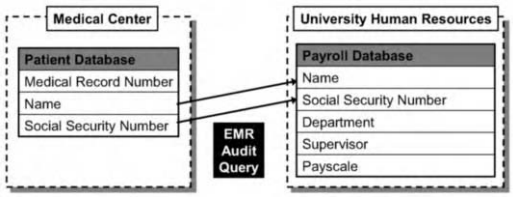


Figure 1- A query across organizational boundaries can violate patient privacy.

How can medical privacy officials determine if a patient is an employee without revealing patient identifiers?

Contributions of this research

In this paper, we propose a novel protocol called CAMRA (Confidential Audits of Medical Record Access) that allows an auditor to access information from non-EMR systems without revealing the identity of those being investigated. CAMRA leverages a secure protocol in which information is encrypted and stored at a third party. The EMR auditor sends queries of encrypted information to the third party to gather information from disparate systems for its investigation. The end result is that the identities of those being investigated are never provided beyond the EMR system. CAMRA simultaneously enhances confidentiality (i.e., no disclosure of person identifiers) and security (i.e., improved audit ability) in EMRs.

Methods

Record linkage and identity protection

Record linkage is an activity that is common to a range of biomedical environments. Record linkage within an organization is often performed via person-specific identifiers, such as name, address, or a unique identifying number (e.g., the Social Security Number is a feature frequently used in the United States). However, disclosing identifiers across organizational boundaries is often legally prohibited.

To satisfy legal constraints, biomedical researchers have applied various techniques to link records on entities without revealing identifiers. Secure one-way hash functions were developed and successfully applied for epidemiological follow-up studies [11]. Hash functions convert an individual's identifiers into a pseudonym; e.g., the name *John* becomes 0sad01a. To link information, each record is hashed and matching hashes are linked. A limitation of the one-way hash; however, is that all organizations linking data must share a hash key. The existence of a common key is a security risk because the hashed identifiers are susceptible to a "dictionary attack". For instance, an HR system administrator can hash identifiers until it matches a received hash from the medical center.

This is deemed acceptable risk in France, where the methods have been standardized for use at the Securite des Systemes d'information [12, 13], but in other countries this model has been contested.

To overcome the use of a common key, alternative methods have been proposed. For instance, Berman proposed a commutative random string technique [14]. In this method, each organization generates a string of random characters. The organizations exchange the strings and add them together to generate a set of common strings. Next, the organizations add identifiers (e.g., names, dates, etc.) to the random strings and compare the results.

One-way hashing renders it impossible to recover the original identities. Once records are linked, the organization must sacrifice the knowledge of the corresponding identities. This is adequate when identity is not needed. However, this model is not acceptable for EMR audits. To complete a privacy audit the EMR administrator must append knowledge to the identities of the individuals under investigation.

Commutative cryptography for record linkage

In this research, we adopt a commutative method. Specifically, we leverage a cryptographic technique called quasi-commutative encryption [15]. Each organization encrypts and decrypts identifiers to satisfy the following commutative property:

$$H(H(John\_Doe, \varepsilon_1), \varepsilon_2) = H(H(John\_Doe, \varepsilon_2), \varepsilon_1)$$

for any ordering of keys $\varepsilon_1, \ldots, \varepsilon_n$ and a function H.

A crucial distinction between the proposed method and prior models is that the function we apply can be converted into an asymmetric keyed cryptosystem. In other words, key ε_i can be paired with a key κ_i so that the original identifying value can be recovered using the same function, but different keys. So, organizations can encrypt and decrypt identifiers, such that

$$H\,(H\,(H\,(H\,(John\_Doe, \varepsilon_1), \varepsilon_2), \kappa_1), \kappa_2) = John\_Doe.$$

Fortunately, a variant of RSA cryptography satisfies these properties, such that $H(x, y) = x \bmod (n)^y$. RSA is an accepted standard for secure messaging in healthcare systems, so the CAMRA protocol can be built on top of existing healthcare information technology infrastructure. Unlike how RSA is used in practice; however, we define a private-key system, such that no organization discloses any keys. Note, the switch from a public to a private-key system does not require changing infrastructure.

CAMRA overview

The CAMRA protocol is designed so that disparate organizations can share encrypted versions of their identifiers with a "semi-trusted" third party. The third party is trusted to correctly execute a record linkage function, but the third party is not trusted to view the original identifiers for which it is performing the linkage. The third party analyzes queries of encrypted identifiers from the EMR auditor and responds with either encrypted identifiers

linked to relevant information for another organizations information system or "No Link Made". The feedback that is provided to the EMR auditor is encrypted information from which only the auditor, not even the third party, can learn the identifiers.

Let us walk through a basic, high-level, implementation of the CAMRA protocol. In this version, there are two organizations, such as the VUMC and the HR divisions in the example above. Figure 2 provides an illustration of the process.

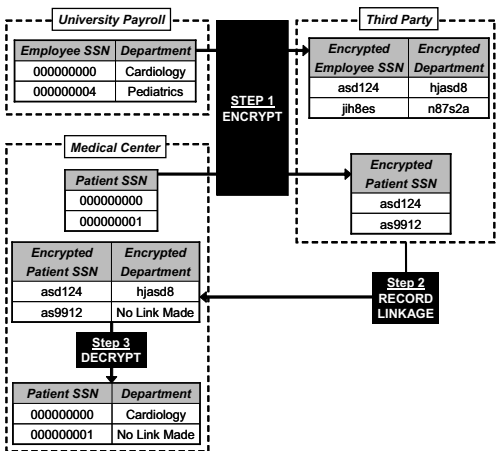


Figure 2 - Enrichment of access logs via the CAMRA protocol.

Step 1: Encrypt. The two organizations commutatively encrypt each other's datasets. Upon completion of this process, the organizations send their encrypted records to the third party. At no point throughout the encryption process can either participant achieve a dictionary attack because the records are not encrypted with the same key. After completion of the encryption phase, all records are comparable because they are encrypted with both of the participants' encryption keys [16].

Step 2: Link. Upon reception of the commutatively encrypted records, the semi-trusted third party performs record linkage on the encrypted identifiers. A list of matches, if any, is produced and sent back to the EMR auditor.

Step 3: Decrypt. The EMR auditor initiates a commutative decryption of the linked records. In the end, this provides the auditor with a list of decrypted identifiers that are linked to relevant information from outside of the EMR.

The above overview describes salient features of the proposed methodology. This example illustrates the architecture of the CAMRA protocol, but it obscures the exact ordering and manner by encryption and decryption process.

The CAMRA protocol

We now describe the CAMRA protocol in more detail, as well as showing how the protocol generalizes to settings with an arbitrary number of participating organizations.

To implement the commutative encryption and decryption required for CAMRA, each of the participating organizations must encrypt/decrypt each organization's sets of identifiers. To achieve this goal, two technical aspects must be in place. The first is a protocol for how to use maintain and use encryption/decryption keys at each organization. The second is a routing scheme is that specifies the order according to which organizations exchange sets of identifiers to eventually provide the semi-trusted third party with a set of encrypted, comparable records.

Protocol keys. Let D_V and D_R be the sets of patient and employee identifiers held by the VUMC and HR, respectively. Similarly, let $\langle \varepsilon_V, \kappa_V \rangle$ and $\langle \varepsilon_R, \kappa_R \rangle$ be the keys for the VUMC and HR, respectively. After commutative encryption the HR has $H(H(D_R, \varepsilon_R), \varepsilon_V)$. Notice, however, that HR is also in possession of $H(D_V, \varepsilon_V)$, which it received from the VUMC. Now, since HR can decrypt commutatively, it can generate $H(H(H(D_R, \varepsilon_R), \varepsilon_V), \kappa_R)$, which results in $H(D_R, \varepsilon_V)$. The consequence is that now the HR can compare its records and the VUMC records as if they were singly hashed by the VUMC: $H(D_R, \varepsilon_V)$ compared to $H(D_V, \varepsilon_V)$. This problem was recognized in [16] and so, to prevent such leaks, it is necessary to use "blinding".

More formally, let K be the set of organizations that share information with the third party. Each organization ($k=1,...,K$) will maintain two pairs of ⟨encryption, decryption⟩ keys,

$$\langle \varepsilon^b_k, \kappa^b_k \rangle \text{ and } \langle \varepsilon^m_k, \kappa^m_k \rangle,$$

for an agreed upon quasi-commutative hash function H as defined above. The function H is made public, however, all keys are kept private to each organization. The first key pair is used for "blinding" purposes (superscript b) by organization k with its own set of records, akin to the blind signature process defined in the original description of untraceable payment systems [17]. Blinding the records prevents leaks of information as illustrated above. The second key pair corresponds to "multi-party" keys, which each organization uses to encrypt/ decrypt each organization's set of records.

Record Routing. Alternative routing schemes share the property that each participating organization *starts* and *ends* each round of commutative encryption (initial and final) with its own set of records. In other words, each organization will blind its records before sending them according to a pre-specified routing scheme for commutative encryption. The organization will then remove the blinding from its records before sending them to the third party after all other organizations have encrypted it with their multi-party key.

Figure 3 depicts two possible routing schemes. The left panel shows a circular routing procedure based on commutative protocols. This procedure is efficient in terms of communication costs, but it requires that a peer-to-peer architecture be set in place. In contrast, the centralized routing model, shown in the right panel, are more computationally expensive, but can take advantage of existing network architectures that are designed for

high-speed environments with dedicated and trusted communication channels

The best routing option will depend on the features of the organizations involved in the audit. Regardless, the security of the CAMRA protocol does not depend on the specific routing scheme that is implemented.

The complete CAMRA protocol is executed as follows.

Step 1: Blinding to Encrypt. Each organization k creates a dataset of *dummy* records and adds them to their own set of EMRs, and encrypts the resulting set of records D_k using ε^b_k. After this initial encryption, a blinded dataset $H(D_k, \varepsilon^b_k)$ exists for, and is in the sole possession of, each organization.

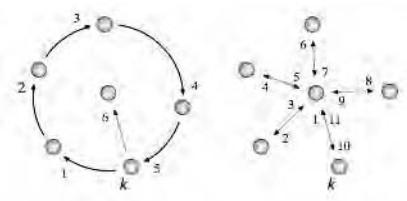


Figure 3 - Two alternative routing schemes from organization k to the third party (in the middle). Nodes are different organizations and numbered edges are routing steps.

Step 2: Commutative Encryption. Each organization encrypts its blinded records with its multi-party key and sends the result to organizations following the pre-specified routing scheme. Each organization encrypts the received sets of records its multi-party key, shuffles the order of the records, and continues to send the records around according to the routing scheme. At the end of this process, each organization k is in the possession of

$$H(H(H(\ldots H(H(D_k, \varepsilon^m_1), \varepsilon^m_2) \ldots \varepsilon^m_{K-1}), \varepsilon^m_K), \varepsilon^b_k).$$

Notice that the blinding key ε^b_K must be removed if the third party is to perform record linkage. This is where commutative decryption is handy. Organization k can remove the blinding simply by decrypting with κ^b_k.

Step 3: Encrypted Record Linkage. Each participating organization sends the resulting dataset to the semi-trusted third party, who performs record linkage over the set of encrypted EMRs. The resulting list is sent to the EMR auditor.

Step 4: Blinding for Decrypt. Upon reception, the auditor selects a new pair of blinding keys and blinds the list.

Step 5: Full Decryption. As in Step 2, the auditor, sends the encrypted list of matches to each organization according to the routing scheme. Now, each organization decrypts the list and sends it back to the auditor. Once every organization has decrypted the list, the auditor decrypts to remove the remove the blinding key.

Results

The CAMRA protocol is based upon our previous work on collusion resistant protocols [16] and inherits some of its desirable properties. We list them below.

Collusion-Resistant. It is possible to show that no set of organizations can successfully collude to learn the identifiers in the encryptions sent by the EMR auditor. Similarly, the EMR auditor can not collude with any set of organizations to learn the contents of another organization's set of records without proceeding through the commutative decryption process.

It is important to note that collusion resistance in CAMRA extends to prevent the third party from colluding with organizations against the EMR auditor. For instance, imagine the third party passes the VUMC records to HR. HR will learn which records the EMR and HR have in common, but will not know which records they correspond to. This is because 1) HR never decrypts records in CAMRA and 2) the VUMC shuffled the encrypted HR records. In doing so, HR cannot determine which of its encrypted records corresponds with any of its unencrypted records. The fact that HR cannot complete decryption without the assistance of the VUMC is the benefit of using a multi-party authentication mechanism with a single organization that is capable of decryption.

Detection of Malicious Actions. Collusion is a primary concern when organizations act according to the specified protocol. However, beyond collusion, there are actions that an organization can take to prevent record linkage from being correctly executed at the third party. For instance, an organization may use a faulty encryption key during the encryption process and the following comparison could be made by the third party:

$$H(H(John, \varepsilon^m_{\text{VUMC}}), \varepsilon^m_{\text{HR}}) \quad H(H(John, \varepsilon^{bad}_{\text{HR}}), \varepsilon^m_{\text{VUMC}})$$

In this case, HR correctly used $\varepsilon^m_{\text{HR}}$ with the name John from the first set of records (on the left), but inserted a "bad" key into the commutative encryption process to encrypt John in the second set of records (on the right). As a result, the names are not properly linked.

Fortunately, the CAMRA protocol can be extended to detect such malicious behaviors both during the encryption process and during the decryption process. These extensions are described in [16] and are important when trust between organizations is low.

Scalable. The CAMRA architecture is extendible in several ways. First, the security protocol is not limited by the number of organizations that are involved in the investigation. To increase the number of organizations that contribute to an audit, we only need an additional set of encryption/decryption keys. Second, CAMRA can be augmented so that each participating organization is provided with differential responses. To do so, the third party can send a response list to each organization. Each organization can use its own set of blinding keys to perform decryption. Yet, this extension must be performed with

caution as it violates collusion resistance property (i.e., the third party can collude with any organization).

Discussion

The CAMRA protocol illustrates that privacy does not have to be sacrificed in healthcare operations. The protocol prevents the disclosure of identifiers during the audit process. Nonetheless, the protocol has certain limitations that we now address.

Towards a fault tolerant model

Typographical errors and variation of personal information are a part of healthcare. For instance, an entity's name may be "Jon" in the EMR, but "Jonathon" at HR. Traditional record linkage methods account for such issues through string comparators and probabilistic matching algorithms [18]. Unfortunately, hashes and encrypted and encrypted versions of identifiers do not retain their similarities. As a consequence, linking encrypted identifiers can trigger false matches and non-matches.

Recently, methods have been proposed to measure the similarity between strings in an encrypted environment [19]. However, existing methods do not scale for use with CAMRA. This is because these methods are based on protocols that are designed for use between two organizations. Moreover, due to the cryptographic basis of these methods, they do not scale beyond two organizations. Yet, the CAMRA protocol is applicable to environments in which two or more organizations are involved. The development of scalable record linkage algorithms for encrypted data will limit the false non-linkage rate in CAMRA.

Third parties in healthcare

In the realm of healthcare, third parties are leveraged for a variety of purposes, including data warehousing, data aggregation, and brokering. Yet, from a data security perspective, the less number of organizations that handle data the better. As such, a more attractive possibility is to remove the third party from EMR auditing. Research in computational theory has shown that third parties can be removed from cryptographic protocols without sacrificing the level of security. However, the application of such theory in the real world is limited because resulting protocols are computationally burdensome and inefficient for daily practice. Nonetheless, minimal information sharing is the key to maintaining privacy in healthcare systems. We intend on developing confidential EMR audits models that limit the involvement of third parties.

Acknowledgments

The authors would like to thank Professors Dario Giuse and Dan Masys of the Department of Biomedical Informatics, and Gaye Smith at the Privacy Office of the Vanderbilt University Medical Center, for helpful discussions during this research.

References

[1] National Research Council. For the record: protecting electronic health information. Washington, DC: National Academy Press. 1997.

[2] Safran C, Rind D, Citroen M, Bakker AR, Slack WV, and Bleich HL. Protection of confidentiality in the computer-based patient record. MD Computing 1995: 12: 187-192.

[3] U.S. Department of Health and Human Services, Office for Civil Rights. Standards for protection of electronic health information; Final Rule. Federal Register, 2003 Feb 20; 45 CFR: Part 164.

[4] U.S. Department of Health and Human Services. Standards for privacy of individually identifiable health information; Final Rule. Federal Register, 2002 Aug 12; 45 CFR: Parts 160-164.

[5] European Commission. Directive 95/46/EC: the protection of individuals with regard to the processing of personal data and on the free movement of such data. 24 Oct 1995.

[6] Diet of Japan. Law on the protection of personal information. 23 May 2003.

[7] Office of the Federal Privacy Commissioner, Commonwealth of Australia. The commonwealth privacy amendment (private sector) act. 2001 Dec 21.

[8] Rehm S and Kraft S. Electronic medical records: the FPM vendor survey. Fam Pract Manag. 2001: 8(1): 45–54.

[9] Ferreira A, Cruz-Correia R, Antunes L, Farinha P, Oliveira-Palhares E, Chadwick DW, and Costa-Pereira A. How to break access control in a controlled manner. Proc IEEE CBMS 2006: pp. 847-54.

[10] Youngstrom N. Nosy employees are a risk, require a wide range of remedies. Report on Patient Privacy, Atlantic Information Services. August 2005: 5(8).

[11] Bouzelat H, Quantin C, and Dusserre L. Extraction and anonymity protocol of medical file. Proc AMIA Symp. 1996: pp.323-7.

[12] Quantin C, Kerkri E, Allaert FA, Bouzelat H, and Dusserre L. Security aspects of medical file regrouping for the epidemiological follow-up. Medinfo 1998: 9 (Pt 2): 1135-7.

[13] Quantin C, Bouzelat H, Allaert FA, Benhamiche AM, Faivre J, and Dusserre L. Automatic record hash coding and linkage for epidemiological follow-up data confidentiality. Meth Inf Med 1998: 37(3): 271-7.

[14] Berman JJ. Zero-check: a zero-knowledge protocol for reconciling patient identities across institutions. Arch Pathol Lab Med 2004; 128(3): 344-6.

[15] Benaloh J and deMare M. One-way accumulators: a decentralized alternative to digital signatures. LNCS 765: Proc Eurocrypt '93. 1994: 274-86.

[16] Malin B, Airoldi E, Edoho-Eket S, and Li Y. Configurable security protocols for multi-party data analysis with malicious participants. Proc IEEE ICDE 2005: pp. 533-44.

[17] Chaum D. Blind signatures for untraceable payments.Advances in Cryptography, Crypto 1982. Plenum Press. 1983: pp. 199-203.

[18] Grannis S, Overhage M, and McDonald C. Real world performance of approximate string comparators for use in patient matching. Medinfo 2004: 11 (Pt 1): p. 43-7.

[19[Churches T and Christen P. Some methods for blindfolded record linkage. BMC Med Inform Decis Mak 2004: 4: 9.

Address for correspondence

Bradley Malin, Vanderbilt University, Department of Biomedical Informatics, Eskind Biomedical Library, Fourth Floor, 2209 Garland Avenue, Nashville, TN 37232 USA.

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Design of a Decentralized Reusable Research Database Architecture to Support Data Acquisition in Large Research Projects

Jimison Iavindrasana [a], Adrien Depeursinge [a], Patrick Ruch [a], Stéphane Spahni [a], Antoine Geissbuhler [a], Henning Müller [a]

[a] *Geneva University Hospitals, Geneva, Switzerland*

Abstract

The diagnostic and therapeutic processes, as well as the development of new treatments, are hindered by the fragmentation of information which underlies them. In a multi-institutional research study database, the clinical information system (CIS) contains the primary data input. An important part of the money of large scale clinical studies is often paid for data creation and maintenance. The objective of this work is to design a decentralized, scalable, reusable database architecture with lower maintenance costs for managing and integrating distributed heterogeneous data required as basis for a large-scale research project. Technical and legal aspects are taken into account based on various use case scenarios. The architecture contains 4 layers: data storage and access are decentralized at their production source, a connector as a proxy between the CIS and the external world, an information mediator as a data access point and the client side. The proposed design will be implemented inside six clinical centers participating in the @neurIST project as part of a larger system on data integration and reuse for aneurism treatment.

Keywords:

Integrated information management systems, hospital information system, multi-institutional research databases.

Introduction

The diagnostic and therapeutic processes, as well as the development of new treatments are hindered by the fragmentation of information which underlies them. Many developments have emerged in the genomics, proteomics and medical imaging domains during the last 10 years. Linking distributed, multi-format, and multi-scaled data from genetics, proteomics, the individual, and epidemiological data for diagnosis and treatment is a new challenge in the biomedical informatics domain [1]. Many research projects are working on the integration of genomics knowledge into clinical practice. The INFOBIOMED[1] project is studying the relationship between bioinformatics and medical informatics in provision of individualization of healthcare. Integrating multi-scale data also means a need to integrate expertise of various participants: clinicians, biologists, statisticians, etc.

To develop new diagnostic or therapeutic processes, prospective clinical research can be conducted across multiple institutions. The electronic patient record can be used as initial data input and the data can be used for more than the treatment of a single patient. A multi-institutional data collection permits to have a statistically significant number of cases in a shorter time period. On the other hand, sharing clinical data for external researchers creates many problems: 1) patient data privacy and confidentiality need to be ensured; 2) there is no standard secure communication protocol and format for data representation as HL7[2] and DICOM[3] are mainly used for communication inside a hospital and do not cover all aspects required, particularly concerning terminologies to be used.

In multi-institutional clinical research, most of the data are collected inside the clinical information system (CIS) or are exported to create databases for one particular task. Specialized data collections such as genetic sequencing are rarely incorporated into routine medical practice. These data need to be entered from the outside of the clinical centers. Moreover, connections with existing public knowledge bases such as SWISSPROT[4] are rarely done. Designing a research database infrastructure for managing heterogeneous data formats, heterogeneous distributed data sources, access by multiple experts to the entire data located in various geographical regions is a big challenge.

Two categories of architectures exist to conduct multi-institutional clinical research: centralized [2] and distributed [3, 4]. The main drawback of a centralized architecture is the maintenance and data security [5]. In a distributed architecture, data can be stored in a local system and all participating institutions use the same data model, query language and the same database management software (distributed database system - DDBS) [6]. In the DDBS architecture, the maintenance complexity of the centralized architecture is reported on the level of each clinical center. In the real world it is also difficult to have the same database management system installed on the

1 http://www.infobiomed.org/

2 http://www.hl7.org
3 http://medical.nema.org
4 http://www.expasy.org/sprot/

same operating system across various participating institutions, leading to a heterogeneous database system (HDBS) [6]. Mediators have been proposed to allow access to heterogeneous data sources [3].

For security reasons, a CIS is generally a closed system. However, more and more clinical research is conducted inside hospitals and an increasing number of research networks are set up to share clinical data among institutions. Generally, in DDBS or HDBS, data are stored in a secured zone outside the CIS for secondary use. Nevertheless, it is risky to leave clinical data outside of the CIS because it is difficult to achieve a complete anonymity of clinical data. Genetic sequences, for example, are unique for each individual. Clinical data are collected over time and a visit date can be linked to other databases and can permit to identify a patient. Moreover, in provision of a patient-centered healthcare, there is a need to go back to the patient to inform him about new discoveries that might arise in a research project and as a new treatment of a disease.

Data integration is defined as the problem of combining data residing at different sources, and providing the user with a unified view of these data [7]. Generally, clinical centers have their own security, access rights management and privacy protection policy according to the role the user [8], and have the know-how concerning data access and communication using standards such as HL7, DICOM, IHE, or business components such as web services [9]. In multi-institutional prospective clinical research, a decentralized HDBS architecture is most flexible. It is also often the safest for the management of multimedia, multi-source, and distributed user data. In this architecture, the data are stored at their source i.e. the CIS of the institution where the data were created. Across institutions this can be in multiple databases, managed by various database management system installed on distinct operating system.

Often, half of the money of large scale clinical studies is paid for data creation and maintenance.

The objective of this paper is to propose a design of a distributed re-usable research database architecture, which permits managing and retrieving heterogeneous data to have real-time and up-to-date integrated data. Interoperability issues, research database maintenance and access are also discussed. The work is being carried out in the context of the European Union research project @neurIST and is part of a larger infrastructure within this project. The architecture proposed concerns “open” CIS: accessible for change and having direct data access, in the context of a multi-institutional research project.

Methods

The article describes the design and first implementations of a reusable research database architecture for a large-scale European Union funded clinical research project. Clinical data acquisition is planned in four countries at the first phase and then in a larger number once all system components are in place.

To do so, many constraints were taken into account from these countries on the one hand side from a legal standpoint and also from a technical standpoint. The architecture is to be implemented between the clinical institutions that will produce the data and several research applications that are the potential users of the data.

A web-services HTTP/XML based approach was chosen for the communication. To model all possible scenarios a large number of use cases were defined among the participants of the research project. Based on these use cases, legal and technical aspects of the functionality were defined and an architecture planned that was proposed to all partners and that is currently under internal evaluation for possibly missing parts.

Results

Requirements

Critical tasks in a database project are: data collection and validation, data storage and communication. In a decentralized database architecture, most of the clinical data collection, data validation and storage are done inside the participating clinical centers.

Data categorization and storage

In a decentralized research database architecture, it is crucial to differentiate the treatment and the research cycle. The treatment cycle produces clinical data and the research cycle provides derived data.

Additional data such as genetic analyses and research results related to a specific patient are in our case collected outside of the clinical centers and have to be stored inside the participating clinical centers’ information systems. Other global research data such as epidemiological study results or other public knowledge useful for the users of the database cannot be stored in any participating clinical center’s information system.

Patient identification

In a patient-centered clinical context, a patient can have access to all data stored in the CIS related to his health. Derived data need to be separated from clinical data by the means of patient identifiers. An internal patient identifier is used to identify a patient in the treatment cycle and an external identifier is used to identify the patient for secondary use of its data. As other clinical data may be collected outside of the hospital, the internal identifier cannot be used to identify the patient during the upload. First, the internal identifier may be created from private information of the patient. Second, external clinical data need to be validated by a clinician before its insertion into the electronic health record. Other specific identifiers need to be set up to identify external clinical data. In prospective clinical research, a patient needs to be recruited and followed in a single center. At this stage, we do not consider mobile patients between centers. This problem could be resolved with a global identification solution but is discarded for now.

Data integration

One of the biggest issues in a multi-institutional research database is data integration. The decentralized research database architecture is implemented as a global-as-view [7] and is able to manage structural, naming, semantic and content differences of the data from different sources. Inside a participating clinical center, data are coded into the local terminological system. The normalization of the data into a common terminological system is done on-the-fly.

Communication and security

Interoperability is also a crucial point in the design of a research database. XML/HTTP is the de facto standard to achieve interoperability. The architecture will use XML for data encoding and HTTP as transport protocol.

Privacy of clinical data needs to be ensured by the architecture: all data leaving the CIS, for secondary use have to be anonymized and the process has to be done on-the-fly before the data leave the clinical center. However, the architecture needs to permit to re-identify the patient to store related research result in the CIS. The architecture also needs to allow identification of the end-user. All communication has to be encrypted.

Design

The architecture we present in this paper has 4 layers: 1) the *data source layer* inside the CIS or at a node of the research network; 2) the *connector* on each participating clinical center which is mainly responsible for the horizontal data integration; 3) the *information mediator,* which is mainly responsible for vertical data integration and 4) the *client* (Figure 1). For security reasons and re-usability, operations are decentralized at the appropriate layer.

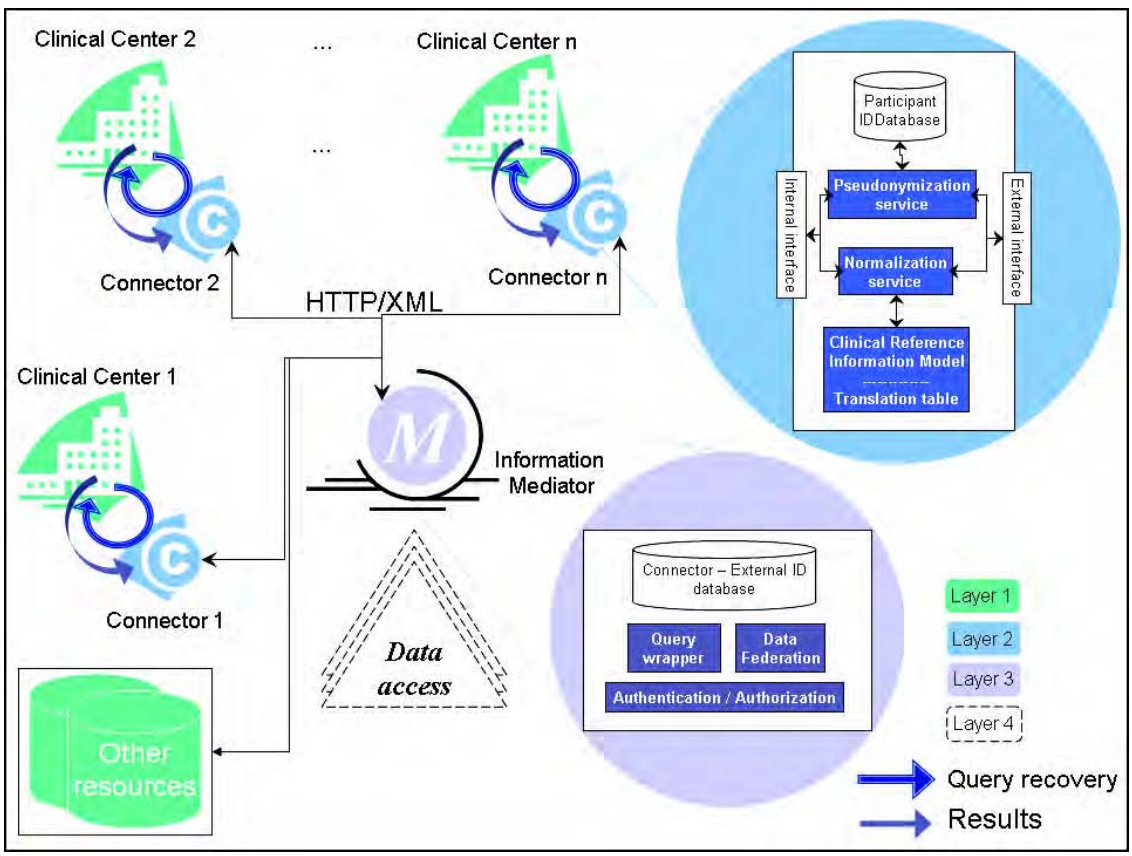


Figure 1 – A 4- layer architecture composed of: 1) the data source in the CIS or at a node of the research network; 2) the connector on each participating clinical center ensuring data normalization and pseudonymization; 3) the information mediator responsible for the query and user management and 4) the client

Data source

A large part of the data are stored inside the CIS and accessed using the existing data access components. Clinical data are stored in the electronic health record and derived data in a research database. The latter is integrated in the CIS because research results may contain private information such as gene sequences. There is no possibility to link clinical and derived data inside the clinical center: querying derived data using private information of the patient is prohibited. Even if derived data related to a patient are stored in the research database inside the CIS, it cannot be queried directly from the inside: it is only accessed through the information mediator. Other global or public derived data, not related to a specific patient are stored at their source.

The source can be queried using traditional communication protocols such as HL7 or DICOM or other components used inside the clinical center. The data access component inside the CIS communicates with the connector using XML/HTTP. For security reasons, the CIS always initiates the communication with the connector and pools the results. The CIS checks for new queries transmitted by the connector at fixed time intervals.

The connector

The connector is installed in the DMZ of each participating clinical center and uses the clinical reference information model (CRIM) for query and data translation. The CRIM contains the semantics of all existing data attributes: name, type and definition. These attributes are mapped to the local data attributes in the translation table. The normalization service is a bidirectional conversion of data format in the local or the global terminological system and it is invoked during the transition of the data inside the connector. The mapping has to be strong enough to ensure data integrity.

The pseudonymization service removes all private information from the local data before their delivery outside of the DMZ. It uses an ID database containing the mapping between various patient identifiers in the provision of re-identifying the patient for update tasks. The pseudonymization service can be extended as a patient identifier generator (external and other identifiers) but does not store any private information concerning the patient other than the internal ID.

The connector has two web service interfaces: the internal interface communicates with the CIS and the external one communicates with the information mediator. The connector does not identify the end-user: this task is centrally done at the information mediator level and the connector accepts only queries from the information model.

The information mediator

The information mediator is the data access point. It manages queries to various data sources: CISs and other public or private resources stored outside clinical centers. The user authentication and authorization are done at the information mediator level. The mediator also contains a database with the mappings between the external patient ID and all connector identifiers having information about the patient. This limits the network usage when a query is performed for a specific patient. It uses the clinical reference information model to formulate queries and it can be extended using an ontology to enable semantic mediation.

Discussion

We detailed in the preceding sections the design of a reusable research database for managing heterogeneous data sources to have real-time and up-to-date integrated data and address interoperability issues. The architecture permits to address data structure, naming, semantic and content differences across various clinical centers. Storing an internal identifier in the DMZ can raise legal issues. It has to be mentioned in the study protocol and the patient consent that the identifier will be stored in the DMZ and that it is never published beyond.

Limits of the approach

Using production databases for secondary use is badly perceived by IT decision makers in the health field because of: 1) technical constraints that the CIS is not open to change; 2) the sensibility of clinical data. However, with the increasing number of clinical and research networks managing and duplicating clinical data for secondary use will become increasingly expensive. For this reason most participating clinical centers prefer to export anonymized clinical data into a database in the DMZ instead of doing on-the-fly data extraction from the CIS. The research database will be hosted in a secured zone as shown in figure 2.

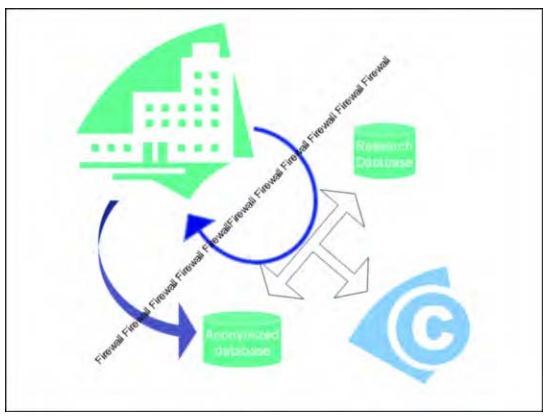


Figure 2 – Second storage architecture

In this second storage solution, the pseudonymization service of the connector is activated to pseudonymize data before their export into the DMZ. When there is a need to go back to the patient's data in the CIS, a message is sent by the CIS to collect queries from the connector. This process can be automated by setting up a service that will check for the existence of new queries at a fixed time interval.

In the proposed design, the communication between the CIS and the connector is unidirectional and intermittent. It can affect the query timing especially for *select* queries. Opening a permanent communication between the 2 layers of the architecture can raise security risks.

Contribution

The main advantage of the proposed architecture is its reusability, scalability and lower maintenance cost for research projects. Adding a new clinical center means installing a standard connector and providing a mapping between local and global data attributes. Defining a new research database project consists of defining the CRIM for this domain and adding clinical centers.

The emergence of GRID network technologies can empower the analysis and treatment of complex and manifold multimedia data: heavy computing tasks can be externalized onto a GRID-computing environment [10] and might allow new possibilities in the reuse of data. This

architecture can be interfaced with grid computing infrastructures that allow access to computational power that many institutions could not afford on their own [11]. In this case, the connector plays more than the role of a "data proxy". It receives computing requests from the CIS, pseudonymizes patient data if needed, and send the queries to a GRID-computing service provider. When the computation is finished, the connector receives the results, re-identifies the patient and sends them to the CIS.

Future developments

In the first part of the project, a patient is recruited and followed in a single clinical center. However, a patient may move between places after being recruited while continuing to take part in the study. There is a need to develop a unique external identifier to link information about this patient across clinical centers.

Conclusion

The design of a decentralized, scalable and re-usable, research database architecture with lower maintenance costs is described in this paper. The architecture has layers to manage heterogeneous data collection, storage and retrieval in a secured way and ensures the patient's information privacy. The architecture is suited for extensible clinical information systems. The structure is part of a larger @neurIST architecture.

Acknowledgement

This work was generated in the framework of the @neurIST Integrated Project, which is co-financed by the European Commission through the contract no. IST-027703.

References

[1] Hunter P, Smith N, Fernandez J, and Tawhai M. Integration from proteins to organs: the IUPS Physiome Project. Mech Ageing Dev 2005:126(1):187-92.

[2] Kerkri EM, Quantin C, Grison T, Allaert FA, Tchounikine A, and Yetongnon K. A virtual intranet and data-warehousing for healthcare co-operation. Medinfo 2001:10:23-7.

[3] Astakhov V, Gupta A, Santini S, and Grethe JS. Data Integration in the Biomedical Informatics Research Network (BIRN). In: Ludäscher B, and Raschid L, eds. Second International Workshop, Data Integration in Life Sciences. San Diego. Proceedings. Lecture Notes in Computer Science 2005: 3615: 317-20.

[4] Saltz J, Oster S, Hastings S, Langella S, Kurc T, Sanchez W, Kher M, Manisundaram A, Shanbhag K, Covitz P. caGrid: design and implementation of the core architecture of the cancer biomedical informatics Grid. Bioinformatics 2006: 22(15): 1910-6.

[5] INFOBIOMED. State of the Art on Data Interoperability and Management. Available at http://www.infobiomed.org / paginas_en/D11_State_of_Art_Data.pdf.

[6] Sujansky W. Heterogeneous Database Integration in Biomedicine. J Biomed Inform 2001: 34(4): 285-98.

[7] Hernandez T, Kambhampati, S. Integration of biological sources: current systems and challenges ahead, ACM SIGMOD Record 2004: 33(3): 51-60.

[8] Lovis C, Spahni S, Cassoni-Schoellhammer N, Geissbuhler A. Comprehensive management of the access to a component-based healthcare information system. Stud Health Technol Inform 2006: 124: 251-6.

[9] Geissbuhler A, Lovis C, Lamb A, Spahni S. Experience with an XML/http-based federative approach to develop a hospital-wide clinical information system. Medinfo 2001:10:735-9.

[10] Müller H, Garcia A, Vallée JP, and Geissbuhler A. Grid-Computing at the University Hospitals of Geneva. Proceedings of the First HealthGrid Conference, Lyon, 2003, pp 264-76.

[11] Müller H, Geissbuhler A, Ruch P. Report on the CLEF Experiment: Combining image and multi-lingual search for medical image retrieval. In CLEF Proceedings: Springer Lecture Notes in Computer Science, 2005:3491:718-27.

Address for correspondence

Jimison IAVINDRASANA
Service d'Informatique Médicale
Hôpitaux Universitaires de Genève
Rue Micheli-du-Crest, 24, Genève, Switzerland
e-mail : jimison.iavindrasana@sim.hcuge.ch

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

The Cancer Biomedical Informatics Grid (caBIG™): Infrastructure and Applications for a Worldwide Research Community

The caBIG Strategic Planning Workspace[a]

[a] *Members and Affiliations Listed in the Acknowledgment*

Abstract

Information explosion and new advances in high throughput experiments have challenged biomedical research, and suggested a future in which inter-institutional and international collaborations will be the norm. The cancer Biomedical Informatics Grid is an ambitious initiative launched by the US National Cancer Institute to develop a network of tools, data, and researchers to support translational and clinical research in oncology, with an ultimate goal to improve cancer care for patients. The three year pilot phase of caBIG ends in 2007, and has engaged over 900 clinicians, scientists, and patient advocates as developers, adopters, and workspace participants. Progress has been demonstrated in creating tools and building prototype grid architecture for collaborative research. Accomplishments in the pilot phase set the stage for extension of the community into other biomedical domains and for federation of the caBIG enterprise with similar initiatives in other scientific areas and in other countries.

Keywords:

biomedical informatics; computer communication networks; diffusion of innovation; medical oncology

Introduction

In recent decades, the information infrastructure requirements of biomedical research have increased exponentially. Clinical trials now reach beyond the academic milieu to include the broader clinical community, while regulatory reporting requirements and concerns about patient confidentiality and data security have grown. Basic and translational sciences now employ high-throughput genomic and proteomic technologies that generate terabytes of data and require new analytical tools. Translational research demands that information from biological samples be merged with clinical and population data, yet rarely are data collected in ways that make this possible. These problems compound when considering research conducted in teams from several universities and research centers. In short, modern medical research requires integrating, evaluating, and searching large volumes of diverse data, generated at multiple institutions. While this integration is sometimes possible, it is often an inordinately time-consuming and difficult process. In this decade informaticians have identified grid computing as offering one set of solutions to enable inter-institutional biomedical research [1].

Grid computing in biomedicine takes several forms. Computational grids assemble processing power from multiple sites to tackle intensive algorithmic problems [2,3]. Data grids allow the assembly of information from multiple disparate sites [4]. Community grids link investigators who share common interests, and allow data from one site to be used in a computational tool resident at another [5,6].

Recognizing that integration is key to accelerating cancer research, the United States National Cancer Institute (NCI) in 2003 announced the Cancer Biomedical Informatics Grid (caBIG™, http://cabig.nci.nih.gov) whose goal is to develop the information infrastructure needed to share data, specimens, tools, and computing power within and across institutions involved in cancer research. caBIG is creating an infrastructure that individuals and institutions can use to connect, nationally and internationally, effectively forming a worldwide collaborative cancer research community. This paper describes the motivation for the caBIG community, the standards and tools used in developing its underlying architecture, some early applications and infrastructure generated in the initial three-year pilot phase, as well as the potential to link caBIG to other biomedical grid projects underway in the USA and worldwide.

The caBIG Pilot: 2004-2007

Development of the community

The NCI Director vested oversight for a new information technology infrastructure in the NCI Center for Bioinformatics (NCICB), and in late summer of 2003, over 30 NCI-designated cancer centers held on-site discussions with NCICB to identify informatics resources that they might be able to contribute to caBIG, as well as elements they would hope to acquire from their colleagues using this newly conceived network (Figure 1)[7]. These discussions led to the establishment of three domain-specific workspaces in which tools would be developed: Clinical Trials Management Systems (CTMS) would develop tools to facilitate the design, management, analysis, and reporting of clinical trials; Tissue Banks and Pathology Tools (TBPT) would develop tools for managing specimens from clinical trials and patient care and for

sharing those specimens and their data across institutions while protecting patient privacy; and the Integrative Cancer Research (ICR) Workspace would develop a set of bioinformatics tools to analyze and integrate data from across the spectrum of clinical and basic research. A fourth domain space to develop tools for in vivo imaging (IMAG) was added in the second year of the project. All domain areas would ultimately make data available to the cancer research community across a computational grid.

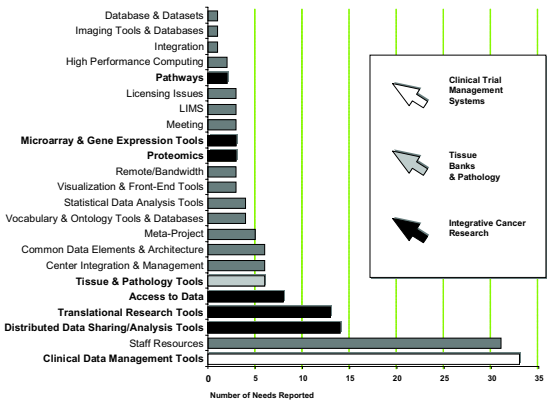


Figure 1 - Responses from Cancer Centers on needs in a biomedical informatics grid

Based on the feedback from the site discussions, NCICB and its general contractor developed a set of contracts to enable many institutions to take roles in caBIG [7]. Cancer Centers were contracted to participate in one or more of these workspaces as software developers, adopters, and/or working group members. In addition to the domain-specific workspaces, five others were formed to develop the standards for vocabulary (VCDE) and architecture (ARCH), to address issues of data sharing and intellectual capital (DSIC), to develop documentation and training (D&T), and to provide input on overall direction (Strategic Planning - SP). Patient advocates were recruited in each workspace to assure that patient issues and concerns were addressed.

The first year: Setting direction and expectations

caBIG was officially launched in February 2004 as a three-year pilot project. First year objectives focused on building the framework for workspace activities and engaging the cancer research community. Each workspace established its own timetable and set of strategic objectives. These were collated by the Strategic Planning Workspace into an overall Strategic Plan. The plan, presented at the first caBIG Annual Meeting in April 2005, set the stage with a scenario:

A motivating example: Scenario 2009[1]

> A researcher involved in a phase II clinical trial of a new molecularly targeted therapeutic for brain tumors observes that cancers derived from one specific tissue progenitor appear to be strongly affected. The trial has been generating proteomic and microarray data. She would like to identify potential biochemical and signaling pathways that might be different between this cell type and other potential progenitors in cancer, deduce whether anything similar has been observed in other clinical trials involving agents known to affect these specific pathways, and identify any studies in model organisms involving tissues with similar pathway activity.

This example includes a number of experimental systems and biological entities: small molecules, cell types, pathways, homologous proteins, animal models, clinical trials and therapeutics. In 2006 such a project would involve immense manual work getting information locally and from other sites. The fundamental idea behind caBIG is to develop the infrastructure and tools necessary to automate much of these steps so that by 2009 such multidisciplinary, multi-institutional research would be becoming routine.

Overall goals of the strategic plan

The caBIG Strategic Plan set three target dates for development: late 2005, 2008, and 2010. From the outset a dynamic tension has existed between delivering tools to satisfy the needs identified in Figure 1, while at the same time making progress on architecture and vocabulary projects that will enable the realization of the 2009 scenario. These were reflected in the 2005 short-term goals:

- Develop first generation caBIG compliant tools, data sets, and supporting data and architecture standards.
- Promulgate Gold, Silver, and Bronze development standards to the national cancer research community.
- Establish a mechanism for engaging the private sector, cancer research organizations not currently contracted under caBIG, and other Federal agencies and divisions into the caBIG community.

First generation tools naturally would arise from existing bioinformatics solutions, data sets would be those available, and standards were those already in adoption as of 2005. An important early goal was the development of three levels of standards that would guide caBIG participants and other interested parties in contributing projects to the grid (see below). Notably an interest in reaching out to other entities beyond the NCICB and the project itself appeared as a short-term strategic goal.

Goals for a year after the end of the pilot phase (2008) emphasized a functional community using the interoperable features of the grid, and a cancer research enterprise that was naturally adhering to caBIG standards:

- Develop sufficient research tools and standards to have a positive impact on the cancer research community, as measured by adoption of relevant caBIG principles in project proposals.
- Ensure widespread adoption of developer standards so that funded developer projects are operating under the Gold standard of compatibility.

1 Scenario courtesy of Michael Ochs, Johns Hopkins University, formerly of the ICR Workspace

- Adopt and use caBIG interoperable tools and data sets within the caBIG community.
- Develop mechanisms for engaging and promoting caBIG compliant technologies and established datasets within the oncology research community.

By the end of the decade the strategic plan envisioned caBIG principles and tools institutionalized within the cancer research community, and evidence of clinical impact:

Each workspace contributed specific goals for 2005, 2008 and 2010 to the strategic plan. The domain workspaces naturally focused on tools, whereas the cross-cutting ARCH and VCDE concentrated on construction of the grid itself, and the data elements that would link tools and datasets.[2]

The second and third years: Early results

In its second year, caBIG started to deliver products in its domain workspaces. More applications and the first working grid architectures appeared in year three. Table 1 shows a sample of these products, organized by maturity and workspace. As an example, the TBPT workspace has developed a comprehensive biospecimen management system to inventory, track, mine and visualize samples from geographically dispersed repositories, and link these specimens to clinical and molecular correlative descriptions. Modules within the biospecimen management environment include caTissue Core, caTIES (Text Information Extraction System), and CAE (Clinical Annotation Engine).

In addition to tools, the community has created critical components that, while not directly addressing researchers' needs, are required to ultimately make the caBIG tools work together. These include caGrid, caCORE, the BRIDG model, and the caBIG Compatibility Guidelines mentioned above in the strategic plan. caGrid, the heart of the entire project, sets the standards and protocols that allow researchers to use the Internet to access data, applications, processing power, and storage capacity at other institutions[8]. Built to international standards, employing the Globus Toolkit, with caGrid version 1.0 an institution can establish *data services*, allowing remote researchers to query the institution's data sources, and *analytical services*, allowing remote researchers to send data for processing by tools housed at the institution. The services already available on caGrid are noted in Table 1 with an asterisk, and a number of products from IMAG and ICR are planned in early 2007[6]. The ultimate success of caBIG in enfranchising multidisciplinary and cross-institutional research will depend largely on the continued evolution of caGrid.

Table 1 – caBIG Products, 12/2006

caBIG Product	**Workspace / Description**
C3D	CTMS: A web-based application for managing data across multiple cancer clinical trials. http://trials.nci.nih.gov/projects/trialmanagement/c3ds_project/c3d
C3PR 1.1	CTMS: A web-based application used to manage participants in cancer clinical trials. http://trials.nci.nih.gov/projects/trialmanagement/c3ds_project/c3pr
CRIX/ FIREBIRD	CTMS: A tool enabling investigators to file FDA Form 1572 online with NCI and other sponsors. http://ncicbsupport.nci.nih.gov/sw/content/crix.html
Lab Integration Hub	CTMS: An open source software tool that is used to collect, process and report laboratory data gathered during a clinical trial. https://cabig.nci.nih.gov/tools/LabIntegrationHub/
Patient Study Calendar	CTMS: A tool to assist in managing and documenting clinical trials protocols through the entire life cycle of the trial. http://cabig.nci.nih.gov/tools/PatientStudyCalendar
caArray*	ICR: A microarray data repository that is MIAME 1.1 compliant, supports MAGE-ML import and export, contains utilities for the submission and retrieval of Affymetrix and GenePix native file formats, and is accessed through a MicroArray and Gene Expression Object Model Application Programming Interface (MAGE-OM API) http://gforge.nci.nih.gov/projects/caarraydev/
geWorkbench 3.0	ICR: A suite of tools for loading, visualizing and analyzing gene expression data that will provide access to data from any repository with (MAGE-OM API), such as caArray. http://gforge.nci.nih.gov/projects/geworkbench/
caIntegrator	ICR: A platform that supports the retrieval, aggregation, analysis and sharing of data from heterogeneous operational repositories -- microarray, genomic, tissue array, imaging and clinical. http://caIntegrator.nci.nih.gov
TrAPSS	ICR: A system of several tools that aid scientists in searching for the genetic mutation(s) that cause a defect or disease. The system allows the researcher to create and prioritize a large list of candidate genes, select order, manage primer pairs, and predict secondary structure. http://gforge.nci.nih.gov/projects/trapss/
Reactome (GKB) Data	ICR: A curated database of fundamental biological pathways in humans that uses strict rules of assertion and evidence tracking to ensure a consistent high quality product. This database is compatible with the pathways exchanges standard identified by caBIG™. http://gforge.nci.nih.gov/projects/reactome/

2 The caBIG Strategic Plan can be viewed at https://cabig.nci.nih.gov/working_groups/SP_SLWG/Documents/index_html/document_view

caBIG Product	Workspace / Description
Rproteomics *	ICR: A system that aids cancer researchers in the processing of matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF) data. These R (a software language for statistical computing and graphics) libraries are incorporated into a caBIG™-compliant system that will aid cancer researchers in post-processing of their mass spectrometry data. http://gforge.nci.nih.gov/docman/?group_id=52
National Cancer Imaging Archive (NCIA)	IMAG: An web portal to a repository of downloadable in vivo images obtained from patients who have or are being evaluated for cancer. http://ncia.nci.nih.gov/
Cancer Text Information Extraction System (caTIES) 2.0	TBPT: A general purpose Text Information Extraction System that automates the process of coding, storing and retrieving data from free-text Pathology Reports. http://gforge.nci.nih.gov/docman/?group_id=17
caTISSUE Clinical Annotation Engine 1.2	TBPT: A clinical data mapping module designed to retrieve data from tumor registries and clinical and anatomic pathology Laboratory Information Systems (LIS) to connect, query, and share data with the caTISSUE system. http://gforge.nci.nih.gov/docman/?group_id=20
caTISSUE CORE 1.0.1	TBPT: A tool for biospecimen inventory, tracking, and basic annotation,.which can be rapidly deployed by cancer centers that have no such system or need to replace an aging legacy system. http://gforge.nci.nih.gov/docman/?group_id=18

CaCORE and BRIDG address issues of vocabulary and modeling. caCORE (Figure 2) is a set of tools created by NCICB for vocabulary management and model-driven software engineering[9]. It is intended to ensure that data will be collected using common data elements (CDEs), so that the data can be shared. caCORE ensures that caGrid can serve as a production environment as the caBIG community matures. The Biomedical Research Integrated Domain Group (BRIDG)[3] protocol model is a domain analysis model representing protocol-driven biomedical/clinical research, developed to be understood by domain experts. BRIDG provides the basis for harmonization among standards within the clinical research domain and between biomedical/clinical research and healthcare. It is being developed in conjunction with other standards organizations, and once it is in place, software developers will be able to write code utilizing this model rather than going through the time-consuming process of developing their own models.

The caBIG Compatibility Guidelines[4] provide a set of definitions that measure the maturity level of caBIGapplications with respect to interoperability. The guidelines

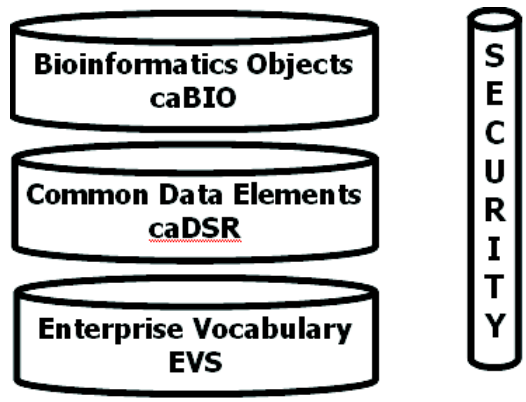


Figure 2 – caCORE Components: Cancer Bioinformatics Objects (caBIO), the Cancer Data Standards Respository (caDSR), and the Enterprise Vocabulary Server (EVS)

define requirements with respect to interfaces, vocabularies, data elements, and information models. The requirements reflect international standards, vocabularies that are firmly established across much of the scientific community, and IT industry "best practices" for technology development. Three levels of compatibility (Bronze, Silver, and Gold) are defined:

- Bronze: the application is built according to accepted standards and vocabularies, and can use external data
- Silver: the application contains caBIG validated application programming interfaces and vocabularies
- Gold: the application incorporates all Silver level requirements plus is fully operable on caGrid

The formal verification process for certifying the caBIG compatibility level of applications is still evolving, but when it is in place, users contemplating whether to adopt a caBIG or commercial product will be able to rely on compatibility level as an indication of how well the product will share data and analytic services with other caBIG™ products.

Workspaces also reached out to involve parties outside of NCI-designated Cancer Centers. Groups so approached include the Specialized Programs of Research Excellence (SPORE), cooperative clinical trials groups, the United States Food and Drug Administration, standards organizations, pharmaceutical companies, commercial software developers, and the United Kingdom's National Cancer Research Institute. caBIG now includes more than 900 participants across over 80 institutions and companies.

Moving beyond and reaching out

As these achievements demonstrate, caBIG is starting to deliver tools that will facilitate activities across the spectrum of cancer research, and connect these tools to caGrid. New projects extend interoperability. For example, the Cancer Translational Research Informatics Platform (caTRIP) will enable clinicians and translational researchers to query data resources in support of clinical decisions. caTRIP will integrate five tools, enabling queries across

3 http://www.bridgproject.org
4 https://cabig.nci.nih.gov/guidelines_documentation/

caTissue CORE, caTissue Clinical Annotation Engine, caTIES, Tumor Registry, and the Clinical Genomics Object Model of caIntegrator. Another integration project is LexBIG, which is strongly anchored in VCDE tools and resources, removes proprietary software dependencies in EVS (Figure 2), extends the API capabilities for vocabularies, affords interoperability with tools at the National Center for Biomedical Ontologies (http://ncbo.us), and addresses the broader use-cases needed in cancer research (relative to HL7).

From the outset, the Strategic Planning Workspace argued that staying within the cancer research domain would ultimately defeat the purposes of caBIG. caBIG is one of several promising projects that explore grid architectures to enfranchise multidisciplinary, multi-institutional research projects. In the United States, the Biomedical Informatics Research Network (BIRN) was initially established in 2001 to develop an inter-institutional collaboration in the area of neuroscience. BIRN's scope is smaller than caBIG, but it has been in existence longer and is significantly more developed. Although the recently launched NIH Roadmap Initiative to develop Clinical and Translational Science Awards (CTSA) does not directly address grids, the initiative does stress the role of biomedical informatics in translational research. The concepts embodied in caBIG are necessary to the collaborative success of the CTSAs. Internationally Biomedical Research Informatics Delivered by Grid Enabled Services (BRIDGES), a core research program within the United Kingdom, applies state-of-the-art concepts to cardiovascular functional genomics. Germany's MediGRID project, just getting underway, has a strong application focus. The main goal of MediGRID is the development of a Grid middleware integration platform enabling eScience services for biomedical life science. They have a bias toward early results, so caBIG's robust infrastructure tools could benefit MediGRID while their application use cases could stimulate caGrid applications.

For the vision of a worldwide web of translational and clinical research, where investigators can tap data and use tools that may "live" anywhere on the network, these and other projects must join forces. The European Healthgrid Initiative is a strong step in this direction-members of the caBIG community are collaborating with the association on "globalizing" the initiative. Extension of this collaboration to Asia/Pacific grid projects must follow [5].

caBIG was envisioned as a community of communities to support cancer research, and a solid contributor to the reduction of the burden of human cancer. It has succeeded in fostering collaboration among a diverse group of biomedical researchers and informaticians, to deliver a range of products and prototypes within its three year pilot phase. Focus on the infrastructure, adoption, training and outreach to other communities must characterize the next phase for the vision to be realized.

References

[1] Silva JS, Ball MJ. Prognosis for year 2013. Int J Med Inform 2002; 66:45-9.

[2] Lee HC, Salzemann J, Dacq N, Chen HY, Ho LY, Merelli I, Milanesi L, Breton V, Lin SC, Wu YT. Grid-enabled high-throughput in silico screening against influenza A neuraminidase. IEEE Trans Nanobioscience 2006; 5:288-95.

[3] Milanesi L. Merelli I. High performance GRID based implementation for genomics and protein analysis. Stud Health Technol Inform 2006; 120:374-80.

[4] Breton V, Medina R, Montagnat J. DataGrid, prototype of a biomedical grid. Methods Inform Med 2003; 42:143-7.

[5] Ranganathan S, Subbiah S, Tan TW. APBioNet: the Asia-Pacific regional consortium for bioinformatics. Appl Bioinformatics 2002; 1:101-5.

[6] Gurcan MN, Pan T, Sharma A, Kurc T, Oster S, Langella S, Hastings S, Siddiqui KM, Siegel EL, Saltz J. GridIMAGE: A novel use of grid computing to support interactive human and computer-assisted detection decision support. J Digit Imaging 2007 Feb 22 [Epub]

[7] Kakazu KK, Cheung LW, Lynne W. The cancer Biomedical Informatics Grid (caBIG): pioneering an expansive network of information and tools for collaborative cancer research. Hawaii Med J 2004; 63:273-5.

[8] Saltz J, Oster S, Hastings S, Langella S, Kurc T, Sanchez W, Kher M, Manisundaram A, Shanbhag K, Covitz P. caGrid: design and implementation of the core architecture of the cancer biomedical informatics grid. Bioinformatics 2006; 22: 1910-6.

[9] Phillips J, Chilukuri R, Fragoso G, Warzel D, Covitz PA. The caCORE Software Development Kit: streamlining of interoperable biomedical information services. BMC Med Inform Decis Making 2006; 6:2.

Acknowlegements

caBIG Strategic Planning Workspace Participants:-

City of Hope National Medical Center & Beckman Research Institute, Joyce C. Niland. Cold Spring Harbor Laboratory, R. Michael Townsend. Duke Comprehensive Cancer Center, Robert Annechiarico, Kimberly Johnson (CALGB). Fox Chase Cancer Center, J. Robert Beck, Frank J. Manion. Fred Hutchinson Cancer Research Center, Robert J. Robbins. Mayo Clinic Comprehensive Cancer Center, Christopher G. Chute. MD Anderson Cancer Center, Lynn H. Vogel. Ohio State University Research Foundation, Joel H. Saltz. Siteman Cancer Center (Washington University), Mark A. Watson. The University of Iowa - Holden Comprehensive Cancer Center, Thomas L. Casavant. UAB Comprehensive Cancer Center, Seng-jaw Soong. University of Colorado Comprehensive Cancer Center, Jessica Bondy. University of Pennsylvania- Abramson Cancer Centre, David A. Fenstermacher. University of Pittsburgh Medical School, Michael J. Becich. USC/Norris Comprehensive Cancer Center, John T. Casagrande. Yale Cancer Center, David P. Tuck.

Address for correspondence

J. Robert Beck, Fox Chase Cancer Center, 333 Cottman Avenue, Philadelphia, PA 19111 USA j.robert.beck@fccc.edu

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

The Integration of Grid Resources into a Portal for Research Collaboratories

Yassene Mohammed, Sabine Rey, Fabian Rakebrandt, Ulrich Sax

Department of Medical Informatics (Professor Dr. Otto Rienhoff), Georg-August-University of Goettingen, Germany

Abstract

Information technology (IT) infrastructures for research collaboratories and virtual research businesses are essential elements of science in medicine. The sustainability of the infrastructure depends on how it becomes a normal element of the scientific process, including its financing as part of a scientist´s workbench that can be easily configured to their needs. The results should also be distributed through a suitable platform to other research networks. The development of such an infrastructure serves all parties – mostly the funding agencies. This objective could be achieved by the integration of several portals of different research networks and the interfacing between these portals to maximize the usability and minimize the cost.

Keywords:

grid-computing, portal, data protection, medical research networks, virtual organization

Introduction

International competition in optimizing collaborative scientific research in medicine has led to a general understanding that smaller research collaboratories [1] (term coined by the US National Science Foundation) have no chance to establish sustainable collaborative infrastructures at a higher level. Therefore, the US and the EU are pumping major funds into the development of appropriate generic infrastructures. Some results of such funding procedures offer software solutions for other users like the caBIG-project in the USA [2]. Others build a horizontal infrastructure for different user communities, including medicine, called national grids or e-Science infrastructures, e.g., D-Grid in Germany [3, 4].

In this context new policy objectives have to be considered in developing IT-infrastructure for medical research networks [5]. The new infrastructure should:

- Link all research modules, facilitate their research, and support translational activities between research, health professionals, patients and citizens
- Use national and international existing concepts and solutions to efficiently achieve maximum support
- Transfer experiences to, and import results from, the different medical research networks.

For new research consortia, the aim is the integration of support functions in a user-oriented portal, if possible with an integrated access function over all applications. The portal is designed as a solution for research networks and/or groups using a standardized and portable technical platform to enhance information exchange and cooperation. Usually, portals are linked to a public website where the users/researchers receive access to their role-oriented, customized portal working environment, and to the different functions of the collaborative working environment after appropriate authentication. The development of the collaborative portal for group research also leverages integration of data and applications.

However, with the different goals, technologies and requirements of the various research groups, and to stay competitive, it will be necessary to develop and implement interfaces among different services-portals from different research networks instead of integrating all services in one portal. This keeps costs low, maximizes the usability and saves implementation efforts.

A successful health care information technology (IT) solution offers the possibility of transferring data from research into the treatment process and vice versa. The conceptual design of the IT-infrastructure and IT-solution methodologies requires new concepts defined in cooperation with the different working groups [6, 7]. In medicine, this applies especially to the legal framework in data protection (e.g., conforming design of pseudonymization process) and standardization of services. Legal demands for quality assurance, such as monitoring and source data verification, also need to be handled with state-of-the-art technology that uses standardized interfaces to export clinical data into analysis systems.

Materials and methods

Since 2001, the University of Goettingen has been developing an IT-portal for research networks in collaboration with several medical competence networks. It is a specialized version of a more general solution provided to Goettingen University, the Max Planck Institute and the German Association of University Heads (Rektorenkonferenz) [8]. For many years, the Department of Medical Informatics in Goettingen has focused on the development of IT systems and IT solutions to support clinical research. This work is supported by major third-party funding in the field of competence networks. Comprehensive communication and infrastructure concepts for supporting eTrial processes were developed and imple-

mented. Particular relevance for the integrative application in the eTrial process is the integration of the content management system and the clinical electronic patient record (EPR) system iXserv. This portal system was acquired by current and earlier third party projects. Synergy effects resulting from the shared resources are invaluable. These projects are:

- Competence Network for Congenital Heart Diseases
- Competence Network for Dementia
- Research Network for Creutzfeldt-Jakob Disease.

The corresponding licenses were already available for these projects and only the maintenance (and additional resources related to the project) needed to be funded. Exemplified by these competence networks, a standardized and portable portal platform based on MS-SharePoint server was introduced to support work and communication processes. It offered role- and process-oriented access for a more efficient use of IT infrastructure.

The services in this infrastructure could be offered as operative infrastructure to new research networks that need these services, such as the Competence Network for Degenerative Dementias (KNDD), on a low-level budget that covers customizing activities to the changing needs of the research projects in the new network.

Thus, the Department of Medical Informatics has already built a research IT infrastructure: the mentioned collaboration research portal that supports clinical research and a variety of methodology procedures. Special requirements of new research networks can be realized by appropriate parameterizing of this system (see Figure 1).

Although this is an example of how new research networks can inherit and reuse services from established networks, difficulties could arise if the new network demands new services or resources that the current system cannot fulfill or integrate. While building a completely new IT solution is expensive, reengineering the existing solution could lead to unneeded complexity that may affect the old network's operation.

To face the challenges of new resource demanding research applications in medicine, the existing IT infrastructure shall be complemented by access to processing power, storage and special applications. These resources will not be available within new projects and it would not make sense to fund single powerful computing nodes for small research networks (such as the KNDD). Instead, access to such resources can be eased through implementation of a grid [9].

'Grid-computing' stands for making computer power, storage resources, data and algorithms easily available to the scientific end-user. It has become a key infrastructural component on the way to broadly applied computational medicine. In a project funded by the Federal Ministry of Education and Research in Germany since September 2005, the University of Goettingen and seven partners from the community medicine sector were asked to cooperate with four other communities (including engineering climate research, high-energy physics and astronomy) to build the national grid-computing infrastructure [3].

Representing the community medicine in D-GRID, the MediGRID project has the task of incrementally developing and providing a grid infrastructure to meet the needs of the biomedical users. The user communities are represented in three research modules for biomedical informatics, image processing and clinical research, and they cooperate closely with the four methodological modules: middleware, ontology, resource fusion and eScience, which have the task of implementing the MediGRID infrastructure and connecting it to D-Grid [3, 4].

Through MediGRID, there are various resources available that are shared with the biomedical community and other communities. Following the grid paradigm, the threshold

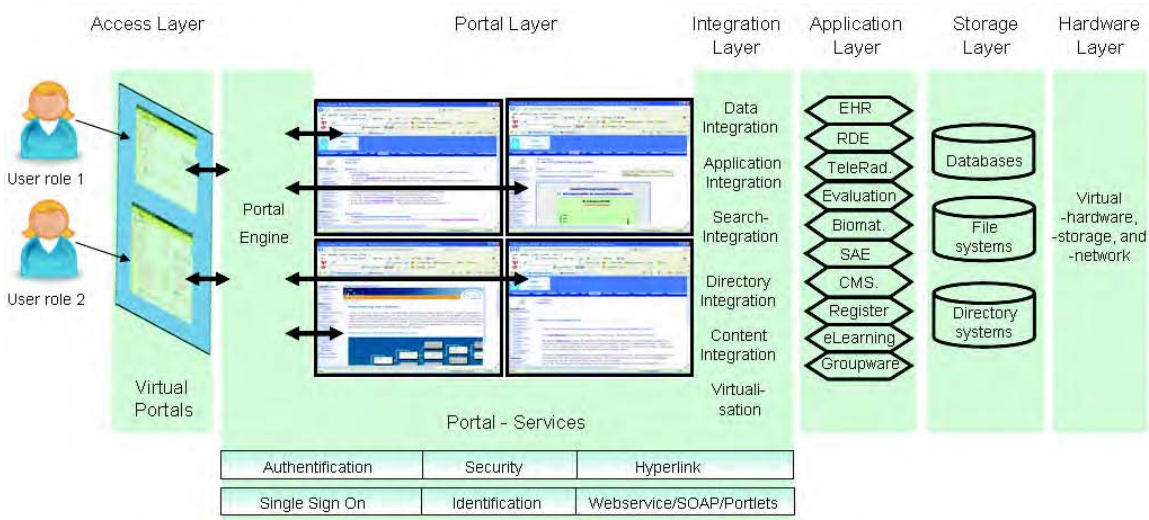


Figure 1 - Overview of the architecture of a portal for research networks. Beyond generic groupware, the services include study management, remote data entry (RDE), content management, clinical guideline management, eLearning module and other functions

to using these resources should be as low as possible for the academic user. Therefore, the grid resources are presented in one portal being set up in the MediGRID project (GridSphere and the specific portlets). Based on the experience in the MediGRID pilot applications current analysis methods as well as new methods and applications can be integrated as additional portlets in the online grid portal [10].

MediGRID is dealing specifically with the privacy and data protection challenges of the grid and distributed computing and storage. Sharing the MediGRID portal with new research networks (like KNDD), the latter could make full use of solutions for auditing, tracking, fine granular access control and privacy control in a distributed resources environment. Moreover, MediGRID policies would be used and applied to all new users/accounts; that is, all the administrative problems related to granting accounts and developing portlets in a distributed environment are solved [11, 12].

Depending mainly on open-source products and making all developments within one community available to other communities in D-Grid, MediGRID takes advantage of technologies developed within the overall D-Grid. Accounting, billing, monitoring, data provenance and long-term archiving are examples of such technologies [12] and will guarantee the sustainability of services.

The users of the MediGRID portal could access the different implemented portlets according to the MediGRID resource usage policies.[1] At the moment, portlets for image processing, bioinformatics and ontology are implemented.

Apart from the IT infrastructure itself, MediGRID also started a strategic cooperation with other grid projects hosted by the Goettingen State and University Library (SUB) and the Goettingen University Computer Center (GWDG). In March 2006, the three institutions set up the local grid-computing community in Goettingen, the so-called Goe-Grid. Goe-Grid is working closely with the kopal Project (Kooperativer Aufbau eines Langzeitarchivs digitaler Informationen – Cooperative Development of a Long-Term Digital Information Archive), also represented by SUB and GWDG, to import the kopal results into the grid environment and offer solutions for long-term archiving for grid users that could also be used by new research networks that have access to the MediGRID portal.

In this context, an approach to an operative infrastructure built on the collaboration portal could be massively extended by two elements that are of great benefit for collaborative research in medicine and need specific attention:

1. Integration of the grid portal for the life sciences into the portal for the research networks and harmonizing the two security infrastructures.
2. Solutions for long-term archiving by the corresponding kopal project.

Table 1 - Functions/services within the portals. While the collaboration portals focus on management, grid portals focus on sharing resources

(Medi)GRID portal [4, 10-12]	**Collaboration portal**
Enhanced security (audit, tracking, fine granular access control …)	Data protection and access control (PID-generator, pseudonymisation service…)
All applications are implemented as portlets in the portal. New application can also be implemented	Information services (accessed from www via browser, forums …)
Sharing resources, monitoring, billing, data and process provenance. Virtualization technology. Virtual organization management	Collaboration services (document management, project management, journal management, video conferencing …)
High-volume data storage and processing. Storage resource broker: data access on distributed platform	Data management (remote data entry system-KIS/KAS, national register-Ixmid/IXserv …)
Use of open-source standards and software	Quality management (evaluation systems-Evasys, transfer of knowledge …)

Results

The portal integration has to take place in the areas of security, the portal GUI and the applications.

Security

MediGRID is working under the umbrella of the German national D-Grid Project. D-Grid currently consists of 11 partners from different communities (engineering, astronomy, climate, libraries). This number will increase in February 2007 with new communities and projects from the second D-Grid Call. Dealing with patient data that is sensitive, MediGRID has the task of ensuring data security and data protection, and helping to develop/steer the technology for the D-Grid IT infrastructure regarding data protection laws and concepts [11].

Currently, MediGRID users need a DFN grid certificate that is harmonized in the EU Grid PMA. A security policy defines the process flow of granting MediGRID portal accounts. In addition, a three-step sign in the process using the DFN grid certificate and MyProxy upload tool enables

1 The MediGRID usage policies present guidance for the use of MediGRID resources. Current MediGRID policy is *Resource Usage Policy Phase One – Development Phase version 1.7.*

the use of grid resources for all authorized users without the need to install any middleware on local PCs (see Figure 2).

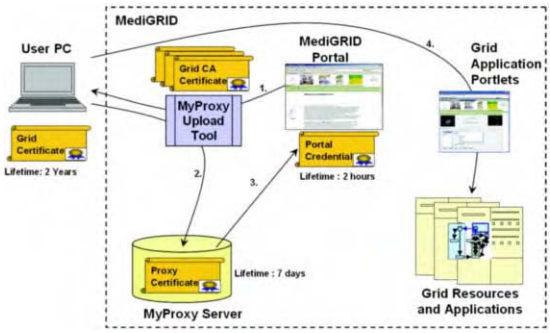


Figure 2 - Secure use of the resources and applications implemented in the MediGRID portal: (1) Signing into the MediGRID portal and downloading the MyProxy upload tool (2) Using the DFN certificate and the private key to create a MyProxy certificate and then uploading it in MyProxy Server (3) Creating own credential for the MediGRID portal (4) Using resources and applications

The kopal Project is a library-oriented solution for long-term archiving. It has its own security management using an SSL connection and users have, as in MediGRID, their own DFN certificate and portal account. However, unlike MediGRID, the access authorization in kopal is granted using an IP-white-lists that allow only certain PCs in certain institutes to access the resources.

The collaborative research portal uses the active directory system from Microsoft to administer user accounts. The security is managed using an SSL connection to the portal, and an access-control mapping process to the local resources using Meta-Directory from Novell and special connectors to these resources.

To be able to use the three infrastructures within one portal, the first step is to harmonize the security policies and the authentication/authorization methods between the MediGRID portal, kopal solutions and the collaborative research portal. Here, the MediGRID resources usage policies and user/developer agreements would be applicable, which means direct importing of solutions for administrative problems such as granting accounts, using the infrastructure for developing applications and so on. The security management for access to archived documents should be transferred from kopal, obtain legal acceptance for use with medical data (including patient data) and, if necessary, it can be redefined before implementation.

Harmonization of the portal GUI

MediGRID uses the open source GridSphere portal software that is implemented in Java according to java portal specifications JSR 168 that is expected to be in version JSR 268 in mid-2007. The GridSphere portal was developed from the GridLab project funded by the EU IST program in 2002–2005 to be similar to the commercial WebSphere portal from IBM. GridShere, according to the developer, was designed to be compatible with WebSphere. The collaboration portal is based on the standard Microsoft Office SharePoint Portal platform. Kopal provides services offered by Java API, which could be imported directly in the applications.

The goal of this step is to conceptualize and implement an interface between both portals, and to import the kopal functionalities given the results of the first step regarding security. To save the principles of a user-oriented portal, special focus has to be placed on:

- Maximum synergy for the scientist to allow the use of the services in both portals without extra effort
- Simple portal principles of single sign on
- User personalization and different contents on one screen aggregation possibilities.

Application integration

While one main task of MediGRID is developing a grid middleware integration platform enabling e-Science services for biomedical life science, the other task is to gain knowledge to implement applications on this infrastructure. An integration pipeline process was planned from the beginning of the MediGRID project. Currently, the knowledge gained in the first 'gridifying' steps is helping to accelerate the implementation of applications (see Figure 3).

The applications within new research networks that intend to access (Medi)grid resources have to be analyzed and categorized. Non-grid applications can be integrated as simple services outside the grid portal. More demanding applications can be integrated as grid services allocating and using grid resources. The new medical focus may demand new analysis methods not implemented previously. Therefore, access to NLM resources as well as to caBIG open-source code and applications is of paramount interest. Some of these applications are ready to be deployed on grid infrastructure, which means a saving in development costs.

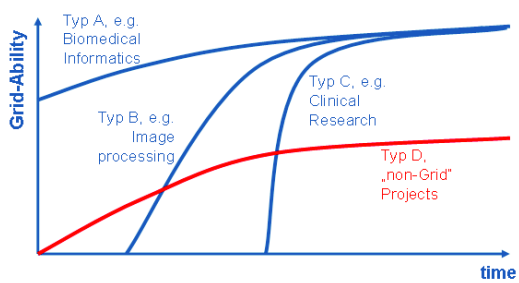


Figure 3 - The pipeline process in MediGRID: type A projects are prepared for grid computing. Other projects can be 'gridified' using the experience of the first projects (type B). Type D projects should be avoided

Discussion and conclusion

In this paper, we have discussed the different projects being conducted by the Department of Medical Informatics at the University Hospital of Goettingen, and the necessary steps to merge the IT infrastructures of these different projects starting from the point that access to resources uses a higher level portal access to the resources and aiming to keep the access process as simple as possible.

Our objectives in this study were to:

- Keep costs as low as possible
- Maximize the usability
- Save implementation time and effort.

These objectives reflect the triangle of cost, quality and resources. Our analysis shows that a three-step procedure should be performed and successfully completed to achieve the mentioned goals. These three steps are:

- Proving and interfacing the security policies and technologies between the different project/research networks
- Harmonizing the different portal-access methods to have one GUI
- Integrating all applications and services as portlets in the suitable portal, where the necessary resources can be accessed.

By building such an interface, the different medical research groups could work together more easily and could use the advantages of the available IT infrastructure more efficiently.

Taking into account the complexity of portal integration, it becomes clear that the integrated solutions have to be developed and tested by major centers. In a second step, end-users can customize their specific solutions in the developed system. However, smaller research consortia have no chance to develop such infrastructure by their own.

Acknowledgments

This work was funded by the MediGRID grant of the Federal Ministry of Education and Research in Germany (BMBF) e-science-initiative program under the promotion number: 01AK803H, the BMBF Competence Network congenital heart diseases (KN AHF) project under the promotion number: 01GI0210, and the BMBF Competence Network Dementia (KND) project under the promotion number: 01GI0428.

References

[1] National Research Council (U.S.). Committee on a National Collaboratory: Establishing the User-Developer Partnership. National collaboratories: applying information technology for scientific research. Washington, D.C.: National Academy Press; 1993.

[2] caBIG, cancer Biomedical Informatics Grid, http://cabig.cancer.gov. Accessed March 30, 2007.

[3] D-Grid Initiative, an e-Science-Framework for Germany, www.d-grid.de. Accessed March 30, 2007.

[4] MediGRID, Networked Computing Rssources for Biomedical Research, www.medigrid.de. 2005. Accessed March 30, 2007.

[5] Rienhoff O. Medical Research Networks – An International Comparison. In: Blobel B. and Pharow P, eds. Advanced health telematics and telemedicine: the Magdeburg expert summit textbook. Studies in health technology and informatics, v. 96. 2003, Amsterdam; Washington, DC, Tokyo: IOS Press; Ohmsha. viii, 291 p.

[6] Külz S and Rienhoff O. Die Relevanz vertikaler Netze in den medizinischen Kompetenznetzen für eine verbesserte Versorgungsqualität – Am Beispiel des Kompetenznetzes Demenzen, In: A. Jäckel, Telemedizinführer Deutschland, 2006, P. 136-140

[7] Rienhoff O. Bedeutung der Kompetenznetze für die innere Medizin. MED KLIN, 99(7): 407-11, 2004.

[8] Koke H. GÖ*Integriertes Informationsmanagement im heterogenen eScience-Umfeld, GWDG-Bericht Nr.64 2004

[9] Rienhoff O, Sehrt D, and Hamer B. Grids, Kompetenznetze, Gesundheitstelematik: Chancen für eine integrierte Infrastruktur. Deutsches Ärzteblatt, 2004(25): p. A-1791 / B-1491 / C-1437.

[10] Sax U, Mohammed Y, Viezens F, Lingner Th, Morgenstern B, Hartung M, Rienhoff O. Medical Grid Computing, in the International Supercomputer Conference, June 27th - 30th 2006, www.supercomp.de/show_exhibitor.php?id=325. Accessed March 14, 2007.

[11] Mohammed Y, Viezens F, Sax U, Rienhoff O. Rechtliche Aspekte bei Grid-Computing in der Medizin. In: Niederlag W, Dierks C, Rienhoff O, and Lemke HU, eds. Rechtliche Aspekte der Telemedizin. 2006. p. 302.

[12] Sax U, Mohammed Y, Viezens F, and Rienhoff O, eds. Grid-Computing in der Biomedizinischen Forschung – Datenschutz und Datensicherheit. Medizinische Informatik, Biometrie und Epidemiologie. Vol. 90. 2006, München: Medizin und Wissen Publ.Comp.

Address for correspondence

Yassene Mohammed
Department of Medical Informatics
Georg-August-University of Goettingen
Robert-Koch-Str. 40, 37075, Goettingen, Germany
E-mail: ymohammed@med.uni-goettingen.de
Phone: +49-(0)551-39 91217

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Biomedical Data Mining in Clinical Routine: Expanding the Impact of Hospital Information Systems

Marcel Müller[a], Kornel Markó[b], Philipp Daumke[b], Jan Paetzold[b], Arnold Roesner[b], Rüdiger Klar[b]

[a] *Department of Dermatology, University Medical Center Freiburg, Germany*
[b] *Department of Medical Informatics, University Medical Center Freiburg, Germany*

Abstract

In this paper we want to describe how the promising technology of biomedical data mining can improve the use of hospital information systems: a large set of unstructured, narrative clinical data from a dermatological university hospital like discharge letters or other dermatological reports were processed through a morpho-semantic text retrieval engine ("MorphoSaurus") and integrated with other clinical data using a web-based interface and brought into daily clinical routine. The user evaluation showed a very high user acceptance – this system seems to meet the clinicians' requirements for a vertical data mining in the electronic patient records. What emerges is the need for integration of biomedical data mining into hospital information systems for clinical, scientific, educational and economic reasons.

Keywords:

hospital information systems;
natural language processing; information storage and retrieval; medical records systems, computerized

Introduction

Hospital Information Systems (HIS)

Hospital Information Systems are widely adopted as a valuable and indispensable tool in the healthcare domain: they integrate patient information and improve efficiency and quality of care [1-3]. Hospital Information Systems usually focus on patient-centered, "intra-patient" issues, like storage of laboratory results, medical reports, discharge letters, coded diagnoses and procedures, and other clinical documents: Waegemann [4;5] defines the Electronic Health Record (EHR) as a *computer-stored collection of health information about one person linked by a person identifier*. Another aspect of the EHR is defined by the Healthcare Information and Management Systems Society (HIMSS) [6]: *The Electronic Health Record (EHR) is a secure, real-time, point-of-care, patient centric information resource for clinicians.* […] *The EHR also supports the collection of data for uses other than direct clinical care, such as billing, quality management, outcomes reporting, resource planning, and public health disease surveillance and reporting*.

According to those definitions and to individual patient care, access to the medical information contained in current Hospital Information Systems is mostly horizontal, i.e., patient-centered (figure 1): by opening a patient's electronic record the doctor gets all information about him. The HIMSS definition suggests more scenarios of use by aggregating information in the vertical view of all electronic patient records. This information relies usually on structured entries like billing information, coded diagnoses and procedures, structured laboratory or microbiology results – it easily can be selected using appropriate and well-known database and data warehouse technologies.

For the clinician, non-structured, heterogeneous information like discharge letters, anamnesis, reports of particular investigations and other narrative data are of high relevance for patient care – the more information is stored in the HIS, the more interesting are its vertical, i.e. inter-patient interdependencies.

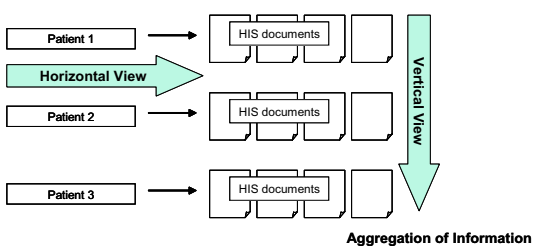


Figure 1 - Views on the HIS

Some questions the doctor could pose:

- "What patients did I treat that had the same disease?"
- "What was the outcome of that treatment?"
- "What were the adverse events of that treatment?"
- "Did I have patients with disease *xy* and symptoms *ab*?"
- "I'd like to get some typical cases with images of disease *xy* with symptoms *ab* for my students."
- "How was the name of the patient with symptoms *ab* that I treated three weeks ago?"
- […]

Although several promising technologies like the Clinical Document Architecture [7-10] and medical terminologies have been developed in order to standardize and structure clinical information, there is still a gap between this clinical need and today's practice. Furthermore, structured information like ICD Codes is often less accurate compared to text.

Narrative data cannot be selected using traditional database-driven technologies due to several limitations. Natural Language Processing (NLP) has developed efficient tools and new approaches to improve text mining mainly due to its paradigm shift from rule based methods to statistical and corpus based approaches [11]. The architecture of a medical text mining system may consist of three core functions [12]:

- Morphological analysis of words
- Structured analysis of sentences
- Semantical interpretation

To achieve basic results of text mining, it is not necessary to completely build all these components including the highly sophisticated semantic interpreter. We started with a medicine specific word analysis and focused on a solution of grammatical phenomena of word building severely impeding user friendly text retrieval:

- Inflection (lesion, lesions)
- Derivation (psoriasis, psoriatic)
- Abbreviations (LE=lupus erythematodes, SSM=superficial spreading melanoma)
- Synonyms (Sarcoidosis, Besnier-Boeck-Schaumann disease, Lupus pernio, benign Lymphogranulomatosis, Uveoparotid fever)
- Composition (leg ulcer, ulcer of the leg)
- […]

Data mining in the healthcare domain

In molecular biology, data mining technology has been successfully adopted to automatically summarize the immensely increasing scientific literature [13-17]. In the context of healthcare related information, natural language processing and data mining technology are often used in radiology findings [18-21].

At the University Medical Center of Freiburg, a comprehensive and flexible cross-language document retrieval engine ("MorphoSaurus") for the medical domain has been developed [22].

Biomedical data mining in dermatology

As a conclusion of the above, the document retrieval engine should be implemented into clinical routine, based on a large set of clinical data from a dermatologic university hospital, presented to those physicians, which actually generated the medical content.

These questions should be answered:

- How shall text mining be presented to the doctors?
- What features are important for clinical routine?
- What additional data can enhance doctors' compliance to this system?
- How useful is text mining for clinical routine as well as for scientific or educational purposes?

Materials and methods

Morpho-semantic indexing

At the core of this approach lies the MorphoSaurus text processing engine (an acronym for MORPHeme TheSAURUS). Its indexing technique is particularly sensitive towards cross-language morpho-semantic regularities. The system is centered around a new type of dictionary, in which the entries are subwords, i.e., semantically minimal, morpheme-style units. Language specific subwords are linked by intralingual as well as interlingual synonymy and grouped in terms of concept-like equivalence classes at the layer of a language-independent interlingua.

Morpho-semantic indexing starts from the assumption that neither fully inflected nor automatically stemmed words – such as common in many text retrieval systems – constitute the appropriate granularity level for lexicalized content description. Especially in the medical domain, a high frequency of do-main-specific suffixes (e.g. "-itis", "-ectomia") and numerous occurrences of complex word forms such as "pseudo | hypo | para | thyroid | ism" or "gluco | corticoid | s". To properly account for these particularities, the notion of subwords, i.e. self-contained, semantically minimal units had been introduced. [23]

Subwords are assembled in a multilingual dictionary and thesaurus, which contain their entries, special attributes and semantic relations between them, according to the following considerations:

- Subwords are listed, together with their attributes such as language (English, German, Portuguese) and sub-word type (stem, prefix, suffix, invariant). Each lexicon entry is assigned one or more morpho-semantic identifier(s) representing the corresponding synonymy class, the Morphosaurus identifier (MID).
- Semantic links between synonymy classes are added.

Figure 2 depicts how source documents (top-left) are converted into an interlingual representation by a three-step mor-pho-semantic indexing procedure:

- First, each input word is orthographically normalized in terms of lower case characters and according to language specific rules for the transcription of diacritics (top-right).
- Next, words are segmented into sequences of subwords or left unchanged when no subwords can be decomposed (bottom-right).
- The segmentation results are checked for morphological plausibility using a finite-state automaton which rejects invalid segmentations (e.g. segmentations without stems or ones beginning with a suffix).

- Finally, each meaning-bearing subword is replaced by an language independent semantic identifier, its MID, thus producing the interlingual output representation of the system (bottom-left).

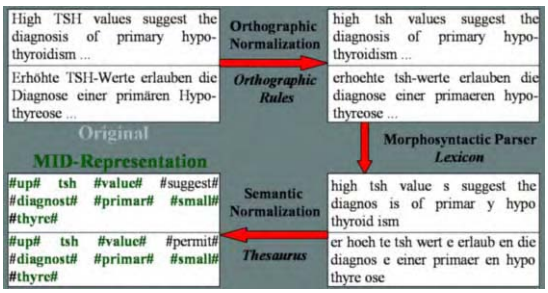


Figure 2 - Morpho-semantic indexing pathway

Having been successfully used for several scenarios [24;25], the impact on purely clinical data had to be tested in clinical routine.

Implementation

Clinical data set

Since the HIS [26] does not provide an interface, which could be used for implementing biomedical data retrieval, we had to extract all relevant information and to integrate it into a web-based user interface: Almost 30.000 clinical documents, which were stored in Rich Text Format (RTF) were extracted from the HIS database. These were mainly discharge letters, but also surgical reports, immuno-dermatological findings and different other narrative reports of clinical results. These documents originated from 2000 to 2006. The RTF-documents then were converted to plain text files that could be processed by the morpho-semantic indexing system.

Building the web interface

A web interface then had been developed based on Apache Lucene [27], a freely available open-source search engine which combines boolean searching with a sophisticated but highly efficient ranking model. It supports a rich query language like multi-field search, including more than ten different query operators. The web interface itself had been programmed in php [28] (figure 3).

Adding image repository

Since patient photographic images play an important role in dermatological record keeping and training [29], a large image repository, consisting of 90.000 photographs, had been patient-specifically linked to the web interface, in order to enhance the overall benefit of the system.

Evaluation

As the system should show the clinical impact of biomedical datamining, an evaluation engine was implemented. Users were intermittently encouraged to assess their benefit of the system and its further potential for their work. Search terms that were applied by the users were recorded anonymously.

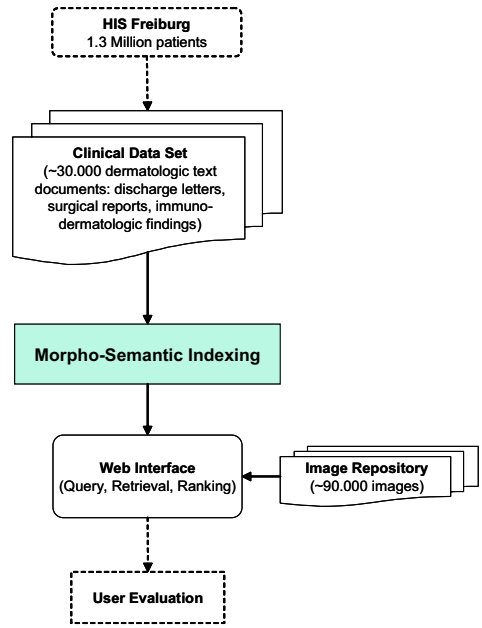


Figure 3 - Implementation

Results

Application interface

The main design principles for the web-based application interface were combining ease of use with specific search options according to clinical needs (figure 4): An edit box allows free-text search in all documents using the MorphoSaurus. Results can be ranked by relevance, date, patient name and date. In the results section, the relevant part of the underlying document is shown depending on the search terms. Search words or their derivations are highlighted. Information about the patient, the author and creation date of the document is shown in the respective context. The original document can be viewed as a PDF file that is linked to the record.

Another link opens the image library of the respective patient (figure 5), where all images are shown together with localization and creation date. This is particularly helpful for dermatologists, so they can analyze the images in the context of the underlying documents and vice versa.

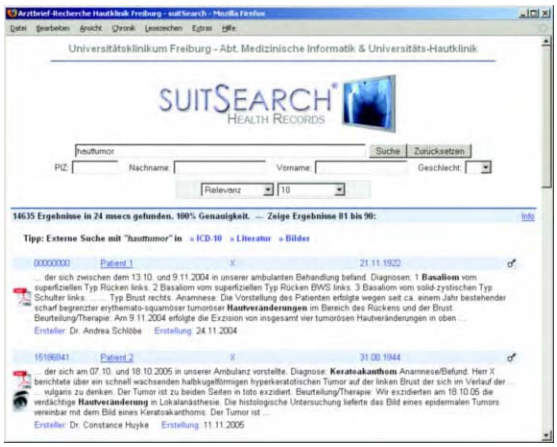


Figure 4- Web-based user interface

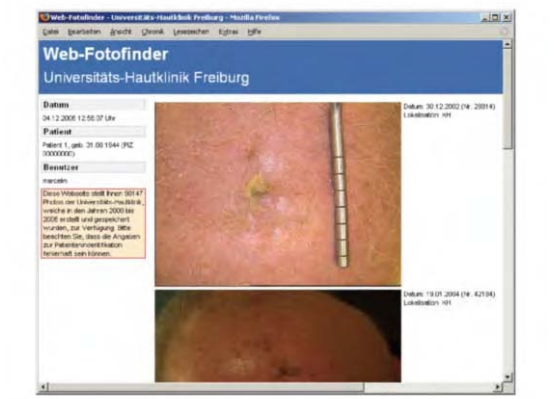


Figure 5 - Image retrieval

User evaluation

General acceptance

Users' acceptance of the data mining engine was remarkably high and enthusiastic from the beginning: For the first time, they had access to thousands of documents and images independently from the individual patient record. Three weeks after the introduction of the system, it was already used by three doctoral students in order to extract information for their doctoral thesis. Some even found unpredicted use scenarios, like searching good discharge letters as an example for their own letters they had to write.

Systematic evaluation

The use of the system has been systematically evaluated in three different major topics (clinical, scientific, educational impact), for each using structured questions as well as open free-text comments. Some preliminary results of that evaluation are given here.

The evaluation revealed 20 user evaluation data sets (15 physicians, 3 students, 2 information specialists). 82% stated that the system could enhance their clinical performance. Almost 89% thought that this kind of biomedical data mining and the integration of narrative text with dermatological images have a very positive impact on their scientific work. The impact on dermatologic education has been less estimated, only 52% of the users saw a potential benefit.

Users generally wished the integration of more clinical data, like radiology images and reports, laboratory results and other clinical findings.

Discussion

As long as Hospital Information Systems mainly contain narrative text, biomedical data mining seems crucial for enhancing clinical benefit in many ways. Although our system is currently successfully and regularly used in clinical routine by approximately 25 users, it has several limitations:

- Clinicians' search terms are not predictable; they vary from a single word to complex phrases, so a thorough analysis of the aggregated search history is necessary.
- A domain-specific dermatological MorphoSaurus dictionary has to be developed in order to enhance the response rate of their searches.
- Clinical data cannot be accessed in "real time", i.e. they have to be exported periodically from the HIS and indexed by the MorphoSaurus, before they can be retrieved.
- A direct integration into the HIS would be highly desirable and is planned in 2007.
- The integration of even more and also structured clinical data like laboratory results, histological findings, diagnosis and procedure codes etc. seems of broad interest.
- User acceptance and expectations towards biomedical data mining clearly has to be further evaluated in-depth.

Conclusion

Our experiences show that there is an emerging need for integration of biomedical data and text mining techniques into hospital information systems for clinical, scientific, educational and economic reasons. Nevertheless, structured data entry and storage mechanisms should be encouraged as well.

Acknowledgments

This work was partly supported by Deutsche Forschungsgemeinschaft (DFG), grant KL 640/5-1 and the European Net-work of Excellence "Semantic Mining" (NoE 507505).

We wish to thank all clinicians of the Department of Dermatology for fruitful discussions and evaluation of the system.

References

[1] Harrison JP, Palacio C. The role of clinical information systems in health care quality improvement. Health Care Manag (Frederick) 2006: 25(3): 206-212.

[2] Pare G, Sicotte C, and Jacques H. The effects of creating psychological ownership on physicians' acceptance of

clinical information systems. J Am Med Inform Assoc 2006: 13(2): 197-205.
[3] Ammenwerth E, Kaiser F, Wilhelmy I, and Hofer S. Evaluation of user acceptance of information systems in health care--the value of questionnaires. Stud Health Technol Inform 2003: 95: 643-648.
[4] Waegemann CP. The vision of electronic health records. J Med Pract Manage 2002: 18(2): 63-65.
[5] Waegemann CP. The five levels of electronic health records. MD Comput 1996: 13(3): 199-203.
[6] HIMSS Electronic Health Record, Definitional Model, Version 1.0: http://www.himss.org/content/files/EHRAttributes.pdf, last accessed: 2006-12-04
[7] Dolin RH, Alschuler L, Boyer S, Beebe C, Behlen FM, Biron PV, and Shabo SA. HL7 Clinical Document Architecture, Release 2. J Am Med Inform Assoc 2006: 13(1): 30-39.
[8] Muller ML, Uckert F, Burkle T, and Prokosch HU. Cross-institutional data exchange using the clinical document architecture (CDA). Int J Med Inform 2005: 74(2-4): 245-256.
[9] Muller ML, Butta R, and Prokosch HU. Electronic discharge letters using the Clinical Document Architecture (CDA). Stud Health Technol Inform 2003: 95: 824-828.
[10] Dolin RH, Alschuler L, Beebe C, Biron PV, Boyer SL, Essin D, Kimber E, Lincoln T, and Mattison JE. The HL7 Clinical Document Architecture. J Am Med Inform Assoc 2001: 8(6): 552-569.
[11] Manning D, Schuetze H. Foundations of Statistical Natural Language Processing. Cambridge, MA; London, UK: Bradford Book, MIT Press, 1999.
[12] Klar R. So hilft die Medizinische Informatik der Dokumentation und Klassifikation. Was heißt und zu welchem Ende betreibt man medizinische Dokumentation? Medizinische Dokumentation zur Optimierung von Qualität, Transparenz und Vergütung in der Gesundheitsversorgung. Köln: Deutscher Ärzte-Verlag, 2005: 41-55.
[13] Ananiadou S, Kell DB, and Tsujii JI. Text mining and its potential applications in systems biology. Trends Biotechnol 2006: 24(12): 571-579.
[14] Natarajan J, Berrar D, Dubitzky W, Hack C, Zhang Y, DeSesa C, Van Brocklyn JR, and Bremer EG. Text mining of full-text journal articles combined with gene expression analysis reveals a relationship between sphingosine-1-phosphate and invasiveness of a glioblastoma cell line. BMC Bioinformatics 2006: 7: 373.
[15] Andrade MA, Bork P. Automated extraction of information in molecular biology. FEBS Lett 2000: 476(1-2): 12-17.
[16] Ponting CP, Schultz J, Copley RR, Andrade MA, and Bork P. Evolution of domain families. Adv Protein Chem 2000: 54: 185-244.
[17] Andrade MA, Ponting CP, Gibson TJ, and Bork P. Homology-based method for identification of protein repeats using statistical significance estimates. J Mol Biol 2000: 298(3): 521-537.
[18] Mendonca EA, Haas J, Shagina L, Larson E, and Friedman C. Extracting information on pneumonia in infants using natural language processing of radiology reports. J Biomed Inform 2005: 38(4): 314-321.
[19] Huang Y, Lowe HJ, Klein D, and Cucina RJ. Improved identification of noun phrases in clinical radiology reports using a high-performance statistical natural language parser augmented with the UMLS specialist lexicon. J Am Med Inform Assoc 2005: 12(3): 275-285.
[20] Imai T, Onogi Y. Extracting diagnosis from Japanese radiological report. AMIA Annu Symp Proc 2003: 873.
[21] Mamlin BW, Heinze DT, and McDonald CJ. Automated extraction and normalization of findings from cancer-related free-text radiology reports. AMIA Annu Symp Proc 2003: 420-424.
[22] Marko K, Schulz S, and Hahn U. MorphoSaurus--design and evaluation of an interlingua-based, cross-language document retrieval engine for the medical domain. Methods Inf Med 2005: 44(4): 537-545.
[23] Marko K, Schulz S, and Hahn U. Automatic lexicon acquisition for a medical cross-language information retrieval system. Stud Health Technol Inform 2005: 116: 829-834.
[24] Marko K, Daumke P, Schulz S, and Hahn U. Cross-language MeSH indexing using morpho-semantic normalization. AMIA Annu Symp Proc 2003: 425-429.
[25] Schulz S, Hahn U. A knowledge representation view on biomedical structure and function. Proc AMIA Symp 2002: 687-691.
[26] Schrader U, Zaiss A, Büchele S, Binder M, Leupolz U, Klar R. Realization of a CPR using HL7 Queries from Communication Server to Information Servers. Medical Infobahn '99. Amsterdam: IOS Press, 1999: 834-837.
[27] Apache Lucene: http://lucene.apache.org/, last accessed: 2006-12-04
[28] PHP: www.php.net, last accessed: 2006-12-04
[29] Scheinfeld N. Photographic images, digital imaging, dermatology, and the law. Arch Dermatol 2004: 140(4): 473-476.

Address for correspondence

Dr. M. L. Müller, Department of Dermatology, University Medical Center Freiburg, Hauptstr. 7, 79104 Freiburg, Germany.
Phone: +49-761-270 6701, Fax: +49-761-270 6829,
E-Mail: marcel.mueller@uniklinik-freiburg.de

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

The Swedish National Pharmacy Register

Bengt Astrand[a], Bo Hovstadius[b], Karolina Antonov[c], Göran Petersson[d]

[a] *Apoteket AB and School of Pure and Applied Natural Sciences, University of Kalmar, Kalmar, Sweden*
[b] *Öhrlings PriceWaterhouseCoopers, Uppsala, Sweden*
[c] *The Association of Pharmaceutical Industry, Stockholm, Sweden*
[d] *E-Health Institute, School of Human Sciences, University of Kalmar, Kalmar, Sweden*

Abstract

*To achieve a safer future prescribing, the Swedish government has introduced a mandatory registration of all drugs dispensed at pharmacies. The medication history in the register may be accessed online by registered individuals, prescribers and pharmacists. After 15 months of action, the prevalence of individuals with dispensed drugs in the Swedish population was 71.0% (6,424,487/9,047,752); women 78.8% and men 63.1%. The incidence rate for individuals with dispensed drugs was estimated as 12.4 (1,000*111,960/9,047,752) per month and 1,000 inhabitants. The mean number of dispensed prescriptions was 12.1 (median 6, Q1-Q3 2-15) per individual. For the elderly (age group 80-89), the mean number of dispensed prescriptions was 27.8 during the study period (median 24, Q1-Q3 13-38); women 28.8 and men 26.1.*

When introducing a National Pharmacy Register, containing personal drug information for the majority of the population, issues on security, confidentiality and ethics have to be taken into consideration. The lack of widespread secure digital signatures in health care may delay general availability. To clinically evaluate individual medication history, the relatively high prevalence of dispensed drugs in the population, seems to justify the National Pharmacy Register.

Keywords:

decision making, computer-assisted, drug utilization, incidence, internet, pharmacoepidemiology, prescriptions, drug, prevalence

Introduction

Since many patients have several health care contacts concomitantly at different institutions, physicians are often reduced to prescribing potent drugs unaware of each other's prescriptions. This might influence the effect of the therapy in an uncontrollable manner. Moreover, the polypharmacy, with accompanied risk for drug interactions [1-4], may be so profuse that a low degree of compliance to the prescribed therapy has been named 'intelligent non-compliance', indicating that it may be beneficial for patients not to adhere to the prescribed therapy [5].

One approach to improve the situation for prescribers has been to store information on the *prescribed* therapy in databases at institutions or within larger regional organisations, mostly as part of an electronic health care record. When prescribing, the physician will be able to check the patient history for polypharmacy, interactions and different kind of contraindications. Automated systems for interaction alert may also be integrated in these systems. Results of the alerts may be to refrain from prescribing, adjusting the dose or monitoring the plasma concentration of the drug. Discontinuation of previous prescriptions may also be relevant.

Another approach to provide prescribers with medication history is the new legislations in Denmark and Sweden allowing nationwide databases with information on *dispensed* prescriptions at the pharmacies. The information is accessible to the patients, and with the patients' conditioned consent, to prescribers and dispensing pharmacists (Fig.1). These databases comprise a large proportion of the national population and are subject to strict security regulations to ensure individual confidentiality.

In Sweden, all prescriptions dispensed at pharmacies, with the exception of drugs dispensed for in-patients at hospitals and OTC (Over-The-Counter) sales have been stored in a National Pharmacy Register since July 2005 [6-7]. Use of the medication history in this register is, by a special law, restricted to clinical use and documentation;

- achieving a safer drug prescribing for the registered person in the future,
- providing the registered person care or treatment,
- supplementing the registered persons health care record,
- assisting dispensing pharmacists and
- facilitating the registered persons drug utilization.

Aim of the study

The present study was conducted in order to describe the information content of the National Pharmacy Register.

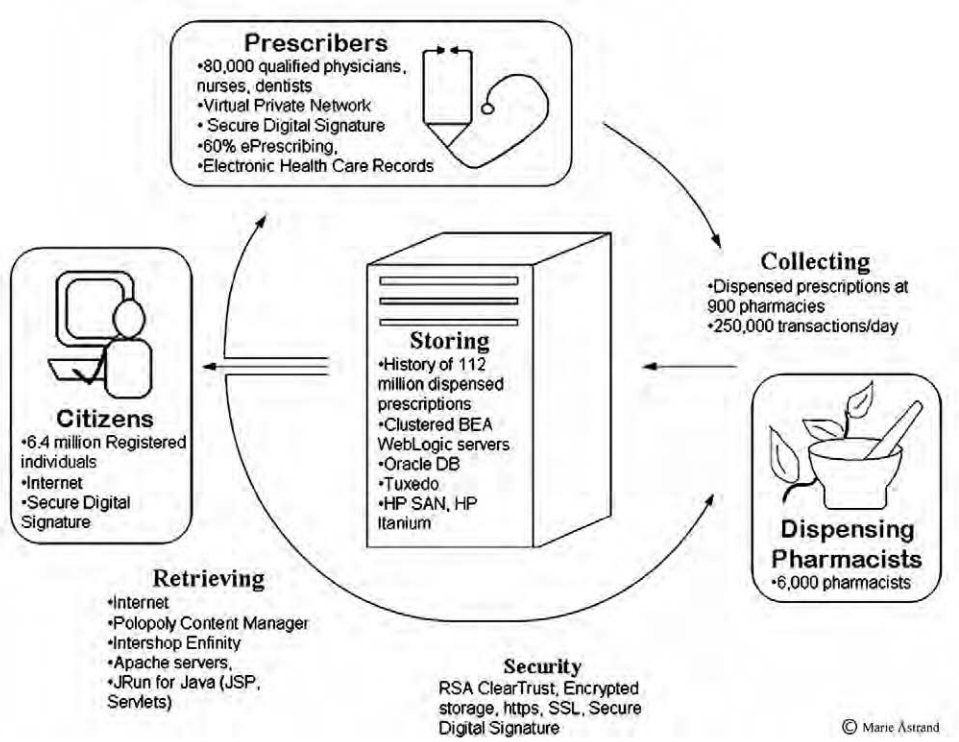


Figure 1 - Information model applied to the National Pharmacy Register

Materials and methods

The National Pharmacy Register

The National Pharmacy Register is individual-based and contains data from all dispensed out-patient prescriptions at all Swedish pharmacies from July 1, 2005, including multi-dose dispensed prescriptions and legal internet sales. Data collection from about 900 pharmacies and the National Pharmacy Register is administered by the National Corporation of Pharmacies, Apoteket AB [8]. The registered data is available for the registered individual at the pharmacy desk, with valid identification, and on the internet, with a secure digital signature. After conditioned consent from the registered individual, the register will be accessible on-line for prescribers and dispensing pharmacists. The registration is mandatory and includes the name and the personal identification number (social security number) of the registered individual along with day of dispensing, drug name, prescribed amount and dosage (Table 1). The information is stored in the register for a period of 15 months and thereafter cleared.

Table 1 - Information content in the national pharmacy register

Object	Variable
Patient	**Name** **Personal identification number***
Drug	**Name** **Amount** **Dose**
Pharmacy	**Data of dispensing**

All individuals (outpatients) filling prescriptions at Swedish pharmacies during the first 15 month period were included in our study. The measures used were number of individuals, number of dispensing dates and number of drugs. Epidemiological measures were used and defined; *prevalence* - the proportion of individuals filling prescriptions in the Swedish population during 15 months; *incidence* - the total number of new individuals filling prescriptions in the Swedish population per month, after a wash-out period of 12 months; *incidence rate* - number of new individuals filling a prescription*1,000/number of individuals at risk (the entire Swedish population); relative risk, RR – the ratio between risks in two groups for filling a prescription during the 15 month study period.

Calculation of sums, frequencies and ratios were aggregated and compiled from production data of the source database. All data processing in our study were done anonymously, without the personal identification number; only gender and year of birth, originally embedded in the personal identification number, were used. The study population was stratified by gender and age (10-year classes) on July 1, 2005. Results were compared to statistics on the number of individuals per gender and age group in the Swedish population on December 31, 2005 [9].

Results

After the first 15 months (July 2005 to September 2006) the register contained 6,424,487 individuals of the Swedish population of 9,047,752 inhabitants. Thus, the prevalence of individuals with dispensed drugs in the entire Swedish population was 71.0% (women 78.8%, men 63.1%) (Fig.2). The prevalence of individuals with multi-dose dispensed drugs was 2.4% (213,395/9,047,752). During the first month, 30.3% (1,946,885/6,424,487) of the registered individuals filled a prescription at a Swedish pharmacy. After four months, 2/3 (4,328,627/6,424,487) of the individuals were registered in the database (Fig.2). After a 12-month wash-out period, the incidence was estimated as 111,960 new individuals per month (Fig.3). Hence, the incidence rate of dispensing drugs on prescriptions per month and 1,000 inhabitants was estimated as 12.4 (1,000*111,960/9,047,752).

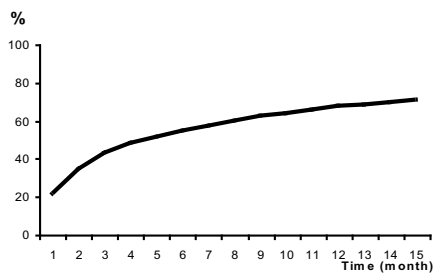


Figure 2 - Prevalence of individuals with dispensed drugs in the National Pharmacy Register during the first 15 months

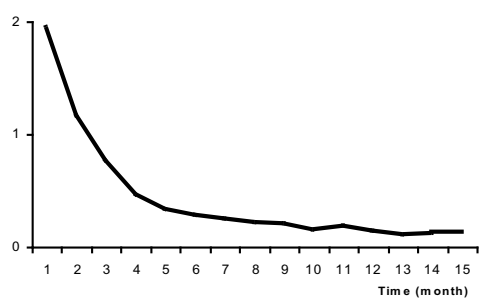


Figure 3 – Number of new individuals in the National Pharmacy Register during the first 15 months.

Most individuals in the register were women (56.0% women, 44.0% men), with a relative risk of 1.25 ((3,595,152/4,561,202)/(2,829,335/4,486,550)). The prevalence of individuals with dispensed drugs showed a positive linear relationship with age, with a higher prevalence for the elderly. Above 70 years of age, the prevalence was 92.1%, 94.9% and 94.3% in the respective age groups. For children 0-9 years of age, the prevalence was 62.8% (women 61.2%, men 64.3%). For women, the prevalence was 77% or more, for all age groups 20 years and above (Fig.4)

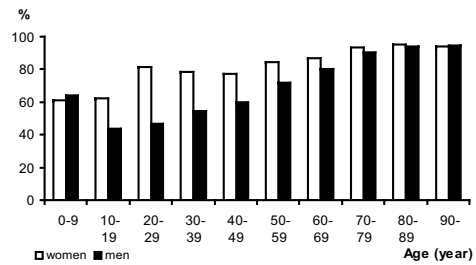


Figure 4 - Prevalence of individuals with dispensed drugs according to age groups and gender in the National Pharmacy Register during the first 15 months

During the 15 month study period the mean number of dispensed prescriptions was 12.1 (median 6, Q1-Q3 2-15) per individual, multi-dose dispensed prescriptions excluded; for the elderly above 70 years of age the mean number of prescriptions was 25.1 (median 20, Q1-Q3 10-34), 27.8 (median 24, Q1-Q3 13-38) and 24.8 (median 21, Q1-Q3 10-35) in the respective age groups, for children 0-9 years of age, the mean number of dispensed prescriptions was 3.7 (median 2, Q1-Q3 1-4). The overall mean number of dispensed drugs at each pharmacy visit (same date) was 1.88 (76,004,043/40,445,263), multi-dose dispensed prescriptions excluded.

Discussion

In the present study, less than 30% of the Swedish population did not fill any prescription, revealing that more than 2/3 of the population filled at least one prescription, during the study period of 15 months. The prevalence of dispensed prescriptions must be regarded as relatively high, since for elderly it was above 92% and for women over 20 years of age the prevalence was never below 77%. Even for children, a high prevalence of 63% was observed. The elderly filled more than twice as many prescriptions as the general population (24.8-27.8 vs. 12.1). Of those who filled a prescription, filled 15 or more prescriptions. As a rule, with the Swedish reimbursement model, the filling of one prescription covers consumption for a three-month period. We did not evaluate to what extent the prescriptions were intended for continuous use, rather than for a shorter treatment period. The relative risk of 1.25 for women may be explained by women's use of sex hormones and modulators of the genital system [7].

The prevalence and incidence of dispensed drugs may be viewed as indicators of the health status of a population.

However, drugs are not only used for curing and alleviating diseases but also for prevention and, more rarely, diagnosing. Moreover, we did not measure the actual consumption of the drugs, only the prescriptions dispensed at the pharmacy. The register does not contain information on OTC sales or drugs dispensed for inpatients in hospitals and institutions, with the exception of multi-dose dispensed drugs at pharmacies.

The Swedish health care system is tax financed and readily available for all and the patients are free to consult an unlimited number of physicians. To clinically evaluate individual medication history, the relatively high prevalence of dispensed drugs in the population, seems to justify the National Pharmacy Register.

Prescribed, dispensed and consumed

The patients' degree of *adherence* to the doctors' prescribed therapy has been described as *compliance* or more recently as *concordance*, indicating a more inclusive and less paternalistic view of the doctor-patient relationship [10].

Prescribing is not always equivalent to the amount of dispensed drugs or the actual drugs consumed by the patient. If the patient does not fill the prescription at a pharmacy a so-called *primary non-compliance* occurs. As a result of experienced or imagined side effects or otherwise unfulfilled patient expectations, the patient may choose, or just forget, not to continue a medication, although the prescriptions have been presented to a pharmacy and subsequently filled. This is referred to as *secondary non-compliance* and may result in early cessation of therapy, prescriptions not refilled or simply a reduced intake of the amount of the medication. Thus, as a measure of true drug consumption, prevalence and incidence estimates in the present study may be both over- and underestimated. Some of the filled prescriptions will never be used by patients, resulting in an overestimation. In contrast, additional sources like OTC sales, herbal remedies, in-hospital medications, and illicit internet sales will result in an underestimation of the true exposure.

Clinical aspects

In order to reduce polypharmacy, drug interactions and side-effects, the expected clinical benefits of the National Pharmacy Register would be to provide a correct and updated medication history for all patients, readily available whenever needed, and independent of who has prescribed the therapy. To what extent does the National Pharmacy Register fulfil these expectations? As a source for a national medication history, the register seems to be rather complete, exceeding 2/3 of the entire national population. While pilot tests are in an early phase, clinical experiences are anecdotic and not yet systematically evaluated. Lack of widespread secure digital signatures in health care may, however, delay general availability. The usability would also be improved if the information from the register were available within the framework of systems for electronic health care records, a technology still under construction.

Epidemiological research

The register is intended for prescribers, pharmacists and the registered individuals for clinical use and documentation. The equivalent information is also collected and distributed to the Swedish National Board of Health and Welfare for statistical, epidemiological and scientific purposes, stored without any time restrictions, with new opportunities to explore drug and disease associations and the risks, benefits, effectiveness and health economical effects of drug use [11].

Quality and security

The development in Sweden of medical information systems for the physician office launched the first electronic transfer of an ePrescription in 1983 [12-13]. In December 2006, 62% (intra-county range 33-81%) of all prescriptions were transferred as ePrescriptions, with a steadily growing trend. Thus, the prescriber is entering prescription information into electronic health record systems that will subsequently be used, not only for the dispensing of new prescriptions and refills, but also as an information source for future prescriptions by other prescribers, which is one of the main objectives of the National Pharmacy Register. In order to prevent medical errors, it will be of utmost importance to monitor and improve the quality of prescriptions.

For authorization of prescribers a secure digital signature is required; in Denmark by law and in Sweden by regulation of authorities. Until now, these were not readily available in health care and will present a constraint for making the register available to all prescribers in the near future. The register is protected against unauthorized access by laws, rules, written agreements, authorities' inspection, technical and physical procedures. The data collection procedure is presently done batch-wise overnight, but should be altered to online and momentarily, ensuring constantly updated and accurate information in the register.

Ethical considerations

Individuals may benefit from the register being complete and correct at all times but may at the same time be cautious about the confidentiality of their personal health information. To solve this dilemma, the approach in the national pharmacy dispensing registers in Denmark and Sweden, has been to mandatory collect and register the information in the national database, but also to give the patient the right to restrict the accessibility of the information to certain individual health care professionals. In addition, the patient is offered full transparency to who has accessed the information. According to national laws, unauthorized access might be subject to a lawsuit. As an exception, prescribers may, in emergencies, have access to the register without a conditioned consent.

Future improvements and studies

If feasible, it would be beneficial to include all drugs in the register, e.g. OTC sales and drugs used for inpatients in hospitals. Also, to increase the clinical usability of the information in the register, information on the actual prescriber of each prescription should be included.

A coordinated nationwide pilot plan is presently being conducted within several Swedish health care organisations and pharmacies. For the registered individuals' access to the register on the Internet, about 40,000 accesses have been registered (March – December, 2006). Evaluations are conducted by the e-Health Institute, University of Kalmar, Sweden.

The National Pharmacy Register has a clinical objective, namely to improve drug utilization. Future studies should evaluate the availability and the deployment in health care as well as to what extent the intended benefits have been reached.

The results in our study raise many research questions on the utilization of drugs on a population level. For further studies the pharmacoepidemiological database at the National Board of Health and Welfare opens new prospects for researchers.

The register can also serve as a model for coming health care services for individuals on the Internet. The Swedish government has published a National Strategy for eHealth, with the objective to facilitate access to information across organisational boundaries; making information and services easily accessible to citizens [14].

Conclusions

The Swedish National Pharmacy Register provides prescription dispensing information for the majority of the population. The medication history in the register may be accessed online, to improve drug utilization, by registered individuals, prescribers and pharmacists in a safe and secure manner. Lack of widespread secure digital signatures in health care may delay general availability. To clinically evaluate individual medication history, the relatively high prevalence of dispensed drugs in the population, seems to justify the National Pharmacy Register.

Acknowledgements

We are indebted to Marie Åstrand for illustration of Figure 1 and to Mats Wickbom for help with statistical procedures. The study was supported financially by Apoteket AB, including support to the e-Health Institute.

References

[1] Bergendal L, Friberg A, and Schaffrath A. Potential drug-drug interactions in 5,125 mostly elderly out-patients in Gothenburg, Sweden. Pharm World Sci 1995: 17 (5): 152-7

[2] Bjerrum L, Andersen M, Petersen G, and Kragstrup J. Exposure to potential drug interactions in primary health care. Scand J Prim Health Care 2003: 21 (3): 153-8

[3] Bjorkman IK, Fastbom J, Schmidt IK, and Bernsten CB. Drug-drug interactions in the elderly. Ann Pharmacother 2002: 36 (11): 1675-81

[4] Merlo J, Liedholm H, Lindblad U, Bjorck-Linne A, Falt J, Lindberg G, Melander A. Prescriptions with potential drug interactions dispensed at Swedish pharmacies in January 1999: cross sectional study. Bmj 2001: 323:427–8

[5] Gross ME. A self-medication administration program. J Geriatr Drug Ther 1992: 6(4), 879-89.

[6] Swedish Ministry of Health and Social Affairs Improved patient safety within the pharmaceutical arena. [In Swedish] Stockholm: Swedish Ministry of Health and Social Affairs, 2004. URL: http://www.regeringen.se/sb/d/108/a/ 38890 [accessed March 8, 2007]

[7] Åstrand B, Åstrand E, Antonov K, and Petersson G. Detection of potential drug interactions – a model for a national pharmacy register. Eur J Clin Pharmacol 2006: 62: 749-56

[8] Westerlund T, Björk T. Pharmaceutical Care in Community Pharmacies: Practice and Research in Sweden. Ann Pharmacother 2002: 40(6): 1162-9

[9] Statistics Sweden. Population in the whole country, counties and municipalities on Dec 31, 2005. URL: http://www.scb.se/templates/tableOrChart____159262.asp [accessed March 8, 2007]

[10] Dickinson D, Wilkie P, Harris M. Taking medicines: concordance is not compliance. Bmj 1999; 319:787.

[11] Wetterberg B, Hammar N, Fored CM, Leimanis A, Otterblad Olausson P, Bergman U, Persson I, Sundström A, Westerholm B, and Rosén M. The new Swedish prescribed drug register – opportunities for pharmacoepidemiological research and experience from the first six months. Pharmacoepidemiol Drug Saf 2006: Published online DOI: 10.1002/pds.1294 (In Press).

[12] Nilsson S, Ockander L, Dolby J, _strand B. A computer in the physician's consultancy. In: van Bemmel J, Ball M, Wigertz O, eds. MEDINFO-83. IFIP-IMIA, North-Holland Pub.Co, 1983; pp. 1185-6

[13] Åstrand B, Doctor's use of VDUs for medical prescriptions In: Peterson HE, Schneider W, eds. Human-Computer Communications in Health Care, IFIP-IMIA. Elsevier Science Publishers B.V., North-Holland, 1986; 267-9

[14] The National High-Level Group for e-Health. National Strategy for eHealth – Sweden. Stockholm: Swedish Ministry of Health and Social Affairs, 2006. URL: http://www.sweden.gov.se/content/1/c6/06/43/24/f6405a1c.pdf [accessed March 8, 2007]

Address for correspondence

B. Astrand,
Apoteket AB,
Box 873,
SE-391 28, Kalmar, Sweden,
Email: bengt.astrand@apoteket.se

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

The Use of Existing Low-Cost Technologies to Enhance the Medical Record Documentation Using a Summary Patient Record [SPR]

Dr S. Bart[a], Dr T.Hannan[b]

[a,b] *Department of Medicine, Launceston General Hospital, Launceston, Tasmania, Australia*

Abstract

The Institute of Medicine has described the Electronic Medical Record [EMR] as an essential technology for health care that improves patient safety and the quality of care when compared to traditional paper-based records. [1,2,3] Despite major financial expenditures on health information technology related to EMRs in developed countries such as North America, Britain and Australia, success rates for successful implementations have been low. One specific domain of information management relates to the communication of health care between those involved in the care process. Summarization of patient histories is a core component of EMR systems and assists in the communication of health care. This paper demonstrates how the utilization of simple technologies, can lead to the successful implementation of the Summary Patient Record [SPR] component of an EMR system.

Keywords:

summarization, communication, technology

Introduction

In 1991 the Institute Of Medicine [IOM] published a review stating that computer-based technologies are the essential technologies for improving health care delivery. [1] Subsequent publications from the IOM on aspects of health care delivery and the quality of care, affirm the critical role of the Electronic Medical Record [EMR] in reforming the health care processes. [2, 3]. No clear definition of the EMR exists but it is essentially a pre-birth to post death electronic record which meets the record-keeping requirements for any setting -whether intensive care or primary care, whether obstetrics or gynecology.

Despite this knowledge on the importance of the EMR, implementation of such systems has been slow. The urgency EMR implementation has recently been emphasised by the American Medical Informatics Association [AMIA] [4]. This organisation has devised a plan to facilitate the implementation of a national EMR in the United States. [4] The impetus for this plan is the continuation of poor quality of care, adverse patient outcomes and expanding health costs. [2, 3, 5]

Review of the Integrated Area [Advanced] Information Management System [IAIMS] by Stead in 1997, showed that implementers of health information systems were more concerned about the cost of computer resources than information management tools. [6] He indicated the issues we needed to address relate to our current costs of information management and how can we obtain timely, reliable and complete clinical information? [6]

EMR definitions describe core functional elements, and one of these is Summarization, [creation of a summary of the patient's medical record] as a tool for effectively communicating health care between professionals. Benefits of Summarisation of the EMR have been shown by Whiting-O'Keefe in 1980 and Fries in 1984. [7, 8]

Utilising the experiences of one of the authors [TH] in the successful implementation of an electronic medical records in resource poor and developed nations, [9, 10, 11] we decided to explore the feasibility of implementing a Summary Patient Record [SPR] in an ambulatory Diabetic Clinic in a regional hospital in northern Tasmania with the aim of improving the documentation and communication of patient care information.

Materials and methods

Setting

The Launceston General Hospital [LGH], is a 300 bed regional teaching hospital in northern Tasmania. The Diabetic Clinic sees approximately 50-100 patients per week. Clinic staffing is by a Specialist, training registrars, diabetes educators, social worker, nurses etc.

Information use prior to the study

The institution does use an EMR. It has a robust intranet infrastructure that links health care institutions within the State of Tasmania. This system provides hospital-wide Internet access, email facilities and in each ward, doctors offices and nursing stations there is computer access to the internet and electronic laboratory data, radiology reports and in 2006 radiology images.

The ancillary patient data is recorded and stored electronically in separate non-communicating systems. All results are printed and stored in the paper record. Radiology has two separate systems, one, which stores digital images, and the other text reports on these images.

Current record documentation in the clinics is by dictated, free-text, typed letters with no clear structural continuity. These letters may be a brief summary of a given visit or an

attempt to repeat and record the whole of that patient's medical history.

The standard process for managing records in the Diabetic Clinic is independent of the SPR system, and all notes are hand written and the and used in conjunction with the current records. This usually includes the paper Diabetic Clinic Record and the separate previous paper In-Hospital Record. For a patient on dialysis or undergoing oncology therapy these records are not available in the Diabetic Clinic. Laboratory and text Radiology reports are available on-line and in printed reports. In general the printed formats are used to assess the patient's results.

Normally, the patient is seen by the clinician who dictates a letter which is delivered to and typed in a typing pool, approximately 2-3 weeks later the letter is returned for editing and signing by the clinician. If corrections are required then the process is repeated. If correct, it is mailed to the referring practitioners. All correspondence is then integrated manually into the Clinic's paper file.

Methods

For this study we performed a prospective study of 100 consecutive patients in the Diabetic Clinic in the LGH. All patients had a SPR Template completed by one author [TH] for both initial and follow up visits. The SPR was stored in MS Word format on a password protected directory on the hospital network. The initial SPR Template was completed using the information recorded at the patient's last visit. This created the core historical SPR. Using this process we could verify the accuracy and content of the information and update any new clinical information that may have been recorded since the previous visit. Information relevant to the current encounter was then recorded in the SPR.

For an initial encounter the SPR Template was completed at the end of the visit and before the arrival of the next patient. If the visit was a follow up and the patient had an existing SPR, then the visit information and management was entered directly into the SPR with the patient in the clinic room. Each Follow Up encounter Table was inserted into the record using macro key functions [two key strokes]. The patient was then given a printed copy or had it emailed to them at the time of the visit. If the patient was travelling near to their Primary Care Practitioner then they would be offered the option of taking it directly to that person. If the clinician did not have an email address or chose not to receive an electronic format of the SPR, a hard copy was sent by post.

For Follow Up visits the record was also printed, posted or emailed in the same manner as the original with the current copy replacing all previous versions of the SPR. Records were emailed using the guidelines described by Delbanco. [12] All email communications between any of the recipients of the SPR were stored electronically in the follow up section by date and time.

Access to radiology or laboratory data during the patient encounter required a single click of the pre-defined SPR *Hyperlink* making these results available in seconds. Use of the Back arrow on the Toolbar returned the user to the contents of the SPR.

Some patients preferred to document and email their Finger Stick Blood Sugar Level [FSBSL] results in a Microsoft Excel file. These were stored in a separate directory under a unique identifier and accessed from within the SPR through the *Hyperlink* tool. One author [TH] had created the Excel with formulae so that all FSBSL readings were averaged over time as were the HbA1c readings. The FSBSL readings were emailed the clinician [TH] and all updated files were stored in a specific computer directory. The standard procedure for FSBSL assessment in the Clinic was a weekly [or longer] review of the FSBSL recorded on a paper from phone conversations and these recordings continue to be hand written.

Outcome measures

We chose the following parameters as outcome measures.

- Turnaround time of documents. This was the time taken to create and transmit the SPR to the relevant people involved in the care process. This included the time to create the SPR at the patient visit and the time taken to send to the appropriate referrers.
- Accessibility of laboratory and ancillary services information. This is the time taken to access the last known data recorded on the patient. We attempted measure the time it took to find the relevant information in the paper record compared to accessing the information on line from within the SPR.
- Accuracy and completeness of information. This based on the content and completeness of the information recorded at the last visit of the patient prior to the use of the SPR. Both authors independently screened the last communication within the paper record for allergies, medications, surgical and diagnostic procedures, the presence of duplicate records and for record validation through the clinicians' signature.
- Patient accessibility to their clinical record. Here we measured how many patients had a copy of any part of their clinical record either the SPR or original Clinic record.
- Timeliness of communication of relevant information. In this assessment we evaluated the availability of the record at future visits or patient encounters with the clinic and also the use of emailed FSBSL readings to the diabetic educators and clinicians.

Results

- Turnaround time of documents.

  The previous record documentation required a minimum of 2-3 weeks. With workloads on the typing pool or with inaccuracies in the typing requiring retyping, this turnaround time for record completion could take up to six weeks. Using the SPR initial history recording by one author [TH] took approximately 10 minutes [usually less] based on clock monitoring. The SPR is printed immediately, given to the patient, and sent

referrers. If email facilities were available the SPR was sent instantaneously to the relevant person.

For Follow Up encounters this process required less than five [5] minutes.

- Accessibility of laboratory and ancillary services information.

  If available the paper-printed reports from laboratories, radiology and other ancillary support systems were filed in the patient's paper folder. Alternatively, Laboratory and Radiology results are accessible in seconds from within the SPR using the *Hyperlink* function.. This occurred during the patient encounter, and did not require sorting through pages of results, and the results could be discussed with the patient while visualising the screen information together. Return to the SPR record required only a click on the Back arrow button on the control panel. If results were required to be stored within the SPR then the *Cut & Paste* function in Word could be used.
- Accuracy and completeness of information.
- We evaluated this by having the authors [SB, TH] review the first 100 records independently using the last formal communication printed in the paper notes. The following list shows the percentage of each of these measurements and represents similar findings for each author. Formal statistical analysis of the evaluations was not performed.
  - Duplicate records-98%
  - Incomplete medication list-63%
  - Incomplete diagnosis-68%
  - Incomplete documentation of procedures-88%
  - Incomplete allergy list-86%
  - Unsigned document-50%
- Patient accessibility to their clinical record.
- Prior to the use of the SPR a patient was rarely provided access to their medical record. In our study every patient was provided with a printed or emailed copy of their SPR at each visit. Updated records were printed or emailed and replaced the previous versions of the document. We found that 35% of patients used email for this communication and 17% of these were over 65 years of age. All these patients chose to receive their SPR by this form of communication. Any comments or corrections made by patients on their record were stored as text in the last Follow Up table. The only negative comments came from a Primary Care practitioner who did not like the format of the SPR.
- Timeliness of communication of relevant information.
- The turnaround for record documentation and distribution time was dramatically shortened from weeks to minutes. 100% of patients had a copy of their record always available immediately. The SPR in electronic format was immediately available for review by authorised persons or transmission to another location e.g. the hospital or to the individual patient if they were travelling long distances. If the patient was subsequently admitted then a copy was in the existing Diabetic Clinic record. Each patient was encouraged to carry the record with them or bring it when admitted. The need for redundancy of copies of the SPR in the notes was because this record format was new and those with access to the patient's record were not familiar with this new format of record documentation. Reports from laboratories and radiology were available from any terminal in the hospital for those who had authorised access to the SPR.

  The emailing of FSBSL readings they would save up to 5 hours in waiting time to make phone contact with the Diabetic Educators in the clinic. We did not evaluate the effect on Diabetic Educators time. Some Educators preferred to use the older hand written charts and record the readings over the phone stating they could interpret the data better in the older format rather than the new graphical displays. Emailed FSBSL measurements were evaluated at the time they were received and a response [included in the SPR] was emailed back to the patient and updated in the existing paper record.

Discussion

The availability of timely, reliable, complete health care information within the medical record is fundamental to good clinical care. The use of paper-based systems in modern health care makes this virtually impossible. [1, 2, 3] Since 1994 it has been shown that the current forms of communication between hospitals and Primary Care practitioners is poor and of low quality. [13]. There has also been a greater focus on technology costs than on information management costs. [14]

Hannan and Fraser have demonstrated that even in resource poor environments complex clinical information systems can be implemented with beneficial outcomes. [9, 11]

Through the use of standard, affordable technologies such as word processing, email and internet connectivity, we have demonstrated that a low technology effective SPR can be implemented in resource rich environment. The SPR in the LGH provides timely, reliable and legible information on patient care. Patients have now become more involved in their health care with up to date information. This fits with the philosophical approach outlined by Weed in 1989. He stated that patients "*must be given the right tools to work with. They are the most neglected source of better quality and savings in the whole health care system. After all, they are highly motivated, and if they are not, nothing works in the long run anyway. They do not charge and they even pay to help. Also there is one for every member of the population.*"[15]

The SPR described here is now used in a General Medicine clinic [TH], another General Physician is trialling it as a model for his records, Outpatient Clinic nursing staff anecdotally report that they have time savings of up two hours per nurse per clinic per day by using this SPR. Inpatient discharges are more timely and accurate on specific patients [TH]. The SPR is used for peri operative manage-

ment for co-morbid states such as diabetes, venous thromboembolism and hypertension. It is hoped that this SPR will become a model for the hospital. Its adaptation with internationally standardized data formats such as ICD10, SNOMED and LOINC, will facilitate data capture for patient care and research.

References

[1] *The Computer-based. Patient Record: An Essential Technology for Health Care.* Washington, DC: National Academy Press, 1991.

[2] Institute of Medicine. *To Err Is Human: Building a Safer Health System*. Washington, DC: National Academy Press, 1999.

[3] *Institute of Medicine. Crossing the Quality Chasm: A New System for the 21st Century.* Washington, DC: National Academy press.

[4] Cooper GF. J Am Med Inform Assoc. 2004 Oct 18; Defining a Workable Strategy to Stimulate Widespread Adoption of Electronic Health Records in the United States.

[5] *National Health Information Infrastructure Act 2003* Nancy Johnson US Congress.

[6] Stead WW. The evolution of IAIMS: Lessons for the next decade. JAMIA 1997; 4, March-April Supplement:S4-S9

[7] Whiting-O'Keefe QW Simborg DW Epstein WV Medical Care 1980;18:842-852.

[8] Fries. J. Alternatives in medical record formats. Medical care. 1984; 12:871-881.

[9] Hannan TJ, Tierney WM, Rotich JK *et al*. The MOSORIOT medical record System [MMRS] phase I to phase II implementation: an outpatient computer-based medical record system in rural Kenya. Medinfo 2001; 10[Pt 1]: 619-22.

[10] McDonald CJ, Overhage JM, Tierney WM, Dexter PR, Martin DK, Suico JG, Zafar A, Schadow G, Blevins L, Glazener T, Meeks-Johnson J, Lemmon L, Warvel J, Porterfield B, Warvel J, Cassidy P, Lindbergh D, Belsito A, Tucker M, Williams B,Wodniak C. The Regenstrief Medical Record System: a quarter century experience. Int J Med Inform. 1999 Jun;54[3]:225-53.

[11] Fraser HS, Jazayeri D, Nevil P, Karacaoglu Y, Farmer PE, Lyon E, Fawzi MK, Leandre F, Choi SS, Mukherjee JS. An information system and medical record to support HIV treatment in rural Haiti. BMJ. 2004 Nov 13;329[7475]:1142-6.

[12] Delbanco T, Sands DZ. Electrons in flight-e-mail between doctors and patients. N Engl J Med. 2004 Apr 22;350[17]:1705-7.

[13] P.J. Branger, J.S.Duisterhout. Communication in Health Care.JAMIA.1994; 69-77.

[14] Prof. Safran C. Keynote Address AMIA Spring Congress. San Francisco. May 1994.

[15] Weed LL. New premises and new tools for medical care and medical education. Methods Inf.Med.1989. Nov; 28[4]: 207-14

Address for correspondence

Dr T. J. Hannan FRACP, FACHI,
FACMI terry.hannan@dhhs.tas.gov.au

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

A New Approach in Nursing Documentation: Community Nursing Case

Uros Rajkovic[a], Olga Sustersic[b], Vladislav Rajkovic[a], Vesna Prijatelj[c]

[a] *Faculty of Organizational Sciences,University of Maribor, Slovenia*
[b] *College of Health Studies, University of Ljubljana, Slovenia*
[c] *Clinical Centre Ljubljana, Slovenia*

Abstract

This paper presents the development of electronic documentation for community nursing using a system approach. Documentation is viewed as an information model for organizing and managing processes. The community nurse plans the nursing process after gathering and evaluating information on the patient's health and his/her family status. Documentation is thus considered to be a basis for the successful work of the health team and as a way of ensuring quality in nursing. The paper describes a prototype software model for e-documentation in community nursing together with its evaluation in practice.

Keywords:

information system, nursing, community nursing, documentation, software solution

Introduction

A system approach in the organizational and informational context brings specific challenges in terms of the complete management of complex systems. Systems in healthcare, of which nursing is a part, similarly belong within this framework [1, 2]. Along with the system approach, it makes sense to use the potentials of contemporary information and communications technology (ICT) and to study the possibility of adding value in managing complex systems, especially in terms of the effective use of resources and quality assurance.

E-documentation of any process is an information model, which uses ICT for organizing and managing the process according to established goals. Nursing documentation consists of patient and family data. Nurses use these data to plan the nursing process, which in short covers assessing patient's nursing problems, making nursing diagnoses, implementing nursing interventions and evaluating the work [2, 3]. The documentation of the nursing process is the basis for the successful work of a nurse, and also represents an element of quality assurance in nursing [1, 2].

We wish to propose a model that will serve for the reengineering of classical documentation into e-documentation. With a suitable object-oriented organization and use of contemporary ICT it is thereby possible to achieve a higher level of quality especially in regard to integral treatment of the patient. The active computer model itself supports the work of the nurse and, at the same time, reduces the possibility of mistakes at work.

This paper is based on findings and models developed within the framework of a Project for preparation of a model tool for establishing quality with the aid of documentation in nursing at the Ministry of Health of the Republic of Slovenia. Below are presented the elaboration and implementation of the proposed model, as well as testing of the prototype software for community nursing.

Analysis of existing documentation in nursing

Using the survey research we first analyzed the current state of documentation in nursing in selected health organizations in Slovenia. The sample included three old people's homes, Ljubljana Health Centre with five units, Clinical Centre Ljubljana and Maribor General Hospital. We distributed 386 questionnaires, of which 286 (response rate of 74.1%) were returned.

From the results of the survey on the use and suitability of nursing documentation we can conclude both the actual state of the documentation itself and the process of documenting, and also the perception (opinions, considerations) of existing problems and possible solutions on the part of those surveyed.

The majority of documents (86%) are prescribed on the level of the institution. The only exceptions are community nursing and old people's homes. Documentation for community nursing is prescribed and unified throughout the country, while old people's homes have a uniform computer supported information system. Rather less than 13% of documents are computer supported. Among the types of documents, the following were most frequently listed: nursing care plan, referral/discharge document, continuation notes and variance report, admission document and report on an undesired event.

It can be concluded that those five most frequently used documents should be unified firstly, taking into account the specificities of individual services. Given that with contemporary ICT we can generally provide effective support to documentation and increase the use of computers.

From the perspective of content, a process method of work is only used in 32% of nursing documentation. Over 52%

use only a fragmented process approach. It appears that existing documentation is to a large extent at fault for this, since the majority uses only those elements that the documentation enables. It is therefore sensible to reengineer the documentation in a way that will enable documenting all phases of the nursing process.

Minimal data set on patient are recorded by three quarters of survey participants. One of the reasons that the percentage is not higher is unsuitability of existing documentation.

Discussions with the patient, observation of the patient and measurements are sources of data for completing documentation in more than half of cases. Slightly less than half have also stated nursing documentation as a regularly used source of data. It is sensible to consider links between other health documentations and nursing documentation.

In the opinion of the surveyed nurses, they see the purpose of documentation or documenting mainly in the continuity of nursing, security for members of the nursing team and patient and an account of the work of individual members of the nursing team. The content thus supports the work, with emphasis on the legal security of members of the nursing team and the patient.

Among reasons for the non-use of nursing documentation, according to a quarter of nurses, are understaffing and insufficient knowledge of the nursing process, and among unspecified reasons, the fact that existing documentation is unsuitable was most often noted.

In terms of the influence of nursing documentation, the following are highlighted: the quality of nursing, uniform doctrine of work and reducing the possibilities of mistakes. With improved documentation we expect most changes in the quality of collaboration inside the health team and in the distribution of work and responsibilities among nurses and other health team members.

The results have shown that reengineering documentation using ICT can and should positively influence on the quality of nursing care. Because of unified documentation in the community nursing we have decided to begin the reengineering of documentation in community nursing.

Process method of work in nursing

The basis for developing e-documentation is the nursing process. Figure 1 shows a schematic presentation of the process method of work in the IDEF1 standard. The division in the figure differs from the literature [1]. A major difference is in the stage of evaluation due to standardization restriction. IDEF1 standard does not allow any process to appear in the scheme more than once. There is also a difference due to the cybernetic feedback loops, which are of crucial importance from a systemic point of view for system management, in this case of nursing.

The user interface of the prototype supports this process method in the nurse's job sequence. Only a few elements must be added, which are specific for community nursing [4]. These are elements such as entering referrals for community care visits to patients or families and for planning dates of home visits. Home visits can only be planned on the basis of referrals received from the general practitioner and on the basis of instructions for implementing community nursing. Later on the same steps apply for each home visit as in the already mentioned process scheme.

Database model

The base for a software solution is a database that enables data archiving and accessing data. Critical analysis of nursing documentation was a starting point in the database design process [5].

In paper form, the documentation is often mainly unstructured. Thus words in sentences can be entered. A problem occurs when seeking data in a longer text. The legibility of the writing often presents additional difficulty. Similarly, the statistical processing of data for research, education and other purposes can become unreasonably difficult [6, 7].

The higher level of data structuring in electronic form enables a higher usability of acquired data. The nurse is reminded with the entry fields, of all data that are desired in the documentation. The nurse is also forced to gather and record important data in the compulsory fields. This results in electronic form of data that enable electronic processing.

Additional fields for entering comments and data that were not envisaged in the original structure are also important. These fields serve in the prototype solution also as information for further development – which data must be additionally structured for electronic processing.

For ease of overview, we have grouped similar data according to their semantic relations. Tables 1 and 2 show tree structured data for describing a patient and a family.

A relational database suitable for storing data in electronic form was developed. It enables simple entry of data, reduces duplication of data and provides fast extraction [8]. The entity-relationship diagram of the proposed database is shown in Figure 2.

It is worth highlighting some particularities in the database diagram. The nursing diagnosis is directly bound to the subject of nursing. In the nursing diagnosis we record to which basic living activity (BLA) it is bound, and we are aware that after the evaluation of the nursing care plan, it may remains in the care plane throughout one or more of the following visits.

The subject of nursing can be a patient or a family. In the case of a visit to a family or a visit to a patient living with other family members, we also see a list of all family members who live together. From this list, we can access data on an individual member or on the whole family.

We see elements of the diagnostic therapeutic program as a list of previously determined nursing interventions, which must be carried out independently in relation to the established nursing problems, nursing diagnoses or nursing goals.

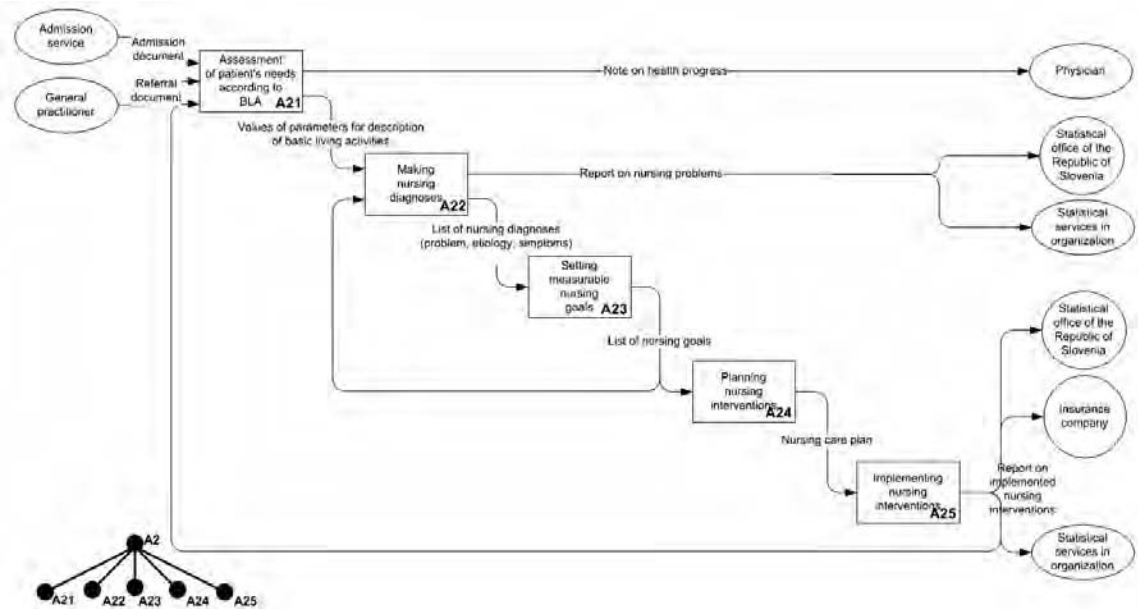


Figure 1- Schematic presentation of the process method of work in nursing according to the IDEF1 standard

In electronic documentation a user uses a password protected log-in. While it was necessary to sign some of documents in paper form, in electronic version data on the user are automatically recorded. For example, when the user records that an individual nursing intervention has been carried out, it is automatically recorded who entered the data for the individual activity and when.

Table 1 - Tree structure of data groups on a family

```
Family
├─General data
│  ├─Demographic data
│  └─List of family members
├─Home visits
└─Document of nursing care
```

Software solution

The steps that comprise the desired course of work of the community nurse (CN) required for each home visit are in accordance with the process method of work in nursing. When the CN selects a patient or a family and one of the planned home visits, a screen picture is shown for the individual home visit. In the upper part are shown data on the selected patient or family, and below the individual steps of the nursing process supported through four tabs: nursing anamnesis, assessment of patient's/family's need, planning and implementation. We will describe later how the evaluation phase is supported.

Table 2 - Tree structure of data groups on a patient

```
Patient
├─Basic data
│  ├─Basic data (name, surname, sex etc.)
│  ├─Addresses
│  ├─Admission data
│  ├─Selected physicians
│  ├─Education and employment
│  ├─Health insurance
│  ├─Nursing attributes
│  │  ├─Chronic non-contagious diseases
│  │  ├─Genetic diseases
│  │  ├─Disability
│  │  └─Other
│  └─Notes
├─Other data
├─Family
│  ├─Patient's data in regard to the family
│  └─Family data
│     └─List of family members
├─Visits
│  └─Home visits' specific data
├─Childbirth and development
│  ├─Newborn infant
│  │  ├─Data on birth
│  │  ├─Measurements at time of birth
│  │  ├─Data at time of discharge from hospital
│  │  └─Other data
│  ├─Infant and pre-school child
│  ├─School pupil and teenage years
│  └─Age independent child's data
├─Pregnancy/postnatal
│  ├─List of pregnancies
│  ├─Data on selected pregnancy
│  └─Data on selected delivery
│     └─List of born children
└─Document of nursing care
```

We have grouped criteria for an overall assessment of patient's need in a tree structure based on the fourteen BLAs [9, 10, 11]. It is a professionally accepted and well-known division as it has been confirmed in our survey. A list of parameters opens for each BLA of which we wish to remind the CN for gathering relevant data. These parameters are taken from the profession, and in nomenclature we

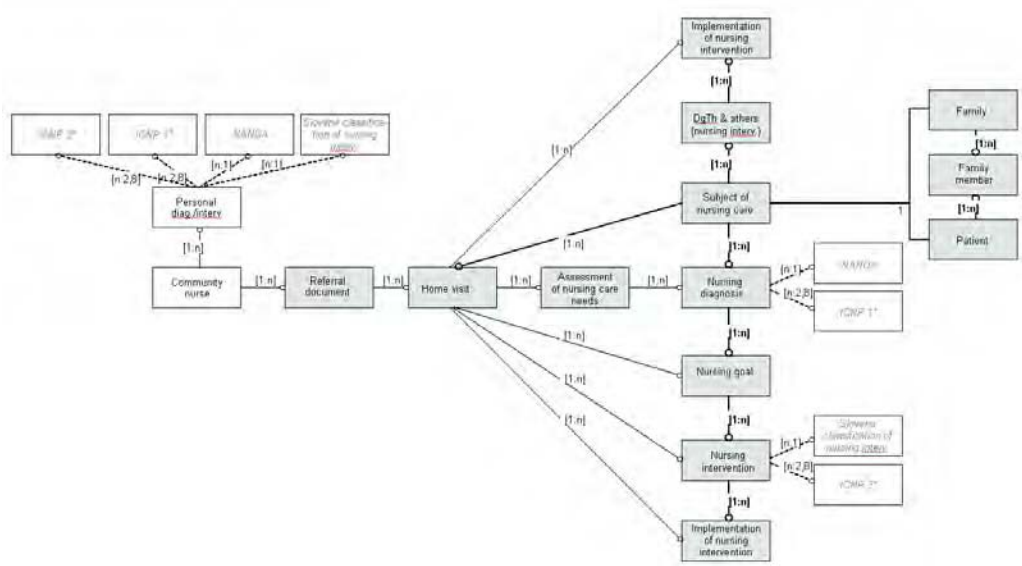


Figure 2 - Entity-relationship diagram of the proposed database

followed the Slovene translation of the International Classification of Nursing Practice, version beta 2 (ICNP) [12].

With each parameter there is a free text field for entering the values (e.g., with the parameter of excessive body weight we can insert the body mass index). In addition, for each parameter a CN also determines the degree of a problem using a five-point scale (no problem, minor problem, medium problem, major problem, very severe problem). Degrees are chosen in relation to the assessed state with individual nursing subjects.

From the values describing degrees of problems for parameters under the same BLA, the degree of problem for an individual BLA is calculated. These calculated values are then shown in the phase of planning. We will later show how we have supported evaluation with these grades. For example, the value of the parameter "appetite" has an impact on a BLA "diet and drinking" which affects the "physical BLA" and, consequently, the overall assessment of the patient. Values of higher-level parameters in the tree structure are calculated. The CN records values only for parameters on the tree leaves. After a simple calculation, we then obtain the grade of nursing problem for individual BLAs and the total overall assessment.

Under the tab planning we compose a nursing care plan in a tree structure. At the first level we see a list of BLAs and with each a calculated degree of a problem. On the second level we can add an arbitrary number of nursing diagnoses to each BLA. To each nursing diagnosis at least one nursing goal must be added, and to each nursing goal at least one nursing intervention (Figure 3).

The nursing diagnosis is made according to the PES system (Problem, Etiology, Symptoms). In denoting a problem, the nurse can use ICNP or the classification of the North American Nursing Diagnosis Association (NANDA) [3, 12].

Nursing interventions are described by name and frequency of performing them. Nurses denote interventions with the aid of the Slovene classification of nursing interventions and the ICNP beta 2.

The nurse can store her most often used nursing diagnoses and interventions in her personal directory. This can be simply supplemented with the catalogues that International Council of Nurses propagates as lists of the most often used nursing diagnoses and interventions for individual fields of nursing.

Under the implementation tab are shown all planned nursing interventions, those carried out need only to be marked. For the needs of evaluation, we can record comments on individual nursing interventions, e.g., ongoing evaluation or values of measurements, materials and time used.

After carrying out nursing interventions it is necessary to reassess the condition at the end of the visits. Assessments

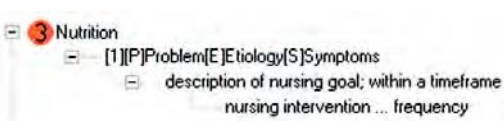


Figure 3 - Schematic presentation of nursing care plan in tree structure with explanation of individual levels

of the condition between two visits can also be compared. The evaluation phase is supported with the following visualization elements: comparison of overall assessments that shows progress for every parameter, a progress graph for a selected parameter throughout all previous visits. CN uses these measurements of changes in patient's needs to evaluate nursing goals and other elements in the nursing care plan.

When comparing assessments, we can compare the grade of nursing problem for the recorded and calculated criteria for two entries of overall assessments. We can thus compare two home visits, analyze the condition in the time between visits, or compare the overall assessment before and after visits are made, and analyze the impact of the intervention carried out on the change of condition. Where we have a number of assessments, it is also possible to show a time series of levels of nursing problem for an individual parameter – what sort of level of nursing problem was shown through all assessments of the condition e.g., appetite.

With these elements we wish to support the evaluation phase. The result of the evaluation phase is reflected in the changed nursing care plan. This means in practice to seek inappropriate elements in the nursing care plan, supplement them, exchange or remove them, and plan a new part of the nursing care plan for the new focus of problems.

The tabs are similar for home visits to the family, differing only in the parameters by which we describe the overall assessment. Instead of BLAs, we divided the parameters of families into the following groups: socio-economic state, health anamnesis, relations within the family and with the wider environment and functions of the family.

Computer support is provided by a completed prototype software solution. The solution used is a type of client-server and enables the use of laptop computers, which the CN can use directly in the field.

User's manual contains organizational and informational instructions for the direct use of the prototype software. It is basically intended as an aid to the work of the CN in using the software.

Testing the model in practice

We wanted to check the following categories in testing the proposed solutions:

- Success in implementing the nursing process in accordance with individual phases,
- Strengths and weaknesses of the use of hierarchical models of BLAs in the process,
- Suitability of the structure of data in nursing documents with an emphasis on the nursing care plan,
- Accordance of the data model and links among data with the current method of work or existing documentation,
- Interface with other processes.

In the alpha phase of testing, after completing the writing of the manual, we checked the operation of the program in compliance with the manual using simulation of real data. This was first carried out by the programmer and then by two working groups.

Beta testing of the software took place in the community nursing units of Ajdovscina Health Centre and Ljubljana-Bezigrad Health Centre. In total 8 nurses were involved. At both locations we placed software of a client-server type. Each participant in the testing, a CN, thus had available her own computer supported worksite. Data inserted at locations was gathered on the server.

At each location we had an introductory seminar covering:

- Presentation of the program in accordance with the process method of work,
- Presentation of the manual and annex and
- Test entry of data.

The CNs were then given a month time to become familiar with the program. During this time their questions related mostly to the terminology, the new model of the process and software solution. During this month they were asked to enter test data each day (at least one visit). This was the introductory period, which was intended to make a significant contribution to CNs being able to carry out the extensive plan of testing.

As a last step, we presented a detailed plan of testing. More than 80 entries in the period of one month provide the framework for testing various subjects with various needs and difficulty of work. At a new meeting, we then discussed possible difficulties and proposals of improvements and examined the entered records.

A SWOT analysis was carried out, which we performed with the help of the participants of testing at a final meeting.

Advantages:

- Providing users with integral nursing of high quality;
- Timely recognition of dangers that threaten the patient;
- Systemically arranged data of a relatively large quantity providing an easy to view information picture;
- Encouragement to the CNs own professional development.

Weaknesses:

- Insufficient ICT equipment;
- Work norms of CNs are often exceeded leaving too little time for entering the data into a computer;
- Lack of professional knowledge;
- The question is raised as to whether we know and can suitably use the available data.

Opportunities:

- To be more attuned to the user by means of the available data and to offer a higher quality of nursing;
- Users can be better informed and educated;
- Timely recognition of conditions;

- Production of guidelines for professional treatment and higher quality services for users;
- Including family and others in the nursing process;
- Motivation of staff.

Dangers:

- Insufficient ICT equipment of community nursing services could hold back the use of the system;
- Lack of permanent professional training and willingness of nurses to change could negatively affect on the use of such system;
- Changes in existing methods of work often trigger resistance in staff;
- Commitment to the computer rather than the patient.

Although the model and its prototype are already suitable for use in practice, we will continue to take certain comments into account and make the necessary changes. Extended testing will follow, which means monitoring use of the model in practice in a larger number of community nursing institutions throughout longer period of time.

Conclusion

The presented model of e-documentation covers the treatment of patients and families both from the aspects of processes and data. On this basis, a prototype organizational and informational solution of nursing documentation for the community-nursing segment was developed and has been tested in practice and critically evaluated.

The added value that contemporary ICT can contribute to nursing was presented, deriving primarily from a structured information picture, which monitors the patient and the nurse in the nursing process. It is worth highlighting in particular the use of hierarchical models in the treatment of BLAs. The model of calculating the grade of nursing problem, which the computer carries out concurrently in relation to the condition of the patient, thus enables an integral overview of the patient and systemically links apparently separate problems. It is thus a direct contribution to reducing the possibility of overlooking something important. E-documentation relies on the nursing record of the patient as a part of the overall health record of the patient [2]. This way we avoid duplication of data and the associated excessive work and obtain an overall information picture, which significantly contributes to a greater security for the patient and members of the nursing team.

We will continue the work not merely by extending testing and analysis of this model, but also by developing a similar model for documenting nursing in hospitals and dispensaries.

Acknowledgments

The authors thank the Ministry of Health of the Republic of Slovenia for financing the research work. We are also sincerely grateful to all institutions and individuals who took part in the project.

References

[1] Taylor C, Lillis C, and LeMone P. Fundamentals of Nursing, The Art and Science of Nursing Care. 4th ed. Philadelphia: Lippincott, 2001.
[2] Van Bemmel JH, and Musen MA, eds. Handbook of Medical Informatics. Berlin: Springer-Verlag, 1997.
[3] Gordon M. Nursing Diagnosis. Process and Aplication. 3rd ed. St Louis: Mosby, 1994.
[4] uteri O, Rajkovi V, Leskovar R, Bitenc I, Bernik M, and Rajkovi U. An Information system for Community Nursing. *Public Health Nurs* 2002: 19: 184-90.
[5] Klein J. Predictions for 2004 in Healthcare and Life Sciences. Gartner Group, 2003. http://www4.gartner.com/ resources/ 119000/119010/119010.pdf
[6] Kroell V, and Birthe Garde A. Strategy for Documentation in Nursing at a National and at a Local Level in Denmark. In: Oud N, Sermeus W, and Ehnfors M, eds. ACENDIO 2005: Documenting Nursing Care. Bern: Verlag Hans Huber, 2005; pp. 139-142.
[7] Handler TJ, and Hieb BR. The Gartner 2004 Criteria for the Enterprise Computer-Based Patient Record. Gartner Group, 2003. http://www4.gartner.com/resource/118300/ 118364/118364.pdf
[8] Madsen I. How to Avoid Redundant Data in an Interdisciplinary Electronic Patient Record. In: Oud N, Sermeus W, and Ehnfors M, eds. ACENDIO 2005: Documenting Nursing Care. Bern: Verlag Hans Huber, 2005; pp. 139-142.
[9] Henderson V. Basic Principles of Nursing Care. Geneve: International Council of Nurses, 1997.
[10] Bohanec M, Zupan B, and Rajkovi V. Applications of Qualitative Multi-Attribute Decision Models in Health Care. *Int J Med Inf* 2000: 58-59: 191-205.
[11] uteri O, Rajkovi V, Kljaji M, and Rajkovi U. Improving Nursing Care Documentation by Computerised Hierarchical Structures In: de Fatima Marin HM, Marques EP, Hovenga E, and Goossen W, eds. Proceedings of the 8th International Congress in Nursing Informatics. Rio de Janeiro: International Council of Nurses, 2003; pp. 212-216.
[12] International Council of Nurses. International Classification of Nursing Practice. Beta. Geneve: International Council of Nurses, 1999.

Address for correspondence

Uros Rajkovic, B.Sc.
University of Maribor,
Faculty of Organizational Sciences
Kidriceva cesta 55a, SI-4000 Kranj, Slovenia
Email: uros.rajkovic@fov.uni-mb.si

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

A Survey of the Effects of the full Computerized Nursing Records System on Sharing Nursing Records among Health Professionals

Yukio Kurihara[a], Naho Asai[a], Eri Ishimoto[a], Shigeyuki Kawamata[a], Satsuki Nakamura[a]

[a] *Department of Nursing Science, Kochi Medical School, Japan*

Abstract

In the last decade computerized nursing records systems (CNRSs) have been implemented at many hospitals around the world. Several effects of the CNRS were expected; the improvement of the quality of medicine and nursing care, the increased efficiency and the reduction of the cost. This study focused on the effects of the CNRS on access and sharing of nursing records among various health professionals. Timely access and availability of nursing records should improve the quality of medicine and nursing care. In 2003, we conducted a survey of the effectiveness of the CNRS on access and availability of nursing records among health professionals. We found that the CNRS contributed to the multidisciplinary sharing of nursing records without increasing the overall time spent on nursing documentation at most hospitals. However, effective sharing of nursing records even among nurses through the use of the CNRS did not occur at many nursing divisions.

Keyword:

Computerized nursing records system, survey, technology assessment

Introduction

Nurses record various observations of patients and make documents of nursing care throughout the patient's entire hospital stay. Although documentation is a large and time-consuming task for the average staff nurse, nursing records are used to share important information about the patient's status and plan of care among nurses, doctors and ancillary staff. However, paper-based nursing records are not simultaneously available for multiple users and sharing important information is sometimes delayed.

A computerized nursing records system (CNRS) [1], in which nursing records are accumulated into a database, might be effective to solve these problems. In the last decade, the CNRSs have been implemented at many hospitals around the world. Besides increasing availability of nursing records, several other effects of the CNRS were expected: the reduction of documentation time, an increased awareness of errors, and the reduction of the costs. There are many assessment reports about the effects of the CNRS on documentation time [2-5]. Although the reduction of time is desirable from the nurses' standpoint, the access and timely sharing of nursing records is much more essential, because one of the primary functions of nursing records is the continuity of care. We investigated the effects of the CNRS on access and sharing of nursing records among health professionals (i.e., pharmacists, nutritionists, medical technologists, and radiological technologists). We also investigated the changes in the amount of time spent o n nursing documentation, because a large increase in documentation time is not desirable.

Methods

In August and September 2003, we surveyed the effects of the CNRS on accessibility and sharing nursing records among health professionals at Japanese hospitals. We sent four sets of questionnaires to the director of nursing at each hospital and requested answers from responsible nurses of four major nursing divisions. We expected that the four major nursing divisions would include the following: internal medicine units, surgery units, combined internal medicine and surgery units, and intensive care units (ICUs).

Selection of hospitals

In a national survey, the status and function of health information systems in all main Japanese hospitals were published every year on the Japanese journal Gekkan Shiniryo. We selected all hospitals listed in 2001, 2002 and 2003 [6] that used the CNRSs; questionnaires were sent to a total of eighty-nine hospitals.

Questionnaires

Demographic data was requested about the respondents (i.e., position and work experience), the hospital (i.e., number of beds) and the division (i.e., type of division, number of beds, and number of staff nurses).

The questionnaire included:

1. What nursing record contents were computerized in the CNRS?
2. Which professionals had access to the nursing records before and after the implementation of the CNRS?
3. How had time spent on nursing documentation changed for the following: assessment data, daily reports, graphs of vital signs, nursing summaries, nursing diagnosis and care plans, and end-of-shift reports?
4. What kind of computerized nursing documentation was done at the bedside?

Data analysis

First, we checked what nursing record contents were being accumulated in the CNRS. We did an in-depth investigation of the answers from divisions where the CNRS contained the main contents of nursing records: assessment data, daily reports, graphs of vital signs, nursing diagnoses, care plans, evaluations and summary reports. We designated this type of CNRS as the "full CNRS". Secondly, we checked the type and size of divisions. We excluded specialty divisions and small divisions which had less than twenty beds.

Finally, we investigated the changes of access and sharing of nursing records among health professionals before and after the implementation of the full CNRS.

At Japanese hospitals, the end-of-shift report is usually performed. If the full CNRS could make nursing records more readily available in a timely fashion the end-shift meeting could be omitted or the meeting time could be decreased. We investigated the effect of the bedside computer documentation on the end-of-shift report times because bedside documentation enables the real-time sharing of nursing records. We also investigated the changes of time spent on various types of documentation.

We used SPSS Version 11.0 for the statistical analysis.

Results

Description of respondents

We received responses from 153 divisions of forty-six hospitals. Since we sent questionnaires to eighty-nine hospitals and requested each hospital to return answers from four different divisions, the respondent rate based on the total number of hospitals was 51.7% and the average number of the respondent divisions at each hospital was 3.3 (3.3/4=83.2%). Twenty-six were mid-sized hospitals, equipped with a hundred to four hundred beds and twenty were large hospitals equipped with more than four hundred beds. Since we requested answers from the nurse manager of the division, The respondents were mainly nurse managers or assistant nurse managers at divisions (Table 1). Most respondents had sufficient length of work experience at the hospital to understand the overall activities of the divisions.

Description of CNRS

Table 1 – Position and work experience of respondents

Position	Number of respondents (divisions)	Averaged work experience at the present hospital
Nurse manager of division	100	17 ± 9 yrs
Assistant nurse manager of division	15	15 ± 7 yrs
Staff nurse of division	29	8 ± 6 yrs
Others	9	11 ± 9 yrs

Half of the divisions used the full CNRS (Table 2). About 20% of the divisions implemented the CNRS-1 (one of the main elements of the content was missing). The CNRS-1 mainly lacked inclusion of nursing diagnoses. One third of the divisions used incomplete or partial CNRS possibly because the development of the system may have been in progress. Since the lack of any one element could affect the effectiveness of the CNRS, we elected to analyze data from only divisions using the full CNRS.

Table 2 - Level of CNRS

Contents	Computerized contents					
	Full CNRS	CNRS-1*				incomplete CNRS
Assessment data	ya	ya	ya	ya	ya	more than two elements are not computerized.
Nursing diagnosis	ya	ya	ya	ya	**no**	
Care plan	ya	ya	ya	**no**	ya	
Daily record	ya	ya	ya	ya	ya	
Graph of vital signs	ya	ya	ya	ya	ya	
Evaluation	ya	ya	**no**	ya	ya	
Nursing summary	ya	**no**	ya	ya	ya	
Number of divisions	77	2	6	1	19	52

* CNRS-1 lacks one element of the contents.

Description of divisions

In Table 3, we show the type and size of divisions. Since the responses from seven divisions lacked data about the type and size of division, we excluded these seven from our analysis. Since the length of stay of patients in the ICU is shorter compared to the stay on general wards, we excluded ICUs from the analysis. Finally sixty-two questionnaires from twenty-eight hospitals remained for further analysis. Eighteen hospitals were mid-sized hospitals equipped with hundred to four hundred beds and ten hospitals were large-sized and equipped with four hundred to seven hundred beds.

Table 3 - Profile of divisions implementing full CNRS

Type of divisions	Number of divisions*			
	1 to 19 beds	20 to 39 beds	40 to 59 beds	more than 60 beds
Internal medicine	0	1	17	4
Surgery	0	2	17	3
Mixed division of internal medicine and surgery	0	4	5	1
Pediatrics	0	0	2	0
Rehabilitation	0	0	2	0
Long care stay	0	0	4	0

Changes of access to nursing records for various health professionals

In Table 4, we show the numbers of divisions sharing nursing records with various health professionals before and after the implementation of the full CNRS. Before the implementation of the full CNRS the nursing records were shared in most divisions among nurses and doctors, and at a limited number of divisions with other professionals. After the implementation of the full CNRS the professionals outside the division could access the nursing records at most divisions. In the Kruskal Wallis's ^{2} test, for the percentage of divisions sharing nursing records among health professionals we could not find any significant difference according to the different type and different size of divisions.

Table 4 - The divisions sharing nursing records among various professions (n=62)

Professional	before implementing full CNRS	after implementing full CNRS
Nurse	55 (89%)	62 (100%)
Doctor	54 (87%)	62 (100%)
Pharmacist	32 (52%)	60 (97%)
Nutritionist	18 (29%)	55 (89%)
Medical technologist	12 (19%)	49 (79%)
Radiological technologist	12 (19%)	47 (76%)

Changes in the end-of-shift report time

In Table 5, we show the numbers of divisions that had changes in the length of time for end-of-shift report (T_m), related to bedside documentation using the full CNRS. If we disregarded bedside computer entry, about half the divisions had Tm that were essentially unchanged. However, if we divided divisions into two groups according to those who used bedside computer documentation and those that did not, many divisions using bedside input decreased T_m or abolished the end-of shift report. In the statistical analysis, we found that the bedside input into the CNRS significantly contributed to decreasing the end-of-shift meeting time; Mann Whitney's U test, p-value 0.001.

Table 5 - The change of the Tm before and after implementing full CNRS*

The change of Tm*	Number of divisions		
	Bedside computer documentation	No bedside computer documentation	Sum
Increase	0	0	0
Minimal change	11	20	31
Decrease	17	6	23
Absence†	4	0	4
Sum	32	26	58‡

However, many nurses pointed out several problems of bedside documentation with the CNRS. The communication with the patients was interrupted because the nurses had to focus on the data entry; some nurses had an increase in documentation time because of poor keyboard skills. Also, there was an insufficient number of mobile computers and bedside PCs.

Changes of time spent on documentation of nursing records

In Table 6, we show the percentage of divisions having changes in the time spent on documentation of various nursing contents before and after the implementation of the full CNRS. There was a small portion of the divisions that had increases in documentation time and 75% of those divisions had only one or two contents increased in documentation time. There were relatively much divisions that had the increased documentation time of the assessment data. Many respondents pointed out that the increase of assessment items in the full CNRS contributed to this increase in the documentation time.

Discussions

In principle the full CNRS increases access to nursing records for all health professionals. As shown in the previous section, sharing nursing records among nurses and

doctors was common practice for all divisions. However, the health professionals outside the divisions still could not easily access the nursing records at about 20% of the divisions. Cost and policy might be two factors that hinder the sharing of nursing records. In order to enable all professionals to share the nursing records, we would have to modify the system and the access control would require increased security, therefore the cost would increase. Another factor is the policy of the individual hospital in determining who should have access to patient information and what type of information should each professional group is allowed to have. In order to enhance the quality of care, we recommend that sharing patient information among health professionals is essential.

Table 6 - The change of the documentation time of various contents before and after implementing full CNRS

The change of documentation time	**Percentage of divisions**						
	Assessment data	Nursing diagnosis	Care plan	Daily record	Graph of vital signs	Evaluation	Nursing summary
Increase	26%	6%	5%	15%	15%	8%	8%
Minimal change	23%	16%	8%	32%	15%	48%	34%
Decrease	51%	76%	84%	48%	70%	40%	56%
No answer	0%	2%	3%	5%	0%	4%	2%

If nursing records were readily available in a timely fashion through the use of the full CNRS, the end-of shift meeting could be eliminated or shortened. However at half of the divisions shortening the time for shift report did not occur. The bedside input of nursing records was conducted at 78% of the divisions abolishing or shortening the meeting. Bedside computer entry seems very effective for facilitating access to nursing records; however, it could be potentially difficult for nurses to enter all data into the full CNRS at the bedside. A small device, such as a PDA is usable for documenting vital signs; however a mobile PC would be more useful for documenting various nursing observations. Unfortunately, the mobile PC currently weighs more than 1 kg, and it is impossible for nurses to carry one as they deliver patient care.

Comments from respondents highlight concerns that the nurse's attention is taken away from the patient and redirected towards the computer. We have to consider various possibilities to overcome the problems of the bedside computer documentation. One possibility is to set a PC outside each ward, and then the nurse could immediately input a memo into the full CNRS. Another possibility is to use a PDA for short comments. Any alternative solution should allow entry of key information into the full CNRS as soon as possible.

Conclusion

The full CNRS clearly contributed to the enhancement of sharing nursing records among health professionals and did not increase the overall documentation time in most divisions. Bedside computer entry of nursing observations also facilitated the sharing of nursing records among nurses. However, the bedside documentation has several practical problems related to the present technology and various solutions that allow efficient and timely bedside data entry are necessary.

References

[1] Kurihara Y, Kusunose T, Okabayashi T, Nyu K, Fujikawa K, Miyai C, Okuhara Y. Full implementation of a computerized nursing records system at Kochi Medical School Hospitals in Japan. Comput Nurs. 2001; 19: 122-129

[2] Larrabee JH, Boldreghini S, Elder-Sorrells K, Turner ZM, Wender, RG, Hart JM, Lenzi PS. Evaluation of documentation before and after implementation of a nursing information system in an acute care hospital. Comput Nurs. 2001; 19: 56-65

[3] Korst LM, Eusebio-Angeja AC, Chamorro T, Aydin CE, Gregory KD. Nursing documentation time during implementation of an electronic medical record. J Nurs Adm. 2003; 33: 24-30.

[4] Smith K, Smith V, Krugman M, Oman K. Evaluating the impact of computerized clinical documentation. Comput Inform Nurs. 2005; 23: 132-8.

[5] Poissant L, Pereira J, Tamblyn R, Kawasumi Y. The impact of electronic health records on time efficiency of physicians and nurses: A systematic review. J Am Med Inform Assoc. 2005; 12: 505-16

[6] The list of hospitals implementing EMR, Gekkan Shiniryo, 2001; 94-101, 2002; 98-107 and 2003; 132-133 (in Japanese).

Address for correspondence

Y. Kurihara: Kohasu Oko, Nankoku 783-8505, Japan.
e-mail: kurihary@kochi-u.ac.jp

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

New Method of Realization of Nursing Diagnosis Based on 3N in an Electronic Medical Record System

Young ah Kim[a], Mijung An[b], Jungyoen Park[b] , Hyensun Jung[b], Yongoock Kim[c], Byungchul Chang[d]

[a]Dept. of Medical Informatics, Yonsei Univ. Medical Center, korea, [b]Dept. of Nursing, Yonsei Univ. Medical Center, korea
[c]Dept. of Plastic & Reconstructive Surgery, Yonsei Univ. College of Medicine, korea
[d]Dept. of Cardiovascular Surgery, Yonsei Univ. College of Medicine, korea

Abstract

An electronic medical recording (EMR) system is enlightened as a solution for deciding which nursing diagnosis was selected on the basis of a computerized system; it can help collect and analyze lots of diverse data in an objective way. But there are few reports of successful electronic nursing diagnosis on EMR systems. This study was to develop the objective decision system prior to nursing diagnosis and to adopt it in the Severance EMR system. We adopt a new concept, situational variables, as the key elements of the nursing process, based on the items of nursing intervention. It enables appropriate nursing diagnosis through complex clinical objective data (laboratory results, vital signs, etc.). Through these elements between nursing intervention and nursing diagnosis, we can create reliable evidence-based nursing diagnoses successfully, and make possible fast settlement of nursing intervention in the various clinical fields.

Keywords:

EMR, ENR, nursing process, situational variables

Introduction

The academic concept of a nursing process is constructed on the serial thread of nursing diagnosis, nursing intervention, and outcome of nursing (Figure 1).

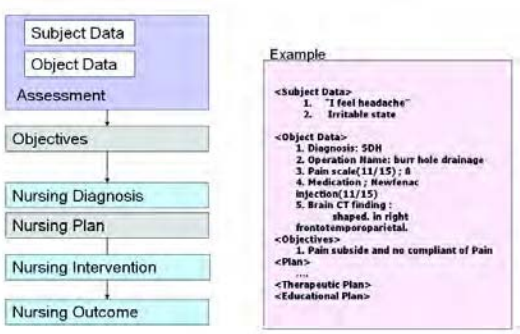


Figure 1 - Nursing process

The first step of this system is the nursing diagnosis, whether it is derived from medical diagnosis or the subjective decision from the experienced professional nurse who can judge the nursing diagnosis (Figure 2).

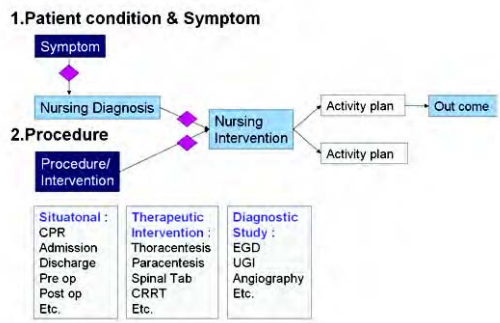


Figure 2 - Algorithm of nursing process

In real situation, Nursing diagnosis in the clinical field has been inevitably limited because of the inavailability of the experienced nursing professional with education and experience in nursing diagnosis.

An electronic medical recording (EMR) system is enlightened as a solution of nursing diagnosis because an electric nursing record (ENR) was expected that will help the implementation of appropriate nursing diagnosis with saving of the recording time and providing the standardization of the nursing records. However, there are few reports of successful nursing diagnosis systems. For the success of nursing diagnosis system, the decision of nursing diagnosis should be decided not in a subjective way but in an objective or evidence-based way. So, authors find out that the key point was based on the items of nursing intervention – so-called *situational variables* – instead of nursing diagnosis itself as a first step of the nursing intervention system.

To develop the objective decision system prior to nursing diagnosis, a new methodology was adopted in the Severance EMR system. Through this method we can create objective data based nursing diagnosis successfully, and it

makes possible fast settlement of nursing intervention in the various clinical fields.

Method

The steps of this study are shown in Figure 3.

Figure 3 - The steps of this study

The pre-existed items of nursing intervention protocol (NIC, 3rd edition) were used. NIC involves 1609 nursing activities with references to a translated version of the 2nd edition. The analysis was done in parallel way between above NIC items (1,800 items) and the items of intervention which was done in the real patient according to the medical diagnosis by doctors.

Through this comparison, the same item was categorized as a original NIC division, and the other related items from the real field was involved into the new created subdivision of the NIC items. This procedure can make a sophisticated set of nursing intervention items based on the specific situation and real activities.

This new set of nursing intervention item makes the relationship between nursing diagnosis(NANDA) and nursing intervention more concrete. From this new set of nursing intervention, we postulate and extract the situational variables. And then we create new set of situational variables which can decide nursing diagnosis automatically through the EMR system.

For processing of above procedure, all units of inpatient gathered the top 20 disease items that the hospitalization is required. Fianlly, total 125 disease items were selected and analyzed.

Results

The 46 nursing plans were postulated from 125 disease items as a prior step before decision of nursing diagnosis and the identification of items of nursing activities (Figure 4).

In the new set of nursing intervention (nursing activity), a total of 14,296 items were involved. Among them, 13,548 items were involved in the substructure of the NIC structure through a new methodology (Figure 5). Situational variables were extracted from the new set of nursing intervention for automatic extraction of nursing diagnosis (Figure 6). Initially, 118 sets of situational valuables were established in the Severance nursing record system for clinical nursing intervention. This set means that one set of nursing activities can determine specific nursing diagnosis, and in turn nurrsing diagnosis can determine more sophisticated nursing activities.

Ex) Gastro-Intestinal Medicine

Dx	Nursing Dx
Stomach Ca	1. Acute pain
	2. Altered Nutrition: Less than body requirement
	3. Fluid Deficit
Esophageal ca	1. Altered Nutrition: Less than body requirement R/T Dysphagia
	2. Risk for Impaired Skin Integrity R/T Mechanical Force about Gastric tube Insertion Site
	3. Deficient Knowledge
Pancreas Ca	1. Acute pain R/T Malignant tumor
	2. Altered Nutrition: Less than body requirement R/T Anorexia
	3. Anxiety R/T Threat of death

Figure 4 - Relation of medical diagnosis and nursing diagnosis

Ex) Stomach Ca

Nursing Dx	Nursing Activity
Acute pain	1. Perform a comprehensive assessment of pain to include location, characterist
	2. Provide the person optimal p prescribed analgesics.
	3. Promote adequate rest of fac
	4. Evaluate the effectiveness of measures used though ongo pain experience.
	5. Encourage patient to discuss expression

Figure 5 - Nursing plan, nursing activity and diagnosis

Ex) Infection, Risk for

- Event : Operation
- Patient Monitoring
  - Body temperature > 37.5
  - BMI < 20
- Laboratory data
  - WBC > 11,000 or WBC <3,500
  - ESR > 15
  - CRP > 0.8
  - Serum Albumin level < 3.0

Figure 6 - Nursing activity and situational variables

- Automatic nursing diagnosis; Successful
- Implementation of ENR system; Successful
- Construction of Nursing Terminology DB;
  - Nursing diagnosis ; 13 domains 92

- Nursing interventions ; 13,909
- Nursing activities with Order Treatment; 2,183
- Situational variables; 63
- Set of situational variables connected with nursing diagnosis; 118

1 year after adopted EMR system, we got those responses from the nurses of clinical field.

Usage of nursing diagnosis, Addition of nursing activities is more increased.

Request of creation of situational variables, request of changes of preset nursing diagnosis is also increased.

Conclusion

The study revealed that standardization of nursing activity based terminology can be established and can be evolved by itself.

An automatic nursing diagnosis system based on situational variables is a solution for realization of a Nursing Process based on 3N(NANDA, NIC, NOC). Automatic nursing diagnosis systems should be based on an objective, evidence based standpoint. Clinical Decision Support is also possible through automatic nursing diagnosis system.

Discussion

The nursing process is one of the most valuable concepts for effective and safe patient care [1-2]. However, its realization has been a questionable one, because of difficulty in adoption in the field for real-world use. Most nursing professionals think that the items of the NIC system can accomplish the thread of the nursing process [3]; we focused on whether the NIC system is able to express the interventional activities of the real practice in the field. If there is limitation of expression of real activities, it would indicate a disconnect between the conceptual items and their related activities.

So the trial of a process was done for the new connection between academic concept and real activities that can be involved in the academic concept. [2][4] All the currently used interventional activities should be gathered from the specified units. [4]

During these procedures we find that the meaning of the items of NIC is much broader than items from the real field. So, many items were found that can not be expressed by the items of NIC system in the real field, so we categorized additional items of nursing activities as a substructure of NIC items from the clinical disease base. This new structuring was the key point of establishing the successful nursing process.

From the set of situational variables, we decide the nursing diagnosis. This can be said as a kind of clinical decision support system in the real patient care field. The nursing process is impossible without first designation of nursing diagnosis in current concept. However the decision of nursing diagnosis is not an easy procedure for all the acting nurses. So, if there is no automatic creation or selection of nursing diagnosis, the realization of a nursing process would be only happen in the far future. To overcome this obstacle, the set of situational variables in Severance hospital was created and it reveals its effectiveness in real field(Figure 7).

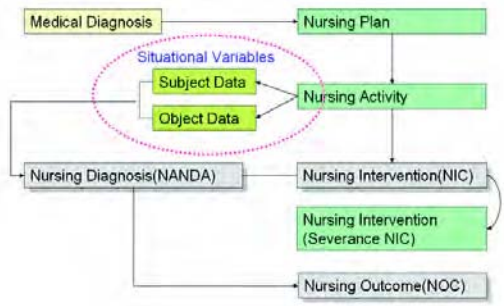


Figure 7 - New methodology of Nursing Process in Severance ENR

125 disease items may be a small percentage among all disease entities, however, we are sure these 125 high-frequency diseases can involve all the nursing intervention items, nursing diagnosis, and nursing plans. This hypothesis should be tested in terms of the rate of participation and reuse of nursing process in the field, and tested though 1 year's usage in Severance hospital. As another advantage of this system, the evaluation and validation of the activity based cost of the nursing activity will be enabled in every clinical activity. This kind of advantages would demonstrate the effectiveness or return on investment of this methodology.

For better nursing process, the contents of substructure of the intervention items should be reinforced with the concrete relationship with the nursing diagnosis and outcome systems. [2][3][4]

Acknowledgment

This study was supported by a grant from the Korea Health 21 R&D Project, Ministry of Health & Welfare, Republic of Korea (Grant No. A050909).

Reference

[1] Yong Oock Kim, et al. Essential elements of EHR system, Seoul (ROK): Koonja, 2005.
[2] Saba VK. Essentials of Nursing Informatics. 2005, McGraw-Hill
[3] Handbook of Informatics for Nurses & Health Care Professionals (3rd Edition), Toni Lee Hebda
[4] NANDA, NOC, and NIC Linkages: Nursing Diagnoses, Outcomes, and Interventions, Marion Johnson, 2001, Mosby: St.Louis.

Address for correspondence

e-mail sgm625@yumc.yonsei.ac.kr
office 82-2-2228-22218

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Integration of Longitudinal Electronic Records in a Large Healthcare Enterprise: The U.S. Veterans Health Administration Experience

Arthur C. Curtis[a], Joseph Gillon[b], Dale C Malmrose[c]

[a] *VHA Health Enterprise Strategy, Boston MA, USA,* [b] *VHA Medical Center, Ann Arbor MI, USA*
[c] *VHA OI Field Office, Salt Lake City UT, USA*

Abstract

The U.S. Veterans Health Administration (VHA) provides care to some 5.2 million patients spread across the continental United States, Alaska, Hawaii, Puerto Rico, and the Philippines. Sites of care include 157 medical centers, nearly 900 outpatient clinics, long-term facilities, and home care. Over the last 10 years, major changes in the nature of VHA healthcare have imposed a requirement for longitudinal electronic health records and integration of those records across the enterprise at the point of care. VHA has now evolved through three generations of applications that support such integration. This paper reports on the VHA experience, points out lessons learned, and outlines future directions for electronic health record integration in VHA.

Keywords:

Electronic Health Records, EHR, Longitudinal Records, Data Integration, VistA, Veterans Health Administration.

Introduction

Veterans health administration

The United States Department of Veterans Affairs (VA) provides benefits and services to the nation's veterans and selected family members ("To care for him who shall have borne the battle and for his widow and his orphan…" - Abraham Lincoln). The Veterans Health Administration (VHA) is the branch of VA that provides healthcare benefits.

VHA includes approximately 1,400 sites of care, including 157 medical centers, nearly 900 outpatient clinics, 135 long-term care facilities, and 88 home-care programs. VHA now treats some 5.2 million patients, with approximately 7.6 million enrollees. In 2005, VA workload included 587,000 inpatient admissions and some 52 million outpatient visits. 60% of US health professionals (70% of physicians) have some training in VHA. VHA has some 193,000 employees (approximately 14,000 physicians, 50,000 nurses, and 33,000 allied health personnel). VHA's budget in 2005 was $32.5B US.

In the last ten years, VHA has transformed itself from a loose federation of independent medical centers focused primarily on inpatient care into a fully integrated health care delivery system that promotes primary and ambulatory care. Between 1996 and 2005, the annual number of ambulatory care visits increased by 84%, the number of acute care hospital beds dropped by 68%, inpatient admissions dropped by 27%, and total bed days of care fell by 47%.

Aging of the veteran population is a major issue confronting the VA. Today, 9.6 million veterans are age 65 or older, representing 38 percent of the total veteran population. By 2030, the proportion of older veterans will increase to 45 percent of the total. As in the general U.S. population, those aged 85 or older (the "old-old") are the fastest growing segment of the veteran population, representing 3 percent of current veterans. The number of veterans age 85 or older is expected to nearly double from 764,000 to a peak of 1.4 million between 2003 and 2012.

In comparison with the U.S. Medicare population, the VHA patient population is older and sicker. 38% are over age 65. In comparison to age-matched Americans, VHA patients carry one additional mental health diagnosis and three additional non-mental health diagnoses. They are also less affluent: approximately 70% with annual incomes of less than $26,000 and some 40% with annual incomes less than $16,000. Demographics of the patient population are changing as older veterans die and veterans of more recent conflicts become eligible for care.

A progressively older population with more chronic diseases and increasing complexity of illnesses has resulted in involvement of more clinical providers (both primary and specialty care), more types of health care programs (inpatient, outpatient, and home care), and more diagnostic procedures done at geographically dispersed sites. In addition, restructuring of VHA into health care networks has increased patient travel and the number of institutions at which a given patient can be expected to receive care, which has in turn increased the medical and financial value of sharing information on patient care across sites on an enterprise basis.

VHA clinical information systems & electronic records

Veterans health information systems and technology architecture (VistA)

The Veterans Health Information Systems and Technology Architecture (VistA) supports day-to-day operations at VA health care facilities [1,2]. It incorporates both administrative and clinical applications; including a full suite of ancillary departmental systems as well as the Computerized Patient Record System (CPRS). A descendant of the Decentralized Hospital Computer Program (DHCP, circa 1982), VistA has been in continual evolution since 1996. It consists of more than 100 corporately developed applications as well as software developed at local medical facilities; it also includes support for integration of commercial off-the-shelf products. Historically, VistA has employed a decentralized architecture operating at the medical center level; applications and national databases derived from locally generated data that reside in VA's centralized corporate data center lie outside the scope of VistA. VistA employs a client-server architecture: applications are written in M (also known as MUMPS) on the server side and in Delphi on the workstation side; VHA is in the early stages of a major technology migration that will shift application development to Java and relational database systems. VistA has been used by a number of private and public health care facilities in the United States and around the world.

Computerized patient record system (CPRS)

CPRS is the electronic "clinical practice environment" for VA clinicians and all members of the healthcare team; it is a single, integrated clinical information system used throughout the organization in all healthcare settings (Inpatient, Outpatient, Long-term care). CPRS provides a longitudinal medical record as well as active functionality covering all aspects of patient care and treatment, including: electronic order entry and processing; entry and retrieval of clinical documents such as narrative notes and discharge summaries; display of results for laboratory, imaging, and diagnostic procedures; consult management and reporting; alerts of abnormal results; clinical reminders; problem list management; inpatient and outpatient medication management. It also serves as an enterprise integration platform, giving clinicians the ability to see electronic records from any other VHA facility where the patient has received care. As shown in Table 1, CPRS is heavily used.

Table 1 - CPRS usage as of December, 2006

Content category	Total count	Average daily volume
Documents (Progress Notes, Discharge Summaries, Reports)	874 M	638 K
Orders	1.6 B	912 K
Images	591 M	884 K
Vital measurements	1.0 B	729 K
Medications Administered - Bar Code Medication Administration	850 M	607 K

VHA has gone through three approaches to enterprise integration: WebTop, CPRS Remote Data Views, and VistA Web.

Materials and methods

WebTop

Motivation: Development of WebTop began in 1996 as a centrally-funded research project to explore the feasibility of extracting and displaying data from a VistA system using browser-based technology, with an emphasis on use of XML, Java technologies, and linkage of patient data to external resources. A follow-on project focused on integration of data across sites was undertaken in 1997, and the resulting application was brought to operational status in VISN 1 (VA New England) in 1998. Other VISNs adopted the technology, and at its peak, WebTop was deployed in 16 of 22 VISNs. WebTop development preceded that of CPRS, and was thus unable to take advantage of some of the foundations of the latter application.

Architecture: WebTop provided browser-based access to VistA clinical data using a combination of client-side and server-side software controlled and organized by a middleware server. (See Figure 1.) The middleware retrieved clinical data from multiple VA Vista systems and formatted it into a consolidated display. The middleware also supported retrieval of data from other intranet or internet information resources, such as knowledge bases, using patient data to construct a search query. Deployment was based on a regional model, with one WebTop instance per VISN; the architecture allowed for cross-VISN connectivity to produce integrated data sets covering multiple regions, but this approach was never generally adopted due to both technical and organizational obstacles at the time.

WebTop middleware was based on Java servlets, which handled log-on, patient look-up, clinical data retrieval, and traversal of links to other network resources. In keeping with the focus on web technology, HTTP was initially the sole choice for connectivity; when CPRS was implemented, the set of transport layer protocols was expanded to include Remote Procedure Calls (RPC).

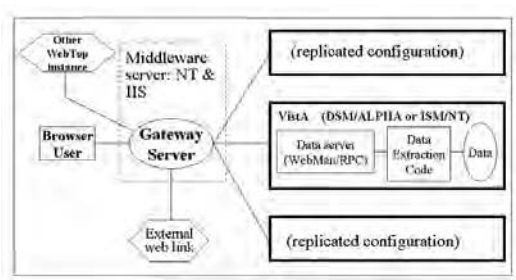


Figure 1 - WebTop architecture

Operational model: Users connected via browser to the WebTop middleware server, which validated the user's access codes against a selected VistA system. Once validated, the user was assumed to have access to data that resided on any Vista system participating in the WebTop domain (see discussion of business rules below). After selecting a patient, the user could browse the patient's record, selecting individually the type of data desired (e.g., lab results, progress notes, etc.). The middleware server processed such user requests, retrieved data from the individual VistA databases and aggregated it for display by applets running on the user's workstation.

CPRS remote data views

Motivation: Development of CPRS Remote Data Views (CPRS/RDV) began in 2001. Its objective was to support on-the-fly remote data views of patient information within VHA's mainstream clinical application. Because CPRS deployment was ubiquitous, CPRS/RDV offered the first opportunity for true enterprise-wide integration of the longitudinal electronic health record in VHA.

Architecture: Traditional CPRS architecture was based on communication between a client CPRS workstation and a VistA server using remote procedure calls (RPCs). Security concerns, desire to avoid additional servers, etc. led to implementation via server-to-server communication rather than client-to-servers or client-to-middleware. (See Figure 2.) To add this new capability to CPRS, changes were made to both the CPRS application and the underlying infrastructure; the RPC Broker was enhanced to provide the Server-to-Server RPC support required, and HL7 messages were encapsulated and exchanged within RPC messages.

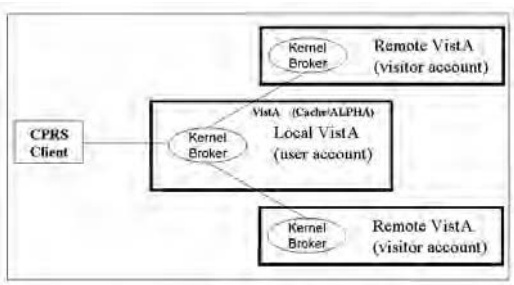


Figure 2 - CPRS Remote Data Views architecture

Operational model: Since CPRS RDV is part of CPRS, user logon, authentication to VistA system, and selection of a patient are all accomplished in the normal CPRS context. CPRS knows the sites at which a patient has data courtesy of the VHA enterprise Master Patient Index; information on the sites applicable to a specific patient is kept as part of the local patient record, and a user asking for remote data can limit the sites queried if desired. Once a type of data is chosen, the sequence is as follows: RDV asks the local VistA for remote data; the local VistA system adds a user identifier to the request and passes it on to the remote VistA systems; the local VistA returns a handle to CPRS; while the remote VistA systems assemble the date, CPRS iteratively asks the local VistA if the data is ready yet; when all the data has been returned from the remote VistA systems, the local VistA answers CPRS's next query with a YES, and CPRS issues the third and final RPC to get the data. The data is then displayed in spreadsheet form or tabs of narrative text, much like the approach taken with WebTop. At the remote sites, RDV uses "visitor" accounts that have no access or verify codes, no menu options, and are aliased as "VISITOR". However, the identity of the user originally requesting the data is passed to each remote VistA system and entered in an access log.

VistA web

Motivation: VistAWeb was initially developed within one of VHA's regional networks as an exploration of alternatives to WebTop and CPRS Remote Data Views. It also explored user interface alternatives, including alternative ways to deliver graphical information such as ECG or DICOM images.

Architecture: VistAWeb uses Medical Domain Objects (MDO) as its data abstraction layer, and MDOs handles all connectivity to VistA. On one side an MDO communicates with various data sources such as VistA, HL7, DICOM, XML, and SQL. On the other side, it delivers to the client a uniform, well-defined suite of objects from the medical domain (e.g., patient, provider, progress note, lab result, etc.) MDO delivers the same object regardless of the data source. So, to the client software, a progress note from VistA looks and behaves exactly the same as a progress note from an arbitrary source via HL7 or SQL or XML. All functionality is read-only.

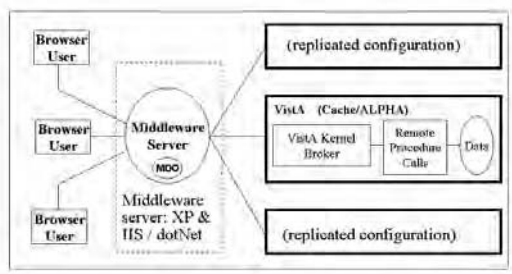


Figure 3 - VistAWeb architecture

Data Access Objects (DAO) are used to carry out communication. In the general case, DAOs would exist for different access protocols: VistA, HL7, SQL, DICOM, or XML. MDO is an ongoing project; the DAOs currently operational are for accessing VistA systems via RPC and the HDR via HL7.

Operational model: Users authenticate to VistAWeb via standard VistA credentials and are then able to browse the clinical record. Connections are made to sites determined by MPI data stored with the patient. If necessary, a new visitor identity is created at remote sites to support auditing of remote data access. Users are warned if they select a deceased or sensitive patient (e.g., patient as employee, patient as high-profile patient, etc.); a check is also made to prevent users from accessing their own records. User

requests for a specific type of data are sent to DAOs, one DAO per site, each of which communicates directly with its associated VistA system. MDO acquires data from VistA using the same RPCs that CPRS uses; the data returned is labeled with a site tag, collated, sorted, and returned to the user's browser as a standard HTML page – data displayed is patient-centric instead of geo-centric.

Results

WebTop: Deployed in 1998, WebTop employed a user interface that was different than that of CPRS and required an independent user authentication step and invocation of a separate application (i.e., a web browser) to view an integrated set of patient data. In addition, as a field-developed product that required local or regional implementation effort over and above mainstream systems, WebTop never achieved enterprise-wide deployment. As a result, some users were reluctant to invest in learning how to use it. However, WebTop provided a compelling proof-of-concept, and demonstrated that clinicians were willing to make compromises on data normalization across sites (e.g., variability in drug or lab test names) in return for the ability to see data integrated across sites. WebTop undoubtedly succeeded at the regional level in meeting its fundamental objective of providing clinicians with the ability to access the equivalent of a single clinical repository containing all data on a patient. It quickly become a part of the daily life of clinicians in the VISNs in which it was implemented, as evidenced by the rapidity with which clinicians were heard from when there was any problem with the system.

CPRS/RDV: Implemented in April, 2002, CPRS/RDV rapidly supplanted WebTop due to its availability within the CPRS user interface as well as its national scope. As a corporate product it had advantages over field-developed products (e.g., inserting a stub in the user file to support tracking of access to patient data). It is interesting to note, however, that until WebTop was finally decommissioned, there remained a core of users who preferred its speed and simplicity in spite of its constraints. Implementation of CPRS RDV was a watershed event in terms of making data available across an entire healthcare enterprise of major size and scope, and it continues as a core function of the clinician interface. While it is clear from observation and anecdotal reports that CPRS RDV is used extensively, the application is unfortunately not instrumented so as to allow reporting on actual usage volume and on the types of data being retrieved.

VistAWeb: Released to the field in March of 2005, VistAWeb combines aspects of WebTop and CPRS, and adds major innovations of its own. What began as a local project to explore alternatives to a mainstream corporate application became, in turn a corporate product itself. In the process, important foundations were established: use of web services, a simplified model for clinical data, and return to a model for access to data from VistA sites not based on VistA-to-VistA communication. VistAWeb is in production use in VHA an alternative to CPRS RDV; it is slated to replace the Reports tab of CPRS in an upcoming release, providing a convenient integrated view of data from within the context of the CPRS interface.

Discussion

The industry press is replete with accounts of both successful and unsuccessful attempts to introduce clinical information systems into clinical care. Consideration of the factors which have led to VHA's successes in this arena over a quarter century is beyond the scope of this paper: suffice it to say that a major factor has been a tradition of large-scale involvement of clinical users in defining requirements and functionality of clinical systems. Another is reflection in spending patterns of belief in clinical systems. Until very recently, VHA never had a budget line item for IT – it was paid for out of health care dollars, thus assuring that investments made were perceived as having high value in clinical care.

Beyond such broad considerations, however, VHA has learned a number of useful lessons from the series of projects described in this paper which should be of use to other organizations taking on the challenge of data integration across a healthcare enterprise.

- It is essential to achieve a compromise between perfection, in terms of normalization of vocabularies and forms of representation, and practical utility. As the saying goes, "Don't let the perfect be the enemy of the good". VHA's experience clearly demonstrates that the human brain remains capable of smoothing over discontinuities that could otherwise prevent a project from getting started at all.
- Clinicians favor working in a single application over multiple applications or multiple browser windows, and value incorporation of functionality such as data integration with applications that are part of their primary workflow.
- Day-to-day operation of an integration framework depends on changing the traditional attitude that access to patient data at one medical center by staff at another is suspect. Thus, an enterprise or consortium intending to implement such a framework must be willing to implement business rules such as the following: a provider who has access to a given patient's data at one medical center has access to that patient's data at all participating sites; authentication of a provider's credentials at one site is sufficient to authenticate that provider's access at all sites; a locally acceptable indication of right to access patient data is an acceptable surrogate for permission to access patient data at all sites.
- Regardless of whether an organization has an inclination toward research, it is important to instrument applications so that usage patterns can be studied and understood. This has not been adequately appreciated in our own organization, and we have regretted it a number of times.
- While long-term success requires that an enterprise eventually establish standards for what data each site must commit to collect electronically, sites just begin-

ning the process must educate providers to have realistic expectations for what data will be found in the base systems, and thus the integrated data set.

- While corporate information technology programs are essential to the widespread deployment and sustained operation of sophisticated systems, investigative projects and local innovation continue to be the well-spring of watershed projects that lead to major advances in the use of information systems in clinical practice.

Future plans

While VistA is a tightly integrated suite of applications, VHA's current facility-level systems represent islands of data in an enterprise sense bridged by a series of integration applications. Desire for a patient-centered, "gold standard" legal electronic record and a more tractable approach to making computable data available for decision support have motivated VHA to undertake migration to a new environment, referred to as "Health*e*Vet", a core component of which is a health data repository (HDR) at a national level. Although the transition from VistA to HealtheVet will of necessity be gradual (akin to building an airplane in flight), the HDR will eventually serve as the source of a patient's electronic health record, providing a superset of the integration available today and obviating the need for applications such as CPRS/RDV and VistAWeb. As a corollary activity, VHA is embarking on a major data standardization and data modeling effort which will support application re-engineering. A major long-term implication of this undertaking is the positioning of VHA as a single national health information organization (NHIO) interacting through a National Health Information Network (NHIN) with other public- and private-sector entities delivering care to VHA patients.

Conclusions

VHA has demonstrated the feasibility of an enterprise-wide integrated patient record in a very large healthcare organization of national scope. While others have reported on clinical data integration projects, most are still in a formative stage and few if any have the breadth or depth of VistA and its current integration platforms [3]. Trends in healthcare and efforts at all levels encouraging the development of partnerships for sharing electronic health records have had major impact on requirements for cross-site data integration: practical technical and economic models are still being developed. It is interesting to observe that the majority of the concepts first proposed in the initial paper from the Institute of Medicine on electronic health records [4] are only emerging now, some 15 years later.

We close this paper with an anecdote that illustrates the value of the applications described here. VHA was severely affected in the coastal Southeastern states when Hurricane Katrina hit on August 29 of 2005. Multiple sites were affected: as an example, the New Orleans VA Medical Center was submerged to the second floor (where the computer room was) and was closed, with most inpatients going to Houston, Texas. Outpatient services in affected areas were dislocated. The "RecoverAll" emergency response team was dispatched on September 1, and the New Orleans computer system was replicated in Houston in twenty-four hours, with all records restored. It is estimated that it required about 100 hours to restore EHR capability for all affected facilities.

In the days following Hurricane Katrina patients started to appear at medical centers outside the affected area, initially at other VA medical centers in the Southeast and later at hospitals as far away as Puget Sound, in the far Northwest. The high number of refugees precluded normal access through CPRS/RDV. The solution was to grant temporary special user privileges to all VistAWeb users which gave users the ability to do patient lookup at other sites, and thus get VHA-wide data on any patient from the affected VISNs regardless of whether they were registered in the provider's local VistA system. Thus, electronic health records – and beyond that the integration of data across sites – played a major role in continuity of health care through a massive natural disaster.

References

[1] Kolodner RM, editor. Computerizing large integrated health networks: The VA Success. New York: Springer-Verlag. 1997.

[2] Meldrum K, Volpp B, Vertigan R. Department of Veterans Affairs' Computerized Patient Record System. Proc AMIA Symp 1999(1-2):1214.

[3] Brown SH, Lincoln MJ, Groen P, Kolodner RM. VistA: The U.S. Department of Veterans Affairs National Scale Hospital Information System. International Journal of Medical Informatics 2003:69(2/3):135-156.

[4] Richard S. Dick and Elaine B. Steen, Eds., The Computer-Based Patient Record: An Essential Technology for Health Care, Institute of Medicine, Washington, DC, 1991.

Address for correspondence

Arthur Clayton Curtis MD
VA Medical Center (WHC), 150 South Huntington Avenue
Boston, MA 02130.
clayton.curtis@va.gov

The opinions expressed in this paper are those of the authors, and not necessarily those of the Veterans Administration or the United States Government.

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

The AMPATH Medical Record System: Creating, Implementing, and Sustaining an Electronic Medical Record System to Support HIV/AIDS Care in Western Kenya

William M. Tierney[a], Joseph K. Rotich[b], Terry J. Hannan[c], Abraham M. Siika[d], Paul G. Biondich[a], Burke W. Mamlin[a], Winstone M. Nyandiko[d], Sylvester Kimaiyo[d], Kara Wools-Kaloustian[a], John E. Sidle[a], Chrispinus Simiyu[d], Erika Kigotho[d], Beverly Musick[a], Joseph J. Mamlin[a], Robert M. Einterz[a]

[a] *Indiana University School of Medicine and the Regenstrief Institute, Indianapolis, Indiana, USA*
[b] *Moi University School of Public Health, Eldoret, Kenya*
[c] *Launceston General Hospital, Launceston, Tasmania, Australia*
[d] *Moi University School of Medicine and Moi Teaching and Referral Hospital, Eldoret, Kenya*

Abstract

Providing high-quality HIV/AIDS care requires high-quality, accessible data on individual patients and visits. These data can also drive strategic decision-making by health systems, national programs, and funding agencies. One major obstacle to HIV/AIDS care in developing countries is lack of electronic medical record systems (EMRs) to collect, manage, and report clinical data. In 2001, we implemented a simple primary care EMR at a rural health centre in western Kenya. This EMR evolved into a comprehensive, scalable system serving 19 urban and rural health centres. To date, the AMPATH Medical Record System contains 10 million observations from 400,000 visit records on 45,000 patients. Critical components include paper encounter forms for adults and children, technicians entering/managing data, and modules for patient registration, scheduling, encounters, clinical observations, setting user privileges, and a concept dictionary. Key outputs include patient summaries, care reminders, and reports for program management, operating ancillary services (e.g., tracing patients who fail to return for appointments), strategic planning (e.g., hiring health care providers and staff), reports to national AIDS programs and funding agencies, and research.

Keywords:

edical record systems, computerized; HIV; acquired immunodeficiency syndrome; developing countries; Africa south of the Sahara; Kenya.

Introduction

More than 38 million persons are currently living with human immunodeficiency virus (HIV) worldwide, 25 million (63%) in sub-Saharan Africa where only 1 million (<10%) of those eligible for antiretroviral drugs (ARVs) are receiving them [1]. To respond to the HIV crisis in Kenya, in 2001 Indiana University, Moi University, and Moi Teaching and Referral Hospital created AMPATH – an Academic Model for the Prevention And Treatment of HIV/AIDS [2,3] to leverage the power of universities to establish systems of care, train multidisciplinary providers of HIV/AIDS care, and research innovative methods for enhancing the quality and outcomes of care. As shown in Figure 1, initial grown in AMPATH's care programs was slow until mid-2004 when major funding was obtained. By mid-2007, more than 45,000 patients have been enrolled, 1500-2000 new patients being enrolled per month.

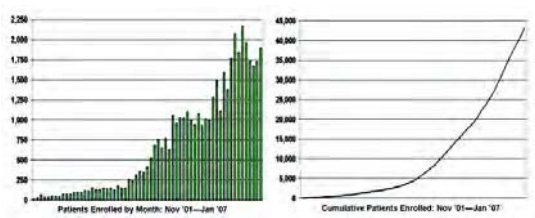


Figure 1 – Monthly and cumulative AMPATH enrollment

Hindering the ramping-up HIV/AIDS treatment programs is the lack of timely and accurate data on (1) the number of HIV-infected patients receiving care, (2) the number eligible for ARVs, (3) the number receiving ARVs, and (4) the positive and negative outcomes of treatment. Funding agencies often deny major program funding unless programs can document the number of patients treated and their outcomes. Paper-based record systems can only meet such information needs for small numbers of patients. The lack of EMRs has slowed the delivery of HIV care despite availability of funds for treatment from the U.S. [4], World Health Organisation [5], and philanthropic foundations [6]. This "Digital Divide" [7] has likely resulted in thousands of unnecessary deaths.

In this paper, we describe how a simple EMR implemented in a single Kenyan rural health centre in 2001 was transformed into a comprehensive, scalable EMR capable of

supporting large (even national) multifaceted HIV/AIDS care programs.

Methods

We describe the process used to initiate the Mosoriot Medical Record System (MMRS) and briefly listing its components, implementation, and use. We then describe how the MMRS was expanded in size and scope into the AMPATH Medical Record System (AMRS). We describe the data model and types of data stored and how the AMRS ultimately failed to support data input/management/output as enrolled patients exceeded 10,000, visits topped 100,000, and the needs of the clinical and funding programs demanded more data flexibility.

We then describe the evolution of the AMRS from its initial implementation as a set of linked spreadsheets into an object-oriented data management system. Finally, we describe how AMRS data serve clinicians, ancillary programs, strategic planning, and reporting to the Kenyan national AIDS control program and international funders.

Results

Initial implementation of the MMRS

In 1999, 3of the authors (WMT, JKR, TJH) visited the six adult and pediatric clinics at the Mosoriot Rural Health Centre to design an EMR as part of the Indiana University-Moi University's NIH-funded medical informatics program. After several days studying local care, redundant data entry in logbooks, and handwritten reports for national reporting, they held discussions with local providers and managers to design the MMRS [8], modeled loosely after the Regenstrief Medical Record System [9]. The MMRS had a 1-page paper encounter form, designed by Kenyan clinicians and used to include a minimum dataset for each visit. After each visit, these data were entered into a computer by a checkout clerk. The MMRS had modules for patient registry, clinical data, reports, and a concept dictionary defining data elements.

The MMRS was programmed in MS-Access® as a set of spreadsheets for clinical observations, laboratory tests, and drugs, all linked by patients' unique ID# (with check digit) and visit date. Data were entered using checkboxes, drop-down menus, or partial name lookup of dictionary terms. The dictionary included term names, synonyms (to ease data entry), term classes (to aid reporting), and charges for drugs and tests. A report module printed standard reports required by the Ministry of Health and for local management activities.

Created and pilot tested in 2000, the MMRS was implemented in 2001 (Figure 2) [10]. Initial use was slow. Adding a second data entry computer (connected by a network cable), closing a clinic exit, and directing patients to the checkout window resulted in use of the MMRS for 100% of visits. Two months later, the logbooks were discontinued. Prior to 2001, creating monthly reports required a half-time clerk. With the MMRS, it took less than an hour. Mosoriot became the #1 health centre in Kenya. Managing drug inventory improved, and quantifying unpaid (charity) increased Ministry of Health funding.

The MMRS also allowed nurses to identify hidden problems: Monthly reports showed a village with few vaccinations (a nurse was sent to educate villagers and vaccinate children). A rabies cluster was identified (the rabid dog was found and shot). A cluster of sexually transmitted infections was noted (the index man was found and treated). A formal time-motion study showed a 10 minute (23%) reduction in patient visits and a tripling of providers' free time from 15% to 46% of the workday. Since 2001, Mosoriot has doubled its patient and visit load but needed no additional staff.

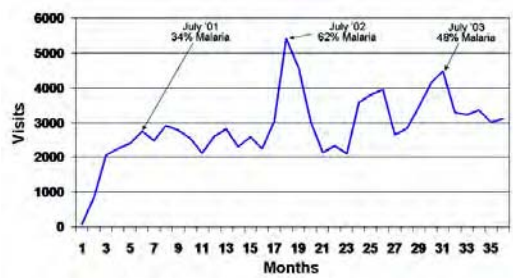


Figure 2 - Visits with data stored in the MMRS, by month

Expanding the MMRS into the AMPATH Medical Record System (AMRS)

In November of 2001, Moi University and Moi Hospital created AMPATH and opened its first 2 HIV clinics. Comprehensive HIV care requires more data and reports to the Ministry of Health and funding agencies. AMPATH directors designed 4 detailed encounter forms for initial and return visits for adult and pediatric patients. As at Mosoriot, data at the point of care were entered onto paper because there was inadequate electricity and funds for computers or hand-held devices. The AMRS' initial visit forms (5 pages long) captured demographics, family/social history, HIV risk, symptoms, prior hospitalizations and medications (ARVs, other antibiotics), alcohol use, physical exam, WHO stage, test results, problem lists and treatments (HIV- and non-HIV-related), and next visit [11]. The 2-page (1 sheet) return visit form collected a subset of the initial data: no symptoms, a limited exam, medication adherence, problems and treatments, and next visit. Importantly, the visit forms were designed by AMPATH clinicians and clinic managers to serve their needs, balancing the amount of data desired and the time it took to record them. Compromises and multiple iterations yielded a consensus minimum dataset [12].

As shown in Figure 3, most AMRS data were entered as checkboxes while the rest used menus and partial name lookup of dictionary terms. Text comments written on the encounter forms were transcribed into AMRS text fields. The data model behind the initial AMRS was the same as the MMRS: a series of MS-Access spreadsheets linked by patient ID# and visit date. The MMRS dictionary was expanded to include additional terms needed to count and describe enrolled AMPATH patients and report what care they received, and report their care and what outcomes occurred [11].

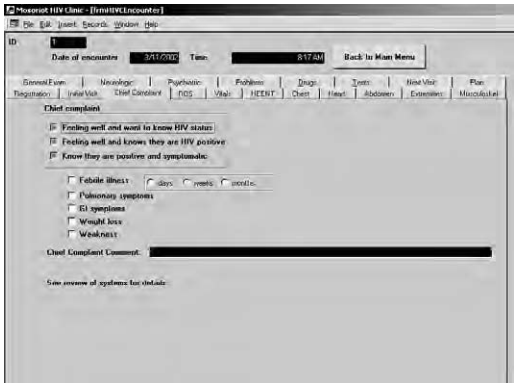


Figure 3 - Initial AMRS data entry screen

Failure of the initial AMRS

Initially, the MS-Access AMRS worked adequately, recording data and providing clinical summaries for providers and reports for AMPATH program managers and funding agencies. However, once the number of patients exceeded 10,000 and visits topped 100,000, the system bogged down. Most symptoms and exam findings were negative, yet an empty field was stored in the AMRS database. The limit of 256 fields per table caused major problems. Variables appearing on multiple forms had to be redefined each time they were used. Patients could only have one AMPATH number, which caused problems when patients visited more than one AMPATH clinic. It was impossible to extract sets of data, such as all patients on ARVs: each drug had to be extracted separately. Thus the MS-Access AMRS became huge, unwieldy, and limiting.

Transforming the AMRS into a Scalable EMR

To deal with these limitations, Regenstrief developers evolved the AMRS data model into one more tightly aligned with the Regenstrief Medical Record System [9]. Written in MySQL, the new, enhanced AMRS [13] has the following components:

Encounter forms

These continue to be defined by the data needs of AMPATH clinicians and managers, led by 3 of the authors (WMN, SK, JJM), They have evolved with changing data and reporting needs. By using checkboxes, menus, and numeric data, the AMRS return visit form takes 1-2 minutes to complete. (Page 1 of the 2-page form is shown in Figure 4, left.)

Patient registry

This defines "who" is providing care. Patients are registered and assigned AMRS numbers. They can have multiple names (e.g., get married, have aliases) and multiple AMPATH numbers: each clinic assigns a local number with a check-digit for ease of filing charts. This database also contains demographics and a history of funding sources. Patients can be linked in families for social, economic, and nutrition interventions.

User registry

This defines "by whom," persons who record, enter, and manage and report AMRS data. Each system user is registered, given a number (with check digit), and provided access and data management privileges needed for his/her required tasks.

Concept dictionary

This is the database of concepts that controls all data entry and extraction. It consists of both terms (e.g., chest xray, diagnosis) and findings (e.g., infiltrate, pneumonia). It defines the format of a term (e.g., coded text, numeric) and the allowable codes or numeric ranges. Terms can be sets (e.g., ARVs) or elements of sets (e.g., AZT) to facilitate data extraction. Term classes include tests, drugs, diagnoses, vital signs, etc.

Encounter site database

This stores "where" care is provided. Each site has a code, and by using the patient identifier, date, and encounter site, an episode of care can be recreated. These codes allow reports to be generated separately for each AMPATH clinic.

Observation database

This stores "what happens" and is the AMRS' heart. Each observation on the encounter forms (other than patient registry data, user ID#, and encounter site) is entered into this database and has 3 components: term ID#, date entered, and result (defined by the dictionary term linked to the data entry screen).Unlike MS-Access, there are no empty fields.

Data lookup

With this program, clinicians and managers look up data on individual patients for making patient and practice decisions.

Data entry

The encounter form is the official medical chart. Once entered into the AMRS, these data serve multiple clinical and managerial needs. Efficient data entry screen mirror the encounter forms which also allows them to be used for direct data entry at the point of care. Figure 4 below shows page 1 of the adult return visit form (filled in) and its AMRS data entry screen.

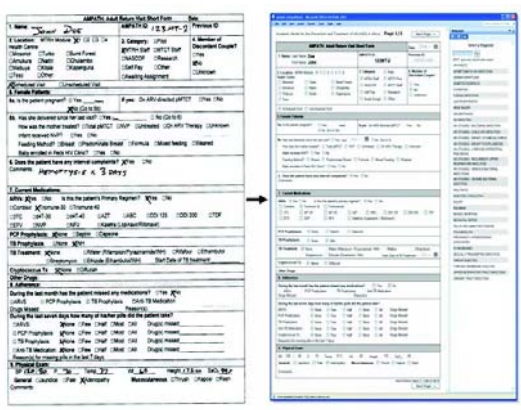


Figure 4 - Paper and electronic AMRS encounter forms

Arden query and reporting system

Using the if/then/else logic of the Arden Syntax [14], this query system provides standard and custom reports such as patient summary flowsheets and reminder reports. Knowledge of clinical medicine, the AMRS database and concept dictionary, and hierarchical data (observations within visits, visits within patients, patients within providers, and providers within clinics) is necessary to generate useful reports [15].

Patient summary and reminder reports

After receiving data for a visit, the AMRS prints out a report that contains diagnoses (with date recorded), WHO HIV stage, drug therapy, drug adherence, serial weight and test results, and reminders (e.g., "Order 6-month CD4 count") [11]. This summary/reminder report is placed in the patient's chart for viewing by clinicians during and between visits.

Non-clinical databases

AMPATH includes programs to prevent HIV infection, provide social and nutritional support, follow-up missed appointments, etc. To enhance efficiency, data security, and avoiding redundant data collection, the AMRS stores data for all ancillary programs and link with clinical data as needed.

Data management

Kenya has limited electricity and Internet access in urban and rural settings. Completed encounter forms are carried by courier to AMPATH's data center where clerks enter data into the AMRS. Data accuracy is enhanced by a financial incentive program. The forms and summary/reminder reports are then returned to the clinics. The AMRS currently contains 10 million data items for 45,000 patients and 400,000 visits to 19 AMPATH clinics. The encounter form is the only medical record, and data are entered into AMRS for virtually all visits. Moving forms between clinics and the data entry centre takes time (~2 days) and petrol. AMPATH will soon move to on-site data entry and printing. Data will be sent to the central database by flash drive or (where available) wireless Internet.

Costs and sustainability

Costs of computers, data entry, and data management are sustained by the agencies that fund HIV/AIDS care. Although originally resistant, these agencies now realize that complete, believable, timely data are essential providing high-quality care and monitoring and evaluating this care and its outcomes.

More constraining is the expertise needed to implement, maintain, and evolve EMRs. For AMPATH, the U.S. and Kenyan partner universities provide sufficient expertise to make this arrangement sustainable. It would be most appropriate for Kenyans to be wholly responsible for managing EMRs supporting care in Kenya. However, the requirements and costs of most medical informatics training programs are currently beyond the reach of most Kenyans. Hence, academic partnerships such as Indiana/Moi can help sustain EMRs in developing countries while training Africans to manage their EMRs.

Uses of AMRS data

Clinicians use the forms and patient summaries and computer reminders [16] for every day patient care. Clinic managers use monthly reports to assess productivity and anticipate personnel needs. For AMPATH, data managers report AMRS data to governmental and funding agencies. Ancillary programs use AMRS data as well: Outreach workers assess no-shows in their homes and facilitate visits. Med nurses assess reasons for non-adherence and counsel patients. Social workers assess financial needs and send patients to AMPATH's economic development program. Nutritionists refer underfed patients to AMPATH's nutrition program, with sufficient foodstuffs from the World Food Program and local farms to feed 30,000 persons/week. Prevention workers help pregnant mothers and newborns to lower HIV risk, caring for those who become infected. Researchers and quality improvement officers assess and improve processes and outcomes of care [17,18].

Expansion beyond Kenya: The OpenMRS Consortium

AMRS developers are committed to improving HIV/AIDS care in developing countries beyond Kenya. In 2004, they teamed up with Partners in Health [19] to expand the AMRS data model into OpenMRS, a free open-source EMR and an international collaboration of implementers [20]. OpenMRS has been implemented in Kenya, Rwanda, South Africa, Tanzania, Uganda, and Lesotho and is being installed in other developing countries. As a result, the Regenstrief Institute is WHO's first Medical Informatics Collaboration Centre. OpenMRS aims to enhance the productivity and accountability of HIV care programs. Key is independence and sustainability of local EMRs. Since early 2006, OpenMRS meetings have been held in Kenya and South Africa, attracting developers from more than two dozen countries. Information about OpenMRS, training, and downloads of programs and encounter forms can be found at www.openmrs.org. Widespread use of OpenMRS, which uses standard messaging formats and data coding, will enhance interoperability between HIV/AIDS providers, enhancing their ability to manage and improve care within and between provider systems.

Discussion

The AMRS has evolved from a simple MS-Access EMR serving a single rural health center in Kenya into a comprehensive, scalable, sustainable, multinational EMR capable of supporting multicomponent HIV/AIDS care. It has succeeded because it meets the needs of clinicians, who need the most data. Others (practice managers, Ministries of Health, funding agencies) need a small subset of clinicians' data. AMRS is the sole medical record; thus clinicians will complete the encounter forms, as they are meeting their own information needs.

The AMRS has succeeded where national HIV/AIDS registry systems have failed, largely because registries–which require completed visit forms to be sent to the Ministry of Health–do not provide any useful information to the clinicians or their practices. Clinicians thus have no incentive to complete these registry forms, and they often do not. Yet to date, many Ministry of Health officials in developing countries are not convinced that the expense of EMRs will

be offset by increased efficiency of care, as was shown with the MMRS [10].

Lack of training in medical informatics for EMR developers and managers will continue to limit the expansion, independence, and sustainability of EMRs in developing countries. Unfortunately, most medical informatics training programs in the U.S. cannot provide stipends to trainees who are not U.S. citizens or permanent residents. This problem needs urgent attention: capacity building will depend on funding from either non-governmental sources in the U.S. (e.g. philanthropic foundations) or governmental sources outside of the U.S.

The AMRS currently uses paper at the point of care. Personal computers, tablets, and other hand-held devices might yield efficient and effective data collection efficiently and accurately. Better Internet accessibility will further enhance connectivity. This should facilitate data collection and management, although even in developed countries, EMRs often use paper forms at points of care [9].

By creating, implementing, and evolving effective data collection and management systems, EMR developers contribute to care by magnifying the abilities of clinicians and managers to care for patients – especially those with HIV/AIDS – who so desperately need it. In no other way can these developers contribute so meaningfully to the battle against HIV/AIDS and other medical miseries that afflict the developing world.

References

[1] UNAIDS. 2006 Report on the Global AIDS Epidemic. Geneva: UNAIDS, 2006.

[2] Voelker R. Conquering HIV and stigma in Kenya. JAMA 2004,292:157-159.

[3] Mamlin J, Kimaiyo S, Nyandiko W, Tierney W, Einterz R. Academic institutions linking access to treatment and prevention: Case study. In: Perspectives and Practice in Antiretroviral Treatment. Geneva: World Health Organization; 2004.

[4] Office of National AIDS Policy. Available at: http://www.whitehouse.gov/infocus/hivaids/. Accessed November 30, 2006.

[5] World Health Organization. Treating 3 Million by 2005. Geneva: World Health Organization, 2003.

[6] Clinton WJ. Turning the tide on the AIDS pandemic. N Engl J Med 2003; 348:1800-1802.

[7] Chandrasekhar CP, Ghosh J. Information and communication technologies and health in low income countries: the potential and the constraints. Bull World Health Organ 2001; 79:850-855.

[8] Hannan TJ, Rotich JK, Odero WW, Menya D, Esamai F, Einterz RM, Sidle JE, Sidle J, Tierney WM. The Mosoriot Medical Record System: Design and initial implementation of an outpatient electronic record system in rural Kenya. Int J Med Informat 2000; 60:21-28.

[9] Overhage J, McDonald CJ, Suico JG. The Regenstrief Medical Record System 2000: Expanding the breadth and depth of a community wide EMR. Proc AMIA Symp 2000; 1173.

[10] Rotich JK, Hannan TJ, Smith FE, Bii J, Odero WW, Vu N, Mamlin BW, Mamlin JJ, Einterz RM, Tierney WM. Installing and implementing a computer-based patient record system in sub-Saharan Africa: The Mosoriot Medical Record System. J Am Med Informat Assoc 2003; 10:293-303.

[11] Siika AM, Rotich JK, Simiyu CJ, Kigotho EM, Smith FE, Sidle JE, Wools-Kaloustian K, Kimaiyo SN, Nyandiko WN, Hannan TJ, Tierney WM. An electronic medical record system for ambulatory care of HIV-infected patients in Kenya. Int J Med Informat 2005; 74:345-355.

[12] Tierney WM, Beck EJ, Gardner RM, Musick B, Shields M, Shiyonea NM, Spohr M. A pragmatic approach to constructing a minimum dataset for care of patients with HIV in developing countries. J Am Med Informat Assoc 2006; 13:253-260.

[13] Mamlin BW, Biondich PG. AMPATH Medical Record System (AMRS): Collaborating toward an EMR for developing countries. Proc AMIA Symp 2005; 490-494.

[14] HL7 (Health Level Seven). Arden Syntax for medical Logic Systems. Ann Arbor, MI: American Society for Testing and Materials (ASTM), 1999.

[15] Tierney WM, McDonald CJ. Practice databases and their uses in clinical research. Stat Med 1991; 10:541-557.

[16] Shea S, DuMouchel W, Bahamonde L. A meta-analysis of 16 randomized controlled trials to evaluate computer-based clinical reminder systems for preventive care in the ambulatory setting. J Am Med Inform Assoc 1996; 3:399-409.

[17] Wools-Kaloustian K, Kimaiyo S, Diero L, Siika A, Sidle J, Yiannoutsos CY, Musick B, Einterz R, Fife KH, Tierney WM. Viability and effectiveness of large-scale HIV treatment initiatives in sub-Saharan Africa: Experience from western Kenya. AIDS 2006; 20:41-48.

[18] Nyandiko WM, Ayaya S, Nabekwe E, Tenge, C, Sidle JE, Yiannoutsos CT, Musick B, Kigotho E, Wools-Kaloustian K, Tierney WM. Outcomes of HIV-infected children on anti-retroviral therapy at AMPATH clinics in western Kenya. J Acquir Immune Defic Syndr 2006; 43:418-425.

[19] Fraser HS, Jazayeri D, Nevil P, Karacaoglu Y, Farmer PE, Lyon E, Fawzi MK, Leandre F, Choi SS, Mukherjee JS. An information system and medical record to support HIV treatment in rural Haiti. BMJ 2004; 329:1142-6.

[20] Mamlin BW, Biondich PG. AMPATH Medical Record System (AMRS): Collaborating toward an EMR for developing countries. Proc AMIA Symp 2005; 490-494.

[21] Biondich PG, Mamlin BW, Hannan T, Tierney WM. A call for collaboration: Building an EMR for developing countries. Proc AMIA Symp 2005; 894.

Address for correspondence

William M. Tierney, MD, e-mail: wtierney@iupui.edu.

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

São Paulo City Health Information System – A Case Report

Cláudio G A Costa[a], Beatriz F Leão[b], Lincoln A Moura Jr[b]

[a] *São Paulo City Department of Health, São Paulo, Brazil*
[b] *ZILICS – Health Information Systems, São Paulo, Brazil*

Abstract

São Paulo is the largest city in Brazil and one of the largest in the world. In 2004, São Paulo City Department of Health decided to implement a Healthcare Information System that would support managing healthcare services and provide an ambulatory health record. The system was designed to build on from national standards that identify healthcare workers, organizations and the relationships among them and reuse as much as possible existing concepts and software. At the same time, the system should reduce fragmentation, not only by integrating existing redundant or competing systems but also by providing a framework in which new modules would be naturally integrated to each other. Today, the web-based system, known as SIGA Saúde, runs in more than 370 healthcare units, processes some 8 thousand appointment scheduling daily and more than 10 thousand high-cost procedure authorizations monthly. This paper profiles the São Paulo City and its needs and describes the project, obstacles and results so far.

Keywords:

healthcare information systems, systems integration, information management

Introduction

SUS – the Brazilian health system

SUS, the Brazilian National Health System was created in 1998 when the Brazilian Constitution was approved in general elections. However, SUS is an evolving project whose history stretches back to as far as the Seventies [1]. SUS is required to provide free-of-charge health services based in principles: universality, integrality, and equity, meaning that all individuals are equally entitled to integral treatment. SUS has a complex organization, but it suffices to say it is funded and run jointly by the three levels of Government: federal, state and municipal. Under SUS, ideally, health services will be delivered and managed at city level, according to pre-defined and agreed programs, which are constantly monitored by the three levels.

Brazil has a long tradition of using Health Information Systems, and SUS data collection is impressive [2, 3]. However, the history of healthcare information systems has been characterized by system fragmentation as most application systems were developed to sort out specific problems, thus using specific vocabularies, patient and workers IDs. As a result of that some reports have counted more than 200 health information systems of different complexities without any integration whatsoever [2, 3].

That scenery started to change when, in 1999, the Ministry of Health proposed the National Health Card Project whose main objective was to define standards for uniquely identifying the individual, the health organization and the healthcare worker as well as to define a core patient data-set for registering the healthcare encounter in the country [3].

In parallel and using the same common ground, SUS started devising means to control patient flow. Patient flow control is currently regarded as the cornerstone to optimizing health resources and ensuring equity of access to health services. Essentially, patient flow control means ***intelligent scheduling*** of specialized consultations, exams and inpatient admission, including emergency. Intelligent scheduling means that allocating resources to answer to an individual's needs takes into account all relevant factors, such as distance, effectiveness, budget and cost. Although all treatment under SUS is offered free of charge, several procedures have to be authorized beforehand. Therefore, part of patient flow control relates to authorizing procedures and inpatient admissions alike.

Patient flow should follow a three-layer model, as depicted below. On the entry layer, lies primary care; on the second specialized treatment and, on the third layer, lies high-complexity treatment. Patient flow to the second and third layer follows recommended regional pathways, thus creating the reference-counter-reference model.

In 2000, the Health Ministry commissioned the development of SisReg – a patient flow control system – that use similar but unfortunately not the same data models of the National Health Card Project.

Both SisReg and the National Health Card projects were moderately successful but their evolution lacked continuity as governments changed hands. Nonetheless, these projects had major impact on how Health Information Systems become to be regarded. By the beginning of 2003 it was clear to most public health services managers (all pertaining to SUS) that national standards are vital, that patient flow control is an essential tool if SUS is ever to become a reality throughout the nation. It had also become

clear that, without proper Health Information Systems, no healthcare service can be operated properly [4].

São Paulo city health system

São Paulo is one of the largest cities in the world, with 10.3 million people in the city and some 18 million in the Metropolitan Area.

In June 2003, São Paulo Public Health came to operate as a "full managed-care" city, which means that resources from the National Health Fund are transferred directly to São Paulo City Department of Health (SPCDH) on a capitation basis. In exchange, the city agrees to meet certain production and quality goals, and has to send monthly reports that allow the Ministry o Health to assess if those criteria are being met.

By becoming a "full managed-care" city, São Paulo Dept of Health came to manage the following yearly figures:

- 10 million primary care consultations
- 8.5 million specialized consultations
- 550 thousand hospital admissions

São Paulo decided that investing in an Information System that would support Patient Flow Control and provide an Electronic Health Record System [5]. The core conditions that underlined the project can be summarized as:

- The information system would not only be fully compliant with all National Standards; but also help promote them;
- Open standards and open source-code should be used at all levels, whenever feasible;
- Whenever possible the project should use the results of previous projects, existing technologies and concepts;
- São Paulo City would receive source-code and consulting support from the Health Ministry and, in return, would send back to the Ministry all deliverables from the project;
- Finally, and most importantly, the system should be fully integrated and provide a framework for continuously embodying new functions in an easy and natural way.

The requirements and concepts summarizes above were described at length in a Term of Reference prepared by São Paulo City Department of Health. The Term of Reference defined 4 major sub-projects that should result in one and only Information System:

Municipal health register, whose objective is to handle and process the identification data for health care users, workers and organizations, as well as the relationships among them. The Register is the prime data source for all other modules, as no operation can be carried out unless its actors are registered.

This subsystem also stores and processes all standard vocabularies in use within SUS. All data within the Municipal Health Register are fully compliant with SUS

Patient flow control handles all requests for health care services (consultations, procedures, inpatients admission and emergency) and finds the best possible match, based on criteria such as budget, distance, availability and waiting time. This module also processes authorization requests for high-cost high-complexity procedures. This is standard-procedure under SUS Also handles exceptions, i.e. whenever resources use exceed predefined limits or are unavailable, an accredited doctor handles the exception, either by extending the budget, finding available resources or holding the request for some time, if that is the case.

It also processes patient flow control within the health care unit and copes with waiting lists, either local or in other reference layer.

Ambulatory electronic health record collects an essential dataset from the encounter and triggers related actions, such as notifying diseases or work-related diseases, when such conditions are met.

Role-based access control system is a single sign-on system that identifies the user and its user profile, thus enabling or disabling access to system's functions. Of course, all users have to be recorded in the Register, before being authorized to access the system. Initially, some 40 profiles were defined. Through system usage the Dept of Health decided to delete some and created other profiles, totaling more than 60 profiles currently.

To develop the project, SPCDH hired several companies that, under its management, focused on specific subprojects or on threads such as hosting, communications, equipment, software development, training and support.

Methods

In order to achieve the proposed goals, five main topics were considered:

Project management

From the very beginning, management was recognized as the most important aspect of the project. In order to cope with that task, three management committees were put together. At the topmost level, the Strategic Committee included the Health Secretary and the top management of all threads. The main purpose of the Strategic Committee was to build consensus among players and made decision on strategic issues such as overall budget, timing and political issues that involved organs or activities that went beyond the Information System's borders. The Managing Committee was formed by SPCDH IT Manager and from representatives from the several threads,. The Committee met together regularly to review progress and take the required action. Finally the Executive Committee formed by project coordinators that would look after every thread's development. Each thread was developed using its own methods that have been submitted and approved by the Managing Committee. Decision has always been taken by consensus. Whenever one level failed to reach consensus, the decision was taken to the upper level, so that eventually it would become a Secretary's decision.

From time to time, workshops and meetings would be held to celebrate major progress or to inform of important decisions or course changes.

Technology

The project was based on open-standards and free-software open-source paradigm. Java Technology was chosen from the first moments for its ability to generate systems that run on any platform; to make extensive use of object-oriented analysis, and to create reusable software components. Although the system should run on a variety of equipment, a basic platform was defined with Linux as the standard operating system, JBoss as the application server and Oracle as the database management system. The only proprietary piece of software chosen was Oracle, as for the foreseen volumes no free-software database management system was considered suitable for the task. The system was developed in three-layers as depicted in Figure 1.

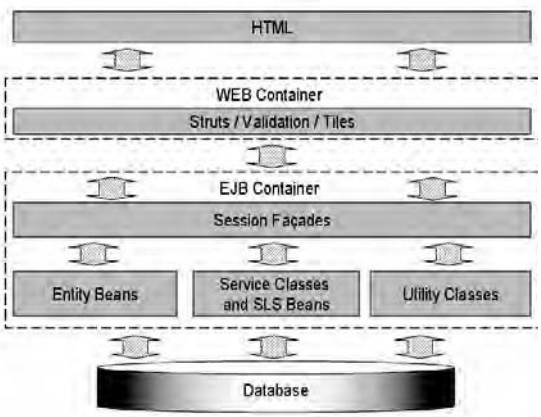


Figure 1 - System architecture overview

System specification

System's functionality was specified using the Unified Process, heavily based on use cases. Users and experts were assembled in forums to review and improve old use cases as well as propose new ones. A trained health informatician led the forum meetings with the aim of identifying and recording all necessary use cases. Throughout the project, use cases' granularity has been chosen to match basic system's functions and associated software components, in such a way that reusability is improved. Once implemented, use cases were validated by comparing their actual operation with use case's description.

In order to allow for full information integration, a Conceptual Information Model was devised to cope with the complexity of the Domain. The Model is described elsewhere, but its schematic representation can be seen in Figure 2.

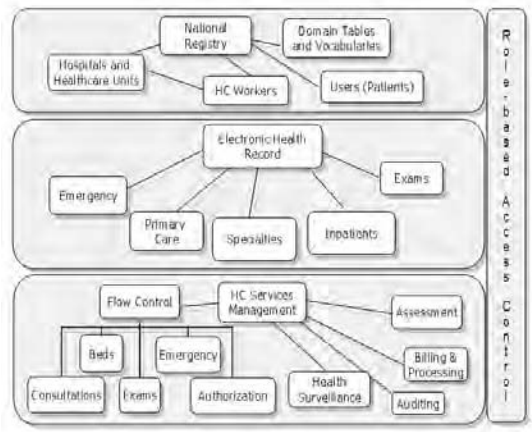


Figure 2 - Health information conceptual model

Human resources

Training was one of the most important aspects to consider, as the number and variety of human resources to be conquered and trained was enormous. The approach to training and HR involvement was based on replication, meaning that the company hired to develop the system gave training to some 20 Dept of Health staff members that, in turn, trained final users. Training was associated with the several subprojects, or modules, described above, so that blocks of similar functions were taught at each time. Additionally, assisted operation took place as equipment was deployed at the health care units. Some 700 computer science students were hired and trained to help end-users at the point of care, for as long as 5 weeks.

System deployment

Deployment of such a huge system was devised as a vertical-to-horizontal process, in which the system was to be fully implemented in one of São Paulo City Health Regions and, as it became stable, it would be rolled out to other regions. However, as a new City administration was elected and took office in January 2005, the deployment method was changed to completely horizontal. In other words, the modules related to registering patients, health care workers and health care units in all health units before moving to new functionalities. As that was a very slow process since it depended on all health units using the system at the same level, the process was reviewed once again, so that a mix of horizontal and vertical process are in place.

Figure 2 shows a schematic representation of the system as it has been deployed. Web-machines at point of care are connected via ADSL to the São Paulo City Datacenter. Via Internet authorized users have access to the system, which is represented in its major modules.

Figure 3 - Schematic representation of system deployment

Results

The system thus developed came to be known as "SIGA Saúde" which means "following health" and is an acronym for "Integrated System for Health Care Management".

As of this writing, SIGA Saúde is in use in 372 primary care units for a) registering patients, workers, health units and their services; b) scheduling local appointments.

The actual numbers of SIGA deployment as of November 6 to 12, 2006 as informed by São Paulo City Health authority are described in Table 1, below.

Table 1 - SIGA Saúde production, October 2006

Item	Amount
Primary Care Units Using SIGA Saúde	372
Primary Care consultations scheduled	1,657,023
Specialized Consultations scheduled	35,250
High-Cost Procedures Requests Processed	223,225
Patients Registered	8,357,863

Figure 4 shows the evolution in number of overal appointmens made using SIGA Saúde for the 5 São Paulo Regions, since October 2004, when the system was launched. Figure 5, in turn, shows the actual number of consultations performed for the same Regions in the same period.

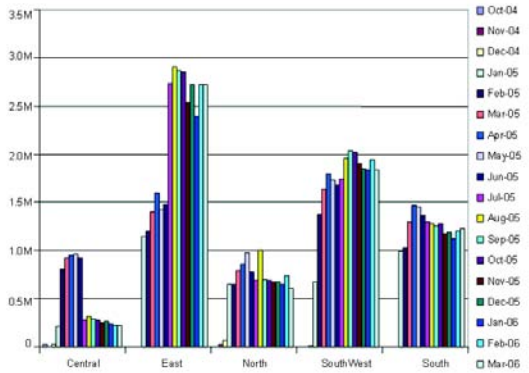


Figure 4 - Evolution of SIGA Saúde use for appointment scheduling, in São Paulo's 5 regions

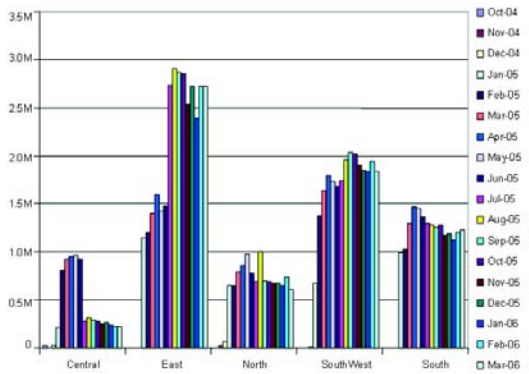


Figure 5 - Evolution of actual number of consultations delivered in São Paulo's 5 regions

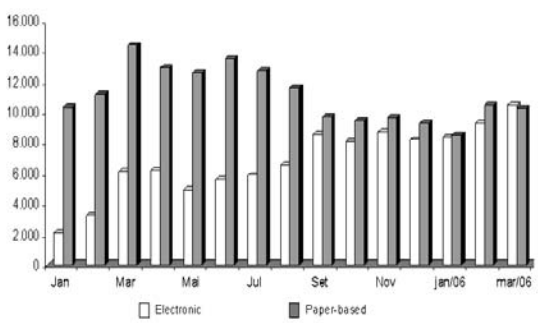


Figure 6 - Evolution of the number of high-cost procedures requested on paper and electronically since SIGA Saúde came to place

Discussion and conclusions

Impact data related to using SIGA Saúde is still under collection. However, a preliminary assessment reveals that outpatient services productivity has increased about 35%.

Patients' perception has changed for better, as patients have stated they can schedule appointments within the same month, against a three-month wait that was typical before SIGA Saúde. The Brazilian Press, usually very critical of public IT services published an opinion article stating their view that SIGA Saúde was a major contribution to the Bazilian Health System [6]. The project has also won a Duke's Choice Awards, granted by JavaOne 2005 Conference Meeting [7].

SIGA Saúde architecure is very innovative. By being fully web-compliant and using public Internet for providing safe role-based access to the system, SIGA Saúde can easily be deployed in other cities or regions. Also, as SIGA Saúde is fully compliant with all Brazilian standards and SUS practices, it is a tool to be used to help SUS itself be implemented throughout Brazil. Finally, São Paulo City's dimensions and complexity are so evident that taking SIGA Saúde to other cities and states requires simplification rather than new functionalities.

SIGA Saúde is a joint property between the Health Ministry and São Paulo City Department of Health, that are willing to share SIGA Saúde with other Public Service organizations, within the country or abroad.

SIGA Saúde architecture has several advantages when compared to conventonal client-server systems:

- The datacenter model allows that all complexity, such as backup, data model maintainance and system evolution, be kept away from the end-user, who accesses the system via a standard browser;
- All the health care unit needs is an Internet access and a browser. No local server is required at health units. When a point-of-care web-computer breaks down it is simply replaced by another one;
- The system can be shared by several cities in what has become increasingly common in Brazil, which are City Consortia. One single server can attend several cities that share services and or patients;

Next steps

SIGA Saúde has also been installed in São Paulo State Dept of Health to help identify and control the flow of hemophilic patients and hemodialisys. According to SUS model, the management of those aspects are responsibilities of States. At the moment, the major effort is to integrate State and Municipal databases. The problem is more of organizational than technological complexity.

São Paulo City Department of Health aims at putting together a communitiy of cities and states that use SIGA Saúde, so that furthere development can be carried out in a shared fashion, thus reducing costs. As SUS is very similar throughout Brazil, that is a feasible task. However managing shared development is a very complex task, as strong methods are required to deal with distributed object-oriented development based-on software components. Although SIGA Saúde has been built using a Development Framework that embodies such concepts, it is not true that cities and states are ready for such an enterprise.

At the moment, SIGA Saúde is being integrated to the information system of 3 clinical analysis labs that provide services to the Public Health System, using HL7 v3 and LOINC. With that, all lab orders and their results will be available on-line at the point of care.

References

[1] Souza RR, El Sistema Público de Salud Brasileño, Proceedings of Internacional Conference on Trends and Challenges for Health Systems in the Americas, São Paulo, Brazil, August 2002.

[2] Leão BF, Bernardes MM, Levin J, Moura L, Bandarra E, Modesto LM, Sousa AM, Cunha RE, Filho MR. The Brazilian National Health Informatics Strategy. Medinfo. 2001; 10(Pt 1):38-42.

[3] Lemos M, Leão BF. Standards adopted in the Brazilian National Health Card Project. IN: Proceedings of the 8th Conference on Nursing Informatics, Rio de Janeiro, June 2003.

[4] Leão BF, Valente GC, et alli. O Desafio de Integrar os Sistemas de Informação em Saúde. Proceedings of the IX Brazilian Conference on Health Informatics, Ribeirão Preto, SP, Brazil, Novemeber 2004.

[5] Moura LA, Vecina-Neto G, Lira ACO, Leão BF, Hirano EM, Troccoli FT, Pacheco R, Castanha LHA. A Construção do Sistema de Informação da Secretaria Municipal de Saúde de São Paulo. Proceedings of the IX Brazilian Conference on Health Informatics, Ribeirão Preto, SP, Brazil, November 2004.

[6] Opinion, A Saúde na Era da TI (Health Care in the Age of IT), O Estado de São Paulo, 23 August 2006, São Paulo, SP, Brazil.

[7] Duke's Choices Award, www.sun.com/javaone/

Address for correspondence

Dr. Cláudio Giulliano Alves da Costa, MD is São Paulo City Department of Health IT Manager and can be reached on +55 11 3218-4244 or by email on cgcosta@prefeitura.sp.gov.br.

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Experience in Implementing the OpenMRS Medical Record System to Support HIV Treatment in Rwanda

Christian Allen[a,b], Darius Jazayeri[a], Justin Miranda[a], Paul G Biondich[c], Burke W Mamlin[c], Ben A Wolfe[c], Chris Seebregts[d], Neal Lesh[a], William M Tierney[c], Hamish SF Fraser[a,e]

[a]Partners In Health, Boston, USA.
[b]Inshuti Mu Buzima, Rwinkwavu, Rwanda.
[c]Indiana University School of Medicine and the Regenstrief Institute, Indiana, USA.
[d]Medical Research Council, Cape town, South Africa.
[e]Harvard Medical School, Boston, USA.

Abstract

The challenge of scaling up HIV treatment in Africa has led to a new emphasis on improving health systems in impoverished areas. One aspect of this is the development and deployment of electronic medical record systems to support HIV and TB treatment. In this paper we describe the design and implementation of a new medical record architecture to support an HIV treatment program in rural Rwanda. The architecture is called OpenMRS and it has been developed to address the problem of configuring EMR systems to suit new sites, languages and diseases. OpenMRS uses a data dictionary called the concept dictionary to represent all the possible data items that can be collected. This allows new items to be added to the system by non-programmers. In addition, there are form creation tools that use drag and drop web technologies to simplify form construction.

The OpenMRS system was first implemented in Kenya in February 2006 and then in Rwanda in August 2006. The system is now functioning well and we are developing extensions to improve the support for the clinic. These include improved, easy to use reporting tools, support for additional clinical problems including nutrition and child health, better database synchronization tools, and modules to collect laboratory data and support the pharmacy. The system is also in use in South Africa and Lesotho and is being deployed in Tanzania and Uganda.

Keywords:

electronic medical records, developing countries

Introduction

The HIV epidemic centered in Africa is probably the greatest medical challenge facing the world today. Currently about 40 million people are infected with the virus and nearly all will die in the next decade without anti-retroviral (ARV) treatment. Real progress is now being made in ARV treatment for patients in impoverished developing countries spearheaded by the WHO 3x5 initiative[1] and supported by the Global Fund to Fight AIDS, TB and Malaria, the US PEPFAR program, EC countries and other government and private funds. Partners in Health (PIH) pioneered the treatment of HIV in the extremely impoverished environment of the Central Plateau of Haiti in 1999 and has now provided ARV treatment to over 3000 HIV patients in that country. Beginning in May of 2005 PIH started an HIV treatment program in some of the poorest rural areas of Rwanda[2]. There are now 3400 patients being monitored and over 2000 patients receiving ARV treatment. All patients on treatment have a baseline assessment of their clinical status, weight and CD4 count. Those patients with CD4 counts below 350 or who have evidence of opportunistic infections are promptly started on triple ARV therapy. Pregnant women who are HIV positive are also started on triple ARV therapy in the 3rd trimester. All treatment is directly observed by a community healthcare worker who visits the patient's home daily. This is to ensure excellent compliance and alert the clinic of any problems that occur between the monthly clinic visits. Patients with malnutrition are given food supplementation monthly.

To support the process of finding patients, initiating therapy and monitoring the patients' treatment we have set up a web-based electronic medical record system in Rwanda. This is used for all the HIV patients, with a focus on those that are on active ARV treatment. It is based on the OpenMRS electronic medical record architecture jointly developed by Partners In Health and Regenstrief Institute in Indiana. The OpenMRS system has now been implemented in Kenya, South Africa and Lesotho as well as Rwanda. We describe here the OpenMRS system, the implementation in Rwanda and the future plans for the OpenMRS system in that country and around the world.

Background

Electronic medical records (EMRs) are becoming accepted as an important tool to support high quality healthcare in the US, Europe and other developed countries. This approach has been driven by evidence that the use of EMR systems improves quality of care[3] and reduces medical

errors[4] and unnecessary medical investigations[5]. Experience in the use of EMR systems in developing countries is much more limited but there is now considerable interest in medical information systems to support the treatment of HIV and TB in Africa, Latin America and Asia. The Regenstrief Institute in collaboration with Moi University in Kenya developed the Mosoriot Medical Record System (MMRS) to record clinical data on general patient visits to clinics in western Kenya[6]. This system was built using Microsoft Access and was subsequently modified to support the care of several thousand HIV patients[7]. However, the design and technology was not easily extendable to cover new diseases or larger numbers of patients. Baobab Health in Malawi have developed an EMR system using innovative, low power touch screens for data entry and display[8]. This system is now used to support the care of more than 7000 HIV patients in the Lighthouse clinic in Lilongwe and is being assessed by the national HIV program for use throughout the country. Careware® an HIV medical information system developed for US patients has now been deployed in Uganda[9] and is planned for use in other African countries and in Latin America. Another system called SmartCare (and based on smart cards) is operational in Zambia. Medecins Sans Frontieres has deployed a system called Fuchsia to clinics where they provide HIV treatment.

The PIH-EMR system

In 2001, Partners In Health developed the PIH-EMR[10] a web-based medical record system to support the treatment of patients with multi-drug resistant TB (MDR-TB) in Peru. That system was based on an Oracle database and uses the Apache web server, Java server pages and the Tomcat servlet engine. It now has records on over 15,000 patients, more than 6000 of who have started or completed treatment for MDR-TB. The system is designed to collect data on the baseline patient assessment, bacteriology test results, drug regimens and patient outcomes. It includes extensive tools for data analysis and drug consumption forecasting as well as communication tools to improve the quality of laboratory data and reduce delays in starting patients on the correct treatment. The PIH-EMR has been evaluated in several studies that have shown that it can reduce: data errors[11], delays in data access and communication, and the workload of documentation and reporting[12]. The PIH-EMR is also used by an MDR-TB treatment program in the Philippines[13].

The HIV-EMR system

In 2002 the PIH-EMR was modified to support the treatment of HIV in rural Haiti[14] as part of the expansion of care funded by the Global Fund. The HIV-EMR is now used in 9 clinics in rural Haiti and collects data on patients' clinical history and examination, drug regimens, lab tests including CD4 counts, and follow-up data including opportunistic infections and medication side effects. The HIV-EMR is accessed by satellite Internet access in each of the clinics in Haiti. It is used to generate reports on patient treatment for the Global Fund and PEPFAR programs and to detect patients whose treatment may be delayed or incorrect. An additional component to the HIV-EMR allows staff to enter data while offline and then synchronize the data when the Internet connection is available. A particularly important extension to the EMR is a web-based pharmacy tracking system that tracks all medications and supplies for the clinics (not just for HIV patients) that handled almost 1.8 million patient visits in 2006.

In 2005, the HIV-EMR was set up to support patient care in our new clinic in Rwinkwavu, Rwanda[15]. The system was modified to match the local conditions including clinic details and Rwandan reporting requirements. A team of data entry staff and a data manager was hired and trained, and data were entered on over 1500 HIV patients taking ARV medication. This implementation was successful in less than one year, though it had the benefit of a full-time programmer on-site in Rwanda. It was clear from the experience in Rwanda and the Philippines that our existing architecture was challenging to set up and modify to suit new projects.

Motivation for a new EMR architecture

As there is a great need to provide software to support new HIV and MDR-TB treatment projects, we needed to develop a system that was very flexible and scalable, but that did not require expert programming to add new forms or tailor it to new sites, languages or diseases. Such a system would ideally use open source code to make it as widely accessible as possible by sites with limited funding. It would also benefit greatly if it could be developed by multiple programmers and groups, to share ideas and programming efforts. This would provide an avenue for other projects to contribute to the system and become part of the team rather than just passive recipients of the code. We believed that the system should retain the clear benefits of a web-based system while allowing local "offline" data entry. It should also provide excellent tools for data analysis and reporting, with the option of exporting raw data for further analysis using standard tools such as statistical packages.

We looked at existing EMR systems and found none that fulfilled these requirements. Most commercial medical record systems are closed, proprietary and, typically, expensive. They are not designed to be extended by the sites that use the system and it is often difficult to extract data for analysis or to upload to other systems. It was essential the system used was open to modification for different languages and local requirements. The small number of open source EMR systems do not have the characteristics required for projects with limited equipment, bandwidth and technology support. A particular requirement was for a data model that explicitly modeled medical concepts and data and that could be viewed and updated by experienced users, not just programmers.

Progress to date

OpenMRS system development

The approach we chose was to create an open source electronic medical record architecture through collaboration between Partners In Health and Regenstrief Institute and,

subsequently, the South African Medical Research Council. The system that came from that collaboration is OpenMRS[16]. The critical difference between OpenMRS and most other systems, such as those previously used by both PIH and Regenstrief (in Kenya), is the use of a comprehensive data dictionary for all clinical data. This allows new data items to be added to OpenMRS without programming or modifying the underlying database structure. All patient data apart from basic name and demographics are coded as concepts in the data dictionary. Data entered into the EMR are added to an observation table along with the date and time of entry. Orders such as drug prescribing are added to an orders table in a similar way. The concepts can be linked to existing coding systems such as LOINC and SNOMED, and can be shared between forms or between different projects. The dictionary also simplifies the process of translating data into different languages and creating libraries for different diseases or clinical problems. Data types supported by concepts include numeric, text and date/time. The OpenMRS concept dictionary allows concepts to be grouped together into concept sets to represent more complex relationships than single concepts alone. We are developing tools to create derived concepts that link together concepts with rules created in Arden syntax[17]. A derived concept could perform calculations on age and weight, for example. In addition, the Arden syntax interpreter will simplify creation of sophisticated decision support rules to drive alerts and reminders.

The OpenMRS system is built in Java using the Spring application framework and the Hibernate object-relational persistence system on top of a MySQL database back-end. (MySQL could be replaced by any RDBMS supported by Hibernate, including Oracle, MS SQL Server, and PostgreSQL.) OpenMRS's web front-end is served by the Tomcat servlet container, and is built in Java Server Pages with JSTL, with a liberal use of AJAX (implemented with DWR), and JavaScript widgets implemented in Dojo. All of these applications and libraries are open-source. We also use an XML-based form creation tool from Microsoft for certain data entry forms. OpenMRS has an API that allows software modules to be added to the core system, simplifying the creation and additional of new functionality. The technologies used for OpenMRS have been described in detail[16] (see also: www.openmrs.org).

The OpenMRS system was first implemented in Kenya in February 2006 with data entered from paper forms by the data entry team. Over 10 million observations have been entered on 48,000 ARV patients and the data are now being used to support the clinical care process and reporting. With their experience in using the web-based HIV-EMR, PIH staff use the EMR in a more interactive way and required extra tools for data entry, viewing, and analysis.

Implementation in Rwanda

We started with the HIV-EMR installation that had been running for one year[15] and developed a tool to migrate the data to the OpenMRS system. The OpenMRS was customized to use a look and feel, and navigation specific to the PIH project in Rwanda. We devised new intake and follow-up forms for use in the clinic, first on paper, and then converting them into the OpenMRS format using the Form Builder tool[16]. We then created the necessary reports in a reporting module so that we could generate output for national-level reporting, clinical support, administrative overviews, and research. Finally, we added custom modules for items such as batch data entry (to speed up repetitive tasks) and one-page clinical patient summaries. We are also creating forms and reports to support related treatment programs not included in our previous systems such as malnutrition and home visits.

The OpenMRS system is now running on a server in Rwinkwavu, Rwanda and is in daily use across six ART sites. Data are entered by a total of 11 data entry staff and reports and patient summaries are printed for doctors, nurses and other staff. While three of the sites (including Rwinkwavu) enter data from their respective sites, data from the three clinics without Internet connectivity are being entered in Rwinkwavu. We next plan to set up replicating servers in the other five sites that report back to the server in Rwinkwavu.

Laboratory data collection application

We have developed a simple laboratory data collection application using MySQL database and programmed in Java for use in Rwanda and similar sites. It allows users to search for a particular patient by name or ID and can show all of that patient's lab orders and results. Results can also be viewed in a register format to assess activity in a given time period. It can run as a "stand alone" system, but it can also be set up to send out alerts about critical results to those who subscribe. Using a simple module that connects to Skype™, these alerts can be sent as an SMS message to a clinician's mobile phone. We are currently refining the HL7 API to allow the data to be automatically uploaded to the OpenMRS system. Third party laboratory systems that can support HL7 data export should also be able to work with OpenMRS.

Reporting tools

To be useful to a wide range of staff it is essential that the reporting and analysis functions of OpenMRS are easy to use and flexible, while also providing access to the full expressiveness of the data model. This involves providing simpler user interfaces and pre-defined reports for some users, and richer user interfaces, user-defined reports, and raw data exports for more sophisticated users. The OpenMRS reporting tools consist of five distinct components: 1) a cohort definition tool, 2) a data set definition tool, 3) a data export tool, 4) a report designer, and 5) report generation modules.

1. *The cohort definition tool* is built into the OpenMRS API and web interface, and allows users to perform analysis on patient groups, based on patient characteristics, observational data, orders, program enrollments, or more complicated rules written in Java (and soon Arden syntax). It also allows users to save cohorts for future use, define them ad hoc, and generate predefined reports such as patient summaries for upcoming visits.

2. *The data set definition tool* is also built into the OpenMRS API and web interface and is closely tied to the OpenMRS concept dictionary. It allows users to specify the data columns (e.g. OpenMRS concepts) needed to implement a report. Data set definitions will be used to specify the data requirements for a report at the report design stage and will also be used by the data export tool at the report generation stage.
3 *Multiple data set definitions* can be defined per report. This allows users to define different data sets for each of the elements in their report. This is useful in scenarios where a single report may require different types of data (e.g. a graph showing enrollment data by age and gender and a table containing data about encounters at a specific clinic). Rather than requiring a user to define a single, monolithic data set which can only be used for that particular report, this feature allows users to define specific data sets for reuse in other reports.
4. *The data export tool* is also built into the OpenMRS API and is similar to the data set definition and cohort definition tool. This tool allows users to choose a cohort and data set definitions and export data in a format that will allow them to perform analysis within a wide-range of reporting and analysis tools. OpenMRS currently supports data export as a CSV (Excel) file, the most commonly used data export format. In the future, it will be extended to provide support for other formats, such as XML and tab-delimited files, as well as data export formats that are implemented through custom modules. The data export tool is also used to interface with the report generation modules (discussed below) to provide seamless data integration at the report generation stage.
5. *Report generation* is implemented within OpenMRS as a separate module. This allows OpenMRS to support a wide range of reporting frameworks, including open-source projects such as Pentaho, JFreeReport, JasperReport, and Eclipse BIRT. The first report generation module implementation will use the BIRT reporting framework. The reason we chose to initially support BIRT was based on its easy-to-use and flexible report designer and support for complex reporting features, such as cross-tabulation, joined data sets, and multiple data set support. In addition, since the BIRT tool is based on the Eclipse platform, OpenMRS developers can implement plug-ins to add new functionality. We also plan to implement an Open Data Access driver that interfaces with OpenMRS and acts as a JDBC-like driver to retrieve concepts and data through the OpenMRS API.

Implementing modules for each reporting framework allows OpenMRS to extend the functionality of the core reporting components without tying it to a particular reporting framework.

Pharmacy data

A critical use of EMR systems is tracking the drugs required for the group of patients for the next six to twelve months. In addition, to forecasting drug requirement from the drug regimen data we also need to track shipments of drugs to our warehouses and pharmacies and the weekly consumption of medications. We added a component to the HIV-EMR for Haiti to allow pharmacy staff to enter medication transactions in electronic stock cards[14]. The system now tracks all movements to and from the warehouses and allows staff to order drugs from the central depot and check on their order status online. Weekly reports are generated automatically that warn of any impending stockouts or excessive stock in each site. This system is separate from the OpenMRS system but is being linked to allow regimen data from the OpenMRS system to drive the dispensing of medications in the pharmacy for each patient. As with the lab system, other pharmacy software should function with OpenMRS if it supports standard data exchange protocols.

Future plans

Data synchronization

A key challenge in using an EMR in a developing country setting is that sites need local copies of data, due to slow and unreliable internet connectivity, while data from all sites needs to be combined in a central repository to allow tracking of transferred patients, project-wide reporting and reliable backup of data. We are currently designing tools to synchronize data and handle data entry conflicts, to allow seamless use of local OpenMRS servers that replicate and provide offline access to a parent OpenMRS repository.

Managing data on a range of diseases

Due to the urgency of this work and the resources available, the first installations of the OpenMRS focused on HIV and TB treatment projects, but it is designed to collect data on any medical care issue. In Rwanda we are already working with our *food distribution program*. Over the course of the next few months, we will continue to expand the use of OpenMRS to other PIH programs that were only indirectly related to HIV care, and were difficult to integrate into our previous EMR system (which lacked OpenMRS's concept dictionary and report designer). These include the *home visit program*, the *malnutrition program*, the *infant formula feeding* program and the *POSER program* (which constructs better homes for patients in desperate need). For each of these programs we will work closely with heads of department to devise effective forms for information intake. We will then work at length to come up with the reports those programs need to generate, both for external funders and for internal program management. The end goals are (1) to improve patient care by automatically providing lists of patients who have missed appointments or need special attention; (2) to make our hard-working staff's jobs easier, particularly by automating reporting that must currently be done manually; (3) integration of data so we can use OpenMRS's analysis tools to easily answer cross-program questions. We plan to study what happens to the clinical status of patients who stop receiving food support (and provide early warnings of patients who deteriorate after leaving the food program), and to assess the effect on patients' health once they have a better roof over their head. In Peru and Lesotho we are setting up OpenMRS to

track patients with TB and MDR-TB with support from WHO. Colleagues in Tanzania have developed concepts and forms for an oncology care version of OpenMRS.

Better integration with the hospital and clinic

We want patient information to be more available and encourage more interaction with the system by clinic staff. In addition, to adding more computers that can access the web-based EMR, we will be implementing a touch-screen interface for usage around the clinic and hospital. This will give clinicians easy access to patient histories, orders and lab data.

New interface tools

Date entry at present mainly occurs via forms implemented with Microsoft Infopath. This tool was used to simplify form design and speed up development but does not work perfectly with Linux, and requires a license fee. We are designing a new open source form creation tool using Java server pages, with help from students at the University of KwaZulu-Natal, South Africa. In addition, we are working to link the OpenMRS system to PDAs and Notepad PCs. Communication of data from the interface to OpenMRS is via the HL7 data exchange standard, simplifying the process of building links to other interface devices and software.

Conclusions

This project and those described earlier illustrate the capability that a system like OpenMRS has to help a clinic in a developing country to manage their data effectively and efficiently. By combining the functionality for patient data recording and display, reporting, and quality control in one system we are able to reduce the cost and complexity of setting up and maintaining the data management processes. The OpenMRS is now in operation in Kenya, Rwanda, South Africa, and Lesotho, with pilot implementations underway in Tanzania and Uganda. We are developing the Spanish language version for HIV and TB treatment in Peru, and there are plans to develop a Portuguese version for use in Africa and Brazil. We believe that the collaborative nature of the development of OpenMRS has not only magnified the resources available to build the software, but has greatly aided our ability to anticipate problems in new sites. We look forward to collaborating with other projects in Africa and other resource poor areas.

Acknowledgements

We thank the WHO and the CDC for supporting the development of the OpenMRS system. We thank the Clinton foundation, the Global Fund for AIDS, TB and Malaria and the US PEPFAR program for supporting clinical care.

References

[1] The WHO 3 by 5 Initiative. http://www.who.int/3by5/en/, 2005.

[2] Rwanda Project Overview, Partners In Health. http://www.pih.org/wherewework/rwanda/index.html, 2006.

[3] Safran, C., D.M. Rind, R.B. Davis, D. Ives, D.Z. Sands, J. Currier, W.V. Slack, H.J. Makadon, and D.J. Cotton, Guidelines for management of HIV infection with computer-based patient's record. Lancet, 1995. 346(8971): p. 341-6.

[4] Bates, D.W., M. Cohen, L.L. Leape, J.M. Overhage, M.M. Shabot, and T. Sheridan, Reducing the frequency of errors in medicine using information technology. J Am Med Inform Assoc, 2001. 8(4): p. 299-308.

[5] Tierney, W.M., C.J. McDonald, D.K. Martin, and M.P. Rogers, Computerized display of past test results. Effect on outpatient testing. Ann Intern Med, 1987. 107(4): p. 569-74.

[6] Rotich, J.K., T.J. Hannan, F.E. Smith, J. Bii, W.W. Odero, N. Vu, B.W. Mamlin, J.J. Mamlin, R.M. Einterz, and W.M. Tierney, Installing and implementing a computer-based patient record system in sub-Saharan Africa: the Mosoriot Medical Record System. J Am Med Inform Assoc, 2003. 10(4): p. 295-303.

[7] Siika, A.M., J.K. Rotich, C.J. Simiyu, E.M. Kigotho, F.E. Smith, J.E. Sidle, K. Wools-Kaloustian, S.N. Kimaiyo, W.M. Nyandiko, T.J. Hannan, and W.M. Tierney, An electronic medical record system for ambulatory care of HIV-infected patients in Kenya. Int J Med Inform, 2005. 74(5): p. 345-55.

[8] Douglas, G., The Lilongwe Central Hospital Patient Management Information System: A Success in Computer-Based Order Entry Where One Might Least Expect. Proc AMIA Symp, 2003: p. 833.

[9] Milberg, J., Adapting an HIV/AIDS clinical information system for use in Kampala, Uganda. Proceedings of Helina2003, Johannesburg, 2003: p. 44-45.

[10] Fraser, H., D. Jazayeri, C. Mitnick, J. Mukherjee, and J. Bayona, Informatics Tools To Monitor Progress And Outcomes Of Patients With Drug Resistant Tuberculosis In Peru. Proc AMIA Symp, 2002: p. 270-4.

[11] Choi, S., D. Jazayeri, C. Mitnick, K. Chalco, F. Pachao, J. Bayona, and H. Fraser, A Web-based Nurse Order Entry System for Multidrug-Resistant Tuberculosis Patients in Peru. Proc.Medinfo2004, 2004. 11: p. 202-206.

[12] Fraser, H., J. Blaya, S. Choi, C. Bonilla, and D. Jazayeri, Evaluating the impact and costs of deploying an electronic medical record system to support TB treatment in Peru. Proc AMIA Symp, 2006: p. 264-268.

[13] Fraser, H., S. Choi, M. Galipot, D. Jazayeri, and N. Mangubat, Successful transfer of a Web-based TB medical record from Peru to the Philippines. Proc AMIA Symp, 2006: p. 924.

[14] Fraser, H., D. Jazayeri, P. Nevil, Y. Karacaoglu, P. Farmer, E. Lyon, M. Smith-Fawsi, F. Leandre, S. Choi, and J. Mukherjee, An information system and medical record to support HIV treatment in rural Haiti. BMJ, 2004. 329: p. 1142-46.

[15] Allen, C., P. Manyika, D. Jazayeri, M. Rich, N. Lesh, and H. Fraser, Rapid deployment of electronic medical records for ARV rollout in rural Rwanda. Proc AMIA Symp, 2006: p. 840

[16] Mamlin, B., Biondich, PG , P. Biondich, B. Wolfe, H. Fraser, D. Jazayeri, C. Allen, J. Miranda, and W. Tierney, Cooking Up An Open Source EMR For Developing Countries: OpenMRS - A Recipe For Successful Collaboration.
Proc AMIA Symp, 2006: p. 529-533.

[17] Hripcsak, G., P. Ludemann, T.A. Pryor, O.B. Wigertz, and P.D. Clayton, Rationale for the Arden Syntax. Comput Biomed Res, 1994. 27(4): p. 291-324.

Address for correspondence

Hamish SF Fraser
hamish_fraser@hms.harvard.edu

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Multilingual Information Retrieval in Thoracic Radiology: Feasibility Study

André Coutinho Castilla[a], Sérgio Shiguemi Furuie[a], Eneida A. Mendonça[b]

[a] *Department of Cardiology and Pneumology, Faculty of Medicine, University of São Paulo, Brazil*
[b] *Department of Biomedical Informatics, Columbia University, New York, USA*

Abstract

Most essential information contained in the electronic medical record is stored as text, and this imposes several difficulties on automated data extraction and retrieval. Natural language processing is an approach that can unlock clinical information from free texts. The proposed methodology uses the specialized natural language processor MEDLEE developed for the English language. To use this processor on Portuguese medical texts, chest X-ray reports were machine translated (MT) into English. The result of serial coupling of MT and NLP is tagged text that needs further investigation for extracting clinical findings. This experiment's objective was to investigate normal reports and reports with device description on a set of 165 chest X-ray reports. We obtained sensitivity and specificity of 1 and 0.71 for the first condition, and 0.97 and 0.97 for the second. The reference was formed by the opinions of two radiologists. The results of this experiment indicate the viability of extracting clinical findings from chest X-ray reports through coupling MT and NLP.

Keywords:

natural language processing, information storage and retrieval, thoracic radiography, machine translation

Introduction

There has been gradually been a change in the storage of medical information from paper to a digital medium in the form of electronic medical record (EMR). One of the greatest advantages of digital storage is the possibility of searching and consulting the stored information. However, a great part of the clinical information is stored as free text in the EMR, which makes searching, analyzing and comparing documents difficult.

Natural language processing (NLP) refers to a set of computational techniques whose objective is narrative text analysis for information extraction, classification and data comparison. This approach is highly effective in extracting and retrieving medical text information in several languages, including English [1, 2].

We decided to use machine translation (MT) associated with NLP in a serialized form for extracting and retrieving medical text information in Portuguese. The machine translates medical texts from Portuguese into English in a controlled way; that is, based on the utilization of a specialized translation dictionary in the chosen domain. Therefore, we would be able to employ a natural language processor, MEDLEE, which had been developed for medical texts in English and already validated for this domain [3]. Consequently, MT would work as a pre-processing stage to normalize and adjust the text before it was submitted to NLP. As the English language is so dominant in medical knowledge, we could assume there would t be expressiveness and representativity loss in the texts content in performing this translation. A general overview of methodology is shown in Figure 1.

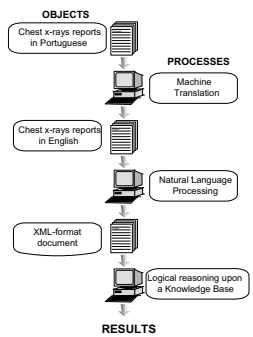


Figure 1 - General overview of proposed methodology

The chosen domain for this project was chest X-ray reports to investigate several clinical findings present in this kind of document. We show the primary results of this methodology in two clinical conditions: radiographic normality and the presence of devices. The results of this processing were binary responses indicating whether these two situations were present, and would be compared to the opinions of two specialist observers (radiologists).

Background

Information retrieval of medical texts using NLP has been successful in languages such as English, Dutch, German, Japanese and French [2]. Recently, a project in Portuguese was described [4]; however, it has obtained limited results. MEDLEE is one of the most successful systems for extracting information from medical texts. Initially, it was developed as a decision tool for X-ray reports but its coverage was extended to a wide variety of medical texts. The processing result is a XML document composed of clinical

Selected for best paper award.

primary findings with modifying elements. One of the most important results was obtained in the processing of 889.921 reports of thorax X-rays [5]. A sample analysis showed a sensitivity and specificity of 81% and 95% respectively on retrieval of documents describing the most important chest radiology diagnoses.

Materials and methods

The MT was performed with the software application SYSTRAN Premium version 5.0 [6]. This application performs rule-based translations and requires a specialized dictionary to process medical texts. Two were configured. The first one was a user's dictionary composed of a unidirectional table of source and destination terms associated with attributes; the most important is category, a mixture of syntactic and grammatical class. The elaboration of this dictionary followed a specific methodology [7]. The terms that constituted this dictionary were taken from chest radiography reports of articles' summaries from specialized literature [8], RADLEX [9], and UMLS [10]. A translation memory was also elaborated that of a pre-translated sentences table suitable to translate those of frequent occurrence. In this case, it was configured to process sentences related to normal findings.

Some 1362 chest radiography reports in Portuguese developed at the Instituto do Coração de São Paulo in 2003 and 2004 were chosen for MT training. Out of these reports, 930 unique sentences were obtained and submitted to MT using the dictionaries described above. Translations were manually revised and adjustments were carried out in the dictionaries. Parallel studies of performance evaluation were also conducted [11,12].

For the experiment, 165 reports performed in January 2006 were randomly selected. In any report, no data was obtained in a way that would identify patients or any clinical information. All incidences were included and there was no distinction if the data referred to radiographies performed with portable equipment on a bed. The only constraint was the presence of a report produced by the radiologist. The professional who developed the report was not identified.

Each report was marked with a unique numeric identifier and its content was located in a text file. Then, it was submitted to MT using the application SYSTRAN configured with the dictionaries described above. The report that had been translated into English was then submitted for processing through MEDLEE [3] and the processing result was stored in a XML file.

The XML-format result needed additional processing to determine the presence of the findings being investigated. To accomplish this, we developed a methodology based on an ontology created with the software PROTEGE [13] and that was submitted to first order logic queries with the Algernon Abstract Machine (AAM) [14].

The ontology was created from the specialized vocabulary in radiology RADLEX. Its hierarchical structure renders the building up of an ontology. Each RADLEX element of the super-classes' 'anatomic location', 'morphologic and physiologic processes', 'diagnoses and etiologies', and 'visual features' was incorporated into the ontology (each one as a class). The RADLEX original hierarchical structure was preserved.

The processing commenced with the inference rules loading. These included lexicon equivalence rules between MEDLEE and RADLEX, and the clinical relation rules shown in Tables 1 and 2.

Table 1 - Simple terms equivalence rule from MEDLEE to RADLEX as pseudocode

if MedLee term is lung then Radlex term is lungs

Table 2 - Example of pseudocode rule that establishes an affect relation between an anatomic structure and a disease

if pneumonia is true then lung instance is true and normal is false

Table 3 - An example of some logical assertions obtained from a processed report, shown as AAM code

(tell ((:add-instance (?x sentence)(tipo ?x "problem") (tipo_value ?x "osteophytosis")(suid ?x "s1.1.1") (value2 ?x :TRUE))))
(tell ((:add-instance (?x sentence)(tipo ?x "problem") (tipo_value ?x "enlarged") (bodyloc ?x "auricle") (suid ?x "s1.1.6") (certainty ?x "no") (value2 ?x :FALSE))))
(tell ((:add-instance (?x sentence)(tipo ?x "finding") (tipo_value ?x "free") (suid ?x "s1.1.2") (region ?x "recess") (value2 ?x :TRUE))))

Then the data derived from the reports were loaded. These data were those contained in the XML file transformed by XSLT into logic instructions in the AAM syntax. To transform a MEDLEE result in XML into logical instructions, a simplification was adopted. Table 3 shows some logical assertions obtained from a processed report after the transformation of the MEDLEE result. These assertions create instances in the ontology that discharge the rules previously loaded, thus creating new instances. The last stage was performing queries through searching instances created in the target classes using a binary response (yes or no) to indicate the presence of the findings. Table 4 shows one of the queries related to devices.

In this experiment, we looked for two fundamental radiological findings. The first was the normal reports that can be defined as those whose sentences report only radiological findings compatible with normal standards. The second related to implants, catheters, probes, equipment, foreign bodies and other implements of surgical nature, such as sutures or clips. Such elements will be generically called devices.

Table 4 - Query performed to search device instances, shown as pseudocode

```
if device is true then print "1" on output file
```

The 165 investigated reports were compared to the reference elaborated by the two radiologists who did not belong to the institution where the reports came from. The reference variability had been measured with the kappa index. System performance was accessed globally and only the final result was evaluated, and not the intermediate stages. Accuracy in the report retrieval was measured by the calculating specificity, sensitivity, and true positive and true negative fractions, including confidence intervals [15].

Results

A personal computer was used in the translation, transformation and logical inference operations, whereas MEDLEE processing was performed by a remote server via the Internet. The whole operation lasted about 90 minutes. The prevalence of the investigated findings according to the employed methodology on the 165 reports is shown in Table 5.

Table 5 - Prevalence of investigated conditions (n=165)

Conditions	Quantity	Prevalence
Normal Report	36	22%
Device	26	18%

True positives were defined when there was agreement between the system and the human reference on positive cases. The following contingency table (6) shows the methodology results compared to human reference in relation to normal reports.

Table 6 - Contingency table related to normal reports

Methodology	Reference		Total
	Positive	Negative	
Positive	36	27	63
Negative	0	102	102
Total	36	129	165

We show the results relating to reports with a description of devices in Table 7.

Table 7 - Contingency table related to reports with device description

Methodology	Reference		Total
	Positive	Negative	
Positive	26	4	30
Negative	3	132	135
Total	29	136	165

Tables 8 and 9 show the accuracy of the methodology compared to human observers on finding normal reports and reports with device description respectively. The inter-observer agreement measured by kappa index is shown in table 10.

Table 8 - Accuracy measures from normal reports

Measure	Observed value	95 % Confidence interval	
		Inferior	Superior
Sensitivity	1	0.88	1
Specificity	0.79	0.71	0.86
True Positive	0.57	0.44	0.69
True Negative	1	0.95	1

Table 9 - Accuracy measures for device reports

Measure	Observed value	95 % Confidence interval	
		Inferior	Superior
Sensitivity	0.9	0.72	0.97
Specificity	0.97	0.92	0.99
True Positive	0.87	0.68	0.96
True Negative	0.98	0.93	0.99

Table 10 - Agreement between observers measured by kappa index

Condition	Observed kappa	95 % Confidence interval	
		Inferior	Superior
Normal Reports	0.9652	0.9172	1.0132
Device	0.921	0.818	0.9862

Discussion

Natural language processing is an effective method of clinical information extraction and retrieval directly from free medical texts, and it is particularly useful when investigating a large number of reports [5]. The unpublished methodology described here connects a machine translator to a natural language processor in English to investigate clinical situations present in reports in Portuguese.

This paper does not intend to evaluate the isolated performance of the various stages of methodology. To achieve**,** this we would have had to evaluate each stage independently and controlled the various influences. We decided to perform a global evaluation of a complete methodology and only evaluated the final result and not the intermediate phases. The results described here are the first using this methodology. To study the feasibility of this approach, we decided to evaluate only normal reports and the presence of device description. Since we obtained favorable results, we will expand our rules and queries to the major chest X-ray diagnoses and conditions.

The main innovation of this methodology is the use of the MT associated to NLP. A similar methodology for information retrieval of medical texts had not been described previously. We hypothesize that a loss of expressivity and representativity of semantic content after translation will not occur. This is based on the fact that English is the most important language in the spread of technical knowledge. Moreover, the semantic content of medical knowledge has a global character based on what has been published in medical literature in English. This makes English the *lingua franca* of medicine [16].

Another aspect determined the elaboration of this methodology. There is no satisfactory medical natural language processor for Portuguese texts The experiment described by Martha [4] obtained sensitivity of less than 50% on information retrieval medical texts. This performance is inferior to that obtained by MEDLEE alone or by coupling with MT as described in this paper.

MT is certainly the most critical aspect of this methodology. The best MT results were obtained in the processing of technical texts [17]. The morphological errors observed in the English texts might be of various types. The first refers to words that present orthographic errors in the Portuguese source text. The machine translator SYSTRAN offers the option of orthographic correction. This correction worked suitably**,** taking into account the fact that accent differences between the terms of the text and the dictionary were ignored and correctly translated by the system. This orthographic corrector thus allows SYSTRAN to eliminate mistakes relating to accent use and other special characters normalizing the text. Another kind of error occurred when the original word did not exist in the MT dictionary. This situation can lead to two situations: the presence of untranslated words in the destination text, or the use of words out of context. An example of the latter kind of error was observed during the training stage when the word *lobo* was translated as *wolf* instead of *lobe*.

The next stage of the methodology involved subjecting the translated text to MEDLEE processing. For instance, in this stage of development and training, an error was observed that was associated with the interpretation of the finding 'linear opacity'. Translated as 'dense line', it was interpreted by MEDLEE as vascular access like 'venous line'.

MEDLEE result present a multilevel hierarchic structure that was simplified to a single level. MEDLEE describes entities in a structured form with various XML elements, whereas the RADLEX-based ontology used in the inference mechanism describes each entity through a single element. In the present experiment**,** this was the major cause of errors and it led to several clinical situations not being identified and jeopardized the normal findings. Improvements will be implemented in the next stage of the project through augmentation of the MEDLEE captured structure.

The next stage was logical reasoning. NLP results are composed of data that need further interpretation for the presence of the investigated clinical states to be established. Here, we utilized a logical inference mechanism based on an ontology that was juxtaposed with the use of the other described methodologies. Specifically for MEDLEE use, an inference system was implemented based on rules to NLP results processing. This approach used Arden Syntax for Medical Modules that accept the coded findings as entries and use a set of rules to determine the presence of selected clinical findings [18]. Our approach used these same data to create instances in an ontology which, in turn, will be queried afterwards.

The hierarchy of anatomic structure classes and findings, as well as their relations rules, is essential to this process of logical reasoning. Queries were loaded to aim at finding instances of investigated classes and subclasses. Table 4 shows the instruction in pseudocode that searched for instances of the 'device' class. This reasoning model was able to check for the presence of investigated clinical situations based on given assertions.

A study compared five types of expert systems in the interpretation of NLP results [19] for the investigation of pneumonia in comparison with three human observers (two radiologists and one clinical). The systems were rules elaborated by specialists, two kinds of Bayesian nets, a decision tree and a search for words. The latter aside (which performed poorly), differences were not observed with the human reference when deciding on the presence of pneumonia in the NLP result.

The accuracy of methodology was higher for devices location than for normal reports. This stemmed directly from the way the logical inference was performed. Normalcy is the initial state of the report. The inference system task was to find out clinical situations described in the report. The system's failure to point out the presence of several findings ensured that no normal reports were construed as normal. This derived from the project development stage. Until that point, only devices findings had had all their rules defined, which was evident from the performance. In the next stage of development**,** rules for other findings will be refined. In doing so, this methodology will be able to investigate a larger range of clinical findings, and therefore reduce the number of false positive normal reports observed in this experiment.

The MEDLEE output has a five-step scale of certainty varying from no to high certainty. The current rules that transform this output into logical assertions capture only two states. When certainty is no**,** it indicates a false assertion. On the other hand, the assertion is considered true, even on low certainty. Another aspect that needs to be mentioned is that the status tag of a MEDLEE output is not captured by this transformation. However, both aspects of logical transformation have contributed to false positive results.

NLP already represents a consistent approach in information retrieval from medical texts. Here, we presented a methodology that permits this information retrieval through a multilingual approach using MT. The feasibility of the hypothesis that it is possible to machine translate texts in Portuguese and submit them to NLP through the MEDLEE tool was demonstrated. This work was innovative in the sense it used MT as pre-processing to realize natural language processing. Another feature that deserves to be mentioned is the use of a logical inference mechanism based upon ontology, which is different from previously described methods for interpreting NLP results from medical texts.

Acknowledgments

We would like to thank to thank Dr. Carol Friedman for permitting the use of MEDLEE. MEDLEE is supported by the grants

R01 LM007659 and R01LM008635 from the National Library of Medicine (USA). This work is also supported by grant 55.2026/2002.6 from CNPQ (Brazil).

References

[1] Hripcsak G, Friedman C, Alderson PO, DuMouchel W, Johnson SB, Clayton PD. Unlocking clinical data from narrative reports: a study of natural language processing. Annals of Internal Medicine. 1995; p.681–8.

[2] Friedman C, Hripcsak G. Natural language processing and its future in medicine: Can Computers Make sense out of Natural Language Text. Acad Med. 1999;p. 890–5.

[3] Friedman C, Alderson P, Austin J, Cimino J, Johnson S. A general natural-language text processor for clinical radiology. Journal of American Medical Informatics Association. 1994;p. 161–174.

[4] Martha AS. Recuperação de informação em campos de texto livre de Prontuários eletrônicos do paciente baseada em Semelhança semântica e ortográfica. Universidade Federal de São Paulo; 2005.

[5] Hripcsak G, Austin JHM, Alderson PO, Friedman C. Use of natural language processing to translate clinical information from a database of 889,921 chest radiographic reports. Radiology. 2002; 224(1):157–163.

[6] Systransoft. SYSTRAN Premium 5.0; 2005.

[7] Dillinger M. Dictionary Development Workflow for MT: Design and Management. In: Proceedings of the Eighth Machine Translation Summit; 2001.

[8] Chang MY, Sun YC, Chang CF, et al. A Free Radiology Dictionary Made From Abstract Corpus Of The Radiology And Radiographics. Chicago; 2003.

[9] RSNA. RADLEX: A Lexicon for Uniform Indexing and Retrieval of Radiology Information Resources; 2003. Available from: http://mirc.rsna.org/radlex/service.

[10] NLM. UMLS Knowledge Sources; 2005.

[11] Castilla AC, Bacic AS, Furuie SS. Machine Translation on the Medical Domain: The Role of BLEU/NIST and METEOR in a Controlled Vocabulary Setting. In: Proceedings of the Tenth Machine Translation Summit. Phuket, Thailand; 2005. p. 47.

[12] Castilla AC. Instrumento de investigação clínico-epidemiológica em cardiologia fundamentado no processamento de linguagem natural. Universidade de São Paulo; 2005.

[13] Protégé. The Protégé Project; 2000. Available from: http://protege.stanford.edu.

[14] Kuipers B. Algernon for Expert Systems. Austin, Texas; 1994. Available from: http://www.cs.utexas.edu/users/qr/algy/algy-expsys/algy-expsys.html.

[15] Woodward M. Epidemiology: Study design and data analysis. 1st ed. Boca Raton: Chapman Hall/CRC; 1999.

[16] Hersh W, Li Zhang MS. Teaching English Medical Terminology Using the UMLS Metathesaurus and World Wide Web. Proceedings of the 1999 AMIA AnnualSymposium. 1999; p. 1078.

[17] Hutchins J. 27. In: Machine Translation: General Overview. Oxford University Press; 2003. p. 501–511.

[18] Hripcsak G, Kuperman GJ, Friedman C. Extracting Findings from Narrative Reports: Software Transferability and Sources of Physician Disagreement. Methods Inf Med. 1998; 37:1–7.

[19] Chapman WW, Haug PJ. Comparing Expert Systems for Identifying Chest X-ray Reports that Support Pneumonia. Proceedings of the 1999 AMIA Annual Symposium. 1999; p. 216–20.

Address for correspondence

André Coutinho Castilla
Rua Cardoso de Almeida 1149 ap 51
São Paulo – SP – BRAZIL – 05013001
castilla@terra.com.br

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Large-Scale Evaluation of a Medical Cross-Language Information Retrieval System

Kornél Markó[a,b], Philipp Daumke[a,b], Stefan Schulz[a], Rüdiger Klar[a], Udo Hahn[c]

[a] *Medical Informatics Department, University Medical Center Freiburg, Germany*
[b] *Averbis GmbH, Freiburg, Germany*
[c] *Jena University Language and Information Engineering Lab, Germany*

Abstract

We propose an approach to multilingual medical document retrieval in which complex word forms are segmented according to medically relevant morpho-semantic criteria. At its core lies a multilingual dictionary, in which entries are equivalence classes of subwords, i.e. semantically minimal units. Using two different standard test collections for the medical domain, we evaluate our approach for six languages covered by our system.

Keywords:

Information storage and retrieval, multilingualism, natural language processing.

Introduction

The main challenges for the architecture and implementation of medical document retrieval systems and their underlying search engines are inherently linguistic. Additional complexity emerges from the multilingual dimension of information retrieval applied in the medical domain [1]. While clinical documents are typically written in the physicians' native language, searches in scientific databases require sophisticated knowledge of (expert-level) English medical terminology which most non-English speaking physicians do not have. Hence, some sort of bridging between synonymous or closely related terms from different languages has to be realized to make use of the information that these databases hold.

Furthermore, the user population of medical document retrieval systems and their search strategies are really diverse. Not only physicians, but also nurses, medical insurance companies and patients are increasingly accessing these resources, with the Web adding an even more scattered group of searchers. Hence, mappings between different jargons and sublanguages are inevitable to serve the needs of such a heterogeneous searcher community. Therefore, the simplicity of the document content representation, as well as automatically performed intra- and interlingual lexical mappings or transformations of equivalent expressions become crucial issues for an adequate methodology of medical information retrieval.

This work presents the MORPHOSAURUS system [2] which is intended to meet the particular challenges of medical language processing, especially for medical information retrieval. Its main component is a dictionary, in which entries constitute equivalence classes of *subwords*, i.e., semantically minimal units. These equivalence classes capture intralingual as well as interlingual synonymy. As equivalence classes abstract away from subtle particularities within and between languages, and reference to them is realized via a language-independent conceptual system, they form an interlingua.

Morpho-semantic indexing

Within the MORPHOSAURUS framework (an acronym for MORPHeme theSAURUS), subwords are assembled in a multilingual lexicon and thesaurus, with the following considerations in mind:

- Subwords are listed with their attributes such as language (English, German, Portuguese, Spanish, French, Swedish) or subword type (stem, prefix, suffix, invariant). Each lexicon entry is assigned one or more morpho-semantic identifier(s) representing its equivalence class, the MID.
- Semantic links between synonymy classes are added. We subscribe to a shallow approach in which semantic relations are restricted to a paradigmatic relation *has-meaning*[1], which relates one ambiguous class to its specific readings, and a syntagmatic relation *expands-to*[2], which consists of predefined segmentations in case of utterly short subwords.

Figure 1 depicts how source documents (top-left) are converted into an interlingual representation by a three-step procedure. First, each input word is orthographically normalized in terms of lower case characters and according to language-specific rules for the transcription of diacritics (top-right). Next, words are segmented into sequences of subwords from the lexicon (bottom-right). Finally, each

1 For instance, {head} {zephal,kopf,caput,cephal,cabec,cefalg} OR {leader,boss,lider,chef}

2 For instance, {*myalg*} {*muscle,muskel,muscul*} + {*schmerz, pain,dor*}

meaning-bearing subword is replaced by a language-independent semantic identifier, the MID. It unifies intralingual and interlingual (quasi-) synonyms, thus producing the interlingual output representation of the system (bottom-left). In Figure 1, bold-faced MIDs co-occur in both text fragments.

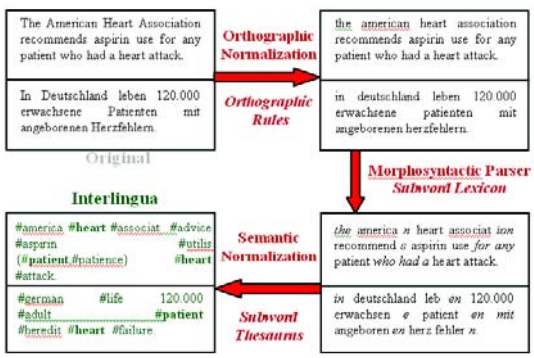


Figure 1 – Morpho-Semantic Indexing (MSI): Ambiguity is represented by the sequence of multiple MIDs in curly brackets. The German sentence can be translated to "In Germany, 120,000 adult patients live with a hereditary heart failure".

The combined subword lexicon currently contains 99,781 entries, with 22,953 for English, 24,357 for German, 15,158 for Portuguese, 13,390 for Spanish, 9,924 for French, and 13,981 for Swedish. All of these entries are related in the thesaurus by 22,707 equivalence classes.

Subword sense disambiguation

In the subword thesaurus, entries can be attributed to different senses (via the *has-meaning* relation). Consider, e.g., the English lexical item "patient", which has (at least) two different meanings. As a noun it refers to a human (*#patient* in Figure 1, bottom left), as an adjective it has a completely different meaning (*#patience*). In order to resolve ambiguities that can occur during the morpho-semantic indexing process, a simple probabilistic model is used [3]. It accounts for co-occurrence information from large textual resources. To illustrate this for the English document sample in Figure 1, the reading *#patient* should be preferred over *#patience*, since *#patient* as well as the context identifier *#heart* also occur in the representation resulting from the processing of the German document (the word *"Patient"* in German is unambiguous).

In earlier experiments on bilingual information retrieval (English and German, cf. [2]), we have shown the usefulness of representing medical documents on an interlingual layer. In this paper, we extend this evaluation by the other languages that are covered by the system. Furthermore, the effect of the disambiguation module attached to the morpho-semantic indexing procedure is analyzed in detail.

Information retrieval experiments

The experiments were run on the OHSUMED corpus [4], which constitutes one of the standard IR testbeds for the medical domain, and the 2006 corpus of IMAGECLEFMED (cf. [5]).

OHSUMED is a subset of the MEDLINE database. Considering the title and abstract field (if available) for each bibliographic unit, the set contains 348,566 documents and 26,705,691 tokens (whitespaces), thus, the average document length is 76.6 tokens. Since the OHSUMED corpus was created specifically for IR studies, 106 queries are available, including associated relevance judgments (actually 105, because for one query no relevant documents could be found). The average number of query terms is 5.2. The following is a typical query: *"Are there adverse effects on lipids when progesterone is given with estrogen replacement therapy?"*

IMAGECLEF is the cross-language image retrieval track which is run as part of the Cross Language Evaluation Forum (CLEF) campaign. IMAGECLEFMED evaluates the retrieval of medical images described by text captions based on queries in different languages. The main goal is to improve the multimedial retrieval of medical images from heterogeneous and multilingual document collections containing images as well as text. In IMAGECLEFMED 2006, the multilingual image retrieval task is based on a dataset containing images from different types (radiology and pathology images, clinical case descriptions related to nuclear medicine). Considering English annotations only, there are 40,709 image descriptions with a highly variable quality within and between the collections. The number of tokens is 1,130,419, thus, the average document length is relatively small compared to OHSUMED (27.7). There are 30 queries for which relevance judgments are available. For English queries, the average number of query terms is 5.8. A typical example is: *"Show me images of a frontal head MRI"*.

The OHSUMED corpus and the IMAGECLEFMED subset considered here contain only English-language documents. This raises the question of how these collections (or, e.g. MEDLINE, in general) can be accessed from other languages as well. It is a realistic scenario, because, unlike in sciences with English as a *lingua franca*, among medical doctors native languages are still dominant in their education and everyday practice. English medical sublanguage capabilities are often quite limited. In order to solve this problem, medical practitioners might resort to translating their native-language search problem to English with the help of current Web technology, e.g., an automatic translation service available in a standard Web search engine. Its operation might be further enhanced by lexical resources as available from the U.S. National Library of Medicine in support of various non-English languages, e.g. the UMLS Metathesaurus (which currently supports, with considerable differences in coverage, German, French, Spanish, Portuguese, Swedish and many others). Relying on the quality of the translation, this procedure

then reduces the cross-language retrieval problem to a monolingual one.

As an alternative, MORPHOSAURUS is used as a cross-language approach. In our experiments, both scenarios will be evaluated on the same query and document set. As the baseline, a retrieval system is provided operating with the Porter stemmer[3] and language-specific stop word lists so that the system runs on English documents with English queries.

In the following experiments, the original English queries were translated into Portuguese, German, Spanish, French and Swedish by medical experts (native speakers, with a very good mastery of both general and medical English).

QTR approach: Machine translation based on bilingual dictionaries

Machine translation based approaches to Cross-Language Information Retrieval (CLIR, cf. [6] for an overview) either translate native-language queries into the target language of the document collection to be searched, or contrariwise, translate the entire set of documents into each (supported) query language [7]. Since the latter is a resource intensive task, query translation (QTR) can be regarded as a standard experimental procedure in the cross-language retrieval community.

For evaluation, the manually translated queries were re-translated into English using the GOOGLE TRANSLATOR.[4] Admittedly, this tool may not be particularly suited to translate medical terminology (in fact, 13% of the German, 7% of the French, and 19% of the Portuguese and Spanish query terms were not translated, while Swedish is not supported at all). Hence, bilingual lexeme dictionaries derived from the UMLS Metathesaurus were used additionally. If no English correspondence could be found, the terms were left untranslated.

Just as in the (English-English) baseline condition, the stop words were removed from both the documents and the automatically translated queries and potential suffixes were stripped off.

MSI-approach: Language independent morpho-semantic indexing

As an alternative to QTR, the approach which is based on the morpho-semantic indexing procedures was probed, as introduced above. Unlike QTR, the indexing of documents *and* queries using MSI (after stop word elimination) yields a language-independent, semantically normalized index format.

Search engine

For an unbiased evaluation, several experiments were run with LUCENE,[5] a freely available open-source search engine which combines Boolean searching with a sophisticated statistical ranking model. Furthermore, this search engine has another advantage: it supports a rich query language like multi-field search, including more than ten different query operators.

Evaluation scenarios

Three different basic test conditions can now be distinguished for the retrieval experiments:

- **BASELINE**: The baseline of the experiments is given by the OHSUMED and IMAGECLEFMED corpus both in terms of their Porter-stemmed English queries, as well as their Porter-stemmed English document collection.
- **QTR**: In this condition, German, Portuguese, Spanish, French and Swedish queries are automatically translated into English ones (using the GOOGLE TRANSLATOR and the UMLS Metathesaurus), which are Porter-stemmed after the translation. These queries are evaluated on the Porter-stemmed OHSUMED and IMAGECLEFMED document collections.
- **MSI**: This condition stands for the automatic transformation of the English, German, Portuguese, Spanish, French and Swedish queries into the language-independent MSI interlingua (plus lexical remainders). The entire OHSUMED and IMAGECLEFMED collections are also submitted to the MSI procedure. Finally, the MSI-coded queries are evaluated on the MSI-coded corpora, both at an interlingual representation level. In this scenario, two different categories can be further discriminated:
  - **MSI**: experiments without incorporating the disambiguation module
  - **MSI-D**: experiments incorporating the disambiguation module

Measurements

Two measurements are taken in comparing the performance of QTR and different MSI. The first one is the average of the precision values at all eleven standard recall points (0.0, 0.1, 0.2, ..., 1.0). While this data is computed with consideration to the first 200 documents under each condition, the exact precision scores for the top 20 ranked documents are also taken into account.

Results

Considering the different test conditions and languages, Table 1 contains the exact numbers for the OHSUMED collection, and Table 2 for the IMAGECLEFMED collection (best results for each language marked bold).

OHSUMED results

For the OHSUMED collection, as depicted in Table 1 (first row), the English-English baseline achieves an 11pt average of 0.19 (Column 3). Running the experiment by MSI-indexing the original corpus and queries, the baseline condition can be exceeded by 16% for English. Additionally using the disambiguation module, 100% of the baseline up to 121% is reached for Portuguese, German and English. For the other languages considered, this scenario (MSI-D) also yields best results, ranging from 79% (French and Swedish) to 84% (Spanish) of the baseline. On the other

3 The stemmer is available on http://www.snowball.tartarus.org
4 http://www.google.de/language_tools
5 http://jakarta.apache.org/lucene/docs/index.html

hand, the QTR approach scores far lower than any MSI condition, reaching 37% of the baseline for Swedish and a maximum of 63% for Spanish. This results in 21 percentage points difference for Spanish up to 53 percentage points for German (QTR compared to MSI-D).

Table 1 - OHSUMED Results: Average and exact precision scores (% of baseline in brackets)

Language	Condition	11pt	Top 20
English	Baseline	0.19	0.27
English	MSI	0.22 (115.8)	0.29 (107.4)
	MSI-D	0.23 (**121.1**)	0.31 (**114.8**)
German	QTR	0.11 (57.9)	0.17 (63.0)
	MSI	0.20 (105.3)	0.27 (100.0)
	MSI-D	0.21 (**110.5**)	0.28 (**103.7**)
Portuguese	QTR	0.11 (57.9)	0.15 (55.6)
	MSI	0.17 (89.5)	0.23 (85.2)
	MSI-D	0.19 (**100.0**)	0.25 (**92.6**)
Spanish	QTR	0.12 (63.2)	0.16 (59.3)
	MSI	0.16 (**84.2**)	0.22 (**81.5**)
	MSI-D	0.16 (**84.2**)	0.22 (**81.5**)
French	QTR	0.10 (52.6)	0.16 (59.3)
	MSI	0.12 (63.2)	0.15 (55.6)
	MSI-D	0.15 (**78.9**)	0.20 (**74.1**)
Swedish	QTR	0.07 (36.8)	0.10 (37.0)
	MSI	0.15 (**78.9**)	0.22 (**81.5**)
	MSI-D	0.15 (**78.9**)	0.23 (85.2)
Average	QTR	0.12 (63.2)	0.17 (63.0)
	MSI	0.17 (89.5)	0.23 (85.2)
	MSI-D	0.18 (**94.7**)	0.24 (**88.9**)

The uneven investment of effort in constructing the different lexicons (see lexicon sizes above) is well reflected in the results. In any case, it seems worth noting that at *no* recall point QTR values were higher than MSI values.

However, there may be considerable variation regarding the actual numbers behind these average values. Medical decision-makers under time pressure are often interested in a few top-ranked documents. Thus, the exact precision scores for these documents are more indicative of the performance. Considering only the top 20 ranked documents (Column 4), precision does not fall below 74% of the baseline for MSI-D. In contrast, QTR does not exceed 63%, which means that MSI-D clearly outperforms QTR in any language condition. Again, focusing on the (English) monolingual retrieval setting, MSI-D gains 15% compared to the baseline.

By averaging over all languages and adding the English baseline condition to the values of the QTR approach for the other languages, query translation has a mean average precision of 0.12, thus reaching 63% of the baseline. MSI achieves 90% while MSI-D performs best with 95%.

IMAGECLEFMED results

Table 2 depicts the results for the IMAGECLEFMED corpus. Unlike the OHSUMED collection, which documents consist of coherent texts (MEDLINE abstracts), IMAGECLEFMED contains short captions of medical images, often only consisting of noun phrases with many acronyms. This might be the reason why the overall-performance is not comparable to OHSUMED for all scenarios, including the baseline condition.

Table 2 - IMAGECLEFMED Results: Average and exact precision scores (% of baseline in brackets)

Language	Condition	11pt	Top 20
English	Baseline	0.17	0.36
English	MSI	0.15 (88.2)	0.35 (**97.2**)
	MSI-D	0.16 (**94.1**)	0.35 (**97.2**)
German	QTR	0.10 (58.8)	0.22 (61.1)
	MSI	0.13 (**76.5**)	0.33 (91.7)
	MSI-D	0.13 (**76.5**)	0.34 (**94.4**)
Portuguese	QTR	0.13 (**76.5**)	0.22 (61.1)
	MSI	0.10 (58.8)	0.31 (**86.1**)
	MSI-D	0.12 (**70.6**)	0.30 (**83.3**)
Spanish	QTR	0.13 (**76.5**)	0.22 (61.1)
	MSI	0.13 (**76.5**)	0.32 (**88.9**)
	MSI-D	0.13 (**76.5**)	0.32 (**88.9**)
French	QTR	0.10 (58.8)	0.18 (50.0)
	MSI	0.11 (64.7)	0.30 (83.3)
	MSI-D	0.12 (**70.6**)	0.32 (**88.9**)

Language	Condition	11pt	Top 20
Swedish	QTR	0.06 (35.3)	0.08 (22.2)
	MSI	0.12 (70.6)	0.32 (88.9)
	MSI-D	0.13 (**76.5**)	0.34 (**94.4**)
Average	QTR	0.12 (70.6)	0.21 (58.3)
	MSI	0.12 (70.6)	0.32 (88.9)
	MSI-D	0.13 (**76.5**)	0.33 (**91.7**)

However, the advantage of MSI compared to QTR is still observable. While, in average, QTR yields 71% of the baseline regarding 11pt average, the MSI-D condition performs best with 77%. On the other hand, considering only a few top ranked documents, MSI-D outperforms QTR by 34 percentage points, reaching 92% of the baseline.

Of course, considering only 30 queries, the IMAGECLEFMED collection has only limited value for the evaluation of information retrieval systems. However, since this collection focuses on an important medical field (medical imaging and picture archiving systems) and results are in-line with those using the more elaborated OHSUMED collection, additional evidence for the excellent performance of the MORPHOSAURUS system in a Cross-Language Information Retrieval setting is available.

Discussion

After more than a decade of intensive research, cross-language information retrieval (CLIR) has produced considerable achievements. From a methodological point of view, the field is divided into dictionary-based *vs.* corpus-based approaches [6]. Since corpus-based approaches depend on the availability of large parallel corpora, which is mostly not the case for domain-specific sublanguages, most efforts in CLIR are centered on either query translation or document translation [7].

The success of dictionary-based CLIR depends on the coverage of the lexicon, tools for conflating morphological variants, phrase and proper name recognition, as well as word sense disambiguation. Within the MORPHOSAURUS system, the lexical coverage is optimized by limiting the lexicon to semantically relevant subwords of the medical domain. This also helps in dealing with morphological variation, including single-word decomposition. Since the latter is a frequent requirement for processing highly compounding languages (but also medical terminology in general), this partially explains the poor results for German in the SAPHIRE medical text retrieval system which uses the UMLS Metathesaurus for semantic indexing [8].

The UMLS is also the lexical basis of the approach pursued by the MUCHMORE project [9]. Although good results are communicated, these are not comparable to those presented here because the authors use a home-grown document and query collection and diverge in the construction of their baseline.

Eichmann *et al.* [10] report on cross-language experiments for French and Spanish using the same test collection as used here (OHSUMED), and the UMLS Metathesaurus for query translation, achieving 71% of their baseline for Spanish and 61 % for French (contrasted to 84% and 79% for MSI-D). With the vector space engine they employ, their overall 11pt performance (0.24) is far above the approach pursued here (0.18). On the other hand, when focusing on exact precision scores that have more explanatory power when thinking of a real world user scenario, the MORPHOSAURUS approach using a standard search engine turns out to be more advantageous. Eichmann *et al.* report precision values of 0.23 for Spanish (0.17 for French) for the top 5 ranked documents. Compared to these scores, the MSI approach (involving disambiguation) reaches 0.32 for Spanish (0.30 for French) for the top 5 ranked documents. Since query translation via the UMLS Metathesaurus was adopted in their work, it is not surprising that the QTR scenario yields comparable results by all means.

References

[1] Hersh, WR. Information Retrieval. A Health and Biomedical Perspective (2nd ed.). New York: Springer, 2002

[2] Markó K, Schulz S, and Hahn U. MorphoSaurus - design and evaluation of an interlingua-based, cross-language document retrieval engine for the medical domain. Methods of Information in Medicine, 44(4):537–545, 2005.

[3] Markó K, Schulz S, Hahn U: Unsupervised Multilingual Word Sense Disambiguation via an Interlingua. Proceedings of the 20th National Conference on Artificial Intelligence: 1075-1080, 2005.

[4] Hersh, WR, Buckley C, Leone TJ, and Hickam DH. Ohsumed: An interactive retrieval evaluation and new large test collection for research. Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval: 192–201, 1994.

[5] Clough, P, Müller H, Deselaers T, Grubinger M, Lehmann JJ, and Hersh WR. The CLEF 2005 cross-language image retrieval track. 6th Workshop of the Cross-Language Evaluation Forum, 2005.

[6] Oard DW and Diekema AR. Cross-language information retrieval. Annual Review of Information Science and Technology (ARIST), Vol. 33: 223–256, 1998.

[7] Rosemblat G, Gemoets D, Browne AC & Tse T. Machine translation-supported cross-language information retrieval for a consumer health resource. AMIA'03: 564–568, 2003.

[8] Hersh WR, Hickman DH, Haynes B, and McKibbon KA. A performance and failure analysis of Saphire with a Medline test collection. JAMIA, 1(1):51–60, 1998.

[9] Volk M, Ripplinger B, Vintar S, Buitelaar P, Raileanu D, and Sacaleanu B. Semantic annotation for concept-based cross-language medical information retrieval. IJMI, 67(1/3):79–112, 2002.

[10] Eichmann D, Ruiz ME, and Srinivasan P. Cross-language information retrieval with the UMLS Metathesaurus. Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval: 72–80, 1998

Address for correspondence

Kornél Markó, e-mail: kornel.marko@uni-freiburg.de, http://www.averbis.de

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

EHR Query Language (EQL) – A Query Language for Archetype-Based Health Records

Chunlan Ma, Heath Frankel, Thomas Beale, Sam Heard

Ocean Informatics Pty. Ltd, Australia

Abstract

OpenEHR specifications have been developed to standardise the representation of an international electronic health record (EHR). The language used for querying EHR data is not as yet part of the specification. To fill in this gap, Ocean Informatics has developed a query language currently known as EHR Query Language (EQL), a declarative language supporting queries on EHR data. EQL is neutral to EHR systems, programming languages and system environments and depends only on the openEHR archetype model and semantics. Thus, in principle, EQL can be used in any archetype-based computational context. In the EHR context described here, particular queries mention concepts from the openEHR EHR Reference Model (RM). EQL can be used as a common query language for disparate archetype-based applications. The use of a common RM, archetypes, and a companion query language, such as EQL, semantic interoperability of EHR information is much closer. This paper introduces the EQL syntax and provides example clinical queries to illustrate the syntax. Finally, current implementations and future directions are outlined.

Keywords:

openEHR reference model, archetype, query language, electronic health record

Introduction

The National E-Health Transition Authority of Australia (NEHTA) defines an EHR query as a formal user or system request for information to the EHR repository/ database that specifies the constraints on precisely what part(s) of the EHR content needs to be retrieved [1]. Due to the lack of any clear standards in EHR query services, NEHTA has identified it as a major area for review [2]. Easy accessibility of data from an electronic health record (EHR) is considered as one of the essential features of the EHRs that can enhance a hospital revenue cycle [3]. There are two major challenges commonly encountered in clinical data accessibility: one is that people who understand the clinical data best, such as health professionals, are not the ones most competent of querying the data; another challenge is that qualified SQL programmers or other query language programmers must spend much time on exploring the data, composing tedious code to provide the data that end user needs [4]. EHR data include thousands of facts to do with a patient's clinical status, which can be highly structured, semi-structured or non-structured, or most commonly, a mixture of all three. The lack of discipline commonly found in EHR data not only increases the difficulties in storing them, but particularly the difficulty in querying them.

The openEHR specifications have been developed to standardise the representation of an international electronic health record (EHR)[1]. Although there are implementations based on these specifications in development, experience with querying archetype-based EHR data is still limited, to the point where it is not clear what kinds of query language(s) are even appropriate for information systems based on archetypes.

Since the openEHR specifications intend to define a standardised EHR infrastructure, a query language for this infrastructure should aim to become an open specification as well. The query language described here will be submitted to the openEHR foundation as a candidate openEHR query language.

There are four requirements for an archetype-based query language. First, the query language should be able to express queries for requesting any data item from an archetype-based system, i.e. data defined in archetypes and/or the underlying reference model . Second, the query language should be able to be used by both domain professionals and software developers. Third, the query language should be portable, i.e., be neutral to system implementation, application environment and programming language. Lastly, the syntax should be neutral with respect to the reference model, i.e. the common data model of the information being queried. Particular queries will of course be specific to a reference model.

The current available query languages that might potentially be used to query openEHR data include the XML Query Language (XQuery) [5] and the Structured Query Language (SQL) . XQuery uses eXtensible Mark-up Language (XML) as its underlying data model. It has rich predefined functions and allows user-defined functions supporting the kind of clinical data requests required in the EHR context. It is platform independent. Nevertheless, its main strength is also its main flaw: it is limited to purely

1 http://www.openehr.org/

XML data environments. Direct use of XQuery for the openEHR EHR would require that all openEHR data must be represented in XML format. However, openEHR is designed as an object-oriented framework, and allows for a multitude of data representations, including as programming language persistent objects (e.g. in the form of Java objects in a product such as db4o[2]); as language neutral objects (such as in a database like Matisse[3]); as relational structures (governed by an object/relational mapping layer), and in various XML storage representations (e.g. XML blob or XML databases). XQuery is therefore problematic, because the query syntax is directly tied to the representational format of the data. Considerable efforts would be required to convert openEHR data in each deployment context to XML just for the purpose of querying; each such transformation may well be custom, requiring special work on the part of the system implementers.

A further disadvantage is that XQuery is a native XML programming language requiring intimate familiarity of both users (in this case, health professionals and software developers) alike with XML and XQuery, something that cannot be assumed.

SQL in its standard form is also not a viable candidate for querying archetype-based EHR data, because it does not support object-structured data, e.g. the data modelled in archetypes. It has been found that considerable intellectual efforts are required when using SQL to search and retrieve clinical data for both individual subject care and clinical research studies [6].

Object Query Language (OQL) is a query language for object-oriented databases. OQL was developed by the Object Data Management Group (ODMG[4]), which was disbanded in 2001. Comparing with XQuery and SQL, OQL would be the best candidate query language used for archetype-based EHR data. However, it is complex and as a result, it has not been widely implemented. For large object models, such as the openEHR RMs, OQL query statements can become extremely verbose. OQL uses an object programming style dot-notation to express object members, while an XPath-based syntax is specified by openEHR to locate archetype data elements. Using OQL with archetype-based EHRs would require the use and translation between these two notation styles.

To satisfy the aforementioned requirements of an archetype-based query language, a new language – EHR Query Language (EQL) is under development. This paper introduces the EQL features and syntax. Example clinical query scenarios are used to demonstrate the use of the EQL expression.

Methods

EQL was developed based on the analysis of a set of clinical query scenarios, the study of the current available query language syntaxes (including XQuery, SQL and Object Query Language), and the study of the archetypes technology, openEHR RM and openEHR path mechanisms.

What is EQL

EQL is a declarative query language developed exclusively for expressing the queries used for searching and retrieving the clinical data found in archetype-based EHRs. It is applied to the openEHR EHR Reference Model (RM) and the openEHR clinical archetypes, but the syntax is generic across applications, programming languages, system environment, and reference model. The EQL is designed as a common language used for expressing clinical data requests across multiple openEHR-based applications.

The EQL has two innovations: 1) utilizing the openEHR path mechanism to represent the query criteria and returned results; and 2) using a 'containment' mechanism to indicate the data hierarchy and constrain the source data to which the query is applied.

OpenEHR path mechanisms

OpenEHR path syntax is used to locate clinical statements and data values within them using Archetypes. An Archetype is a computable expression of a clinical concept in the form of structured constraint statements, based on some reference model [7], such as the openEHR RM which provides the support for clinical archetypes. Each archetype has a global unique identifier and each node of this archetype has a unique archetype node identifier. The openEHR architecture has a path mechanism that enables any node within a top level structure to be specified from the top of the structure using a "semantic" X-path compatible path. The availability of such paths radically changes the available querying possibilities with health information, and is one of the major distinguishing features of openEHR [8]. Consequently, it is possible to locate any node in an archetype, including leaf data elements by using the archetype and archetype node identifiers within openEHR paths.

Features of the EQL

The EQL features are listed below:

- Neutral expression syntax. EQL does not have any dependencies on the underlying RM of the archetypes. It is neutral to system implementation and environment. This is one of the distinguishing features of the EQL.
- Allows setting query criteria using archetype and node identifiers, data values within the archetypes, and class attributes defined within the openEHR RM.
- Allows the returned results to be top-level archetyped RM objects, data items within the archetypes or RM attribute values.
- Supports naming returned results.
- Support queries with logical time-based data rollback.
- Supports query criteria parameters.

2 http://www.db4o.com/
3 http://www.matisse.com/
4 http://www.odmg.org/

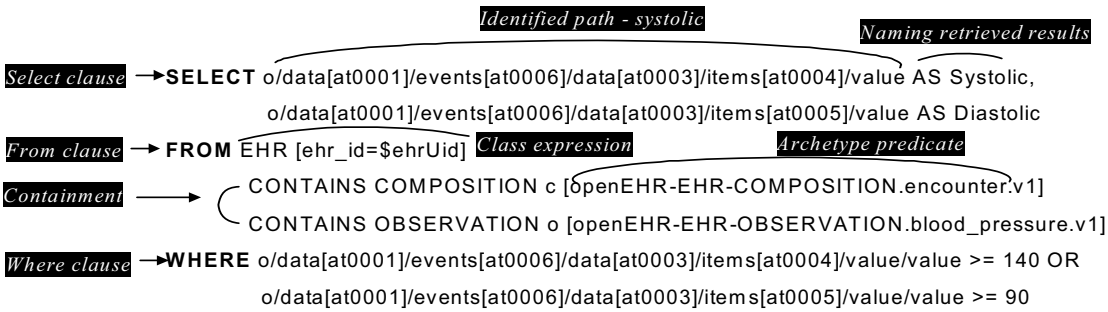


Figure 1 - A typical EQL statement

- Supports arithmetic operations (such as count, addition, subtraction, multiplication, and division), relational operations (>, >=, =, !=, <=, <) and Boolean operations (or, and, xor, not).
- Supports some functions that are supported in XQuery, such as current-date().
- Users could specify their preference on the retrieved data, such as ordering preferences, or total number of retrieved results.
- Supports the queries for individual clinical subjects at the point of care, administrative purposes and clinical research purposes.

EQL Syntax

Figure 1 shows a typical EQL statement, which would return all blood pressure values where systolic value is greater or equal to 140 or diastolic value is greater or equals to 90 within a specified EHR. EQL syntax is a synthesis of SQL structural syntax and openEHR path syntax. The SQL clauses SELECT, FROM, WHERE, and ORDER BY provide the basic structure of EQL. The SELECT clause specifies the data elements to be returned. The FROM clause specifies the result source and the corresponding containment criteria. The WHERE clause specifies data value criteria within the result source. The ORDER BY clause indicates the data items used to order the returned result set.

The openEHR path syntax [7], which is XPath compatible, is used in EQL to identify data items. Apart from the path mechanisms, EQL has a containment constraint, which specifies the hierarchical relationships between parent and child data items. The details of the EQL syntax are introduced below.

openEHR paths

The details of the openEHR path can be found elsewhere [7]. Herein, we briefly introduce its basic syntax.

A general pattern of the openEHR path syntax starts with a slash and followed by an attribute name of an object defined by the RM, and followed by another slash and attribute name and so on if there are multiple attributes involved, e.g. /ehr_id.

Another pattern of the path is a subset of the XPath syntax for predicates with a small number of short-cuts, i.e. archetype path. This path utilises an archetype ID or archetype node ID (i.e. at code, such as at0006) to identify an object. The example shown in Figure 1 uses the openEHR path syntax to locate both systolic and diastolic value in the blood pressure archetype.

Predicates

A predicate pattern is delimited by square brackets ([]). The path predicate has two operands and an operator. The operands are either an RM class attribute, an openEHR path or data value, such as a string or integer. A parameter name prefixed with a $ symbol can be used instead of a data value, which is substituted for a real data value at run time, e.g. ehr_id=$ehrUid shown in Figure 1. The operator can be =, >, <, >=, <=, or !=.

The archetype node ID predicates are a shortcut of a path predicate, which include either an archetype ID or archetype node ID. The archetype node ID predicates can also use a parameter name for the archetype node ID value, e.g. [$compositionArchetypeId]. The archetype node ID predicate shortcut is equivalent to a long-form path predicate with a left operand of archetype_node_id, and an operator of =. For example COMPOSTION c[openEHR-EHR-COMPOSITION-encounter.v1] shown in Figure 1 is equivalent to COMPOSTION c[archetype_node_id = 'openEHR-EHR-COMPOSITION-encounter.v1'].

EQL variables

All variables must have a RM class type assigned and they must be defined in the FROM clause, e.g. character c and o shown in the FROM clause of Figure 1 are EQL variables. EQL variables need to be declared when other clauses need a reference to them.

FROM clause

A FROM clause represents the data source for the query. It starts with the keyword – FROM, followed by a class expression indicating the RM classes, containment constraints and object identifying criteria used to constrain the data source for the query (refer to Figure 1).

For the openEHR RM, the classes declared in the FROM clause may include EHR, COMPOSITION, entry classes (e.g. OBSERVATION), and data structure class (e.g. ITEM_LIST).

Class expressions

An EQL class expression consists of a RM class name, an EQL variable (optional), and an optional class predicate. The class predicate further constrains the objects used as the query source and is normally applied to the archetype node ID or unique object ID class attributes.

Containment is indicated in EQL using the CONTAINS keyword between two class expressions. The left class expression indicates the parent object of the right class expression.

Identified paths

Identified paths are used to locate data items within an archetyped RM class. The identified path starts with an EQL variable that is declared within the class expression. This is followed by a slash and an openEHR path. The expression – o/data/[at0001]/event/…/value – shown in Figure 1 is an example of an identified path.

WHERE clause

A where clause is used to represent further criteria applied to the data items within the objects declared in the FROM clause. A WHERE clause starts with the keyword – WHERE, followed by a criteria expression. It consists of two operands and an operator. The operands may be an identified path or data value, such as a string or integer. A parameter name prefixed with a $ symbol can be used instead of a data value and substituted at run time. The operator can be =, >, <, >=, <=, or !=. Multiple criteria expressions can be combined using Boolean operators: AND, OR, XOR, and NOT.

SELECT clause

A SELECT clause starts with the keyword – SELECT, and followed by a set of identified paths. The set of paths are separated using a comma (see Figure 1).

Results naming

EQL allows users to rename the returned items specified in the SELECT clause. The keyword is AS, followed by the specified name, e.g. AS Systolic shown in Figure 1 SELECT clause.

ORDER BY clause

An ORDER BY clause starts with the keywords – ORDER BY, followed by a set of identified paths indicating the data items used to sort the returned result set. The keywords ASCENDING and DESCENDING (and abbreviations, ASC and DESC) can be used after each identified path as per SQL.

Arithmetic functions

A set of arithmetic functions, such as addition, subtraction, multiplication, and division can also be used in EQL. The use of these functions is the same as SQL, and is not described here.

TIMEWINDOW clause

TIMEWINDOW is an addition query clause used in EQL to constrain the query to data that was available in the system within the specified time criteria. This supports a time-based logical system rollback allowing a query to be executed as though it was performed at that specified time, which is essential for medico-legal reporting. It starts with the keyword – TIMEWINDOW, and followed by a string compatible with the ISO 8601 representation of time interval. The first example below constrains the query source to data committed to the system before 2006-01-01. The second example constrains the query source to data committed within the period of two years before 2006-01-01.

1) TIMEWINDOW /2006-01-01

2) TIMEWINDOW P2Y/2006-01-01

Clinical scenarios

To illustrate the use of the EQL syntax, this section provides the EQL expressions for two typical clinical scenarios.

Scenario one

Scenario description

Get the number of all patients with diabetes who have HbA1c results greater than 7.0 in last 12 months.

EQL expression

```
SELECT COUNT(e/ehr_id)
FROM EHR e
  CONTAINS
  (COMPOSITION c
       [openEHR-EHR-COMPOSITION.problem_list.v1]
    CONTAINS EVALUATION e
       [openEHR-EHR-EVALUATION.problem-
diagnosis.v1]
   AND
   COMPOSITION c1
       [openEHR-EHR-COMPOSITION.report.v1]
    CONTAINS OBSERVATION o
       [openEHR-EHR-OBSERVATION.laboratory-
hba1c.v1])
WHERE
  e/data/items[at0002.1]/value/value='diabetes
mellitus' AND
  c1/context/other_context/items[at0006]/
    items[at0013]/value > current-date() -P1Y AND
  o/data/events[at0002]/data/items[at0013.1]
   /value/numerator > 7
```

Scenario two

Scenario description

Get all HbA1c observations that have been done in the last 12 months for a specific patient.

EQL expression

```
SELECT o
FROM EHR e[ehr_id=$ehrId]
  CONTAINS COMPOSITION c
    [openEHR-EHR-COMPOSITION.report.v1]
    CONTAINS OBSERVATION o
     [openEHR-EHR-OBSERVATION.laboratory-hba1c.v1]
WHERE
c/context/other_context[at0001]/items[at0006]/
items[at0013]/value > current-date()-1PY
```

Scenario three

Scenario description

Get a patient's current medication list

EQL expression

```
SELECT c
FROM EHR e[ehr_id=$ehrId]
 CONTAINS COMPOSITION c
  [openEHR-EHR-COMPOSITION.medication_list.v1]
WHERE c/name/value='current medication list'
```

Discussions

This paper has introduced EQL – a declarative query language developed for querying openEHR-based EHRs by utilising the openEHR path mechanisms and unique containment syntax. However, the EQL syntax is not specific to the openEHR RM and can be used for any archetype-based information system. It could be used as a common query language for disparate archetype-based applications. The EQL will be submitted to the openEHR foundation as a candidate openEHR query language.

EQL implementation

Ocean Informatics Ptd Lty has implemented the components to process the EQL within the OceanEHR suite of EHR tools. These components include an EQL parser, the EHR query object model and a query processor. The implementation does not currently support all features of the EQL. However it has demonstrated the power and flexibility of using a common RM, archetypes and EQL, independent from the underlying system implementation, to retrieve any data set from an EHR.

The EQL is not easily understood by health professionals due to the computer-oriented openEHR path syntax used in EQL. An EQL query editor, which allows users to generate and edit EQL statements, has been developed to empower clinicians with fine-grained access to their EHR data. The query editor provides users with access to an archetype repository to build FROM containment constraints and a tree representation of RM attributes and archetype structures used to set WHERE criteria, SELECT data items and ORDER BY preferences. The tool can then execute the generated query and display the returned results without the user seeing or knowing how to write an EQL statement.

Future directions

The EQL continues to be developed based on requirements from additional clinical query scenarios. New EQL features may need to be provided, such as statistical, string pattern matching and user-defined functions. Existential (∃) and universal (∀) quantifiers may be also required.

Other research areas may include 1) exploring how the EQL supports clinical decision support technologies, e.g., clinical guidelines presentation; 2) investigating the integration of EQL with clinical terminology servers; and 3) conducting field trials using the EQL to represent common clinical queries to retrieve openEHR-based EHR data sets.

Conclusion

The use of a common RM and archetypes supports the sharing of EHR data, and with the addition of a companion query language, such as EQL, achieving semantic interoperability of EHR information is much closer.

Acknowledgements

We extend our thanks to Lisa Thurston for her support, including the development of the EhrView component – a generic EHR viewer – that uses EQL for retrieving EHR data.

References

[1] National E-Health Transition Authority (NEHTA). Acronyms, Abbreviations & Glossary of Terms in Publications. In: eds. National E-Health Transition Authority (NEHTA), 2005.

[2] National E-Health Transition Authority (NEHTA). Review of Shared Electronic Health Records Standards in Publications. In: eds. National E-Health Transition Authority (NEHTA), 2006.

[3] Amatayakul M. When EHRs are a-ok. Healthcare financial management, 2006: 60(2): 146-148.

[4] Nadkarni PM, Brandt C. Data Extraction and Ad Hoc Query of an Entity-- Attribute-- Value Database. J Am Med Inform Assoc, 1998: 5(6): 511-527.

[5] W3C.2006. XQuery 1.0: An XML Query Language.http://www.w3.org/TR/xquery/, accessed on 17 Oct. 2006

[6] Johnson S, Chatziantoniou D. Extended SQL for manipulating clinical warehouse data. Proceedings / AMIA ... Annual Symposium, 1999: 819-23.

[7] Beale T, Heard S. Archetype Definitions and Principles in The *open*EHR foundation release 1.0. In: Beale T and Heard S, eds. The *open*EHR foundation, 2005.

[8] Beale T, Heard S. openEHR Architecture: Architecture Overview in The *open*EHR foundation release 1.0.1. In: Beale T and Heard S, eds. openEHR Foundation, 2006.

Address for correspondence

Dr Chunlan Ma
Clinical Informatics Consultant
Chunlan.ma@oceaninformatics.biz

Heath FrankelProduct Development Manager
Heath.frankel@oceaninformatics.biz

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Design and Evaluation of a Temporal, Graph-Based Language for Querying Collections of Patient Histories

Ole Edsberg[ac], Stein Jakob Nordbø[c], Erik Vinnes[bc], Øystein Nytrø[ac]

[a] *Department of Computer and Information Science (IDI)*
[b] *Faculty of Medicine (DMF)*
[c] *Norwegian EHR Research Centre (NSEP)*
Norwegian University of Science and Technology (NTNU), Trondheim, Norway

Abstract

Giving clinicians and researchers the ability to easily retrieve and explore relevant fragments of patient histories would greatly facilitate quality assurance, patient follow-up and research on patient treatment processes. Established database query languages are inconvenient for such exploration, and may also be too complex for users with limited backgrounds in informatics. We believe that understandability can be increased in return for a sacrifice of some of the power of expression found in general query languages. In order to design a specialized query language, we have collected and synthesized a tentative list of requirements. Based on these requirements, we have designed and implemented Practice Explorer, a prototype for visual query of collections of patient histories, and evaluated the understandability of its query language by testing with medical students. The results indicate that parts of the language are intuitive enough for users to understand without demonstrations, examples, feedback or assistance. They also provide some lessons for future work in this area.

Keywords:

information retrieval. Data display. Medical record

Introduction

Clinicians and health researchers have a need for querying their patient records for relevant history fragments. Four tasks where this need may arise are:

1. Retrospective study of guideline compliance [1].
2. Re-consideration of treatment plans for groups of patients possibly affected by the discovery of new medical knowledge, such as the connection between H. pylori and peptic ulcers.
3. Selection of patients for scientific studies.
4. Development of research hypotheses through explorative search of patient records.

The text-based database query languages available in today's patient record systems are either too complex for users without informatics competence or has insufficient power of expression. As one of the sources in our requirements collection put it, "(...) I lack basic search functionality, mostly because of my own aversion against learning a programming language for searching (...)".

The research question this article addresses is: How can we design a system for formulating temporal queries against patient history databases that is easily understandable for users with little competence in informatics, but still satisfies most of their query needs? The contributions of this article are 1) a tentative list of requirements for the expressiveness of patient history query languages, 2) an outline of our design for a query system satisfying the requirements and 3) results, observations and lessons learned from understandability testing of a prototype implementing our design.

Related work

The system most closely resembling Practice Explorer is the very recent PatternFinder [2]. The most important differences are that Practice Explorer visualizes queries as hierarchical, directed acyclic graphs, whereas PatternFinder uses linear chains of forms, and that Practice Explorer's query language, prompted by the requirements described below and enabled by the more flexible query model, is also able to express queries with parallel and alternative branches. It would be interesting to compare a form-based and a graph-based visual query system to see which is more intuitive for users without a background in informatics. Also related is TVQL [3], a visual query language where binary interval relations are specified via sliders, and where multiple such relations can be combined via neighborhood and disjunction. Another related study [4] proposes three new notations (elastic bands, springs and paint strips) for interval relations and experimentally compares their understandability. It also provides a table comparing different approaches to visualizing temporal relations and specifying combinations of such relations with logical expressions. In our case, the query graph both specifies local temporal relations and conjunction and disjunction of sub-queries through the parallel and alternative branching constructs.

Our visual representation of branching and joining constructs is borrowed from UML activity diagrams [5], a flowchart notation for defining workflows. In fact, our

query graphs can be viewed as flowcharts, extended with some new constructs and given an alternative semantics suitable for matching against histories.

Materials and methods

Requirements collection

Through discussions with two general practitioners, a rheumatologist and a health researcher interested in clinical processes, and through a pilot study applying an early prototype to a general practitioner's database and letting him verbally specify queries to be executed by the developers, we collected example queries and synthesized the following list of requirements for patient history query languages. Queries should be able to find patterns consisting of:

1. A primitive history element, such as a patient encounter, lab test, prescription or correspondence event.
2. Results limited by the date or the age of the patient at a specific point in a pattern.
3. A time interval in which a medication has been prescribed, including overlapping prescriptions, or the start or end of such an interval.
4. Time periods of variable length, with the possibility of specifying that a specified pattern should, or should not, occur during the period.
5. Repetitive occurrence of a specified pattern.
6. Sequences of specified patterns.
7. Parallel occurrence of specified patterns.
8. Alternative occurrence of specified patterns.
9. The first occurrence of a pattern in the whole history.

It should also be possible to:

10. Specify encounter events and medication intervals at various points of abstraction in relevant coding hierarchies.
11. Perform union and intersection set operations on query results.
12. Save and re-use queries or their components.

Table 1 shows a natural-language specification of a query need that exemplifies many of the requirements.

Table 1 - Natural-language specification of a query need

Find all patients who initiated medication with an ACE-inhibitor for the first time in their histories without having any encounters coded as angina, myocardial infarction or heart failure in the preceding two-year period, and who had an en-counter coded as hypertension some time between the history's start and the initiation of the ACE-inhibitor medication.

Find all patients who initiated medication with an ACE-inhibitor for the first time in their histories without having any encounters coded as angina, myocardial infarction or heart failure in the preceding two-year period, and who had an encounter coded as hypertension some time between the history's start and the initiation of the ACE-inhibitor medication.

We have no illusions that our list of requirements is complete. In our experience, potential users are often not fully able to understand the possibilities offered by a temporal query system without having such a system available for use on their own data. Therefore, it will be necessary to iterate between developing prototypes and collecting more requirements.

Data source and data model

Our test case is a patient record database from a general practitioner's office. For the sake of query performance, Practice Explorer extracts relevant data from the database at startup and represents the patient histories in main memory as lists of events. Each event has a time stamp and is considered to last for 24 hours because the test case database does not contain accurate time information at smaller granularities than days. The types of events include patient encounters (with diagnosis codes), lab tests, prescriptions and correspondence. For prescriptions, cessation dates are, where possible, heuristically deduced from fields in the prescription. Events have various attributes, such as codes, values and text, depending on their type. Practice Explorer is currently only able to extract data from Profdoc Vision, a patient record system widely used by Norwegian general practitioners. In this system, encounter diagnoses are coded according to the International Classification of Primary Care (ICPC), and prescriptions are coded according to the Anatomical Therapeutic Chemical Classification (ATC).

Query language

The main idea behind Practice Explorer's query language is to visualize queries as directed acyclic graphs, with each vertex describing a part of the history and with the edges always directed to the right, towards the present time. A query defines what a segment of a history must be like to constitute a match for that query. The informal interpretation of a query is that, for a query to match a segment of a history, it must be possible to simultaneously walk from left to right through the query and the history segment, encountering a matching history part for every query element passed. An edge between two elements indicates that the match for the element on the right hand side must begin at the exact same time that the match for the element on the left hand side ends. In other words, edges do not represent passage of time, but connects temporally juxtaposed events. Figure 1 shows two very simple queries. The bottom query specifies that a contact with diagnosis code K86 must occur, immediately followed by a period of medication with beta blockers. The top query specifies the same situation, except that an unlimited amount of time, represented by the spring-like middle element, is allowed to pass between the contact and the start of the medication period.

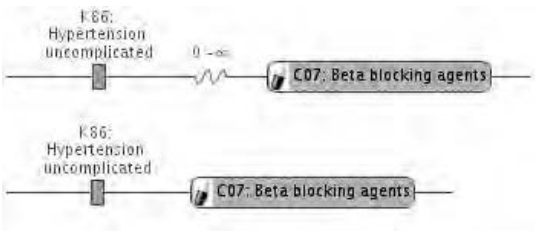


Figure 1 - Simple queries.

Table 2 - The central structure of the query language

```
<Q>        ::= <Point>
             | <Interval>
             | sequentialComposition(<Q>, <Q>)
             | parallelComposition(<Q>, <Q>)
             | alternativeComposition(<Q>, <Q>)
             | firstOccurrence(<Q>)
<Point>    ::= encounter(<IcpcCode>)
             | prescription(<AtcCode>)
             | medicationStart(<AtcCode>)
             | medicationEnd(<AtcCode>)
             | test(<TestType>, <TestValueRange>)
             | correspondence(<CorrType>)
             | dateControl(<DateRange>)
             | ageControl(<AgeRange>)
<Interval> ::= medicationInterval(<AtcCode>)
             | timeWindow(<Dur>, <Dur>)
             | timeWindowWith(<Dur>, <Dur>, <Q>)
             | timeWindowWithout(<Dur>, <Dur>, <Q>)
             | repetition(<Q>, <Int>, <Int>, <Dur>, <Dur>)
```

The user builds a query by dragging components from a menu to a panel containing the query graph under construction. Dialog boxes ask for necessary parameters as components are added. Figure 2 shows a moderately complex query graph.

The query language has a textual syntax, the central structure of which is defined by the grammar in table 2. A query is any derivation from <Q>. The visual representation of many of the elements can be seen in figure 2.

Space does not permit giving a full formal definition of the query language and its semantics. We will now informally describe the semantics of the query components given by the grammar.

Point queries, except for date controls and age controls, match events of the corresponding types satisfying the criteria in the parentheses. Date controls and age controls match all points in histories where the date or patient age could be successfully verified to belong to the specified range.

- Of the interval queries, medication intervals match any continuous time period where the patient is deduced, from prescription events, to be taking medication of the given code. The three types of time windows match time periods of duration between a given minimum and a given maximum, with a possible additional requirement that a given query must, or must not, match within the period. Repetition queries require that a given query gives repeating matches with upper and lower limits for the count and the duration allowed between repeats.

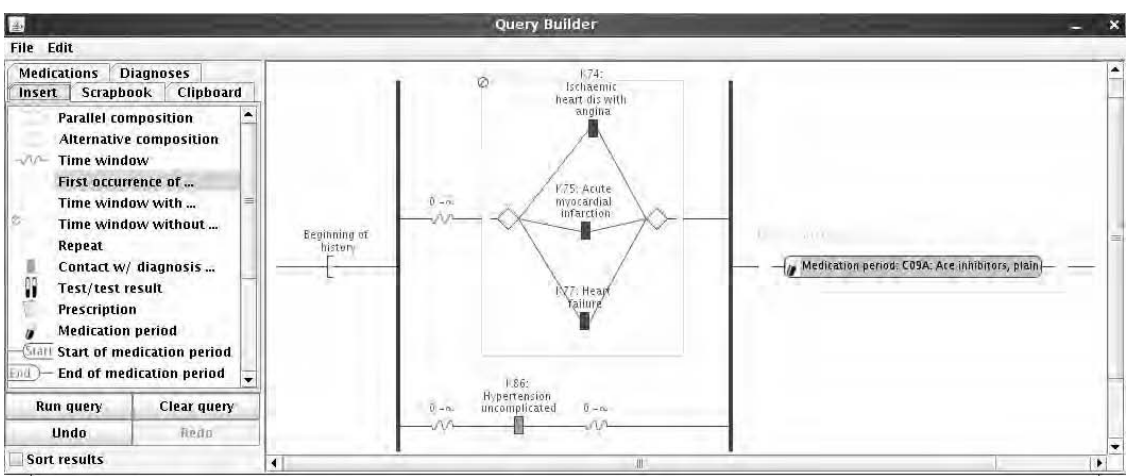


Figure 2 - Query building window showing a query satisfying the query need from table 1. From left to right, we see the following: an element matching the beginning of the history, followed by a branching into two parallel threads, along the upper of which the last two-year-period must not contain a contact with any of three codes and along the lower of which a contact with a given code must occur somewhere, followed by a re-joining of the parallel threads, followed by an interval of medication with a given type of drug, which must be the first of its kind in the entire history. (The query builder was constructed using the prefuse toolkit [6].)

- Sequentially composed queries require that the queries match, in such a way that the end of match of the first query coincides with the beginning of the match of the second query. This construction gives rise to the edges in the query graph.
- Parallelly composed queries require that the queries match, and in such a way that the start points of the matches coincide and that the end points of the matches coincide.
- Alternatively composed queries require that at least one of the queries matches.
- First-occurrence queries merely require that there is a match of the given query and that this is the first such match in the entire history. (This effect can be achieved with a combination of other elements, but the requirement was important enough to warrant support as a simpler formulation.)

The grammar gives the query graph an underlying tree structure where the matches of a node depends on how the matches of its children cohere with the rules and parameters for the node itself. The matches of a query graph are the matches of its root node. Execution of a query is performed with the leaf nodes scanning the history sequentially and the internal nodes iterating through the matches from their children.

The language elements described above straightforwardly satisfy requirements 1-10. Intersection and union of query results can easily be achieved by parallel and alternative composition of the queries, thus satisfying requirement 11. The recursive definition of the query language means that parts of queries, down to primitive components, are queries in their own right that can easily be collapsed, given names, saved and re-used, thus satisfying requirement 12.

Result visualization

Practice Explorer consists of two main windows. The query builder was described in the previous section. The history explorer, is shown in figure 3. It displays a number of vertically stacked horizontal bars, each providing a compact, very simple, LifeLines-like [7], explorable visualization of a history above a common time axis. The history explorer dynamically limits its view to the histories containing matches for the query given by the current state of the query builder, marks the hits of the query with red boxes and synchronizes the histories so that they are aligned on the first match.

Understandability testing

We performed 12 two-hour understandability tests, each followed by a brief questionnaire. The goal was not to test general usability, but to investigate the intuitive understandability of the underlying principles. We therefore simplified the query builder, keeping only the following query elements: encounter, history start, history end, date control, age control, time window, time window without ..., sequential, parallel and alternative composition. We applied the system to 2066 general practice patient histories of lengths up to 12 years. The test subjects were 4th- and 5th-year medical students.

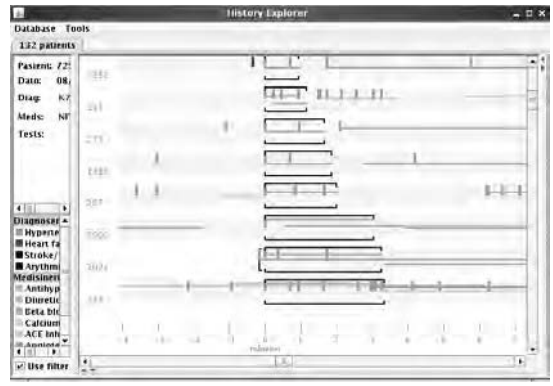


Figure 3 - Visualization of a query result

Tests 1-3 were informal, with demonstration and explanation provided during the test. In tests 4-6, the subjects were given a leaflet containing instructions and examples, as well as a number of query construction tasks to be performed. During tests 1-6 we made the observation that the learning process was greatly enhanced by the availability of examples, demonstration or interactive assistance. In particular, getting feedback on the correctness of one's queries improved ability to accomplish further tasks. We suspected that these factors could obscure issues related to our stated goal, which was to investigate the intuitive understandability of the underlying principles of the query system. Therefore, we devised a testing framework with the following rules: 1) The user will receive written instructions and query construction tasks to solve within an allotted time, 2) There must be no demonstration or assistance and 3) The instructions must contain no example queries. We carried out tests 7-12 in compliance with these rules. The subject was left alone for 105 minutes to read the instructions and attempt to solve the tasks, occasionally prompted via a loudspeaker to explain his or her thinking. Of the 25 tasks, the first 5 were point queries, the next 8 also involved intervals, the next 4 added branching constructs, and the final 8 required complex combinations of different types of elements, In the final 15 minutes, a developer interviewed the subject about the tasks that the test subject had failed to solve, and explained how those tasks should have been solved. Screen, video and audio were captured for further investigation.

Results

Figure 4 summarizes the correctness scores of test subjects 7-12 on the 25 query construction tasks. For an query to be classified as correct, it had to give the exact intended result.

From studying the users' actions on the screen, listening to them thinking aloud, and interviewing them afterwards, we made the following observations:

1. When a time window was specified as having a duration between an upper and a lower limit, for example 0 and 12 months, test subjects would often think of this as a fixed-length window from relative time point 0 to

relative time point 12 months, even though the instructions explicitly stated otherwise.

2. The test subjects made many errors where they seemed to assume that the query described the whole history and not just a fragment of it.
3. Test subjects frequently and successfully used the match-aligned result visualization to check if their queries were correct.
4. The test subjects made many errors related to not understanding that matches of parallel queries must cover the same time period.
5. On tasks requiring nested branching constructs, test subjects seemed to strain under the mental effort required. Some ceased serious efforts to find a solution.

Most of these observations either did not occur before tests 7-12, or occurred much more strongly in tests 7-12.

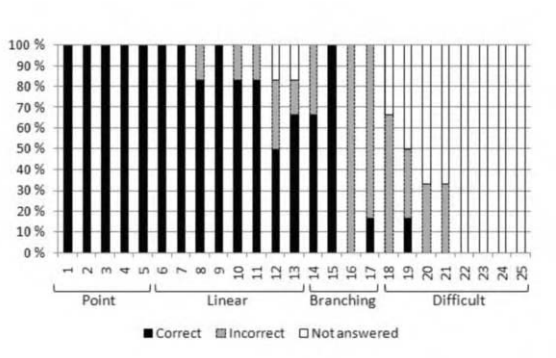


Figure 4: Scores for test subjects 7-12 on the 25 tasks.

On the questionnaire, which used a 5-point Lickert scale, 3 out of 12 marked “agree” or “fully agree” on the proposition “I think the visualization and building of the queries was easy to understand based on the instruction”. 9 out of 12 marked “agree” or “fully agree” on the proposition “I think the visualization and building of the queries was easy to understand based on the explanation given afterwards.”

Discussion

Studying figure 4, it appears that test subjects 7-12, even without examples, demonstration or assistance, arrived at a reasonably good understanding of the linear parts of the query language. The branching constructs appear to have been poorly understood. The survey answers indicate that explanation did help a lot on understanding. From the observations, we arrive at some lessons that may be helpful when designing this kind of system:

1. It seems much more natural for users to interpret a minimum and maximum number of time units as time points defining a constant-length interval rather than as bounds on the duration of a flexible interval.
2. Users may find it more natural to build a query describing the whole history rather than just a fragment.
3. Match-aligned result visualization can help users correct their own thinking and build correct queries.
4. Attention must be paid to make the visualization of branching constructs reflect their properties and reduce the effort required in reasoning about them.
5. Refraining from giving examples, demonstration, assistance or feedback may make testing more effective in uncovering problems.

Another question, partly addressed by our requirements collection, is whether the language is sufficiently expressive. On this topic we also note that, out of Allen's 13 primitive interval relations [8], our query language does not support *overlaps* or *overlapped-by*. The other 11 can be constructed with parallel composition and time windows. Supporting *overlaps*/*overlapped-by* would probably make the language more complex. Since we can match the start or end of a medication interval occurring during the match of another interval, we have not yet seen any query need requiring *overlaps*/*overlapped-by*.

Conclusion

We have designed, implemented and evaluated a temporal, graph-based query language based on our collected tentative list of requirements. Our understandability tests indicate that domain users relatively easily can construct point and interval queries, but not branching queries. Based on observations done during testing, we arrived at a list of lessons of potential relevance for the design of similar systems.

Acknowledgments

The authors wish to thank Arild Faxvaag, Anders Grimsmo, Joe Siri Ekgren, Berit Brattheim, Dag Svanæs, Yngve Dahl and Terje Røsand for valuable assistance and comments. Part of the work was funded by The Norwegian Research Council.

References

[1] Quaglini S, Ciccarese P, Micieli G, Cavallini A. Non-compliance with guidelines: motivations and consequences in a case study. Stud Health Technol Inform. 2004;101:75-87.

[2] Fails JA, Karlson A, Shahamat L, Shneiderman B. A Visual Interface for Multivariate Temporal Data: Finding Patterns of Events across Multiple Histories. IEEE Symposium on Visual Analytics in Science and Technology 2006.

[3] Hibino S, Rundensteiner EA. A visual multimedia query language for temporal analysis of video data. MultiMedia Database Systems, pages 123--159. Kluwer Aca, 1996.

[4] Chittaro L, Combi C. Visualizing queries on databases of temporal histories: new metaphors and their evaluation. *Data & Knowledge Engineering* 2003;44:239-264.

[5] Fowler M. UML Distilled, 3rd edition. Boston: Pearson, 2004.

[6] Heer J, Card SK, Landay JA. prefuse: a toolkit for interactive information visualization. Proceedings of CHI’05, ACM Press, 2005.

[7] Plaisant C, Milash B, Rose A, Widoff S, Shneiderman B. Lifelines: Visualizing Personal Histories. Proceedings of CHI’96. ACM Press, 1996.

[8] Allen J. Maintaining knowledge about temporal intervals. *Communications of the ACM* 1983:26(11):832-843.

Address for correspondence

Ole Edsberg, edsberg@idi.ntnu.no

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Evaluation of a Simple Method for the Automatic Assignment of MeSH Descriptors to Health Resources in a French Online Catalogue

Aurélie Névéol[a], Suzanne Pereira[b, c, d], Gaetan Kerdelhué[b], Badisse Dahamna[b], Michel Joubert [d], Stéfan J. Darmoni[b, c]

[a] *U.S. National Library of Medicine, National Institutes of Health, Bethesda, USA*
[b] *CISMeF, Resarch Department, Rouen University Hospital, France*
[c] *GCSIS, LITIS EA 4051, Institute of Biomedical Research, University of Rouen, France*
[d] *LERTIM, Marseille Medical University, France*

Abstract

Background: *The growing number of resources to be indexed in the catalogue of online health resources in French (CISMeF) calls for curating strategies involving automatic indexing tools while maintaining the catalogue's high indexing quality standards.* ***Objective****: To develop a simple automatic tool that retrieves MeSH descriptors from documents titles.* ***Methods:*** *In parallel to research on advanced indexing methods, a bag-of-words tool was developed for timely inclusion in CISMeF's maintenance system. An evaluation was carried out on a corpus of 99 documents. The indexing sets retrieved by the automatic tool were compared to manual indexing based on the title and on the full text of resources.* ***Results****: 58% of the major main headings were retrieved by the bag-of-words algorithm and the precision on main heading retrieval was 69%.* ***Conclusion****: Bag-of-words indexing has effectively been used on selected resources to be included in CISMeF since August 2006. Meanwhile, on going work aims at improving the current version of the tool.*

Keywords:

abstracting and indexing/methods, algorithms, catalogs, library, information storage and retrieval/methods, evaluation study, France, medical subject headings, Natural Language Processing

Introduction

Since 1995, the catalogue of online health resources in French (CISMeF)[1] has been selecting institutional and educational resources for patients, students and health professionals. The resources are described with a set of metadata and Medical Subject Headings (MeSH® descriptors[1]). Faced with a growing amount of resources to be indexed and included in the catalogue, the CISMeF team has recently adopted a new curation policy. The use of automatic indexing tools was deemed necessary to reduce the indexing back-log of about 7,000 resources. Considering the limitations of automatic indexing methods, it is necessary to distinguish clearly which resources in the catalogue are indexed manually and automatically [3]. In fact, the question arises at the time resources are considered for inclusion in the catalogue: the curation policy must define which resources should be indexed with the higher quality indexing produced manually and which resources may be given less attention and indexed automatically. At CISMeF, the decision was mainly based on two criteria: (a) the depth of indexing required and (b) the level of coverage of a given topic. High quality (manual) indexing must be available for all the topics covered in the catalogue, but when a reasonable level of coverage has been reached (about a dozen resources according to the curator) additional resources may be indexed automatically. Besides, we assume that automatic indexing is more suitable to cover only the central main concepts discussed in a resource whereas manual indexing will be necessary for in-depth indexing. In any case, in answer to an information query, manually indexed resources will always be displayed before automatically indexed resources. Moreover, the type of indexing (manual *vs.* automatic) will be shown to the user.

Teaching material (N=3629) and clinical guidelines (N=2978) are two vast categories of resources indexed in CISMeF. The average number of descriptors used to index a teaching resource is 9.89 +/- 10.87 vs. 13.64 +/- 15.69 for a clinical guideline. The difference is significant according to a Student test ($p<0.0001$) and illustrates the fact that clinical guidelines are indexed more in depth than teaching resources. In fact, for clinical guidelines, the curation policy for CISMeF is to be as exhaustive and as minute as possible: all clinical guidelines are to be indexed manually regardless of the current coverage in the catalogue. Teaching resources, on the other hand, may be given less attention once sufficient coverage has been reached, as they do not require in-depth indexing. To accommodate the difference between these resource types, a longer time is typically spent for indexing clinical guidelines[2] (about

1 MeSH is the US. National Library of Medicine (NLM)'s controlled vocabulary thesaurus. It consists of sets of terms naming biomedical related descriptors in a hierarchical structure. MeSH is used worldwide for indexing articles from the biomedical literature.

60 minutes *vs.* 15 minutes for teaching resources). Furthermore, CISMeF's curation policy is to use automatic indexing on teaching resources[3].

The VUMeF[4] project aims at increasing the amount of material available for French in the Unified Medical Language System. In particular, VUMeF participants agreed to make the development of automatic indexing tools a priority task of the project. It was addressed in two steps: first, automatically extracting a set of MeSH descriptors from online resources in French (which is discussed below) and secondly, in this set, selecting the major descriptors denoting central concepts discussed in the resource [2]. In the framework of VUMeF, the CISMeF team has been consistently researching and evaluating advanced automatic MeSH indexing techniques [3]. Although promising results have been obtained from a research prototype (see [4] for a comparative evaluation of CISMeF's MAIF and the NLM's Medical Text Indexer [5]), the integration of an operational system into the catalogue workflow is a lengthy process. For this reason, more readily available techniques are also investigated in order to speed-up the availability of automatic indexing tools and eventually complement other tools when they become available. Recent advances in Information Retrieval in CISMeF have resulted in the development of a query analysis algorithm designed to map free text to MeSH [7]. Moreover, new partnerships with health information providers accommodate the reception of resources to be included in the catalogue along with some metadata information such as the title, which previous research has shown to be significantly informative of the document content [3], [6]. In this paper we (a) assess the extent of the information contained in resource titles and (b) evaluate an automatic MeSH indexing approach based on bag-of-words indexing of resource titles in the specific context of teaching resources.

Materials and methods

Bag-of-words indexing algorithm

The algorithm used to extract keywords from the title of documents is similar to the query interpretation algorithm currently used in CISMeF [7]. We decided to use this algorithm for indexing based on the assumption that document titles for teaching resources are similar to information retrieval queries with respect to length and information content. In this experiment, teaching resources titles are processed in the same way as queries to extract MeSH descriptors which will then be used as candidate terms to index the teaching resource.

After the title has been normalized (accents are removed, all words are switched to lower case…) and stop words have been removed, a bag of words containing all the content words is formed. Each word is stemmed in order to account for some cases of word flexion and derivation. The "bag" thus obtained is sorted alphabetically and matched against a database of MeSH terms that have been processed in the same way. If no term is found, subsets of the original bag of words are processed. The size of the bags is decreased by one at each step of the process. In an effort to retrieve the most specific keywords, when a match is found, the corresponding content words are taken out of the bag before the next iteration. For a given bag size, when more than one match is found all candidates are kept. For example, the title "prevalence of hepatitis A and B" would yield both *Hepatitis A* and *Hepatitis B* when size-2 bags are considered, but not *Hepatitis* which would have been removed before size-1 bags were processed. If both MeSH main headings and subheadings are retrieved, all the legal[5] pairs are formed from both main heading and subheading sets.

Figure 1 illustrates the processing of a sample title[6] in the test corpus (see next section for corpus description). A bag of seven content words is obtained from the title "Tumeurs cérébrales chez l'enfant: particularités épidémiologiques, diagnostiques et thérapeutiques" (*Brains tumors in children : epidemiologic, diagnostic and therapeutic specificities*). No single MeSH term can be matched to it. Therefore, the size of the bag is gradually reduced. The two-word bag containing "cerebrales" and "tumeurs" yields the main heading "tumeurs du cerveau" (*Brain Neoplasms*). The two corresponding content words ("cerebrales" and "tumeurs") are taken out of the bag. Size-one bags containing the remaining words yield the main heading "enfant" (*Child*) and the subheadings "épidémiologie" (*epidemiology*) and "thérapie" (*therapy*). Finally, as both subheadings are allowable qualifiers for *Brain Neoplasm* but not for *Child*, a list of three indexing candidates is produced: *Child, Brain Neoplasms/epidemiology* and *Brain Neoplasms/therapy.* This particular example will be further commented on in the discussion section.

Test corpus

The algorithm was evaluated on a corpus of 99 teaching resources to be included in CISMeF that were selected for automatic processing. A professional indexer was asked to provide MeSH indexing sets for the corpus resources. First, the indexer was only shown the title of the resource and the indexing set automatically produced using the bag-of-words algorithm. His task was to revise the indexing set

2 Other cataloguing institutions may adopt different policies regarding time issues – e.g. at NLM, the average time spent indexing an article for Medline is 15 minutes regardless of the publication type.

3 e.g. a resource discussing drugs available for the treatment of thrombosis was selected for automatic processing because it consisted of lecture notes and the query "thrombosis/drug therapy" retrieved 31 manually indexed resources on 11/08/06

4 *Vocabulaire Unifié Médical Français* (French Unified Medical Vocabulary). Project sponsored by the French National Research Agency. http://www.vidal.fr/vumef/

5 For each main heading, MeSH defines a set of subheadings called "applicable qualifiers" that can be coordinated with it (e.g. */metabolisms* is applicable to *Amino Acids* but not *Hand*).

6 The original title ("Cancer de l'enfant: particularités épidémiologiques diagnostiques et thérapeutiques") was edited to illustrate an additional feature of the algorithm.

obtained automatically. Then, the indexer had access to the full text of the resource and was asked to index the resource as he would usually do, i.e. he selected MeSH descriptors to index the resource, and assigned to each a "major" or "minor" weight depending on how substantively the concept represented by the descriptor was discussed in the resource. As a result, each resource in the test corpus was manually annotated with two different sets of indexing terms: one based on the title only, and one based on the full text..

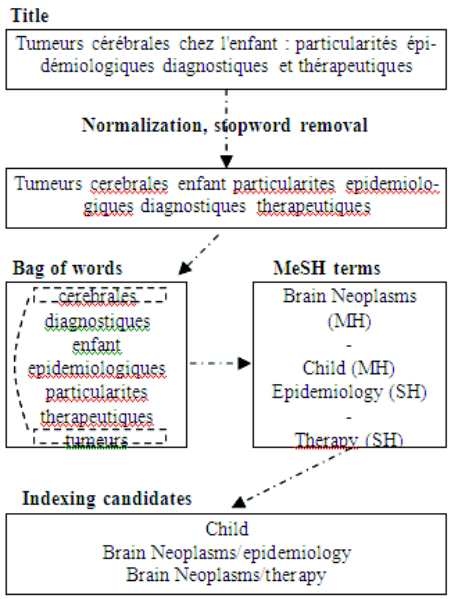


Figure 1 - Bag-of-words indexing of a sample title

It is important to stress that the annotations were obtained through an iterative process in which the quality of the indexing improved with every step, while being consistent [8] with the previous step: the automatic bag-of-words recommendations were revised to obtain manual title annotations. In turn, these were revised to obtain manual full text annotations.

In this evaluation, the indexer is not blind to the recommendations produced by the automatic tool. However our goal was to conduct an experiment reproducing real-life indexing settings where the recommendations of the automatic tool will be available to the human indexer when indexing a resource.

The revisions made by the indexer include adding a main heading, deleting a main heading, modifying the list of subheadings attached to a main heading. Table 1 shows the successive steps for a sample resource in the test corpus.

Evaluation measures

In this study we used precision and recall to compare a set of candidate indexing terms to a set of reference indexing terms. *Precision* corresponds to the number of terms present in both the candidate and reference sets over the total number of terms in the candidate set. *Recall* corresponds to the number of terms present in both the candidate and reference sets over the total number of terms in the reference set.

Table 1 - Processing of a sample corpus resource: overview of the indexing revision cycle

Title	Cancer de l'enfant: particularités épidémiologiques diagnostiques et thérapeutiques (*Cancer in children : epidemiologic, diagnostic and therapeutic specificities*)
Bag-of-words indexing (Title)	Neoplasms/epidemiology Neoplasms/therapy Child
Manual indexing (Title)	Neoplasms/diagnosis Neoplasms/epidemiology Neoplasms/therapy Child Pediatrics
Manual indexing (Full Text)	*Neoplasms/diagnosis *Neoplasms/epidemiology *Neoplasms/therapy Child Continuity of Patient Care Pediatrics/education

In addition, we considered three categories of terms:

- Descriptors: MeSH main headings or main heading / subheading pairs. In this category, subheading coordination is taken into account (e.g. *Pediatrics* does not match *Pediatrics/education*)
- Main headings: Any MeSH main heading. In this category, subheadings or stars (indicating major concepts) are not taken into account. (e.g. *Pediatrics* matches **Pediatrics/education*)
- Central-concept main headings: MeSH main headings that were marked as "major" using the star symbol "*" in the manual indexing based on the resource full text. In this category, subheadings are not taken into account. (e.g. **Pediatrics* matches **Pediatrics/education*)

Results

Information content of resource titles

Table 2 - Information content of resource titles

	Information Content Precision (%) – Recall (%)	
Descriptors	71	24
Main Headings (MH)	81	37
Central-concept (*MH)	78	78*

* Precision and recall figures are the same for central concepts because the indication of whether a term is central (major) is only available for manual indexing on the full text of the resource.

The extent of the information contained in teaching resource titles was assessed by comparing the manual indexing obtained from the title of the resource to the manual indexing obtained from the full text of the resource (see lines 3 and 4 in Table 1 for a specific example – in this case, the precision for *descriptors* indexing was 4/5=80% whereas the recall was 4/6=67%).

Performance of the bag-of-words indexing

The performance of the bag-of-words indexing was assessed by comparing the automatic indexing recommendations obtained by applying the bag-of-words algorithm on the resource title to the manual indexing:

- obtained from the resource title (overall results are shown in table 3; see lines 2 and 3 in table 1 for a specific example. In this case, the precision for *descriptors* indexing was 3/3=100% whereas the recall was 3/5=60%).

Table 3 - Performance of bag-of-words indexing (Title)

	Performance Precision (%) – Recall (%)	
Descriptors	62	56
Main Headings (MH)	69	64
Central-concept (*MH)	58	58

- obtained from the full text of the resource (overall results are shown in table 4; see lines 2 and 4 in table 1 for a specific example – in this case, the precision for *descriptors* indexing was 3/3=100% whereas the recall was 3/6=50%).

Table 4 - Performance of bag-of-words indexing (Full Text)

	Performance Precision (%) – Recall (%)	
Descriptors	54	16
Main Headings (MH)	66	30
Central-concept (*MH)	58	58

Discussion

Information content of teaching resources titles

According to table 2, overall, only 24% of the MeSH descriptors to index a teaching resource may be inferred from the resource title, however including 78% of the central concepts. This shows that, in our corpus, teaching resources titles contain explicit information as to the central content of the resource. In 81% of the cases, the main headings inferred from the title by the indexer were kept after reviewing the full text of the resources. For descriptors, this figure goes down to 71%. Some descriptors were discarded for not denoting concepts that were substantively discussed in the resource. The other descriptors were in fact main headings to which a subheading had to be attached – in the example presented in table 1, the subheading *education* had to be attached to *Pediatrics*, which was inferred from the title.

Bag-of-words indexing

Performance

According to table 4, more than half (58%) of the major main headings were retrieved by the bag-of-words algorithm and the precision on main heading retrieval was 69%. These results show that the algorithm is able to retrieve central concepts, while generating a reasonably low noise. However, the difference between the performance on descriptors and main headings (lines 2 vs. 4 in tables 3 and 4) indicates that the more difficult task of assigning subheadings to the main headings is lacking. The low recall for descriptors and main headings compared to the full text (table 4) reflects the amount of information that may be extracted from the sole title of the resource. For example, for main heading retrieval recall could not be higher than 37% (as shown in table 2), so 30% is comparatively a good performance.

Error analysis

Looking at sample revisions of the bag-of-words recommendations made by the indexer helps identifying the issues that need to be addressed in order to improve the automatic tool. Typical errors fall into the following categories:

- Stemming errors – in the example shown in table 1, the word « diagnostiques » was not stemmed properly and could not be mapped to the subheading *Diagnostic (Diagnosis)*. As a result, the pair *Neoplasm/diagnosis* was not retrieved
- Generic main headings – some MeSH terms that may appear in a resource title are so generic that they are rarely used for indexing. *Syndrom*, *Patient*, *Life* or *Health* are examples of such descriptors.
- Implicit indexing rules – through daily indexing practice, the indexers are able to infer descriptors that do not explicitly appear in the title of the resource. In the example shown in table 1, no cue from the title prompts the use of the main heading *Pediatrics*. However, *Pediatrics* is typically an appropriate descriptor to index a teaching resource discussing a particular disease onset in children (here, cancer).
- Level of specificity – some of the descriptors retrieved by the algorithm were sometimes related to descriptors selected by the indexers, although not identical. For example, *Hemorrhage/therapy* was retrieved instead of *Gastrointestinal Hemorrhage/therapy.*

Based on these observations, a « stop list » of common generic descriptors is currently being compiled and shall be used to reduce the noise of the algorithm. As such, the bag-of-words algorithm cannot be expected to retrieve descriptors that an indexer would *infer* rather than see in a resource title. For this reason, applying post-processing rules (such as described in [9]) to a set of main headings retrieved by the algorithm would be desirable in order to improve the automatic tool.

Although we anticipated that the loss of word order inherent to the "bag-of-word" approach may yield some noise, no errors related to word-order were observed on the test

corpus. This may be explained by the fact that teaching resources titles are generally short and to-the-point. Different results may be obtained when applying the algorithm to more lengthy and complex sentences as can be found in full text.

Limitations of this study

The bag-of-words indexing algorithm presented here was evaluated on a set of 99 teaching resources. The small size of the evaluation corpus is due to the amount of manual labour needed to produce the annotations of the title and full text of the resources. Moreover, the fact that the indexer was shown the automatic recommendations while he was producing his own set of descriptors may yield a bias in favour of the automatic tool.

Impact on CISMeF indexing procedure

The positive results of this study conducted in May 2006 prompted the effective use of bag-of-words indexing in the CISMeF catalogue as of August 2006. The original backlog of 6,832 resources was automatically processed with the bag-of-words algorithm. The automatic indexing for 1127 resources (including the 99 resources of our corpus) have been revised by an indexer and included with the manually indexed resources of the catalogue. For another 557 resources, manual revision is pending. Finally, the remaining 5148 resources are included with the automatically indexed resources of the catalogue (these resources had originally been classified as “low priority” and did not require in depth indexing).

Perspectives

Automatic indexing for CISMeF: In the near future, the automatic indexing of selected low-priority resources to be included in the CISMeF catalogue will no doubt consist of integrating the recommendations produced by the different methods studied by the CISMeF Team: MAIF (i.e. a combination of Natural Language Processing and Nearest Neighbors approaches) and the bag-of-words indexing presented here.

Other uses of bag-of-words indexing: Based on the results of this study, we are planning to adapt the bag-of-words algorithm to the coding of patient discharge summaries with ICD-10 [10], MeSH and SNOMED [11] and to the indexing of French FDA notices with the Unified Vidal Thesaurus (TUV). In this case, documents would be segmented at the sentence level and the algorithm would be applied to each sentence. Ultimately, our goal is to integrate these applications to produce a multi-terminology indexing tool able to process medical documents and extract concepts belonging to several terminologies (MeSH, SNOMED, ICD-10 and TUV).

Conclusion

We have presented a simple bag-of-words indexing method that retrieves MeSH descriptors from resource titles. An evaluation on a corpus of teaching resources shows good performance, in particular for central concepts. For this reason, bag-of-words indexing has been in use for selected CISMeF resources since August 2006.

Acknowledgments

This research was supported in part by VIDAL (http://www.vidal.fr/) and by an appointment of A. Névéol to the NLM Research Participation Program sponsored by NLM and administered by the Oak Ridge Institute for Science and Education. The authors would like to thank CISMeF indexers for their help in the study design and result analysis.

References

[1] Darmoni SJ, Leroy JP, Thirion B, Baudic F, Douyère M, Piot J. CISMeF: a structured Health resource guide. *Meth Inf Med* 2000; 39(1): 30-5.

[2] Joubert M, Peretti AL, Darmoni SJ, Dahamna B, and Fieschi M. Contribution to an Automated Indexing of French-language Health Web Sites. *Proc. AMIA Symp 2005*, pp 409-13.

[3] Névéol A. Automatisation des tâches documentaires dans un catalogue de santé en ligne. PhD dissertation. INSA de Rouen 2005.

[4] Névéol A, Mork JG, Aronson AR, Darmoni S.J. Evaluation of French and English MeSH Indexing Systems with a parallel corpus. *Proc. AMIA Symp. 2005*, pp 565-9.

[5] Aronson AR, Mork JG, Gay CW, Humphrey SM, Rogers WJ. The NLM Indexing Initiative's Medical Text Indexer. *Proc. Medinfo. 2004*, pp. 268-72.

[6] Barthes R. Analyse textuelle d'un conte d'Edgar Poe, l'aventure sémiologique. Paris, Seuil, 1985.

[7] Soualmia LF. Etude et évaluation d’approches multiples d’expansion de requêtes pour une recherche d’information intelligente. PhD dissertation. INSA de Rouen, 2004.

[8] Funk, ME, Reid, CA, McGoogan, LS. Indexing consistency in MEDLINE. *Bull. Med. Libr. Assoc.* 1983;2 (71):176-83.

[9] Névéol A, Shooshan SE, Humphrey SM, Rindflesh TC, Aronson AR. Multiple approaches to fine-grained indexing of the biomedical literature. *Proc. PSB*., 2007.

[10] World Health Organisation. ICD-10 International Statistical Classification of Diseases and Related Health Problems: Tabular List v. 1. Material, 1992.

[11] Lussier YA., Rothwell DJ., Cote RA. The SNOMED model: a knowledge source for the controlled terminology of the computerized patient record. *Methods Inf Med.* 1998;Jun;37(2):161-4.

Address for correspondence

Pr. Stéfan Darmoni - CHU de Rouen, DIR
1, rue de Germont - 76031 Rouen cedex - FRANCE

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Template-based Data Entry for General Description in Medical Records and Data Transfer to Data Warehouse for Analysis

Yasushi Matsumura[a], Shigeki Kuwata[b], Yuichiro Yamamoto[a], Kazunori Izumi[c], Yasushi Okada[c], Michihiro Hazumi[c], Sachiko Yoshimoto[a], Takahiro Mineno[a], Munetoshi Nagahama[a], Ayumi Fujii[a], Hiroshi Takeda[a]

[a] *Department of Medical Informatics, Osaka University Graduate School of Medicine, Japan*
[b] *Department of Medical Informatics, Tottori University Medical School, Japan*
[c] *Medical Systems Division, NEC Corporation, Japan*

Abstract

General descriptions in medical records are so diverse that they are usually entered as free text into an electronic medical record, and the resulting data analysis is often difficult. We developed and implemented a template-based data entry module and data analyzing system for general descriptions.

We developed a template with tree structure, whose content master and entered patient's data are simultaneously expressed by XML. The entered structured data is converted to narrative form for easy reading. This module was implemented in the EMR system, and is used in 35 hospitals as of October, 2006. So far, 3725 templates (3242 concepts) have been produced.

The data in XML and narrative text data are stored in the EMR database. The XML data are retrieved, and then patient's data are extracted, to be stored in the data warehouse (DWH). We developed a search assisting system that enables users to find objective data from the DWH without requiring complicated SQL.

By using this method, general descriptions in medical records can be structured and made available for clinical research.

Keywords:

medical records system, template, structured data entry, data warehouse, XML

Introduction

Recently, electronic medical record systems have been implemented and are used in many hospitals [1,2]. The medical staff expects data entered into electronic medical records (EMR) to be linked with a knowledge base for decision support, or used for assessment of medical activities and clinical research [3-5]. Medical records include patient profile data, diagnosis, chief complaints, patient histories, physical examinations, progress notes, records of orders, examination reports, operation reports, summaries and so on. Some of the patient profile data, diagnosis, records of orders, and laboratory test results are originally structured data, and have thus been stored in the clinical data warehouse (DWH) and previously used for clinical research [6-8]. However, most general descriptions in medical records are diverse and complicated, thus they are entered as free text. The valuable data for clinical research (e.g. stage of cancer, tissue type, side effects of medicine etc.) are usually entered as free text. Analysis of data in free text format is difficult. If these data are not available, the data in EMR declines in value.

In order to analyze the entered data, it must be structured, i.e. the description should be expressed by the assembly of data elements that consist of item and values with codes [9-11]. To get data into this format, an input template is the most practical method [12]. By using input templates, the entered data can be stored in the database and made available for data analysis. The descriptions in medical records are so diverse and complicated, however, that the template method is difficult to adopt. To overcome this problem, we developed a tree-structured template named 'dynamic template', whose basic concept was reported in 1998 [13].

It is usual to adopt a strategy of making templates based on their own database files [14,15]. This strategy makes it possible to analyze the data easily. Because of the lack of productivity and flexibility, however, it is difficult to produce templates for any possible description that might arise within broad specialties.

In this paper, we report on the dynamic template module and the method for analyzing the entered data by the template. In our method, the contents of the templates are independent of the database schema, thus the template can be easily produced and freely revised. We have produced many templates in broad specialties in every department. Furthermore, any data entered by the templates can be searched through DWH with the assistance of the searching system.

Methods

Structure of descriptions in medical records

In this paper we term a cluster of descriptions a 'describing object', which is a record about a certain property of a patient at a certain point in time. A medical record can be

considered as an aggregate of describing objects about one patient. A describing object has the attributes of patient ID, object name, observation time, description time, writer, department, and contents. In our model, contents of a describing object are expressed by the assembly in tree structure of data elements consisting of items and values.

Template module

We developed the input template module to get the data in this structure. The template contents can be considered as a union of the possible descriptions of patients regarding a certain describing object, i.e. each description of a patient is a subset of the template contents. Thus the structure of the template must be the same as that of the describing object.

The template content is the assembly of data elements consisting of items and options or text boxes, which are arranged in a tree structure. The template module displays all the elements in the same layer at once. When a value that has some subordinate elements is selected, these elements are displayed under the selected value. Because the template form changes according to the selected values, we call it 'dynamic template' (Fig. 1).

Although the structured data is processed easily by computer systems, it is difficult for users to understand. Therefore, we adopted a strategy of converting the structured data into narrative form, using the following rule: Each item and value has 4 types of character string: prefix, body, conjunction, and suffix. Linefeed can be set after each character string. If several values are selected for one item, conjunction is set between these values. If there are several items linked with one value, conjunction is set between these pairs of item and value, and the suffix of the value is set after those [16].

We adopted XML for the expression of the template contents master. This master includes the information about a structure of describing element, a way for expression on template of these elements, and a way for conversion of entered data to narrative form.

The main part of the elements in DTD of the template master is as follow.

<!ELEMENT ATOM-LIST (ATOM-STYLE,ATOM*)>

<!ELEMENT ATOM (ATOM-TYPE,DISPLAY-STRING,DOCUMENT-STRING,VALUE-LIST?)>

<!ELEMENT VALUE-LIST (VALUE-STYLE,VALUE*)>

<!ELEMENT VALUE (VALUE-TYPE,DISPLAY-STRING,DOCUMENT-STRING,ATOM-LIST?)>

<!ELEMENT DOCUMENT-STRING (DOC-BODY?,DOC-PREFIX?,DOC-CONJUNC-TION?,DOC-SUFFIX?)>

<!ELEMENT DOC-PREFIX (#PCDATA)>

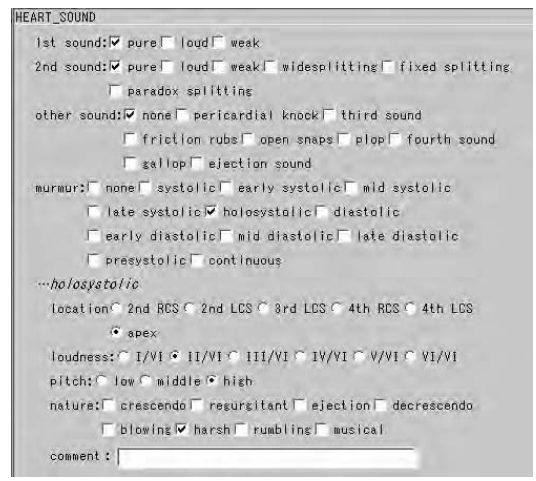
<!ELEMENT DOC-BODY (#PCDATA)>

<!ELEMENT DOC-CONJUNCTION (#PCDATA)>

<!ELEMENT DOC-SUFFIX (#PCDATA)>

DISPLAY-STRING is the character string expressed on the template. DOCUMENT-STRING is the character string on the narrative form, which consists of DOC-PREFIX, DOC-BODY, DOC-CONJUNCTION and DOC-SUFFIX. ATOM is an item of the describing element. ATOM-LIST includes ATOM, and ATOM includes VALUE-LIST. VALUE-LIST includes VALUE, and VALUE includes ATOM-LIST. In this format, the nested structure is presented.

Because the entered patient data is sometimes recalled with the original template and then revised, we adopted a method in which entered patient data is included in the template contents master. We put the SELECTED attribute of VALUE-LIST element in the template master. Before the value is selected, no is set to SELECTED attribute. When the value is selected, it changes to yes. When a character string is entered in a text box, it is set as the element of DISPLAYSTRING and DOCUMENT-STRING of VALUE-TYPE.



Narrative form

1st sound pure, 2nd sound pure, no other sound

holosystolic murmur (apex Levein II/IV high pitch harsh)

diastolic murmur (2nd LCS Levein I/IV high pitch regurgitant)

Figure 1 - An example of the dynamic template.

When the user selects "holosystolic" in " murmur", then the portion under "...holosystolic" appears. After data is entered by this template, the narrative form is generated.

Storage of data into the EMR database

In the EMR database, all kinds of patient data are stored. We call a cluster of information a 'medical event', which is stored in one record of the database. The describing object is one of the medical events. The data in XML and narrative text data are stored in each field, respectively.

EMR database is not suitable for data analysis, because the procedure for such analysis affects the response of the daily online transaction procedure. Furthermore, the patient data is included in the XML. Thus, in order to

search the objective data, all the records have to be checked. To overcome this problem, the data in EMR is transferred to the DWH. In this process, the XML data is parsed to extract patient data, which is stored in the database of DWH (Fig. 2). The structure of the database of DWH is quite simple. One record is made to correspond to one value. The main fields of this database file are as follows: patient ID, date, template name, xPath name, xPath code, value code, value name, suffix. The xPath code and the xPath name indicate the traced items and values in the tree structure. The xPath code (or the xPath name) is expressed by the traced items and values codes (or names) connected with "\", e.g. \A01\V0101\A02.

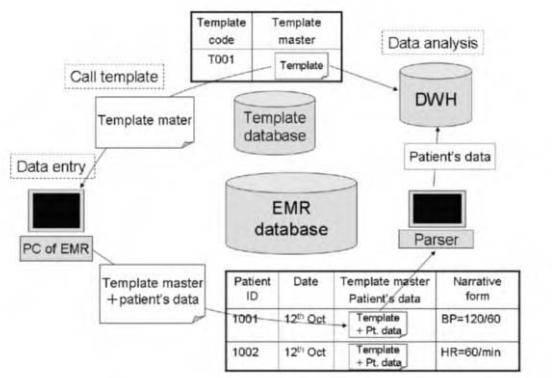


Figure 2 - Flow of template master and patient's data

Search assisting system

We developed an application system for searching the objective data from this database of DWH. In this system, users easily designate the objective data using a GUI. When users select an objective template, the system shows the names of items and values in the template. Thus, users can easily set the combination of search conditions in one template. The character string entered in the text box can also be the object of a search. In the case of numerical values, the conditions "more than" or "less than" can be set. The application system makes a SQL request according to the user's designation and searches the objective data from the DWH database. When the system finds the description that meets the condition, a progress note including the description is shown. Additionally, the system shows the progress notes of other days for the same patient.

Results

Availability of template

Preparation of input templates that are useful and convenient for every member of the hospital staff is essential before a hospital initiates an EMR system. The task is so large for each hospital that we formed an organization to support the preparation of template contents and their supply. The template contents that are developed for hospital use are then stored in the database, from which the staff of the hospital can select useful ones for their practice. The dynamic template module is used in 35 hospitals as of October 2006. Different styles of templates were sometimes produced for a concept. As of that time, 3725 templates (3242 concepts) had been produced and stored in the database. The numbers of template in each class are shown in Table 1. In the description of progress notes, the number of templates for "physical findings" is much greater than the numbers for "symptom", "assessment" or "plan". Not only doctors, but also nurses, pharmacists and nutritionists used templates. Templates are frequently used in examination reports, operation reports, and summaries.

Table 1 – The classification of templates

Classification of Template	No. of Concepts	No. of Template
Symptom	37	46
Physical finding	399	494
Assessment	157	168
Plan	24	26
Disease relating description	618	766
Patient history	184	211
Summary	62	64
Examination report	267	313
Operation report	55	62
Medical checkup	79	92
Description for rehabilitation	134	157
Description by social worker	23	23
Description by nurses	956	1013
Description by pharmacists	12	12
Description by nutritionists	44	44
other	191	234
Total	3242	3725

We examined 100 templates randomly selected from the database. When we count the same item linked with multiple values as 1, the average number of items in one template is 34.8. The average number of layers in one template is 8.1. We examined the maximum depth of layers in one template, and found 1 layer in 17%, 2 layers in 43%, 3 layers in 24%, 4 layers in 25%, and 6 layers in 1% of templates. The total number of items in the database is estimated as about 112,800.

Data warehouse and search assisting system

The DWH and search assisting system were implemented in Osaka University Hospital, and their usefulness was

evaluated. Without the search assisting system, users have to examine the xPath code of the value they want to search beforehand. Furthermore, it is quite difficult to write SQL for a combination of search conditions in one template. Contrasted with this, users can now search the objective data easily by this searching system.

Discussion

General descriptions in medical records, such as progress notes, examination reports, operation reports and summaries, are so diverse and complicated that these data are generally entered as free text in EMR. In order to use these data for research, clinical evaluation and so on, natural language processing is one of the possible methods [17,18]. However, to achieve good results by this method, all of the words in the entire medical field, including abbreviations and frequent typing errors, have to be entered into this system beforehand. This would entail tremendous amount of work.

The strategy of template-based data entry is a practical method from the viewpoint of data analysis [12]. The simple templates that are generally adopted, however, limit what users wish to express, and users have no choice but to tolerate these limitations.

Descriptions in medical records are the observation records of patients. When the observation object is normal, the user just enters “nothing particular”. However, when something abnormal is found, they describe the object in detail. For the observation records, a tree structure is suitable: an abnormal object is described by several features, which may be further described by other properties [19,20]. Thus, a tree structure is necessary for the template. If the structured data is expressed as it is in EMR, it is not acceptable for users because it is quite different from the expression in free text. A person prefers an expression in which obviously understandable words are omitted, rather than a strictly redundant expression. To overcome this problem, we convert the structured data into narrative form [16]. By using this template, users can enter what they want to record in shorter time without the risk of typing errors. This strategy is quite acceptable for users.

The dynamic template is implemented in the EMR system produced by NEC and 35 hospitals actively use this module. Before starting to use EMR system, hospital staff prepared templates useful for their daily practice. During operating the EMR, templates were revised and new ones were added according to the users’ requirement. More than 3725 templates have been produced and more than 100,000 items were entered in the template master. More templates about physical findings or examination reports have been produced than those about symptoms. Template-based data entry is suitable for actively acquired data. On the other hand, it is not suitable for passively acquired data such as symptoms. The concept, which has many options (e.g. portion of skin in dermatology), is difficult to handle with templates. Although the dynamic template is not effective for every type of description in medical records, its popularity and practical accomplishments show that the strategy of dynamic templates, i.e. tree-structured template and conversion of structured data into narrative form, is acceptable in many fields.

For general descriptions in medical records, many types of template have to be prepared. To bring this about, trial and error is inevitable, even after active use in daily practice [21]. Thus, it is necessary to have a system that enables users to produce templates easily and revise them even after releasing. Generally a template is made based on its own database file [14,15], which must be set on each occasion to produce a new template. If the template is revised, the corresponding database file must also be revised. Because this method disperses a patient data in many database files, it is not suitable for EMR system, which must enable users to refer quickly to any patient data.

In our method, the template content master that regulates the content of a template and the patient’s data are simultaneously expressed in XML. After entering patient data by the template, the XML data and the narrative form of patient data are stored in the EMR database. Although this schema is practical for EMR systems, data analysis is virtually impossible. Thus, we developed a parser system that retrieves the data in XML from the EMR database and extracts patient data from it. The parser system then stores the patient data in the DWH database, in which one record corresponds to one value. Because patient data is originally in tree structure, xPath code and name is used to designate each property. This database schema is independent from the template contents; therefore, the user can produce and revise templates freely without thinking of the database schema.

Using this DWH, data entered by the template can be used for analysis. It is, however, difficult for users to find the xPath code of the objective data. Furthermore, in this database schema, it is difficult to execute combined search conditions in one template. Accordingly, we developed a search assisting system which assists in finding the xPath code in the template, and thus enables the user to easily set combinations of search conditions. Although this database schema is not the best for data analysis, the user can search the objective data from the database with the support of the search assisting system.

By using this method, general descriptions in medical records can be structured and analyzed. Although it has been said that structured data entry is essential for advanced functioning in EMR, there are a few reports that it succeeded when limited to a few a fields [14,15,22] or a few department [23,24]. We achieved active use of the structured data entry system in every department in many hospitals.

Acknowledgments

This work was supported by a grant-in-aid for scientific research from the Ministry of Education, Culture, Sports, Science and Technology of Japan **(7)** to Y.M. (15500341)

References

[1] Delpierre C, Cuzin L, Fillaux J, Alvarez M, Massip P, Lang T. A systematic review of computer-based patient record systems and quality of care: more randomized clinical trials or a broader approach? Int J Qual Health Care. 2004 Oct: 16(5):407-16.

[2] Matsumura Y, Kuwata S, Kusuoka H, Takahashi Y, Onishi H, Kawamoto T, Takeda H. Dynamic viewer of medical events in electronic medical record. Medinfo2001: 648-52.

[3] Hunt DL, Hayness RB, Hanna SE, Smith K. Effects of computer-based clinical decision support systems on physician performance and patient outcomes: a systematic review. Jama 1998: 280:1339-46.
[4] Van Ginneken AM. The computerized patient record: balancing effort and benefit. Int J Med Info 2002: 65: 97-119.
[5] Bates DW, Pappipus E, Kuperman GJ, et al. Using information systems to measure and improve quality. Int J Med Infom. 1999: 53(2-3): 115-24.
[6] Wisniewski MF, Lieszkowski P, Zagorski BM et al. Development of a clinical data warehouse for hospital infection control. J Am Med Inform Assoc. 2003: 10(5): 454-62.
[7] Grant A, Moshyk A, Diab H, et al. Integrating feedback from a clinical data warehouse into practice organization. Int J Med Inform. 2006: 75(3-4): 232-9.
[8] Zhang Q, Matsumura Y, Teratani T, Yoshimoto S, Mineno T, Nakagawa K, Nagahama M, Kuwata S, Takeda H. Evaluation of the observed cisplatin nephrotoxicity in adult cancer inpatients: a historical cohort study by using clinical data warehouse. Journal of Health Science 2006: 52(2) 192-7.
[9] Lee FC, Chong WF, Chong P, Ooi SB. The emergency medicine department system: a study of the effects of computerization on the quality of medical records. Eur J Emerg Med. 20018(2):107-15.
[10] Henry SB, Douglas K, Galzagorry G, Lahey A, Holzemer WL. A template-based approach to support utilization of clinical practice guidelines within an electronic health record. J Am Med Inform Assoc. 1998: 5(3): 237-44.
[11] Los RK, van Ginneken AM, van der Lei J. OpenSDE: a strategy for expressive and flexible structured data entry. Int J Med Inform. 2005: 74(6): 481-90.
[12] Rose EA, Deshikachar AM, Schwartz KL, Severson RK. Use of a template to improve documentation and coding. Fam Med. 2001: 33(7): 516-21.
[13] Matsumura Y, Takeda H, Okada T, Kuwata S, Nakazawa H, Hazumi N, Inoue M. Devices for structured data entry in electronic patient record. Medinfo. 1998: 9 Pt 1:85-8.
[14] Green JD Jr, Postma DS, Giddings NA, Sapp KR, Skinner T. Computerized medical record in a private neurotology practice. Am J Otol. 2000: 21(4):589-94.
[15] McCullagh PJ, McGuigan J, Fegan M, Lowe-Strong A. Structure data entry using graphical input: recording symptoms for multiple sclerosis. Stud Health Technol Inform. 2003;95:673-8.
[16] Matsumura Y, Kuwata S, Okada Y, Emichi H, Kawamoto T, Shanmei J, Chen Y, Takeda H. Method of transforming of structured data into narrative form. Medinfo 2004;1746.
[17] Meystre S, Haug PJ. Natural language processing to extract medical problems from electronic clinicaldocuments: performance evaluation. J Biomed Inform. 2006: 39(6): 589-99.
[18] Hyun S, Bakken S, Friedman C, Johnson SB. Natural language processing challenges in HIV/AIDS clinic notes. AMIA Annu Symp Proc. 2003: 872.
[19] Hanzlicek P, Spidlen J, Nagy M. Universal electronic health record MUDR. Stud Health Technol Inform. 2004: 105: 190-201.
[20] van Ginneken AM. Considerations for the representation of meta-data for the support of structured data entry. Methods Inf Med. 2003: 42(3): 226-35.
[21] Hobson JC, Khemani S, Singh A. Prospective audit of the quality of ENT emergency clinic notes before and after introduction of a computerized template. J Laryngol Otol. 2005: 119(4): 264-6.
[22] Sistrom CL, Honeyman JC, Mancuso A, Quisling RG. Managing predefined templates and macros for a departmental speech recognition system using common software. J Digit Imaging. 2001: 14(3):131-41.
[23] Bleeker SE, Derksen-Lubsen G, van Ginneken AM, van der Lei J, Moll HA. Structured data entry for narrative data in a broad specialty: patient history and physical examination in pediatrics. BMC Med Inform Decis Mak. 2006: 136:29.
[24] Los RK, van Ginneken AM, van der Lei J. OpenSDE: a strategy for expressive and flexible structured data entry. Int J Med Inform. 2005: 74(6): 481-90.

Address for correspondence

Yasushi Matsumura MD, PhD.
Department of Medical Informatics, Osaka University Graduate School of Medicine.
2-15 Yamada-oka, Suita, Osaka 565-0871, Japan
Tel: (+81)6-6879-5900, Fax: (+81)6-6879-5903
E-mail: matumura@hp-info.med.osaka-u.ac.jp

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Challenges and Methodology for Indexing the Computerized Patient Record

Frédéric Ehrler[a,b], Patrick Ruch[b] , Antoine Geissbuhler[b], Christian Lovis[b]

[a] *Artificial Intelligence laboratory, University of Geneva, Geneva Switzerland*
[b] *Service of Medical Informatics, University Hospitals of Geneva, Geneva Switzerland*

Abstract

Patient records contain most crucial documents for managing the treatments and healthcare of patients in the hospital. Retrieving information from these records in an easy, quick and safe way helps care providers to save time and find important facts about their patient's health.

This paper presents the scalability issues induced by the indexing and the retrieval of the information contained in the patient records.

For this study, EasyIR, an information retrieval tool performing full text queries and retrieving the related documents has been used.

An evaluation of the performance reveals that the indexing process suffers from overhead consequence of the particular structure of the patient records. Most IR tools are designed to manage very large numbers of documents in a single index whereas in our hypothesis, one index per record, which usually implies few documents, has been imposed. As the number of modifications and creations of patient records are significant in a day, using a specialized and efficient indexation tool is required

Keywords:

information retrieval, patient records

Introduction

The ideal patient record should be unique and persistent over time, consolidated within a distributed computerized patient record (CPR). Whatever its technical, architectural and conceptual organization, accesses should only be constrained by access policies. However, not all the information contained in the records is relevant for each encounter. Care providers waste precious time at searching and browsing the patient record to collect all information pertinent to the actual situation. The problem is emphasized when specific information is scattered in numerous documents. Therefore, creating an efficient tool that allows retrieving relevant information from patient documents should improve the efficiency of care providers.

This paper presents the methodological approach for using an Information Retrieval (IR) tool in order to index the patient records and allowing fast retrieval. The specificity of our approach arises from the specific structure of the patient records. Contrarily to common IR task where all the documents are stored in a unique and large corpus, we have to work with numerous small corpora that must be indexed independently.

Two key measures have been used to evaluate performance, efficiency and effectiveness. As no gold standard was available, an ad-hoc method of automatic query generation has been used to generate the required data. The structure of the dataset has been modified to analyze the consequence of the patient records corpus structure on the indexing process.

Background

Patient records in Geneva University Hospital

The Geneva University Hospitals (HUG) is a consortium of hospitals in four campuses and more than 30 ambulatory facilities in the state. The HUG have about 2,000 beds, 5,000 care providers, over 45,000 admissions and 850,000 outpatients' visits each year. Over than 20,000 computerized records are open daily. More than 4,000 care providers do use this system, including physicians, nurses, medical secretaries, social care providers, physiotherapists, nutritionists, music therapists, etc. Beside order entry, clinical documentation is an important pillar of the CPR. More than 50,000 images, 25,000 lab results, 8,500 documents are stored every day. The clinical database contains more than 130 millions patient facts. All data is available online, there are no archives.

Information retrieval

The purpose of a "document retrieval system" is to select, from a relatively large collection of documents, a manageable number of documents that is likely to satisfy an expressed need for information [1]. It deals with the representation, storage, organization and access to information items. To accomplish the retrieval task, the system needs to build, usually prior to queries, a representation of each document in the collection. This representation consists of a table containing links that allow knowing which terms occur in the documents with which frequency [8].

During the search process, the system typically computes the degree of match between the terms contained in the index and a corresponding set of terms derived from the query [6]. This degree of match based on the frequency of the words in the document itself (TF) and in the corpus

(IDF) provides the basis for deciding whether a document should or not be retrieved.

IR evaluation

Two very different aspects of an IR system can be measured: efficiency and effectiveness [2].

Efficiency can be measured in terms of the resources required by the system for the whole IR process, including storage space and computing resources. Effectiveness attempts to measure the success of an IR system. It should provide a domain neutral measure of the ability of the system to retrieve all relevant documents [5].

The most widespread method of evaluating effectiveness of an IR system involves providing precision–recall values for a set of queries posed on a specific documents collection [4]. Recall measures the proportion of relevant documents retrieved and precision measures the proportion of retrieved documents which are relevant.

Test collection

The final application of our system is dedicated to index the patient records; however, experiments done in this paper are performed on OHSUMED dataset to facilitate measures and reliability.

In this paper, two different kinds of corpus are considered, the source and the experimental corpus. The source corpus denotes the collection from which all the documents supporting the experiments are extracted. The experimental corpus refers also a collection of documents but this collection possesses a special structure. Indeed, in the experimental corpus, the entire documents collection is split into several groups. One group can therefore be assimilated to a given patient record. Note that the total number of documents is independent from the number of groups.

The choice of OHSUMED as the source corpus has been motivated by the easiness of manipulation of the documents for the creation of the different datasets required for our experiments. This corpus is homogenous, meaning that its documents share similar properties like length, word distribution, and word frequency. This homogeneity avoids biasing the experiments and allows focusing only on the quality of retrieval. Additionally, using a standardized corpus accessible to other groups fosters comparisons of alternative methodologies.

As the OHSUMED corpus and the patient records corpus share important common properties regarding the IR process, we believe that we will be able to predict the behavior of our system on the patient records by looking at the results obtained on the OHSUMED corpus.

Patient records

The documents used for the study covers all aspects of clinical documentation, such as admission notes, discharge letters, activity reports, nurses' data entries, consultations and interventions reports. Over 1,500 different categories of documents are used.

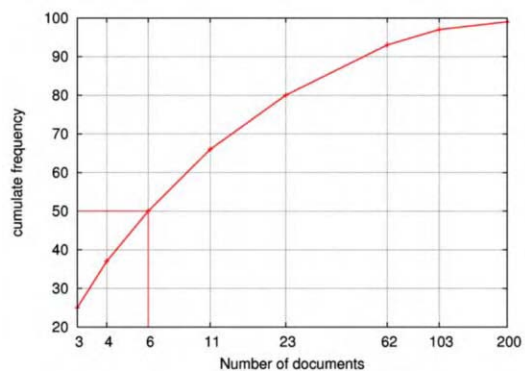


Figure 1 - Distribution of the records size in the patient records corpus

In regard to this distribution, we can stress two issues. First, as most of the records contain few documents (50% of the records have less than 7 documents), there will be little interest in using a tool not returning the exact answer. The second concern is related to records containing numerous documents. In this situation, constraint applied on small record is relaxed as returning a small subset of documents containing the answer could be sufficient. Even if the first answer is not the good one, reducing the possible research space brings a significant gain of time for the care providers.

OHSUMED

The choice of an alternate dataset has been focused on OHSUMED. OHSUMED is a bibliographical document collection, developed by William Hersh and Colleagues at the Oregon Health Science University. It is a subset of the Medline database consisting of 348,566 references from 270 medical journals from the years 1887 to 1991. All of the references have titles, but only 233,445 have abstracts [7].

We refer to the title plus abstract as a document.

Comparison OHSUMED - Patient record

In order to study the similarity of the two corpuses, we selected 8,192 documents from each corpus and compared the frequency of the words of the subsets. Our comparison focused on the frequency distribution of the words contained in the two corpuses as it is the most influential feature on the IR process. Indeed, the observed variation on the inverted index size induced by the modification of the number of patient records is strongly influenced by the distribution of the words in the corpus. While increasing the number of patient records indexed independently, the low frequency words have a different effect on the inverted index size compared to the high frequency words. The former will not be retrieved in independent index whereas the latter will be part of several ones.

We notice in Figure 2 that OHSUMED and the patient record corpora are mainly similar in number of words per document. However, patient record corpus contains less unique words than the OSUHUMED. It is certainly due to

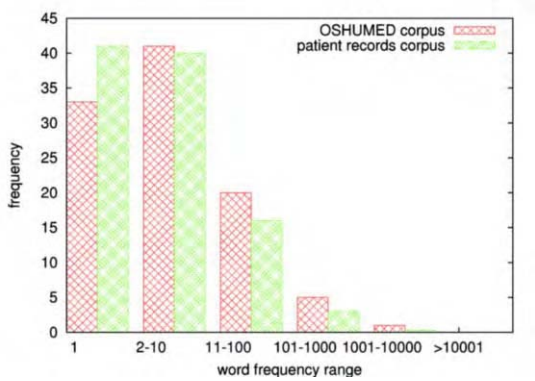


Figure 2 - Frequency of the words contained in the corpus for 8,192 documents

the fact that an important part of the OHSUMED corpus is related to genetics. As there is little consensus on naming genes and proteins, different spelling and acronyms are used and artificially increase word diversity. Beside these the patient records cover mainly the same subjects as the ones addressed in OHSUMED.

Test collection construction

The test collection is composed of three parts, the collection of documents, the queries and the relevance assessments.

The test documents selected from the OHSUMED corpus are structured in order to mimic the existing structure of the patient records corpus. We vary the number of groups while keeping a constant number of documents in order to study the impact of the corpus structure on the performance [3]. As the number of documents is fixed, each increase in groups is accompanied by a proportional decrease in the number of documents per group.

Queries and relevance assessments are dependent of the field of the application. Usually, experts are required to construct relevant and interesting queries for specific domains. We avoid this requirement by automatically generating the queries and their related relevance assessments by considering "known-item-search". The goal in a "known-item search" is to retrieve just a unique document that has been used to build the query. It simulates a user seeking for a particular, partially remembered document in the collection by using a part of a document as a query

The care providers behave quite similarly as the way "known-item-search" works. They have a precise idea of what they search and are often interested in only one particular document that contains the relevant information. Given these similarities, the effectiveness obtained using "known-item-search" is well suited to estimate the usefulness of the tool in a real situation.

The queries are generated by selecting random parts of the documents. Building queries using randomness has the advantage of creating more complex queries than the ones built selecting simply a subpart of the text. Queries built through this process reflect quite properly the ones used in real life situation where they are never exactly similar to the content of the text. Moreover using randomness prevents to match too easily the queries in the initial text.

Methods

Indexing and retrieval strategies

The indexing of the documents is performed with EasyIR, an existing tool [9-10]. This tool has shown good performance on indexing and retrieval in the genetic and biomedical field and should be suitable to be applied on medical terminology.

The specificity of the application arises from the necessity of processing the generated data in real time. In our situation, numerous documents are daily added in the patient records and queries can be performed immediately after. Given these constraints various indexation triggering strategies have been suggested.

a) Indexation launched when a query is performed
b) Indexation performed at fixed, but short, intervals
c) Indexation triggered by a notification when a document is saved

Indexation at fixed time has been dismissed, as it would lead to a lot of useless processing without ensuring proper indexation when needed. Indexation at query time could be interesting, as it requires less indexation process. However, as it inducing a delayed answer when queries are performed, it has a severe impact on perceived performance. Consequently, the last strategy, notification-based triggering appears to be the best solution.

Evaluation metrics

The most basic evaluation of a "known-item search" task, consists in looking at the rank at which the target documents are retrieved. This measure called the "mean-reciprocal-rank" is the mean of the reciprocal of the rank at which the known item was found, averaged over all the queries, and using 0 as the reciprocal for queries that did not retrieve the known document.

Another evaluation measure used is the recall at first retrieved document. It consists in computing the recall of the system when the first returned answer of the system is returned. In our application, recall at first retrieved document is a very useful method to evaluate the utility of our tool. When care providers perform searches in patient records containing few documents, they expect to obtain directly the correct document; therefore the tool must have a high recall at first retrieved document to be really useful.

Trial

Indexing and retrieval - efficiency

The objective is to measure the time cost induced by an increasing number of groups given a fix number of documents. The number of groups grows according to an exponential scale with a parallel decrease in the number of documents in each group.

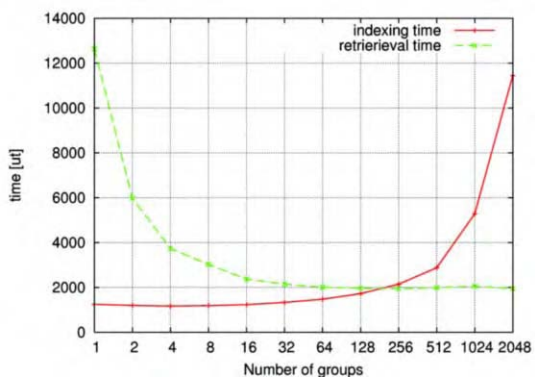


Figure 3 – efficiency given an increasing number of groups for 8,192 documents

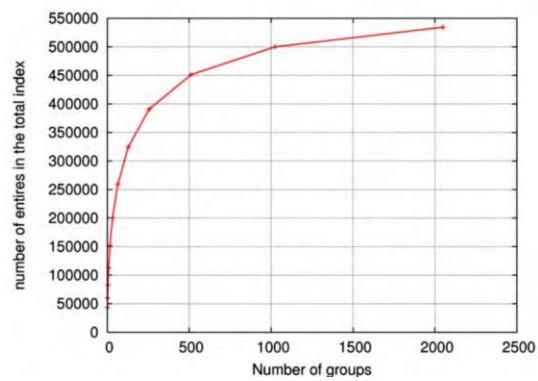


Figure 4 - Total number of entries in the database given the number of groups for 8,192 documents

The measures have been performed using the 8,192 documents with the number of groups varying from 1 to 2,048.

The results show a linear increase in time required for indexation given the number of groups (Figure 3). As the total number of document does not change, this increase in time can be attributed to the indexation initialization overhead.

The time required for retrieval follow a totally inverted tendency in comparison to the indexing process. During retrieval process, as the numbers of queries are constant, it is faster to perform queries on small indexes (Numerous groups). Indeed, if we perform 1,024 queries on an index containing 8,192 documents, it will last longer than performing these 1,024 queries on 1,024 different indexes of 8 documents.

Explanations of the indexing overhead

In this paper, the total index of an experimental corpus refers to the merge of all the individual entries of the index table of each group that compose the corpus. As during the indexing process every group of documents is indexed separately, a separated index table is generated for each group and therefore the total index will contain redundant entries. Indeed, if an entry belongs to two individual indexes it will be present twice in the total index. As the total index contains all the entries which have to be stored on the database, its size is a good estimator of the space required in memory.

As shown in the Figure 4 the number of created entries grows in a logarithmic manner regarding to the number of groups. This is the consequence of the distribution of the frequencies of the words in the corpus that follows a logarithmic decrease (there are many infrequent words and few very frequent words). A word can not occur in a larger number of groups than its document frequency. Therefore, the closer we approach this threshold, the lower the probability of this word occur in an additional separated group and increase the size of the total index. Once a word occurs in a number of groups equal to its frequency, increasing the number of group will not bring any further increase in size of the total index.

In order to analyze the consequences of the size variation of the total entry table on the program overhead, we decomposed the indexation process in three main tasks.

The first step consists in extracting all the words from the documents using tokenization in order to build the vocabulary. This does not lead to significant overhead variation according to the number of groups, as the complexity of this task is only dependent from the total number of words in the whole corpus.

After word tokenization, extracting the term frequency and inverse document frequency values of every term is required to compute the weights. As the inverse document frequency is dependant of the documents contain in every group, this computation is not done once per word but once per entry. Therefore, this process brings clearly an overhead.

The last step consists in storing the indexes in the database. Obviously, there is an overhead linked to the number of groups for this task. As the size of the total index is bigger with a large number of groups, the necessary time to store the total index in the database is bigger too.

Retrieval - effectiveness

Effectiveness evaluation has been done with the mean reciprocal rank and also with recall at first retrieved document.

Table 1 - Recall at top returned documents for different size of experimental corpus given 1,024 queries

Number of document per group	recall at first retrieved document	MRR
8	96%	0.964
16	95%	0.961
32	94%	0.956
64	92%	0.935
128	88%	0.901
256	86%	0.885
512	82%	0.847
1,024	80%	0.819
2,048	74%	0.769
4,096	73%	0.758
8,192	69%	0.722

The Table 1 sho*ws that the fewer documents contained in groups, the better the recall. It is encouraging to see that the tool is effective with small groups. However, results are less impressive with larger groups. Even if the analysis of the data shows that it is extremely rare to have more than 1,000 documents in a patient record this case must be take in consideration.

Mean reciprocal rank results confirm the tendency observed with recall at first returned document measure: the top returned documents of small groups contain most of the time relevant results whereas queries in larger groups must be improved.

Situation with patient records

Given the structure of the patient records corpus one can expect that a significant computational power will be required to offer acceptable efficiency. Indeed, experiments show that when the structure is similar to the one of the patient record the indexing time suffer of a consequent overhead. However given this structure there will be no problem for the retrieval as groups containing few documents allow very quick answer

For effectiveness the 3% of the patient records containing more than 100 documents will be problematic, however this problem should be marginal, at least in the beginning.

Conclusion

Indexing patient records is an unusual information retrieval task as it implies to work with numerous corpuses of relatively small size. Working with such data has revealed weaknesses of systems initially dedicated to manage large corpuses. Highly dedicated tools are needed to answers requirements for real-time and sensitivity. As performing information indexing in patient records is time-consuming, the finest tuning possible should be done in order to increase the efficiency.

The increasing time required when the documents are split into several groups has been a surprising result. However by analyzing the process, we have identify the sources of this overhead and are now able to apply the best solution to solve it.

On the retrieval side, we show that the tool offers satisfying time response and offer good retrieval effectiveness, at least for small patient records.

Experiments done are sufficient to identify weaknesses and strengths of traditional information retrieval tools and define the improvements needed.

Acknowledgements

This work has been funded by the Swiss National Science Foundation (SNF 632-066041).

References

[1] R. Bazera-Yates and B. Ribeiro-Neto. Modern Information retireval. New York: McGraw-Hill, 1983.

[2] D. C. Blair and M. E. Maron. An evaluation of retrieval effectiveness for a full-text document-retrieval system. Communications of the ACM 28 (3): 289-299, 1985.

[3] S. Buttcher and C. Clarke. Efficiency vs. Effectiveness in Terabyte-Scale Information Retrieval. Text REtrieval Conference (TREC) 2005, 2005.

[4] O. Frieder, D. A. Grossman, A. Chowdhury and G. Frieder. Efficiency considerations for scalable information retrieval servers. Journal of Digital information 1 (5), 2000.

[5] D. Grossman and O. Frieder. Information Retrieval: Algorithms and Heuristics: Kluwer Academic Press, 1998.

[6] W. Hersh, C. Buckley, T. Leone and D. Hickam. OHSUMED: an interactive retrieval evaluation and new large test collection for research. ACM SIGIR '94: 192-201, 1994.

[7] M. Martynov and B. Novikov. An Indexing Algorithm for Text Retrieval. Third International Workshop on Advances in Databases and Information Systems - {ADBIS'96} 171-175, 1996.

[8] P. Ruch. Automatic Assignment of Biomedical Categories: Toward a Generic Approach. Bioinformatics 22: 658-664, 2006.

[9] P. Ruch, C. Chichester, G. Cohen, F. Ehrler, P. Fabry, M. J, H. Muller and A. Geissbuhler. Report on the TREC 2004 Experiment: Genomics Track. The Thirteenth Text Retrieval Conference, TREC-2004, 2004.

[10] P. Ruch, A. Jimeno Yepes, F. Ehrler, J. Gobeill and I. Tbahriti. Report on the TREC 2006 Experiment: Genomics Track. TREC Notebook paper, 2006.

Address for correspondence

Frederic Ehrler,
Centre Universitaire d'informatique,
Rue du Général Dufour 24, 1211 Geneva 4, Switzerland.
E-mail: ehrler@cui.unige.ch, TN: +41223797643

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Experiments with Hierarchical Concept-Based Search

Robert Moskovitch[a], Roee Sa'adon[a], Eytan Behiri[b], Susana Martins[c], Aviram Weiss[d], and Yuval Shahar[a]

[a]*Medical Informatics Research Center, Ben Gurion University, Beer Sheva, Israel*
[b]*E&C Medical Intelligence Inc., 41 Madison Ave, New York, NY 10010*
[c]*Medical Corps, Israel Defense Forces, Israel*
[d]*Stanford University, Stanford, VA Palo Alto Health Care Center, Palo Alto, CA, USA*

Abstract

Many digital libraries use hierarchical indexing schema, such as MeSH to enable concept based search in the retrieval phase. However, improving or outperforming the traditional full text search isn't trivial. We present an extensive set of experiments using a hierarchical concept based search retrieval method, applied in addition to several baselines, within the Vaidruya search and retrieval framework. Concept Based Search applied in addition to a low baseline is outperforming significantly, especially when queried on concepts in the third level and using disjunction within the hierarchical trees.

Keywords:

medical text retrieval, concept based search

Introduction

Many digital libraries are indexed using a hierarchical conceptual structure; examples include PUBMED, in which documents are classified along the Medical Subject Headings (MeSH) concepts, and the National Guideline Clearinghouse (NGC[1]) library, each of whose documents is classified using multiple concepts from MeSH and the Unified Medical Language System (UMLS) [1]. Several sites allow browsing through the concepts using the hierarchical structure from the root to the most specific concepts (leaves), which forces the user to navigate the hierarchy. Others enable to query for concepts from MeSH relying on the pre-indexing of the documents along the concepts. Some studies proposed limiting the search to a specific concept (category) [2, 3] and its subconcept contents. In the medical domain, unlike the web, documents are often classified by a multitude of concepts, often as many as a dozen or even tens of concepts, a property which can be further exploited for better retrieval.

The NIH had invested huge amount of money during the past decades in building a set of controlled vocabularies and accessory tools to enable the implementation of concept indexing and retrieval. However, no study had shown that using the conceptual structures outperforms or improves the traditional full text search. One of the famous studies was made by Hersh, in which an attempt to adopt Salton's *tfidf* approach to the conceptual resulted unsuccessfully [4]. Recently we presented *Vaidurya*, a concept based and context sensitive search engine, developed originally, within the Digital Electronic GuidEline Library (DeGeL) [5], to search for textual and marked up clinical practice guidelines, however, we extended it recently to handle general clinical documents. A detailed description of Vaidurya is provided in [6], as well as an extensive and rigorous evaluation, in which a small portion of this study results appear, however, in this study a wider evaluation is provided, in which both concepts based logic operators are evaluated as we will elaborate later. Vaidurya enables the user to query explicitly for concepts given a logic relation between them. In this study we present a novel hierarchical concept based retrieval method including a wide and detailed evaluation of the approach.

We start with a background review of concept based search (CBS) and MeSH. Then we describe briefly the search methods implemented in Vaidurya. We describe our research hypotheses, the experimental plan, and the results. Eventually we discuss the results and conclude.

Background

Concept based search

In the medical domain, CBS refers to a text retrieval approach, in which documents are mapped to concepts, representing a meaningful abstract subject, based on its contents. Hersh's SAPHIRE system [4] uses an approach, in which concepts used for indexing, are automatically extracted from the document. Commonly both documents and queries are mapped, in the case of the biomedical domain into vocabularies such as MeSH and UMLS. However, users are not always familiar with the concepts in these vocabularies, rendering it somewhat limited. It has been previously noted, that the particular implementation of concept-based search described above does not necessarily improve the retrieval performance [7] compared to traditional text retrieval methods. Other studies have tried to exploit the UMLS meta-thesaurus to expand queries, thus extracting the concepts from the query terms [7]. The authors found that query expansion degraded aggregate retrieval performance, but some specific instances of syn-

1 www.ngc.org

onyms and hierarchy-based query expansion improved individual query performance. Aronson [8] compared his methods to Srinivasan's [9] and showed an improvement by expanding text-based queries with both phrases and concepts from the UMLS Metathesaurus. Neither used the hierarchical relationships of the Metathesaurus, yet reported an improvement over a non-query-expanded baseline. Rada [10] developed an algorithm to estimate the conceptual distance between documents and queries using MeSH and suggested that MeSH can be utilized to improve retrieval performance.

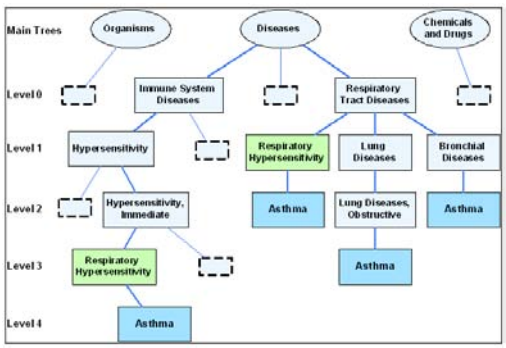


Figure 1 - The concept "Asthma" in MeSH, decomposed to tree-numbers, appearing in several locations in different concept trees

Medical subject headings

The MeSH thesaurus is organized hierarchically and includes 15 concept "trees", consisting of descriptors (concepts). Examples of such trees include *Anatomy* and *Diseases*. At the most general level of the hierarchical structure there are very broad concepts (e.g, 'Anatomy', 'Mental Disorders'). More specific headings are found at lower levels of the eleven levels hierarchy (e.g., 'Ankle' and 'Conduct Disorder'). The same concept may appear as multiple *tree-numbers* within MeSH, thus, in effect, having multiple ancestors. For example, the concept *Asthma* has *four* tree-numbers in MeSH, having three different parents (figure 1). Note that intermediate concepts such as *Respiratory Hypersensitivity* can also appear more than once as different *tree numbers*, possibly at different levels (figure 1).

Vaidurya

Vaidurya is a concept based and context sensitive search engine, developed originally, within the Digital Electronic GuidEline Library (DeGeL) [5], to search for textual and marked up clinical practice guidelines, including three types of implemented search methods: (1) full-text search, using standard key terms; (2) context-sensitive search, which exploits the semantic markup performed on guidelines in *DeGeL*, can be used also for search in a structured document; and (3) concept-based search, consisting on hierarchical concepts indexing structure. The documents can be classified manually or automatically using machine learning based method as we proposed in [11].

Full text search in Vaidurya

Documents are represented and indexed using the *vector space model* introduced by Salton [12], commonly used in free text search, in which each document is represented by a *bag-of-words*. After the document terms are extracted, stop-words (e.g., "and", "the", "are") are removed and the left terms are stemmed to their root using porter algorithm,, a vector of terms is created, such that each index in the vector represents a term frequency in the document, known as ***term-frequency*-inverse-document-frequency** (**TF*IDF**). Term frequency (*tf*) represents the term appearances in the specific document, normalized by the most appeared term (tf^{max}), while the *idf = log(N/n)* represents the appearance of the term in the entire documents collection, where N is the size of the collection, and n is the number of documents in which the term appears. The free text retrieval is based on the *cosine similarity* [12] which measures the distance between a query and a document within the Euclidean space of the terms. More details about the free text retrieval in vaidurya is at [6].

Context sensitive search in Vaidurya

To perform a context-sensitive search, Vaidurya assumes the existence of an internal hierarchical structure of the document (i.e., an ontology that defines an internal contextual model). The internal structure enables the user to query for keywords appearing only in specific contexts (i.e., within segments of text labeled only by the specified tag), thus potentially improving the search and retrieval accuracy. Thus a contextual query includes a set of keywords for each queried context element. In the retrieval process each document gets a rank according to the match for each queried context. This study focuses on the concept based search in Vaidurya, however, more details are provided at [6].

Materials and methods

Concept based search in Vaidurya

In general, our study focused on a very broad class of search methods, which we refer to as *double-operator* methods. Our assumption is that a conceptual hierarchy is always composed of one or more *concept trees*, determined by the roots at level 1 of the hierarchy (level 0 being the root concept). For example, the 15 concepts trees in MeSH.

To perform a concept-based search, Vaidurya includes optional specification of one or more concepts or subconcepts, using the logical operators *conjunction* (AND) and *disjunction* (OR), defining the constraints on the desired relations between the queried concepts, explicitly specified by the user. The first, called *outer-op*, defines the relation among *different* concept trees; the second, called *inner-op*, is defined within the *same* concept tree (Figure 2). Formally, a concept based query $Q^{cb} = \{[t^1<inner^1, c^1_1, c^1_2,.., c^1_{n1}>, t^2<inner^2, c^2_1,.., c^2_{n2}>,..,t^m<inner^m, c^m_1,.. c^m_{nm}>], outer^{op}>$ is denoted by a pair, in which the first element specifies a collection of the queried concepts trees, $T = \{t^1, t^2,.., t^m\}$, in each one of the m concept trees t^1 to t^m. Each queried tree ^{i} is defined by a set of queried concepts

c^i_k, in which *i* is the tree id and *k* is the queried concept id, and a local $inner^i$ operator AND or OR. The second element defines the *outer* logic operator.

During the retrieval process, first, the documents, classified along each queried concept c^i_j and *its descendents*, are retrieved. Then, based on the application of *inner-op*, a set of documents are retrieved for each queried hierarchy, in the case of AND the documents of each concept are intersected and in the case of OR they are unified. Eventually, the application of the *outer-op* logical operator on the documents retrieved for each tree, intersecting in the case of AND, and unifying in the case of OR, results in the final set of documents retrieved for the CBS having the same rank, which is later integrated with additional types of queries, such as full text or context sensitive search, using a weighted average formula.

Four concept based retrieval algorithms were used in this study, defined by the *outer* and *inner* logical operators: (1) *OR-OR*, in which both logic operators set to disjunction. (2), OR-AND, in which the outer is set to disjunction and the inner to conjunction (3) *AND-OR*, in which the outer operator is set to conjunction and the outer operator is set to disjunction, and (4) *AND-AND*, in which both logic operators set to conjunction. In this study the inner operator in all the trees were set to the same value.

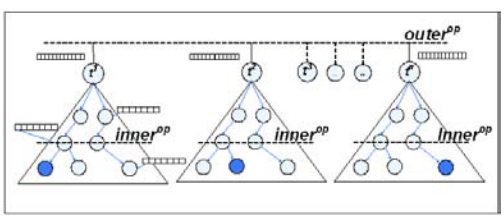


Figure 2 - The inner-op is located within the queried concepts tree and the outer-op defines the relation between the trees, in the final retrieval stage

Evaluation

Test Collection

For our detailed evaluation of the relative value of concept-based and context-sensitive search required a test collection in which documents are structured and classified along hierarchical concepts, we have therefore used the NGC CPGs collection. The NGC website is a repository of CPGs classified by MeSH concepts. The CPGs are classified along two tree-like hierarchical concepts, Disorders and Therapies. Each concept tree has roughly 2,000 unique concepts in a tree-like structure; overall, 5407 concepts were used at the time of the evaluation. In several regions, the concept trees were 10 levels deep, but the average depth was around 4-6 levels. There were 1136 CPGs; each CPG might have multiple classifications indexed by both of the concept trees, including indices that belong to the same tree but at different levels (nodes), not necessarily leaves. CPGs have on average 10 classifications per guideline.

In order to use the NGC collection for an information-retrieval evaluation, we created a set of information needs, queries, and corresponding judgments. Several physicians at the Palo Alto Veterans Administration Health Care System, E&C-Medical Intelligence Ltd Company and a physician from the Medical Crops of the Israeli Defense Forces defined 13 information needs (e.g., treatment of hypothyroidism in a particular population subset). Altogether, *six* physicians participated in the creation of the information needs, queries, and judgments. Each time, a subset of the physicians defined an information need; that group then agreed on the final information need. Eventually, these physicians scanned (*manually*) all of the guideline collection and identified CPGs that were relevant to the information needs (i.e., judgments). In order to evaluate the concept-based search and context-sensitive search methods, in addition to the full-text search, each information need was queried through a combination of a full-text query (FTQ), i.e., a list of keywords searched within the whole document, a context-sensitive query (CSQ), i.e., a query that searches for terms within three predefined context elements (knowledge roles that exist in the NGC ontology), and a concept-based query (CBQ), i.e., a list of concepts, all at the k^{th} level of the concept tree. Thus, each query consisted of three components, each in a different format, each of which could be used, on its own or in combination with one or more other components, to query the guideline database. The typical FTQ consisted of two or three terms after stop-word removal.

We selected three elements from the NGC document structure: "Target Population," "Intervention and Practices considered," and "Diseases and condition(s)" for the CSQ and full-text query. These elements are particularly meaningful when searching for a guideline to answer a clinical question that applies to one's patient, and thus were suggested by the participating clinicians. For the CBQ two types of queries were formulated: (1) concepts from the 2nd level of the conceptual hierarchy, and (2) concepts from the 3rd level of the conceptual hierarchy. In this study we queried the CBQ using the four combinations of the *outer* and *inner* logic operators.

Evaluation measures

In order to evaluate the retrieval performance, we used the traditional precision and recall metrics. *Precision* is the proportion of relevant documents (defined for a specific query within the entire collection, also called judgments) within the set of the retrieved documents, and *recall* is the proportion of relevant-retrieved documents (judgments) retrieved from the set of relevant documents (total judgments), for a specific query. We also interpolated the averaged precision at eleven-points of recall levels 0.0, 0.1,..,1.0.

Experimental plan

In this study we wanted to examine the contribution of the CBQ to the FTQ and CSQ. We also wanted to examine the best level of querying, as well as estimating the best logic operators settings.

Hypothesis I – *adding concept-based search to full-text or context-sensitive search will improve the respective baseline performance.*

Hypothesis II – *Querying concepts from the third level will outperform querying at the second level*.

Hypothesis III - *There are significant differences among the results of the four types of searches*.

To examine these hypotheses three main experiments were designed, in which we used CBQ in addition to a given baseline: (1) *FTQ*, (2) *single CSQ*, and (3) *three CSQs*. In each experiment we evaluated eight combinations of the CBQ resulting from three variables: (1) the queried level of the hierarchy second or third, (2) the *outer* logic operator having *AND* or *OR*, and (3) the *inner* logic operator having *AND* or *OR*, resulting in the four options: *OR-OR*, *OR-AND, AND-OR and AND-AND*. Table 1 presents all the eight combinations including an acronym which we will use in the report of the results.

Table 1 - The eight concept based queries combinations

Level	Outer	inner	Acronym
2	OR	OR	2OO
2	OR	AND	2OA
2	AND	OR	2AO
2	AND	AND	2AA
3	OR	OR	3OO
3	OR	AND	3OA
3	AND	OR	3AO
3	AND	AND	3AA

Experiments and results

We present the results of three experiments, in which the eight CBQs were applied in addition to a given baseline. As a result of the limited length of the paper we present only the four best CBQs in the figures, while report the order of the others performance in the text. We sorted the CBQs according to their average precision at 0, 0.5 and 1 recall level.

Experiment 1 - CBQ in addition to FTQ

In experiment 1 we evaluated each of the eight combinations, in addition to full-text queries. Figure 3 presents the four outperforming CBQs, in addition to the FTQ baseline. Generally at most of the recall levels all the four CBQs, including 3OO, which was significantly greater at the 0.05 significance level when compared at 0.5 recall level, 3AO, 3OA and 2OO in decreasing order, outperformed the FTQ baseline, while beyond the 0.6 recall level the FTQ outperformed. In addition the additional four CBQs (not appearing in figure 3) were 2OA, 2AO, 3AA and 2AA. The four outperforming CBQs can be characterized by querying the 3rd level of the hierarchy, which outperformed the 2nd level. Three of them have OR set to the *outer*, and to the *inner* logic operators. Note that the 3OO outperformed at most of the recall levels while the 3AO outperformed at the low recall levels.

Experiment II - CBQ in addition to single CSQ

In experiment 2 we evaluated each of the eight combinations, in addition to a single context query. Figure 4 presents the four outperforming CBQs, in addition to the single CSQ baseline. Generally, at most of the recall levels all the four CBQs, including 3OA, 3OO, 3AO and 2OO, in decreasing order, outperformed the single CSQ baseline (1CSQ), while beyond the level of 0.3 recall the 2OO was bellow the baseline. In addition the additional four CBQs (not appearing in figure 4) were 2OA, 2AO, 3AA and 2AA. The four outperforming CBQs can be characterized by querying the 3rd level of the hierarchy, which outperformed the 2nd level. Three of them have OR set to the *outer*, and to the *inner* logic operators. Note that while all the four outperforming CBQs were the same, as in experiment 1, but in a slight different order, the best CBQ in the experiment two was 3OA.

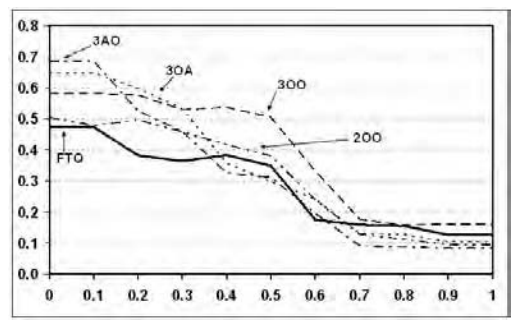


Figure 3 - CBQs applied in addition to the full text queries baseline, in which 3OO, 3AO, 3OA and 2OO outperform the baseline

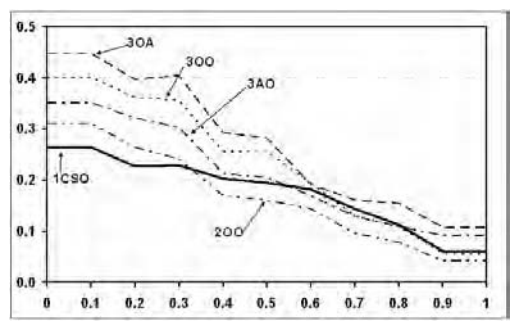


Figure 4 - CBQs applied in addition to single context queries baseline, in which 3OA, 3OO, 3AO and 2OO outperform the baseline at most of the recall level

Experiment III - CBQ in addition to three CSQs

In experiment 3 we evaluated each of the eight combinations, in addition to three context queries. Figure 5 presents the four outperforming CBQs, in addition to three CSQs baseline. Generally, along all the recall levels all the four CBQs, including: 2OO, 3OA, 2OA and 3OO, in decreasing order, performed lower than the three CSQs baseline (3CSQ). In addition the additional four CBQs (not

appearing in figure 5) were 3AO, 3AA, 2AO and 2AA. Unlike the previous experiments, in which CBQs improved the baseline, here it decreased. We will refer to this in the discussion and conclusion section; however, again three of the four outperforming CBQs appear in the first and second experiments. In this experiment while there were two CBQs querying at the 2nd and 3rd hierarchy level, as well as the inner operator, the outer operator is OR at all the four top CBQs. Note that while 2OO was the fourth at both previous experiments, it is the first here (within the CBQs).

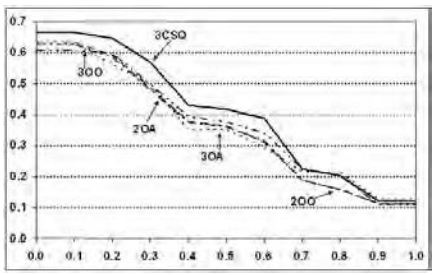


Figure 5 - CBQs applied in addition to three context queries baseline, in which the CBQs (2OO, 3OA, 2OA and 3OO) constantly decreased the baseline (3CSQ)

Discussion

We presented Vaidurya, focusing on its concept based search methods, the hypotheses of the research, the corresponding experiments, in which CBQs in eight settings were applied to varying baselines. Testing hypothesis 1, in which we expected to have an improvement when applying CBQs in addition to varying baselines, shown a significant improvement when applied in addition to FTQ. An improvement was found in addition to a single CSQ, while decreased when used in addition to three CSQs. The reason for the difference between these three scenarios might be that the initial baseline precision in the last case (i.e., when using the three CSQs) was much higher than that when using FTQ or a single CSQ (see figures 3, 4 and 5). In addition, note that the improvement was *greater* when the baseline was *lower* (see figures 3 and 4). Referring to hypothesis 2, querying at the third level outperformed the second level, especially in experiments 1 and 2, while in the third it was even. Referring to hypothesis 3, three CBQs 3OA, 3OO and 2OO appeared within the four outperforming CBQs repeatedly within the three experiments. While 2OO was the last (within the four outperforming CBQs) in experiments 1 and 2, it was the first in experiment 3, which may be explained as well by the high level of the baseline performance.

To summarize, an improvement by applying CBQ in addition to a textual search can be achieved, especially when querying at the third level, setting the *outer* logic operator to OR or AND, and setting OR to the *inner* logic operator. Note that using AND-AND achieved the lowest performance.

Previous studies examining CBS have shown that it does not necessarily improve, and might even worsen, the search-engine's performance [4]. This phenomenon might be due to the fact that current automated extraction modules are not yet sufficiently accurate, and users usually are not familiar with all the concepts when entering keywords. In the case of our CBS the user can manually specify queried concepts from a given predefined ontology of concepts. The limitations of this study are mainly caused by the size of the test collection, which is relatively small. However, in contrast to huge test collections, in which judgments are specified automatically, based on an ensemble of search engines, in our test collection the reviewers browsed manually each of the CPGs and indicated their relevance to the query. We are currently in the process of extending this method to enable a user enter *simply* a textual query which will be converted to a conceptual representation and will be queried in addition to the FTQ. Preliminary evaluation results on the Trec-Genomics are encouraging.

References

[1] Humphreys BL and Lindberg DA. The UMLS project: making the conceptual connection between users and the information they need. Bulletin Medical Library Association, 81(2): p. 170-177, 1993.

[2] Concard J, Yang C, and Claussen J. Effective collection metasearch in a hierarchical environment: Global vs. localized retrieval performance. Proc of SIGIR, 2002.

[3] Wang W, Meng W, and Yu C. Concept hierarchy based text database categorization in a metasearch engine environment. Proc of WISE '00. 2000.

[4] Hersh WR, Hickam DH, Haynes RB, and McKibbon KA. A performance and failure analysis of SAPHIRE with a MEDLINE test collection, JAMIA, Vol 1: p. 51-60, 1994.

[5] Shahar Y, Young O, Shalom E, Galperin M, Mayaffit M, Moskovitch R and Hessing A. A framework for a distributed, hybrid, multiple-ontology clinical-guideline library and automated guideline-support tools. Journal of Biomedical Informatics, 37: p. 325-344, 2004.

[6] Robert Moskovitch, Martins SB, Behiri E, Weiss A, and Shahar Y. A Comparative Evaluation of Full-text, Concept-Based, and Context-Sensitive Search. JAMIA, 14: p. 164-174, 2007.

[7] Hersh WR, Price S., and Donohoe L. Assessing thesaurus-based query expansion using the UMLS Metathesaurus. In Proceedings of AMIA, 2000.

[8] Aronson AR and Rindflesch TC. Query expansion using the UMLS Metathesaurus. In Proceedings of AMIA Annual Fall Symposium, p. 485-489, 1997.

[9] Srinivasan P. Retrieval feedback in MEDLINE. JAMIA, 3(2): p. 157-167, 1996.

[10] Rada R, Bicknell E. Ranking documents with a thesaurus. Journal of the American Society for Information Science, 40: p. 304-310, 1989.

[11] Moskovitch R, Cohen-Kashi S, Dror U, Levy I, Maimon A, and Shahar Y. Multiple hierarchical classification of free-text clinical guidelines. Artificial Intelligence in Medicine, 37: p. 177-190, 2006.

[12] Salton G and McGill M. Introduction to Modern Information retrieval, New York: McGraw-Hill, 1983.

Address for correspondence

robertmo@bgumail.bgu.ac.il

Chapter 3.

Sharing Data

This page intentionally left blank

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

StructConsult: Structured Real-Time Wet Read Consultation Infrastructure to Support Patient Care

Craig Morioka[a], John David N. Dionisio[b], Alex Bui[a], Suzie El-Saden[a], and Hooshang Kangarloo[a]

[a] *Medical Informatics, University of California, Los Angeles, United States*
[b] *Department of Electrical Engineering and Computer Science, Loyola Marymount University, United States*

Abstract

Our research addresses how to improve physician to physician communication of patient information, and how to prevent lapses of patient care as they are referred to other clinicians within the healthcare system. The wet read consultation is defined as a rapid response to a clinical question posed by a referring physician to a clinical specialist. This research involves the development of an imaging-based wet read consultation system called StructConsult (SC), which facilitates communication between non-imaging specialist (i.e., primary care physician (PCP), emergency room (ER) physician, or referring physician), and an imaging specialist-radiologist. To facilitate data mining and effective recall, SC utilizes a data model based on the Digital Image Communications in Medicine (DICOM) standard for grayscale presentation state and structured reporting. SC requires information from four sources: (a) patient-specific demographics, clinical hypothesis, and reason for exam, (b) sentinel image capture from a DICOM image study, (c) direct capture of radiologist's image operations and annotations, and (d) radiologist's response to the chief compliant, and the reason for examination. SC allows users to add additional functionality to a Picture Archiving System to improve patient care.

Keywords:

summarization of patient data, DICOM presentation state, DICOM structured reporting, wet read consultation, teleradiology

Introduction

Radiology is the one specialty that is inherently suited for the practice of telemedicine. In particular, the most common application of telemedicine in radiology is teleradiology. Teleradiology is the transmission of medical images to a remote location for the interpretation of radiographic images. Emergency departments, off-site outpatient imaging centers, and even rural hospitals are utilizing picture archiving and communication systems (PACS) to send images to radiology departments for immediate interpretations and/or second opinions [1-5]. The primary scenarios for the use of teleradiology involve off-hour and weekend interpretation of imaging studies from emergency departments. Second opinions and overreads are also performed via teleradiology when the junior radiologist is unsure of the diagnosis and requires an expert opinion. A survey by Hunter et al. indicated that 60% of 287 academic radiology departments provide teleradiology services [3]. Kalyanpur et al. published an article on private radiology groups that have stationed board certified radiologist in India, in order to have 24 hour read outs of CT imaging studies [4]. This new emergent business model is known as NightHawk or Global Radiology Practices [5]. This allows the radiology group the ability to increase the number of procedures performed as well as their revenue. Primary care physicians usually treat patients for most maladies without consultation. For more difficult cases, the rate of referral by PCPs can be as high as 28.1% per 1000 patients [6]. The consultation process involves communication between physicians. Unfortunately, the communication process can cause serious problems for the patient. The Applied Strategies for Improving Patient Safety collaborative states that communication errors between clinicians, staff, and patients account for over 70% of all errors made in physician practices [7]. Breakdown in communication between clinicians accounts for 80% of all malpractice lawsuits [8]. The Physician Insurers Association of America (PIAA) report that communication errors between referring physicians and radiologists create the fourth most common compliant lodged against radiologists [9]. The PIAA reviewed 144 communication claims, and found that 10% of the written reports never reached the correct physician or patient. In another 10% of the claims, the radiologist's issuance of a delayed written report affected the patient outcome in 75% of the cases. For 60% of the claims against radiologist, the most egregious mistake involved failure to deliver urgent or significant unexpected findings to the referring physician. In a recent study on generalist-subspecialist communication concerning children with chronic conditions, Stille et al. describe specific improvements in order to enhance communication between physicians [10]. The study advocated timely communication, understanding the reasons for the referral and the nature of the child's condition, or appropriate definition of what role the generalist and specialist should play in the treatment of the child. Physicians described numerous examples where communication had direct effects on patient outcomes. Efforts to

improve communication between pediatric generalists and specialists in the care of children with chronic conditions should emphasize the importance of timely information transfer, in order to avoid worsening the patient's compromised health. The content of messages between physicians is also important, but lack of response when needed is more of a problem. Improving generalist/subspecialist communication has great potential to improve the quality of care. Forest et al. found that the top reason for referral was advice related to diagnosis and/or treatment [11,12]. Non-radiology medical clinicians — primary care, internal medicine, and emergency room physicians — typically do not have the proper training nor experience to diagnose difficult radiographic cases [13]. In research studies evaluating general clinicians versus radiologists, radiologists outperformed general clinicians in accuracy of image diagnosis [14,15]. A radiology consult provides two benefits: 1) the consultation helps the referring physician understand the patient's present condition, and 2) the consultation provides the referring physicians with new medical knowledge. The goal of SC is to provide structured clinical consultations that address the specific clinical hypothesis posed by the referring physician. SC stores both the chief complaint, reason for the imaging exam, and the imaging evidence which substantiates the radiologist's response to the reason for exam.

Materials and methods

SC addresses the communication barrier between referring physician and radiologist by enhancing the clinical workflow. The key components of the workflow that are improved include: the image order process, radiologist's wet-read and response to reason for exam, and summarized structured report for the referring physician. The major software components of SC involve: image order, image capture, image order reconciliation, radiologist's wet-read and response to reason for exam, and summarized structured report for the referring physician. Figure 1 illustrates the component or "wiring" structure of the SC software. The software consists of four end-user applications. Three of the applications are based on Java Swing, distributed via Java Web Start in the *op-client.war* archive, while one application is Web-based, deployed through the *op-ros-web.war* Web archive.

- The *referral order system* is the Web application used for entering image orders and viewing their resulting images and annotations.
- The *imaging study reconciler* binds image orders to acquired imaging studies.
- The *reading/annotation workstation* displays the current worklist and corresponding imaging studies for reading, markup, and reporting. When an imaging study has been read and signed, it is "prepped" by the pre-genearation (*pregen)* component for presentation by the referral order Web application.
- The *administration utility* provides a user interface for configuring and setting up the SC installation, and user access and privilege.

The three server-side components are shown near the center of the diagram:

op-ros-ejb.jar is the central set of services provided by the software, as shown by the interfaces that it implements: Admin, UserAdmin, Info, Login, Order, Reconciliation, Station, Image Server, and Results. Other components rely on one or more of these interfaces to accomplish their respective tasks. The four end-user applications rely *only* on this component for all server-side interactions.

pregen is the *results generation daemon* that prepares an imaging study for presentation on the Web application. When an imaging study is signed, the daemon converts the study's selected key images from DICOM to a Web-compatible format such as JPEG or PNG, at the same time compositing any annotations that were made by the study's reader.

image server is the bridge between DICOM devices and SC. It is the gateway that receives incoming studies from these devices and notifies the other SC components of the studies' availability. Received studies are stored on a file server that is mounted over the network on both the reading/annotation workstation and results generation daemon's file systems.

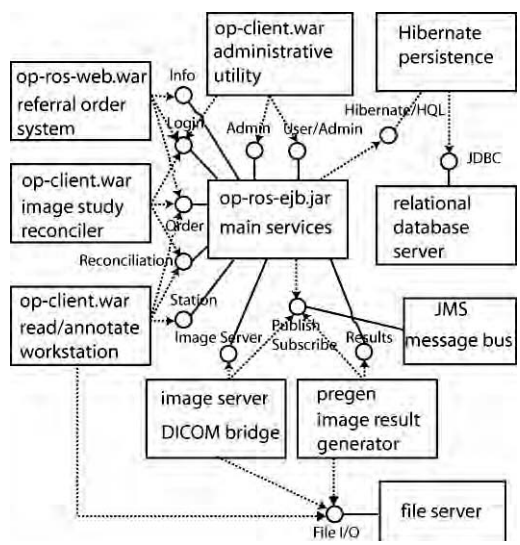


Figure 1 - Component diagram of StructConsult software

The overall design of the SC software follows recommended Enterprise Java Beans patterns, resulting in appropriate separation of classes, excellent portability, and flexibility of configuration and deployment. Every component of the suite has been tested on Windows XP, Mac OS X, and Linux. For optimal security, reliability, and stability, server-side components are typically hosted on a Unix-based operating system such as Linux or Mac OS X. Client machines tend to run end-user-centric operating systems such as Windows XP or Mac OS X. The components can also be deployed completely using established open source software (e.g., Apache, JBoss, PostgreSQL, Samba), thus

reducing the overall cost of an SC installation, especially when compared to existing commercial solutions.

StructConsult Organization

Image order

One of the major problems in medicine is improving physician communication between a referring physician and radiologist. The most prevalent criticism from referring physicians was that the radiology report did not answer the clinical question that they expected based on the patient's chief complaint. Referring physician consultations with subspecialists can be more effective when the referring physician provides a specific question, clinical hypothesis, or specific reasons for exam concerning the patient condition [12]. Once the chief complaint and reason for exam are captured, the structured information is then used by the radiologist at the review station to determine which images in the image study will provide evidence linked to each particular reason for exam. The user interface for the referring physician is intended to capture two important pieces of information: 1) chief complaints, and 2) clinical hypothesis. The chief complaint is entered — *abdominal pain*. The reason for exam is also entered as a list of "Rule-out the following:" *heart attack, ulcer, appendicitis*, etc. Figure 2 shows the referring physician user interface. The referring physician enters the chief complaint (upper right panel) and finally enters the reasons for exam, or clinical hypothesis for the patient (lower right panel).

Image capture

After the image order is complete, the patient is sent to an imaging center to receive their exam. As part of SC, there is a DICOM image server that receives images from the patient scanning devices, or from the PACS server. DICOM image object definitions that can be received by our image server include the following: computed radiography, computed tomography, magnetic resonance, ultrasound, secondary capture, presentation state, and structured report objects.

Image order/image study reconciliation

The image order/image reconciliation process allows the administrator the ability to match up an image order with an image study. The SC infrastructure makes no assumption about the healthcare infrastructure. As healthcare within the United States can be fragmented, it cannot be assumed that the entire healthcare process is within the same healthcare entity. A patient may visit their PCP, or referring physician, who is unaffiliated with the image scanning facility. SC provides a manual method to reconcile a particular image order and image study. This reconciliation process can only begin once the image order is complete, and the patient's image study has been received by SC's image server. A worklist is provided to the SC administrator to identify outstanding imaging orders that need to be reconciled with the image study. The user selects the image order in one column, and then matches the appropriate image study in the other neighboring column. Patient demographic information is provided to help user ascertain the likelihood of a match between the order and the study.

DICOM presentation state

The image viewer developed by our research group reads DICOM compliant radiographic image studies. When reviewing the images, the radiologist can capture the viewing conditions and image operations performed as a DICOM presentation state (PS). The DICOM PS captures the *viewing state* of the radiologist while the image study was dictated. The operations necessary to reproduce the same viewing conditions involve the exact image(s) reviewed, look-up table (LUT) values of the image displayed, scale size, rotation, linear measurements, and annotations. The objective of the PS is to capture the radiologist's viewing conditions at the time the image study was reviewed. Once the user saves the DICOM PS, the DICOM PS object is sent to the image server for storage. The DICOM standard allows DICOM PS objects to be transmitted, stored, and retrieved from a DICOM compliant server.

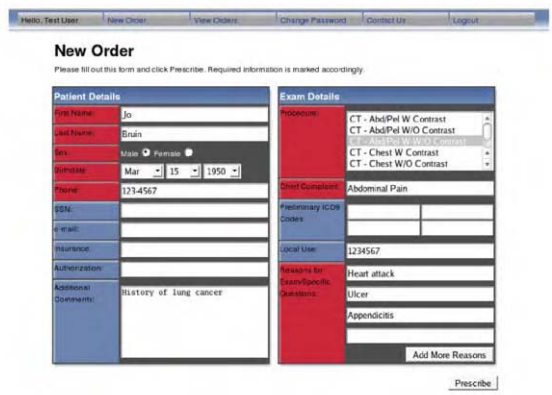


Figure 2 - Referring physician order entry Web page

Radiologist review workstation

The radiologist review workstation allows the user to see patient demographic information, chief complaint, and reason for exam (see Figure 3). The reason for exam is shown in the upper panel in the right corner. The radiologist indicates for each reason for exam the following response: rule-in, rule-out, or unsure. When a reason for exam response is selected, the images that correspond to a particular reason for exam are captured as a presentation state. The response to a reason for exam is also captured as part of the structured report. The structured report can also capture free-text data keyed in by the radiologists as additional comments. Once a particular image study is chosen, the SC image review workstation allows the radiologist to examine the entire image study (see Figure 4). The toolbar along the top of the viewer allows basic image operations: horizontal and vertical flip, window and level, pan, scale, and rotation. The viewer also allows the user to draw line annotations, text, and different types of measurements (point, line, and area). Finally, the image viewer can present studies using different image layouts.

DICOM structured report

The referring physician's clinical hypothesis, DICOM PS of the image study, and the clinical findings that rule-out a clinical hypothesis are stored as a DICOM structured report (SR). The structured report is the underlying framework for capturing the clinical questions and responses for the patient. The primary components of a DICOM SR document consists of meta-information that describes the creator of the document, date of creation, time of creation, institution, and a unique identifier for this DICOM SR object. Other structured information captured by DICOM SR include: patient's chief complaint, referring physician's reason's for exam, radiologist's response to reason for exam, reference to specific DICOM image(s) referred to by DICOM PS, and reference to DICOM PS.

Summarized view of structured text and image

Once the radiologist has structured their response to the chief complaint and reasons for exam by capturing the image(s), text comments, and structured responses, the information is stored in our image server as a DICOM SR object and within the PostgreSQL database. The results view of SC allows clinicians a summarized view of the patient. Patient demographic information, previous clinical documents, responses to the reason for exams, and the key images are stored as a Web-accessible page summarizing the patient's condition. Figure 5 depicts our Web-accessible DICOM SR summarized view. The patient's summary is a Web page that can be accessed anywhere given a user ID and password. The upper left panel in Figure 5 depicts the patient demographics with name, medical record number, gender, and patient's birth date. The worklist to the right of the patient demographics is a list of available documents for the respective patient. All available clinical reports and labs can be selected from this panel. The text panels below the worklist show the most recent oncologist, radiology, and pathology reports. For this lung cancer patient, all responses to the reason for exams were positive and indicated by the red check mark shown on the right side of in the image box. The small thumbnail images along the top of the image panel are the DICOM PS key images captured by the radiologist. The user can select the smaller thumbnail image, which will appear in the larger image viewing panel. This summarized view of the radiologist's responses to the referring physician's clinical hypothesis completes the communication between clinicians.

Results

The SC system has been deployed at a primary care clinic/imaging center. Harris Family Medical Center and University Center Imaging (UCI) in Melbourne, Florida have registered 48 referring physicians. The number of SC image orders from 01-01-2006 to 11-01-2006 was 11,093. This accounts for 32.4% of all images ordered. A majority of the image orders are still paper-based, but as the referring physicians become more comfortable using the Web-based system, a larger number of online image orders is expected. A survey was administered to the referring physicians using the SC system. Five of the referring physicians completed the survey. There were 5 questions asked of the users:

1. What was the availability of the Web-based image report?
2. Compare the Web-based system to the traditional phone order/fax report system.
3. What was the accuracy of the report content of the Web-based report compared to the faxed version?
4. When accessing the Web report from a browser, how long did it take?
5. What was the turnaround time of the Web report compared to the traditional phone order/fax report?

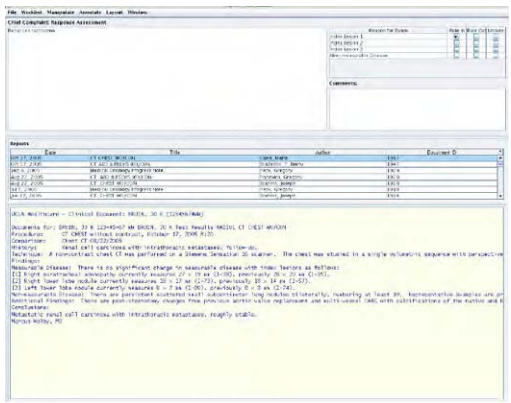


Figure 3 - Radiologist review workstation depicts structured capture of reasons for exam (i.e., index lesion 3 is checked rule-in, upper right corner). Previous radiology reports are also available for review

Each question was based on a 5-point rating scale (1 = superior performance, 2 = better performance, 3 = performance was equal, 4 = poor performance, and 5 = very poor performance). Questions 1 (*avg*=1.2, *std*=0.4, p=0.034, n=5), 2 (*avg*=1.5, *std*=1.0, p=0.046, n=4), and 4 (*avg*=1.2, *std*=0.5, p=0.034, n=5) were statistically significant using a Wilcoxon signed-rank test. Questions 3 (*avg*=1.8, *std*=1.0, p=0.052, n=4), and 5 (*avg*=1.6, *std*=0.9, p=0.052, n=6) were very close to significance.

Discussion

The trend in the replies indicated that the users chose a score of 1, 2, or 3 indicating equal or better performance for the Web-based system for all questions. There were no user responses that indicated that the Web-based system was poor or very poor when compared to the conventional paper order/fax report. The goal of our SC system is to provide structured clinical consultations between referring physicians and radiologists. The SC infrastructure provides improved communication support for the referring physician, particularly for patients with complex medical conditions requiring close monitoring to insure proper quality of care. Utilizing our SC system insures that the communication between clinicians is accurate and justifi-

able. SC stores both the reason for the exam, and the imaging evidence that substantiates the radiologist's response. Our system is not limited to only radiology consults; potentially one could provide accessible services to other subspecialists. Another benefit of the SC infrastructure is the ability to create teaching files, as the radiologist captures key images through DICOM PS. A long-term goal of this research is to generate summarized views of patient data accumulated over long period of care. The SC infrastructure will provide the structured image data necessary to complete this task.

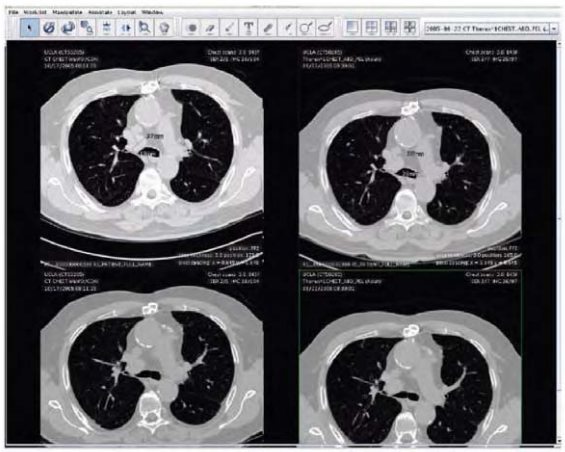


Figure 4 - Image viewer component of radiologist review workstation shows DICOM PS capture of top 2 images. Each image has a 2D tumor measurement

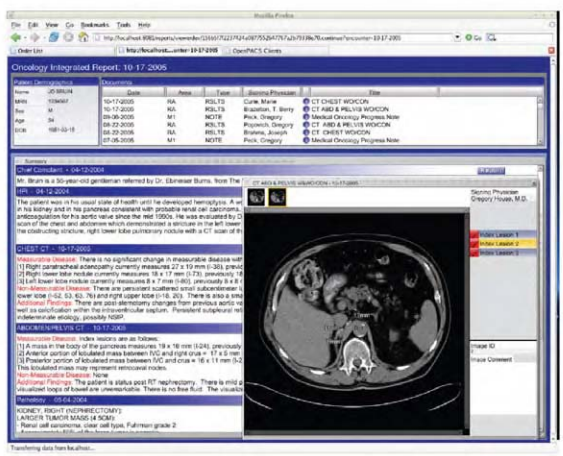


Figure 5 - Web page with summarized state of patient showing DICOM PS (key images) and SR responses (red check boxes)

Acknowledgments

We would like to thank Greg Tashima from Medaxis Corporation for his assistance in gathering the user survey and utilization data.

This research is being supported in part by the National Institutes of Health Grant P01-CA51198-09A1.

References

[1] Franken E, Harken K, and Berbaum K, Teleradiology Consultation for a rural hospital: patterns of use. Acad Radiol 1997;4:492-96.

[2] Lee J, Renner J, Saunders B, Stamford P, Bickford T, Johnston R, Hsaio H, and Phillips M, Effect of real-time teleradiology on the practice of the emergency department physician in a rural setting. Acad Radiol 1998;5:8533-38.

[3] Hunter T, Krupinski E, Hunt R, and Erly W, Emergency Department Coverage of Academic Departments of Radiology, Acad Radiol 2000;7:165-70.

[4] Kalyanpur A, Weinberg J, Neklesa V, Brink J, and Forman H, Emergency radiology coverage: technical and clinical feasibility of an international teleradiology model, Emerg Radiol 2003;10:115-18.

[5] Mun S, Tohme W, Platenberg R. and Choi I, Teleradiology and emerging business models. Journal of Telemedicine and Telecare 2005;11:271-75.

[6] McPhee SJ, Lo B, Saida GY, and Meltzer R, How good is communication between primary care physicians and subspecialty consultants? Arch Intern Med 1984;144:1265-68.

[7] Fernald DH, Pace WD, Harris DM, West DR, Main DS, and Westfal, JM, Event reporting to a primary care patient safety reporting system: a report from the ASIPS collaborative. Ana Fam Med 2004;2:327-32.

[8] Levinson, W, Physician-patient communication. A key to malpractice prevention. JAMA 1994;272:1619-29.

[9] Brenner RJ, Lucey LL, Smith JJ, and Saunders R, Radiology and medical malpractice claims: a report on the practice standards claims survey of the Physician Insurers Association of America and the American College of Radiology. AJR 1998;171:19-22.

[10] Stille CJ, Korobov N, and Primack WA, Generalist-subspecialist communication about children with chronic conditions: an analysis of physician focus groups. Ambul. Pediatr. 2003;3:147-53.

[11] Forrest CB, Glade GB, Starfield B, Baker AE, Kang M, and Reid RJ, Gatekeeping children and adolescents to specialty care. Pediatr. 1999;104:28-34.

[12] Franken Jr. EA, Gergus GR, Koch TJ, Berbaum KS, and Smith WL, Added value of radiologist consultation to family practitioners in the outpatient setting. Radiology 1995;197:759-62.

[13] Schriger DL, Kalafut M, Starkman S, Krueger M, and Saver JL, Cranial computed tomography interpretation in acute stroke. JAMA 1998;279:1293-97.

[14] Kouri BE, Parsons RG, and Alpert HR, Physician self-referral for diagnostic imaging: review of the empiric literature. AJR 2002;179:843-50.

[15] Potchen EJ, Cooper TG, Sierra AE, Aben GR, Potchen MJ, Potter MG, and Siebert, J.E., Measuring performance in chest radiography. Radiology 2000;217:456-59.

Address for correspondence:

Craig A. Morioka, Ph.D.
Assistant Professor
UCLA Medical Informatics,
Department of Radiological Sciences,
924 Westwood Blvd. Suite 420, Los Angeles, CA 90024

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Refining DICOM for Pathology – Progress from the IHE and DICOM Pathology working groups

Christel Le Bozec[a,b,c], Dominique Henin[a], Bettina Fabiani[a], Thomas Schrader [d], Marcial Garcia-Rojo [e] , Bruce Beckwith [f]

[a]ADICAP, France; bINSERM, UMR_S 872, eq.20 Paris, F-75006 France; Univ Paris Descartes, Paris, F-75006 France; [c]APHP, Hôpital Georges Pompidou, Paris, F-75015 France; [d]Department of Pathology, Charite, Berlin, Germany, [e]Pathology Department at the Hospital General de Ciudad Real Spain; [f]Department of Pathology, Harvard Medical School and Beth Israel Deaconess Medical Center, Boston, MA, 02215, USA

Abstract

For making medical decisions, healthcare professionals require that all necessary information is both correct and easily available. We address the issue of integrating anatomical pathology department information into the electronic healthcare enterprise. The pathology workflow from order to report, including specimen processing and image acquisition was modelled. An integration profile - pathology general workflow- was created in the framework of the Integrating the Healthcare Enterprise (IHE). This Integration Profile relies on 8 transactions based on HL7 or DICOM standards. An important issue was to define information entities (order, imaging study and report) and real-world objects (specimen, tissue sample, slide, etc). Joint efforts between IHE and DICOM WG26 has resulted in a proposed common model for "specimen" usable for both HL7 and DICOM transactions related to anatomic pathology.

Keywords:

pathology, standards, imaging, integration

Introduction

Information systems in anatomical pathology departments gather medical data (text, images, etc.) throughout various procedure steps from specimen processing to report editing. Since information systems are not typically integrated, information acquisition is time consuming with double data entry. Orders, images and reports are spread out over different systems which do not interoperate. Although standardization efforts conducted by HL7 [1] and DICOM [2] are progressing to provide integration solutions, HL7 or DICOM messages contain many optional data fields so that being DICOM or HL7 compliant does not imply direct integration. The goal of the Integrating the Healthcare Enterprise (IHE) initiative is precisely to specify how data standards should be implemented to meet specific healthcare needs and to make systems integration more efficient and less expensive [3]. Based on working groups including users and manufacturers, IHE defines "Integration Profiles", that are real-word situations describing exchange of information called "Transactions", from various functional components of a distributed healthcare environment, called "Actors". IHE provides implementation guides for "Transactions", using established standards as DICOM or HL7. IHE has developed in North America, Europe and Asia. The annual definition cycle of new profiles by users and suppliers, ending in the organization of international platforms of interoperability tests (called "connectathons"), confers its unique efficiency, transforming basic standards into "plug and play" solutions.

In some cases, IHE recommends selection of specific options supported by these standards; however, IHE does not introduce technical choices that contradict conformance to these standards. If errors in or extensions to existing standards are identified, IHE's policy is to report them to the appropriate standards bodies for resolution within their conformance and standards evolution strategy.

In Europe, in 1996, the Association for the Development of Informatics in Cytology and Pathology (ADICAP) with the collaboration of different software vendors proposed a European de facto standard for image folders [4].

In the US, Laboratory Digital Imaging Project (LDIP) began in 2005 with a goal to establish an open access, voluntary use specification that will permit images generated in pathology and in clinical laboratories to be widely shared across different applications, databases, and operating systems for the purpose of enhancing image annotation, integration, archiving, publication, and analysis. This effort is intended to be complementary to DICOM, rather than as an alternative to it.

Relying on this experience, the aim of ADICAP is now to promote the use of international standards (DICOM, HL7) in the development of information systems in anatomical pathology. In 2005, ADICAP and SEAP (Spanish association of Anatomic Pathology), with the collaboration of the Group promoting the Modernization of Hospital Information Systems in France (GMSIH), launched the IHE pathology initiative in Europe.

Although specific DICOM objects are defined for pathology, modification and/or extension are necessary for two main reasons. First, the current DICOM specimen module is not sufficiently detailed to capture the complexity of pathology practice, and second, some pathology-related image formats (whole-slide images, multispectral images, flow cytometry, etc) do not have applicable DICOM Information Object Definitions. A specific DICOM working group (DICOM WG26) has been recently created to address these issues. A specific working group (HL7 Pathology Special Interest Group) has been also recently created within HL7 to address the pathologists' needs, in synergy with DICOM WG26, focusing on the orders and reports aspects of the pathology workflow.

The objective of this paper is to present a methodology to integrate anatomical pathology department into the healthcare enterprise. We first describe the IHE-Pathology efforts to model the anatomical pathology workflow in order to define new Integration Profiles. We also describe IHE-Pathology and DICOM 26 joint efforts to define a common model of "specimen" usable for both HL7 and DICOM transactions. DICOM Information Object Definitions for whole slide images is out of the scope of this paper.

Methods

Modeling anatomical pathology workflow

ADICAP, with the collaboration of the GMSIH, solicited participants to work on the Pathology Technical Framework: 12 pathologists and haematologists, 6 professional associations and 12 vendors. 7 working sessions were organized between September 2005 and January 2006. The working group first defined the pathologists' needs and then created a corresponding Integration Profile for anatomic pathology. They identified the Actors and Transactions involved in this profile. Then, they reviewed the literature about order forms and reports in anatomic pathology in order to describe the main requirements for the structure and content of orders, imaging folders and reports.

Defining HL7 and DICOM based transactions

DICOM currently has Information Object Definitions dedicated to anatomical pathology, namely VL Photographic Image (XC) for gross imaging and VL Slide-Coordinates Microscopic Image (SM) for microscopic imaging. Following creation of the DICOM Pathology Working Group (WG-26), five IHE-Pathology-DICOM working sessions were organized between September 2005 and November 2006 to define modification/extension needed for the DICOM specimen module.

Results

General pathology workflow (pathology technical framework - volume 1)

The diagnostic process in anatomical pathology (figure 1) differs from that in the clinical laboratory since it relies on image interpretation. It also differs from that in radiology since it is specimen-driven and when digital imaging is performed many types of imaging equipments (gross imaging, microscopic still imaging, whole slide imaging, multispectral imaging, etc.) may be involved for a single examination. Moreover, images of the same study may be related to different specimen (parts and/or slides) from one or even different patients (e.g., Tissue Micro Array). Finally, slides are always available to acquire more images, if needed. In radiology, the diagnostic process is patient-driven, an examination (study) usually involves a single image acquisition modality and all images of the study are related to one and only one patient.

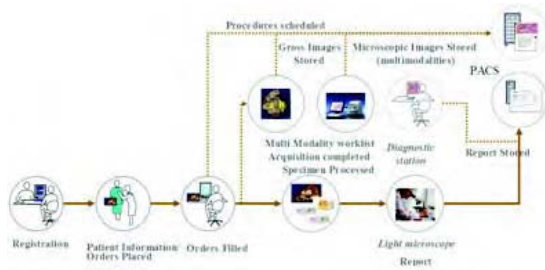


Figure 1 - Anatomic pathology workflow

The Integration Profile "Pathology General Workflow" was proposed for a first IHE cycle: (Specimen, Order and Report Management Workflow) and Pathology Image Workflow. This Integration Profile involves 8 Transactions exchanged by 8 Actors: 3 transactions dedicated to Order Management (Order Placer Management, Order Filler Management, Procedure Scheduled and Update), 4 transactions dedicate to Image Management and one procedure for Report Management. For each transaction, the workgroup proposes the use of the most suitable format (HL7 version 2, - HL7 version 3 (Clinical Document Architecture (CDA)) and DICOM).

General principles were defined with respect to information entities: Order, Image Study and Report. The order for the pathological examination is communicated between the Order Placer (of the Order Entry system) and the Order Filler (of the PIS). In the pathology department environment, the Order Filler also identifies the set of procedures and sub-procedures (procedure steps) that have to be performed in the process of fulfilling the order.

Order

Quality assurance about order forms provides a list of mandatory items [5]: order identification (Order ID), order date & time, identification of the ordering physician and the ordering care department (including call back telephone number), patient identification (PID, name, visit number, etc.), identification of the care unit of the patient (if different from the ordering care department), priority of the order, (date & time when the results are expected to be available), etc. Each order may contain one or more Requested Procedure possibly reported by different pathologists. A Requested Procedure is a unit of work resulting in one report with associated codified, billable acts. For each Requested Procedure, the basic or special

techniques involved in the processing of the corresponding specimen(s) may require different devices (automatons, image acquisition modality, etc). Each Requested Procedure may contain one or more Procedure Steps. A Procedure Step is the smallest unit of work in the workflow that is scheduled (work to do) and/or performed (work done) by a person or a machine (automaton, image acquisition modality, etc) on an object (specimen, tissue sample, tissue section, etc.) Figure 2 depicts an example of cases resulting from an order and of the breakdown in Requested Procedures and in Procedure Steps.

Report

Since 1993, Association of Directors of Anatomic and Surgical Pathology has published recommendations for pathologic reporting [6]. A generic model of structured report can be derived from these templates. In complement, studies about quality assessment of reports provide lists of mandatory items and stress the positive role of checklists to enhance the reporting process [7,8]. According to "evidence-based pathology", only features that are reproducible and relevant – with a demonstrated diagnostic or prognostic signification – should be reported in description and corresponding evidence available"[9,10]. A crucial issue is to identify a technical solution to handle templates of structured reports including findings and their evidences.

Image

In pathology, the image folder (study) is defined at the level of the pathological examination or case. For each case, images acquisition may require different modalities (gross imaging, microscopic imaging, etc). A new series is created whenever an imaging procedure step is performed on a new specimen or slide or when a new type of image is created of the same specimen or slide.

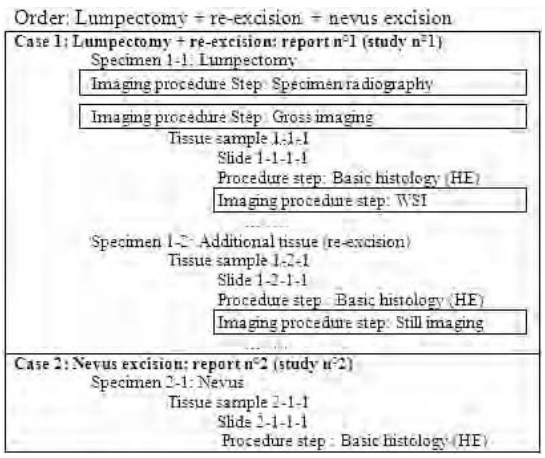


Figure 2 - Extract of the structure and content of an order: the breakdown in cases (corresponding to reports) and procedure steps and the organization of the corresponding images in studies and series

Actors and transactions

Figure 3 - General pathology workflow

Six specific transactions exchanged between 7 actors are required to perform the General Pathology Workflow and are schematically represented in Figure 3. Placer Order Management (PAT1) contains all the messages required between the Order Placer and the Order Filler for the management of the life cycle of the order. Its main goal is to keep a consistent vision of the order, (content and status), between the two actors. Filler Order Management (PAT2) contains all the messages required between the Order Filler and the Order Placer for the notification of a new filler order, as well as the creation of the placer order that reflects it.

Modality Worklist Provided (PAT3) is based on a query entered at the Acquisition Modality. In case of a general query, the list of Scheduled Imaging Procedures with selected demographic information and information about specimens is returned to the Acquisition Modality. In case of a query using the barcode identifying a given specimen being source of an imaging procedure (either part or slide), the specific information about this specimen is returned.

The Procedure Scheduled and Update (PAT4) allows the Order Filler to send the Image Manager and Report Manager information about the scheduled procedure or procedure update.

Report Management (PAT7) carries changes of the observation results and order status from Order Filler to the Enterprise Report Repository i.e. corrections, cancellations.

Using the Image Availability Query (PAT9), the Order Filler or the Report Manager asks the Image Manager if a particular image or image series is available. The worklist provider informs other interested actors of the on-going status and completion of performed work.

The transaction Modality Image Stored is based on two transactions already defined in the radiology domain (RAD8, RAD10) and allows an Acquisition Modality or an Evidence Creator sending acquired or generated images to the Image Archive. Acquisition Modality or Evidence Creator requests that the Image Manager confirm ownership for the specified DICOM objects (images, Key Image Notes, Evidence Documents or any combination thereof)

stored in the Image Archive, thus allowing the sender to delete those objects.

HL7 and DICOM based transaction (pathology technical framework – volume 2)

Prior to the specification of the transaction a common model of "real world objects" (specimens, containers (blocks, cryomolds or cryotubes, slides, etc) was defined.

Specimen identification mechanism

The specimen identification mechanism must allow keeping track of all these "real world objects" (figure 4).

"Specimen" is the role played by any discrete physical object that is the subject of pathologic examination. This includes objects at all levels of processing, including fresh tissue, dissected organs, tissue embedded in paraffin, and sections made from embedded tissue. This extends the common definition of a specimen beyond the object itself received for examination (e.g., from surgery).

"Container" is a physical object which can be labeled (eg with a pathology department accession number, or a block label or a slide label, etc) and contains a tissue sample which may be analyzed or further processed.

"Box" refers to a container which contains Part(s). "Part" refers to a separately identified physical object (tissue) collected in care department upon which a pathologic or laboratory procedure is requested to be performed.

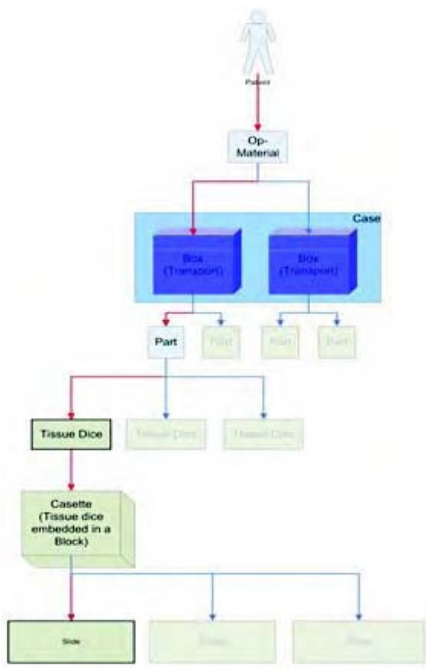


Figure 4 - Typical : Specimen can be identified by containers' ID

"Cassette" refers to a "container" which contains Tissue dice(s) that are embedded in Block (one cassette corresponds to one block that corresponds to one or more tissue dice(s)). "Tissue dice(s)" within a block are tissue samples resulting from sampling processes within the pathology department obtained from one or more parts. They have been processed and embedded in a medium to allow the creating of tissue sections which are mounted on slides for imaging. Tissue dice(s) may be identified (Tissue dice ID, optional). The corresponding Cassette is mandatory identified (Cassette ID)

"Slide" refers to a container which holds a tissue section, a smear, a touch prep, etc. The slide is typically glass for visual light microscopy, but is usually a grid for electron microscopy. "Tissue sections" are created from Tissue Dice(s) embedded in blocks. "Touch preps" are prepared by placing a slide into contact with unprocessed tissue. "Smears" are created from taking a liquid containing cells (peripheral blood or other bodily fluids such as ascites) and spreading this liquid into a thin layer.

Tissue item(s) within a tissue section or a smear may be identified (Tissue item ID, optional). The corresponding Slide is required to be identified (Slide ID)

"Tissue Microarray" (TMA) is a composite specimen which is typically created by taking a small core of tissue from many different tissue blocks and re-embedding them in a new block in an organized manner. Slides created from this TMA block thus have small fragments of many different tissues all of which may be processed at the same time, under the same conditions by a desired technique. These are typically utilized in research (figure 5).

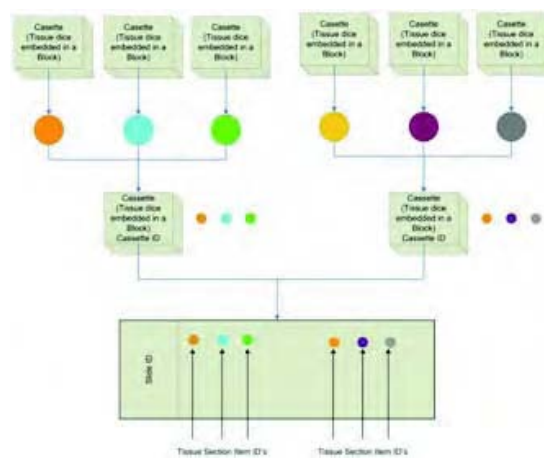


Figure 5 - TMA : Tissue items ("spots") on slide come from tissue dice(s) ("cores") sampled from different tissue blocks (from different parts and from different patients)

Specimen description

Specimen(s) and container(s) description requires a controlled vocabulary. A first step consisted in aligning DICOM Specimen Module and HL7 v2.5 Specimen (SPM) and Container (SAC) segments. Eleven DICOM tags of the Specimen Module correspond to items of the HL7 v2.5 SPM segment: Specimen ID; Specimen Type, and Type Modifier; Specimen Collection Method, Site and Date/Time; Specimen description; Specimen Handling

Code; Specimen Received Date/Time; Specimen Condition; Container Type and Condition and Parent Specimen ID. Four DICOM tags of the Specimen Module could be expressed using items of the HL7 v2.5 SAC segment: Container length, Container width, Container thickness and Container length, width, thickness units.

We haven't found any existing item in HL7 v2.5 to express specimen state identifier and specimen state description. We have no more found a satisfactory solution to solve the specimen identification issue in case of TMA.

Discussion

Quality assessment studies in anatomical pathology show that each of the different steps from specimen processing to report editing may be source of errors and that information systems integration supports error reduction [11-13]. This work is done in the framework of the IHE-pathology initiative to define the requirements of systems integration in anatomical pathology.

The results show that a first significant IHE cycle in anatomic pathology could involve the new integration profile "General Pathology Workflow". The main contributions of this work were to analyze the specificity of the anatomical pathology workflow with respect to laboratory and radiology workflows and to define the structure and content of cases, orders, image folders and reports. There was an issue to make explicit the links between information entities (orders, image folder, reports, etc.) and real-world objects (specimens and containers, etc).

The anatomic pathology specimen model that we have described was developed to be consistent with the Specimen Information Module included in HL7 version 3, which is based on the HL7 Reference Information Model (RIM). The HL7 model was designed for clinical laboratory specimens, but is sufficiently generic to be useful for anatomic pathology specimens. We consciously attempted to maintain consistency with the HL7 v3 Specimen Common Message Element Type (CMET), and in our definition of a specimen, we consciously adopted the terminology of the focal class specimen as a "Role" with an identifier, in accordance with the v3 model.

Although the main output of the anatomical pathology workflow is a timely and clear report of a diagnostic opinion, images will be more and more associated as evidence to textual reports. DICOM seems to be a convenient format for image archiving and communication within the anatomical pathology department. For integration into the Electronic Healthcare Record, HL7 Clinical Document Architecture (CDA) seems to be the suitable format but solutions available to link textual items to DICOM images in an HL7 CDA document must be clarified.

Our perspective is that vendors implement IHE-pathology principles. Thanks to an on going collaborative work involving IHE-Pathology, DICOM WG26 and HL7 pathology SIG, an implementation guide (Pathology Technical Framework) will be available for a first IHE cycle in 2008.

Acknowledgments

This work has been supported by the ADICAP and the College of American Pathologists.

References.

[1] HL7 organization : Health Level 7, http://www.hl7.org
[2] DICOM 3 Standard, NEMA and Global Engineering Group: http://www.nema.org.
[3] IHE - http://ihe.univ-rennes1.fr/whatisihe/. Accessed January, 2006
[4] Klossa J, Cordier JC, Flandrin G, Got C, Hémet J. An European de facto standard for image folders applied to telepathology and teaching. Inter. J. Med. Information. 1998, 48:207-216.
[5] Burnett L, Chesher D, Mudaliar Y. Improving the quality of information on pathology request forms. Ann Clin Biochem, 2004. 41(Pt 1): p. 53-6.
[6] ADASP. Consultations in surgical pathology. Association of Directors of Anatomic and Surgical Pathology. Am J Surg Pathol. 1993 Jul;17(7):743-5.
[7] Campbell, F. and D.F. Griffiths, Quantitative audit of the content of histopathology reports. J Clin Pathol, 1994. 47(4): p. 360-1.
[8] Imperato PJ, et al., Results of a cooperative educational program to improve prostate pathology reports among patients undergoing radical prostatectomy. J Community Health, 2002. 27(1): p. 1-13.
[9] Fleming KA. Evidence-based pathology, J Pathol. Jun 1996; 179(2): 127-8.
[10] Fleming KA. Evidence-based cellular pathology, Lancet, 2002; 359(9312): 1149-50.
[11] Cowan DF. Quality assurance in anatomic pathology. An information system approach. Arch Pathol Lab Med, 1990. 114(2): p. 129-34.
[12] Goldstein NS. Diagnostic errors in surgical pathology. Clin Lab Med, 1999. 19(4): p. 743-56.
[13] Zarbo RJ, Meier FA, Raab SS. Error detection in anatomic pathology. Arch Pathol Lab Med, 2005. 129(10): p. 1237-45.

Address for correspondence

Dr. Christel Le Bozec. UMR_S 872 Eq20, Centre des cordeliers, 15 rue de l'école de médecine, 75006 Paris, France.
Christel.lebozec@spim.jussieu.fr

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

A Generic, Web-based Clinical Information System Architecture Using HL7 CDA: Successful Implementation in Dermatological Routine Care

Thilo Schuler[a,b], Martin Boeker[a], Rüdiger Klar[a], Marcel Müller [b]

[a] *Department of Medical Informatics, University Medical Center Freiburg, Germany*
[b] *Department of Dermatology, University Medical Center Freiburg, Germany*

Abstract

The requirements of highly specialized clinical domains are often underrepresented in hospital information systems (HIS). Common consequences are that documentation remains to be paper-based or external systems with insufficient HIS integration are used. This paper presents a solution to overcome this deficiency in the form of a generic framework based on the HL7 Clinical Document Architecture.

The central architectural idea is the definition of customized forms using a schema-controlled XML language. These flexible form definitions drive the user interface, the data storage, and standardized data exchange.

A successful proof-of-concept application in a dermatologic outpatient wound care department has been implemented, and is well accepted by the clinicians. Our work with HL7 CDA revealed the need for further practical research in the health information standards realm.

Keywords:

medical record systems, computerized, generic architecture, systems integration, hospital information systems; CDA, openEHR, HL7, semantic interoperability

Introduction

The potential advantages of computerized patient records (CPR) versus conventional paper-based files have been known for many years [1]. CPRs were envisioned to improve both routine care and research in a multiple ways. The possible benefits range from obvious ones, like readability or availability, to visionary decision support scenarios. Studies were able to support many theoretical CPR advantages, while increased experience with CPRs and clinical information systems also reveals deficiencies. Ginneken [2] identifies low user-acceptance as a barrier that slows down the adoption of CPRs. This low user-acceptance results from lack of integration and missing flexibility in content and use. Other impediments mentioned are an inability to adapt to change, the lack of data exchange standards, or an insufficient financial return on investment.

The aim of this paper is to share our experiences from designing and implementing a generic clinical information system architecture. This project was initiated to implement a special-purpose application in clinical dermatology. In order to make our decisions comprehensible and to prove our concept, the context and realization of this implementation will be illustrated.

Problem

The outpatient clinic for patients with chronic wounds is a subunit of Freiburg university hospital's dermatology department and needed an application to document their treatment. Up until this point paper files had been used to record visit findings and procedures. As the chronic wound clinic is a multi-provider facility, the main problems with the paper records were difficulties with legibility and conventions about how and what should be recorded. Further issues were the limited data reusability for research (no standardized recording practice and primarily no electronic format) and the non-existence of discharge letter writing assistance (e.g. partially pre-filled letter template).

Within the hospital an in-house developed information system is in use [3]. The system provides hospital wide services such as a central document and picture archive (e.g. discharge letters, radiology images), access to lab data, and management of administrative patient data via a range of client applications. The communication between distributed components is based on HL7 v2 [4]. Besides these general functions a number of custom extensions have been developed to reflect the information needs of single sub-/departments (e.g. a module to document the vein status).

An external solution had also been considered for the chronic wound clinic. After determining scope (2 forms with around 200 data items) and basic requirements (analyzing existing paper recordings, conducting a series of three observation and interview sessions, and review of the existing products in the market) it became clear that no commercial product in the small market of chronic wounds documentation systems could fulfill the special needs. Lacking customisation flexibility, unsolved integration problems, incomplete features, and unacceptable conditions of use by one vendor were the main reasons for the decision to develop an in-house solution.

Health information interoperability standards

Establishment of semantic interoperability between distributed health information systems is probably one of the most important challenges of health informatics today [5]. Most expected CPR advantages [1] are based on this premise. It is widely accepted that only well designed health information standards can solve this problem.

There are currently two major standard initiatives that aim to achieve semantic interoperability of medical information: CEN 13606 [6, 7] & *open*EHR (http://www.openehr.org/, accessed 29 Nov 2006) and HL7 v3 (http://www.hl7.org/, accessed 29 Nov 2006). While it is undisputed that standardized information structures have to be exchanged, the approaches to about how this can be done differ.

HL7 Clinical Document Architecture: HL7's central design artefact is the Reference Information Model (RIM) from which message specifications are derived via a cascade of intermediary models. The idea is that this approach secures shared semantics. The HL7 Clinical Document Architecture release 2 (CDA r2, [8]) is an ANSI-approved exchange standard for medical documents. It has been developed according to the HL7 methodology and is fully based on the RIM. CDA r2 documents are XML instances with two main parts: a *header* setting the document context and the *body* containing the clinical report in a semantically enriched HTML-like markup. The CDA r2 specification purposely has a wide scope to be able to express any clinical document. Further constraint mechanisms are needed to enforce a particular structure. Currently, further constraints are defined by narrative 'implementation guides'. In the future formal constraint expressions, called 'HL7 Templates', are envisioned for this task. In order to ease the adoption of the standard an incremental approach regarding the semantic enrichment is supported. The concept of levels reflects this design, while only a CDA r2 Level 3 document is envisioned to guarantee full semantic interoperability. CDA r2 Level 1 only expects a standard conforming document header while there are few restrictions for the body. In CDA r2 Level 2 the coarse body structure (sections) needs to be understood by the receiving system through definition of meaning in the form of terminology codes. Level 3 can be achieved by adding semantic markup (entries) for every narrative clinical statement. To adhere to the CDA's human readability principle, level 3 markup can't contain more information than the narrative.

CEN 13606 and openEHR: The *open*EHR Foundation is a not-for-profit company behind an open community effort to produce specifications (requirements, technical and clinical models) and reference implementations. The aim is to achieve an "open, interoperable health computing platform, of which a major component is clinically effective and interoperable electronic health care records (EHRs)" [9]. While not a standards body itself, *open*EHR is dedicated to work with standards organizations. The revised European standard CEN 13606 is influenced by *open*EHR. Like the *open*EHR architecture, it promotes a stable reference model, whose classes can be aggregated and further constrained by standardized, formal clinical content models called *archetypes* [10]. This so-called two-model approach separates medical knowledge from technical knowledge to achieve semantic interoperability and future-proof health information systems. Part two of the CEN 13606 standard has adopted the *open*EHR Archetype Description Language (ADL).

Design decisions

As a consequence of the depicted situation the following in-house solution design goals were rated with the highest priority:

- Customisation – to fit to clinical information requirements and workflow needs
- Integration – into the existing hospital IT environment
- Data sustainability and semantic interoperability – through standard conformance
- Generic methodology and components – to foster reusability in other clinical areas
- Pragmatism – to develop a solution that takes the limited resources into account and allows early utilization

Material and methods

The design goals mentioned above determined the choice of development technologies and the applied methodologies. To allow the possibility of reusing the system for similar documentation problems within the department of dermatology (e.g. the clinic for dermatological autoimmune diseases) it was decided to build a generic architecture that can be adapted flexibly according to new needs. XML instances of a form description language implemented in RELAX NG were pictured to drive a GUI generator. Flexibility regarding data entry items implies flexibility of the underlying database. This requirement excluded the use of the central hospital database facilities and we used a variable XML format instead. In order not to create an isolated "data island" we decided to build a XSLT transformation mechanism that could export our data to the standardized CDA r2 format. Integration with the hospital information system (HIS) was to be achieved by regular sending of CDA documents as payload in HL7 v2 messages. Administrative patient data should be similarly imported from the administrative data module of the HIS. We conceived the architecture as a web application framework whose core features (form definition language, data storage) are based on XML technologies and tools. The presented initial version also uses technologies such as PHP5 (form generator) and MySQL (4 tables: patient data, user data, XML form definition and XML form data). The necessary security is achieved as the application can currently only be accessed from within the firewalled hospital intranet. Additionally LDAP authentication and https encryption were installed.

After the development of the generic architecture, analysis and deployment of the chronic wound application were the next steps. A number of form definition iterations were planned to achieve a customized solution. This approach follows the idea of rapid prototyping [11] and allows early

user involvement to secure a high-level of user-acceptance.

Results

Overall architecture

Following the classic web paradigm the system architecture consists of a web server and database containing the application (PHP5) and a web browser on the client side to interact with the application. The application consists of: a form generator that creates HTML forms based on a formal XML form definition and modules for tasks such as user authentication, transformation to CDA r2, or integration with the HIS.

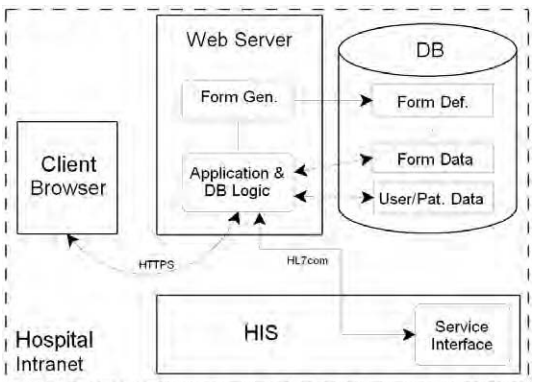


Figure 1 - Architecture of the system framework

HTML form generation

The form generator component was written in PHP5 and dynamically generates HTML forms based on XML instances of a formal form definition language. These forms are used for editing and viewing. There is no restriction in the number of forms.

Form definition language: A RELAX NG schema grammar restricts the form definition elements to 6 types that can be arranged using 3 layout patterns. Additionally the form can be divided in sections and subsections. Each sub-/section XML tag has optional attributes to specify a terminology (e.g. LOINC or SNOMED) and the suitable code. The element types comprise the normal HTML input fields (text, checkbox, dropdown, upload) and widgets for date input (calendar) and for image display with planimetry. Each element has a mandatory label and an optional suffix. The available layout patterns are 'beneath' (field elements below each other), 'float' (field elements next to each other), 'table' (configurable tabular layout).

GUI features: The Graphical User Interface (GUI) is rendered according to the form definition file. By default, every section is displayed in its own tab (see Figure 2 for a screenshot from the chronic wound application).

Data storage

The structure of the form content data depends on the form definition. To store submitted form data, a simple XML file is generated by mapping element names to their current values. This file is stored in the MySQL database together with metadata for retrieval.

Transformation to the HL7 CDA level 2

Based on former experiences with the first release of CDA [12, 13], we developed the framework according to the CDA r2 specifications. Provided correct code mappings exist in the form definition, an XSLT script that creates a CDA r2 Level 2 document can be derived automatically from the form definition. The CDA body content is created from the flexible form definition and the corresponding form data, while the metadata in the CDA header is based on information that is invariably required by the stable module.

Integration with the HIS

Integration with the existing HIS infrastructure is guaranteed by import of master data and export of CDA documents once they are approved by the supervising physician. Technically, import and export works by transferring XML documents via a HL7 v2 service interface provided by the HIS.

Implementing the chronic wounds application

Using the results from the scope and basic requirements analysis (conducted prior to the framework realization), an initial form definition was built by one author with good medical and IT knowledge. Consequently, 5 half-hour sessions with one medical user and one author were sufficient to reach the final definition.

In order to cover the use cases of the chronic wound clinic it was decided to have two types of forms: an 'initial assessment' form and a 'follow-up' form. The 'initial assessment' form contains the detailed patient history and information about special tests, that aren't performed during every visit. It is primarily filled during the first visit, but it can also be updated later. A new 'follow-up' form is completed during each subsequent visit, documenting the current status and determining the further procedure.

Discussion

Proof of concept

The first application of the presented framework has been successful. It is used daily by 3 nursing staff and 4 physicians. During the first 2 months of use about 500 visits of 150 patients have been documented. Although the system doesn't

Figure 2 - Chronic wound web-application GUI

provide all possible features yet (see the section 'Future'), early positive user reactions show that the system provides major improvements to the former solution. A study to analyse the system usability and the added value of the solution is planned.

The user involvement during requirements and implementation iterations was very valuable for both parties (medical and technical) and we can confirm similar experiences mentioned in the literature [2, 14]. Having the possibility to make quick adaptations to the form definition and displaying the result instantly helped the clinicians immensely to assess the current form and recommend improvements. The generator driven evolutionary approach described by Lenz and Kuhn [15] is similar in the rapid prototyping respect but manages system integration differently.

Integration

The presented architecture is potentially autonomous, but an integration with the central HIS components proved to be possible. The decentralized integration strategy is based on the standardized messaging exchange format HL7 v2. An direct extension of the HIS would have meant a prolonged development cycle and a high degree of inflexibility (limited existing data structures), which possibly could have lead to compromises in design.

CDA standard

Standard support is crucial for inter-organizational information exchange and must be a goal of every modern clinical information system. We decided to use the HL7 CDA r2 standard. Its document orientation suits the form-based framework well, and more importantly, it was a pragmatic choice that could be implemented relatively easily on top of ubiquitous XML tools. Currently, the generated CDA documents conform only to Level 2. The generation doesn't impose major difficulties provided suitable codes are found and set in the form definition. First experiments with Level 3 markup showed that it will not be easy to automatically create semantically correct CDA entry statements. Especially the necessary combination of several entries or the expression of post-coordinated SNOMED terms showed much arbitrariness.

Total semantic interoperability means that a receiving system can derive the same meaning from a standardized information unit (e.g. a CDA document) as the sending system. This must be true for any CDA document. In our opinion for this long-term goal the CEN 13606/*open*EHR approach using coherent, standardized content models (archetypes) seems currently better suited. HL7 is aware of this "gap", which HL7 templates are supposed to fill. Further research regarding CDA r2 Level 3 compared to CEN 13606/*open*EHR is needed.

Future

Besides the necessary research mentioned in the preceding discussion sections, various practical and technical improvements to the framework could be explored and implemented to further address issues like the ones mentioned in the introduction (e.g. data reuse). Routine care would benefit greatly from a letter writing assistance tool or a history feature where certain entries can be viewed over time (e.g. numerical values as a chart). An "XML2relational" data dump tool would allow analysing the collected data with conventional statistics programs. At the moment, only one version of each form is valid. A variable versioning mechanism would overcome this restriction. Technically, a re-implementation of the architecture based on pure XML technologies such as XForms, the Apache Cocoon Framework, and a XML database would be interesting.

Conclusion

This article describes a generic system architecture framework for health care applications. Through an autonomous concept special-purpose, form-based solutions that are customized to the needs of clinical users can be developed.

The HL7 CDA r2 standard was used to ensure system integration. A proof-of-concept implementation for routine care shows its applicability.

To what extend HL7 CDA or alternatives like *open*EHR can enable full semantic interoperability needs to be explored in further practical trials.

Acknowledgments

We would like to thank the medical and the nursing staff of the dermatological chronic wounds unit for their assistance and feedback during development.

References

[1] Dick RS, Steen EB. The Computer-Based Patient Record: An Essential Technology for Health Care. Rev ed. Washington DC: National Academy Press, 1997.

[2] van Ginneken AM. The computerized patient record: balancing effort and benefit. Int J Med Inform 2002: 65(2): 97-119.

[3] Klar R, Zaiss A, Binder M, Schrgendorfer I. The minimum basic data set as the core of the Freiburg University Hospital information system. Medinfo 1995: 8 Pt 1: 586-589.

[4] Schrader U, Zaiss A, Büchele S, Binder M, Leupolz U, Klar R. Realization of a CPR Using HL7 Queries From Communication Server to Information Server. In: Kokol P et al., eds. Medical Informatics Europe '99. IOS Press, 1999; pp. 834-837.

[5] Norgall T. Interoperability--a key infrastructure requirement for personalised health services. Stud Health Technol Inform 2005: 117: 125-133.

[6] Health Informatics – Electronic health record communication - Part 1: Reference Model: CEN/TC251; 2005 March 2005. Report No.: prEN13606-1:2005:E.

[7] Health Informatics – Electronic health record communication - Part 2: Archetypes: CEN/TC251; 2005 December 2005. Report No.: prEN13606-2:2005:E.

[8] Dolin RH, Alschuler L, Boyer S, Beebe C, Behlen FM, Biron PV, Shabo AS. HL7 Clinical Document Architecture, Release 2. J Am Med Inform Assoc 2006: 13(1): 30-39.

[9] *Introducing openEHR* [cited 29 Nov 2006]. Available from: http://svn.openehr.org/specification/TRUNK/pulishing/openEHR/introducing_openEHR.pdf

[10] Beale T. Archetypes: Constraint-based domain models for future-proof information systems. In: Baclawski K and Kilov H, eds. Eleventh OOPSLA Workshop on Behavioral Semantics: Serving the Customer (Seattle, Washington, USA, November 4, 2002). Boston: Northeastern University, 2002; pp. 16-32.

[11] Fitton D, Cheverst K, Kray C, Dix A, Rouncefield M, Saslis-Lagoudakis G. Rapid Protoyping and User-Centered Design of Interactive Display-Based Systems. IEEE Pervasive Computing 2005: 4(4): 58-66.

[12] Muller ML, Uckert F, Burkle T, and Prokosch HU. Cross-institutional data exchange using the clinical document architecture (CDA). Int J Med Inform 2005: 74(2-4): 245-256.

[13] Muller ML, Butta R, and Prokosch HU. Electronic discharge letters using the Clinical Document Architecture (CDA). Stud Health Technol Inform 2003: 95: 824-828.

[14] Wu RC, Abrams H, Baker M, Rossos PG. Implementation of a computerized physician order entry system of medications at the University Health Network--physicians perspectives on the critical issues. Healthc Q 2006: 9(1): 106-109.

[15] Lenz R, Kuhn KA. Towards a continuous evolution and adaptation of information systems in healthcare. Int J Med Inform 2004: 73(1): 75-89.

Address for Correspondence

Thilo Schuler
Department of Medical Informatics, University Medical Center Freiburg, Stefan-Meier-Str. 26, 79104 Freiburg, Germany
thilo.schuler@gmail.com

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Analyzing the Key Variables in the Adoption Process of HL7

Alejandro E. Flores, Khin Than Win

Health Informatics Research Centre, University of Wollongong, Australia

Abstract

The intention of this paper is to provide an analysis of variables that could affect the adoption of HL7 message standard. Based on the review of 33 cases where HL7 was successfully implemented the authors present relevant evidence related to inherent limitation of HL7. The result from this study indicates that it is necessary to enhance the standard to overcome particular limitations and facilitate the implementation of inter-institutional software interfaces based on HL7.

Keywords:

Health Level Seven, adoption, connectivity and standardization

Introduction

Health Level seven (HL7) consortium was founded in 1986 to research and develop a set of standards for electronic data exchange in the heath care domain. The HL7 standard is a structured specification that can be used for interconnection and exchange of health records [1]. HL7 is the most widely used messaging standard for clinical and administrative data exchange among health care applications in the information technology industry [2]. HL7 has established a set of information and message models for the development and implementation of interfaces for communication and transmission of medical data among heterogeneous health information systems [2][3]. The aim of HL7 is to produce standards for a particular health care domain considering the holding of a strict and well-defined framework that ensure consensus, openness and balance of interest, and allow the development of specifications for the implementation of messages model and software interfaces[1].

At the actual level of development, the most intractable barrier for the use of HL7 has been the lack of standards for exchanging fine-grained, highly heterogeneous, structured clinical data among information systems that had been implemented under different platforms [2][3]. Moreover, the additional consideration of specific health information domain and inclusion of new information and communication technology adds levels of complexity to the initial hitch [4]. Therefore, it is necessary to explore the HL7 message standard and its application to different health care domains to establish limitations of HL7 standard, and to outline alternative courses of action that permit the overcoming of those restrictions [4].

The aim of this paper is to identify advantages and disadvantages of HL7 standard by the analysis of 33 specific cases. It will permit to identify benefits of the use, understand the adoption process and recognize limitation and barriers that should be considered during the implementation of HL7 standard. This will also offer an initial answer to why it is important to consider the development of new set of communication models based on HL7 in order to overcome the new requirement of connectivity and communication for specific application in the health care domain.

Methods

As the main purpose of this research is to identify the key attributes that facilitate or limit the adoption of HL7. The methodology used in this study was based on an analytical generalization [5] based in the analysis of 33 articles related to the use of HL7 during the design and development of software interfaces for electronic health information systems.

Framework

The Technology Acceptance Model (TAM) [6][7] was used as a framework for this analysis. The variables considered for this analysis were: (1) benefits, (2) adoption and (3) barriers. The variable cost was not considered because of the limited information provided by the authors of the studied papers.

Scope

The search was limited to articles published since 2000. The articles selected include implementation experiences in: (1) Electronic Health Records (EHR) Systems, (2) Cardiology Information Systems, (3) Electro Physiology Information Systems, (4) Radiology Information Systems, (5) Administrative, Support and Knowledge Information Systems, (6) Tele-medicine Services, (7) Regional Health Information Systems, and (8) Home Monitoring Health Information Systems.

Search strategy

The search was realized in March 2006 and repeated in July and October 2006. The second and third searches were conducted to include recent cases where HL7 has been used as a framework for the development of commu-

nication interfaces in the health care domain. The following key words were used during the search process: (1) HL7, (2) Health Level Seven, (3) health information standard, (4) interconnectivity, (5) software interfaces and (6) communication. The databases ScienceDirect, Proquest 5000, IEEE Xplore and SpringerLink were used as knowledge sources during the search process due to the immediate access for the authors.

More than 150 articles were identified during the search process. However, only 63 of them were considered for review based on the relevance over the adoption of HL7 in the health care domain. Three elements were considered as relevant to pick up the list of 63 articles: (1) the article presents information related to the development and implementation of software interfaces for data exchange among health care system, (2) the main standard used for the development was HL7, and (3) the information presented by the authors has been obtained during the implementation of projects since the year 2000.

The Table 1 shows the criteria used to select the final list of 33 articles. The scope of these criteria was to narrow the articles selected to those that present the most recent information about HL7 implementation.

Table 1 - Selection Criteria

Criteria	Considerations
Version	HL7 v. 2.5 and HL7 v. 3.0
Implementation	Modeling, Implementation of Communication Interfaces, Software, and Web-based software and interface.
Message Format	XML
Publishing Date	Year 2000 and after.
Elements Implemented	HL7 messages, new vocabulary, and model representation based on the Reference Information Model (RIM) and/or Clinical Document Architecture (CDA).

Results

The articles selected were classified into the four implementation categories described in table 1. The analysis was conducted considering the following variables: (1) Benefits/Limitations (performance and time saving, adaptability, and extensibility, modeling and implementation support; (2) Adoption of HL7 in health care; (3) Technical barriers for adoption of HL7

Benefit and limitation of HL7

Performance and time saving

The performance and time saving during the exchange of data is a main decision variable for the adoption of international information and communication standards. Moreover, the inclusion of international standards for exchange of information among health information systems, such as HL7, may have a direct and positive impact in the time processing and performance of medical data exchange.

Langer [8] indicated that the implementation of HL7 standard enhances the performance and interoperability of health information systems and diminishes the time required for data exchange and access between hospital units and departments. Moreover, they consider that the incorporation of HL7 messages allows the friendly and easy access to multiple instances of patient's records and clinical information, which permit to enhance the quality of the health care delivery service. Ko et al. [9] agreed that the use of HL7 improves performance and efficiency of connectivity and interoperability among health information systems. However, they pointed out that large implementations, such as integrated institutional web-based application or inter-institutional health information systems, increase the risk of informatics attacks and could provoke loss of robustness, security and flexibility of the systems.

Müler et al. [10] indicated that the implementation of communication standards is a cost factor in modern health care systems. This implies that a better and faster exchange of medical information could ameliorate medical care services. They also agree that HL7 framework, in special the Clinical Document Architecture (CDA), allows the development of efficient and well defined interfaces that enhance the exchange of medical documents in local health information structures and diminishes the time of transference.

In almost all the cases analyzed, the implementation of HL7 has permitted the enhancement of performance and diminution of time during the transaction and data exchange among electronic health information systems. However, the inclusion of security schemes should be considered to ensure the safe delivery of messages and information and overcome possible information threats and loss of robustness of the systems.

Adaptability

Since its beginning HL7 has been developed as a standard for software interface that should be able to connect different and highly heterogeneous software environments. For this reason adaptability has become a keystone attribute that must be accomplished by HL7.

Most of the authors state that HL7 has presented high adaptability to the domains actually included in the standard scope [11][10]. In this sense, the HL7 Reference Information Model (RIM) has been successfully adapted to electronic health records systems, department information system and administrative and financial applications [11]. However, it has presented limited adaptability to nursing information systems, inter-institutional application

and specific health care systems such as general practice and radiology [4]. The Act[1] class of the RIM has been unable to represent nursing activities and the HL7 vocabulary is limited for nursing information [4]. The actual data structure and vocabulary definition is unable to map complete data information for General Practice's information systems [12] and, even though, HL7 has demonstrated excellent performance during the exchange of radiology clinical information, it is still limited for large image exchange [13].

According to Müler et al. [10] CDA can be easily adapted to overcome local health system requirements such as electronic health records systems, decision support systems and knowledge applications [14]. However, some issues, related to adaptability requirements, efforts and cost to meet inter-institutional needs, should be addressed to provide a better support in the implementation of software interfaces for the exchange of clinical document among different actors of the health care domain [15][10]. They also suggested that CDA can be adapted to different inter-institutional scenarios by including additional data structures and vocabulary that allow overcome the two basic limitations: (1) CDA is limited to the scope of HL7 definitions [16] and any additional extension to the data or vocabulary definition is limited to local solutions [10]. Finally, Müler et al. pointed out that the inclusion of CDA permits diminish the cost associate to the exchange of clinical documents and could enhance the delivery of primary and secondary health care services.

Bicer et al. [17] discussed about the necessity to develop message exchange frameworks that provide support semantic interoperability between different versions of HL7 message standard. This is one of the most common problems during the exchange of data between software interfaces developed under both version 2.x and version 3 of HL7.

In conclusion, HL7 has few adaptability issues for health information domains such as electronic health record, administrative and financial systems, and departmental information systems. However, due to limitations in the referential data representation and vocabulary, HL7 has restrictions that have to be considered for implementing health information systems in particular domains such as nursing system, general practice and inter-institutional application.

Modeling, extensibility and implementation support

HL7 allows the incorporation of information structures and extension to the vocabulary and messages specifications. Moreover, the information model and vocabulary should be adapted during the process of software implementation to achieve local needs [4][11]. However, those extensions are limited in exchange of information among different health providers [10].

According to Fernandez and Sorgente [14], the ad hoc variation of the HL7 Unified Modeling Language (UML) is incompatible with existing standards. Moreover, the HL7 documentation is wide and complex making it difficult to understand. In addition, entities, roles and associations had been represented and structured for implementation not for abstract representation, e.g. roles are job descriptions without security specifications and associations that do not have names or semantic values. Indeed, they had been replaced by separate class representations. Moreover, to add extend class representations HL7 uses arbitrary names based on prefixes of the original classes and not stereotypes as usual in UML representations. In general, HL7 artifacts do not completely follow the UML patterns and software engineering rulers. This issue makes the standard unnecessarily complex for the elaboration of model representations and model extensions.

The RIM provides explicit semantic and lexical representations of messages and fields. Additionally, it facilitates the data integration among health care applications, providing structural information models and a health base vocabulary [14]. Moreover, the RIM facilitates the mapping process over basic health care informational representations and model [20]. However, these representations are relatively limited if the standard is applied to some particular health care domains such as nursing systems [4], general practice's information systems [12] or radiology information systems [21].

The HL7 CDA gives a framework for design and interoperability of clinical documents [20]. Furthermore, the CDA provides support and representation for messages based on text, image, sound and multimedia contents, and allows the enhancement of vocabulary and information structure to reach particular requirement. However, it does not provide guides or recommendation for the development of structural or vocabulary extensions and most of the representations require adjustment and modification to local needs [22]. Additionally, any local extension to the model, vocabulary or documents structures must be considered as optional data or field during the interchange of data among health care providers [4]. This implies that, due to the message definition, some relevant information could not be interpreted by the destination bode.

Lebak, Yao and Warren [23] suggest that HL7 should provide a better support for large scale system implementations that consider interconnection among different actors of the health care. It implies the consideration of a framework that includes support for the development and deployment of integrated, interconnected and secure software interfaces. [23], the integration of electronic health records, and provides a wide set of elements that support the modeling and implementation of robust inter-institutional software interfaces based on HL7 standard [15].

In Summary, the had hoc UML model representation used by HL7 artifacts do not follow the object oriented standards, this makes more complex the standard for develop-

1 The RIN Act class is used to represent intentional acts that are performed to benefit the patient and associated clinical activity.

ers and increases the time and cost associated to the development of HL7 message interfaces. HL7 provides extensibility capabilities. However, extensions of HL7 standard are limited to local implementation. Moreover, inter-institutional implementations, such as Regional Health Information Systems, should consider this limitation to include homogenized message structures and message interpretations.

Adoption of HL7 in health care

The use and adoption of HL7 allows the implementation of integrated health care systems. In addition, HL7 provides a native and robust interoperability framework for software development and deployment. Moreover, HL7-CDA reduces the cost of moving existing documents to new standards [10] and enhances the work flow between health information systems. For these reasons most authors explicitly agree that HL7 is a recommended and required standard for information exchange among health care applications. However, the adoption of HL7 should consider several issues that should be addressed to the implementation plan. Some of them are adoption limitation over ad hoc UML modeling of HL7 [18], complexity of the implementation over large information systems, high cost, restrictions of vocabulary and the consideration of other communication standards that provide better support over specific domain, e.g. The Digital Imaging and Communications in Medicine (DICOM) for radiology exchange of information [13].

Technical barriers for adoption of HL7

HL7 provides a wide range of guidelines and specification for implementation of data structures and messages for software interfaces among health informatics applications [2]. Instead, HL7 has several technical limitations related to information model specifications, message definitions, document structures and vocabulary applied to specific health care domains.

First, according to the definitions of HL7 standard, the message should contain a basic set of fields, which must hold the critical information required for exchange; additional information should be provided using the optional fields [1]. This fact does not represent a real inconvenience for local implementations [22][11]. However, this issue could increase the costs and efforts required during the development and deployment of HL7 messages for inter-institutional applications [10].

Second, the RIM has presented issues during the data mapping and development of messages in some health care domains [4][12][20]. According to Danko et al. [4], the RIM class Act is unable to represent complete model structures for nursing information systems. Moreover, they suggested the additional attributes to the RIM-Act class and the enhancement of the HL7 vocabulary to include nursing information. Furthermore, these limitations also affect the development of software solutions in other domains such as general practice [12], and the exchange of referral and discharge letters [11].

Thirdly, the CDA has been developed to provide a framework for document representation and message elaboration based on HL7 standards. However, the CDA framework is in a development stage. This implies that CDA does not provide a complete data representation for some specific health domains or local requirements [11]. Moreover, limitations of HL7 vocabulary and data structure make necessary the development of local solutions, which are not totally compatibles for inter-institutional information exchange [4][11]. In addition, actual vendor's software does not provide complete support for integration of certain external data, and local implementations are restricted to internal needs [10]. These issues add levels of complexity to the development process and increase the cost if implement HL7-CDA messages interfaces over inter-institutional health information systems.

Finally, additional limitations are related the cost and time required for implementation and the complexity of the existent HL7 artifacts [13]. The implementation of HL7 messages based on CDA over XML requires an important amount of time and cost of development. In addition, the deployment of large health information system makes the development and implementation process highly difficult and requires additional resources [14].

In conclusion, HL7 has provided a helpful framework for developing and implementing health information messages interfaces. However, there still exist some issues to address in order to improve the standard.

Discussion

This paper has presented an analysis based on HL7, both version 2.5 and version 3.0, implementations experiences over different health care domains since the year 2000. Those experiences ratified that HL7 provides a wide range of capabilities for the enhancement of message communication among health information systems. However, those experiences also had made manifest that HL7 has structural and technical limitation that could make difficult the adoption process. Those limitations are related to the message modeling and implementation, data structure representation and vocabulary presented in the RIM and CDA.

HL7 standard provides a basic framework for message modeling and implementing. However, in most of the cases analyzed the authors recommend to update the standard to overcome local needs. Furthermore, it is necessary to consider the development of a framework that provides guidelines for the development of inter-institutional message solutions. Additionally, it is also necessary to develop a message exchange framework that enhances the compatibility between version 2.x and version 3.0 HL7 messages.

Both the RIM and CDA have been implemented and used to enhance the data, information and document exchange among local and inter-institutional health information systems. At the local level both provide a framework for message design and extensibility implementation, and system deployment [1][10]. However, the data definition and vocabulary incorporated to the HL7 standard is limited to

the existing in the RIM and CDA definition [20][10][19]. Moreover, the existing data and vocabulary definitions limit the mapping of data and generation of messages. These limitations can be overcome by incorporating extension to the models and vocabulary at the local level. Nevertheless, it is important to consider that according to the HL7 message standards [1][4], extensions must be considered optional attributes or optional values during the exchange of messages. These issues increase the complexity and cost of the development and implementation of health information software solution at inter-institutional level.

Conclusion

HL7 has demonstrated to be an important advance in the development of health information software for medical data exchange. However, the implementation of HL7 in specific cases, requires the development of new information models, message model and vocabulary that allows the implementation of those interfaces [4][12][13]. Moreover, the development of communication interfaces on specific scenarios would permit to enhance the actual information structure of HL7 standard. On the other hand, the definition and specification of HL7 message information models, for specific health domain software, allows the implementation of robust software interfaces. These interfaces would enhance the information exchange and interoperability among different local and inter-institutional health care software applications.

Acknowledgements

The author acknowledges the Government of Chile and University of Talca (Chile) for the support of this research.

References

[1] Beeler G. HL7 Version 3 – an object-oriented methodology for collaborative standards development. Int. J. of Med. Info. 1998: 48 (1-3 Suppl): 151-161.

[2] Henderson M. HL7 Messaging. Texas: Otech Inc, 2003.

[3] Hinchley A. Understanding Version 3: A premier on the HL7 Version 3 Communication Standard. Munich: Alexander Mönch Pub., 2005.

[4] Danko A, Kennedy R, Haskell R, Androwich I, Button P, Correia C, Grobe S, Harris M, Matney S, and Russler D. Modeling nursing interventions in the act class of HL7 RIM Version 3. J. of Bio. Info. 2003: Inf. 36 (4-5 Suppl): 294-303.

[5] Yin RK. Case Study Research: Design and Methods. 3rd Ed. Thousand Oaks: Sage Publication, Inc, 2993.

[6] Davis FD. User acceptance of information technology-Systems characteristics, user perceptions and behavioral impact. I. J. Man-Machine Stud. 1993: 38 (3 Suppl): 475-487.

[7] Davis FD. A critical Assessment of potential measurement biases in the technology acceptance model: three experiments. Int. J. H-C Studies 1996: 45 (1 Suppl): 19-45.

[8] Langer S. Open RIMS: An open architecture radiology informatics management system. J. of Digital Imaging 2002: 15 (2 Suppl): 91 – 97

[9] Ko L, Lin j, Huang Ch, Chang J, Lai F, Hsu K, Yang T, Cheng P, Wen Ch, Chen J and Hsich S. HL7 middleware framework for healthcare information system. Proceedings of the HEALTHCOM 2006

[10] Müller M, Uckert F, Burle T and Prokosch H. Cross institutional data exchange using the clinical document architecture (CDA). Int. J. of Med. Info. 2005. 74 (2-4 Suppl) 245 – 256.

[11] Heitmann K, Schweiger R, and Dudeck J. Discharge and referral data exchange using global standards – the SCIPHOX project in Germany. Int. J. of Med. Info. 2003: 70 (2 Suppl): 195 – 203.

[12] Liaw S, Sulaiman N, Pearce C, Sims J, Hill K, Grain H, Tse J, and Mg C. Falls Prevention within the Australian General Practice Data Model: Methodology, Information Model, and Terminology Issues, JAMIA 2003: 10 (5 Suppl): 425-432.

[13] Um K, Kwak Y, Cho H, and Kim I. Development of an HL7 interface engine, based on tree structure and streaming algorithm, for large-size messages which include image data. J. of Comp. Methods and Programs in Biomedicine 2005: 80 (2 Suppl) 126-140.

[14] Katirai H and Ulrich S. Unlocking the value of clinical information: What you need to do now to enjoy the benefits in the future. Althoff KD et al. eds. WM 2005, LNAI 3782. Berlin Heidelberg: Springer-Verlag, 2005; pp. 330-338.

[15] Berler A, Pavlopoulos S, and Koutsouris. Design of an interoperability framework in a regional health system. Proceeding of the 26th Annual conference of the IEEE EMBS: 2004: 3093-3096.

[16] Dolin R., Alchuler L., Boyer S., Beebe C., Behlen F., Biron P. and Shavo A. HL7 Clinical Document Architecture, Release 2, JAMIA 2006: 7 (4): 30-39.

[17] Bicer V, Laleci G, Dogac A, and Kabak Y. Artemis Message Exchange Framework: Semantic interoperability of exchange message in health care domain. SIGMOD Records 2005:34(3 Suppl): 71-76.

[18] Fernández E and Sorgente T. An analysis of modeling flaws in HL7 and JAHIS. Proceedings of the 2005 ACM Symposium on Applied Computing.

[19] Orgun B. and Vu J. HL7 ontology and mobile agents for interoperability in heterogeneous medical information systems. Computers in Biology and Medicine 2006: 36 (7-8 Suppl): 817-836.

[20] Lyman J, Boyd J, Dalton J, and Egyhazy C. Applying the HL7 Reference Information Model to a Clinical Data Warehouse. IEEE 2003.

[21] Tellis W., Androiole P., Jovais C., and Avrin D. RIM Minus PACS Equals Film. Journal of Digital Imaging: 2002: 15 (1 Suppl): 20-26.

[22] Bilykh I, Jahnke J, McCallum G, and Morgan P. Using the Clinical Document Architecture as open data exchange format for interfacing EMRs with clinical decision support system. Proceeding of the 19th IEEE Symposium on Computer Based Medical Systems: 2006: 3093-3096.

[23] Lebak J. Yao J, and Warren S. HL7 – Compliant Healthcare Information System for home monitoring. Proceedings of the IEEE EMBS, 2004.

Address for correspondence

Alejandro E. Flores
Health Informatics Research Centre,
University of Wollongong, Australia.
Phone: +61 4221 3103
Email aefz871@uow.edu.au.

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

An XML Model of an Enhanced Data Dictionary to Facilitate the Exchange of Pre-Existing Clinical Research Data in International Studies

Stephany N. Duda[a], Clint Cushman[b], Daniel R. Masys[a]

[a] *Vanderbilt University, Nashville, USA*
[b] *University of California, San Diego, USA*

Abstract

Pre-existing clinical research data sets exchanged in international epidemiology research often lack the elements needed to assess their suitability for use in multi-region meta-analyses or other clinical studies. While the missing information is generally known to local investigators, it is not contained in the files exchanged between sites. Instead, such content must be solicited by the study coordinating center though a series of lengthy phone and electronic communications: an informal process whose reproducibility and accuracy decays over time. This report describes a set of supplemental information needed to assess whether clinical research data from diverse research sites are truly comparable, and what metadata ("data about the data") should be preserved when a data set is archived for future use. We propose a structured Extensible Markup Language (XML) model that captures this information. The authors hope this model will be a first step towards preserving the metadata associated with clinical research data sets, thereby improving the quality of international data exchange, data archiving, and merged-data research using data collected in many different countries, languages and care settings.

Keywords:

programming languages, software design,
knowledge representation (computer),
database management systems

Introduction

The meta-analysis of clinical research and community health data from different regions of the world is the focus of an increasing number of international collaborations [1]. Merging such diverse data is particularly challenging in medicine, since geographic differences in patient and disease characteristics, medical resources, language, and cultural beliefs can have a profound effect on the comparability of clinical information [2]. Varying regional standards for measurement units, "normal" ranges, treatments and procedures, and medical definitions may prevent data from being interpreted correctly. These differences are magnified when each participating site has collected the data independently as part of routine patient care. Successful harmonization of diverse, global data requires additional information that is known locally but not generally included in study records or patient care databases. Such metadata, or "data describing data," may help answer the following questions that arise when already-existing data are considered for combined analysis:

- Are the observations stored in a given variable in one data set equivalent to those stored in a variable in another data set, or are the variables similar but not identical for the purposes of a particular form of analysis?
- What computer program has been used to store and format the data?
- What is the "normal" range for this observation?
- What are the valid values that this variable may take?
- Who owns the data, and under what circumstances can it be used for analysis?
- From what geographic region are the data derived?
- In what language are the text-based data represented?
- What quality control measures have been applied to the data set to ensure that it is accurate and complete?

Historically, limited metadata related to these questions have been recorded in narrative form or made available by person-to-person communication with the owners of the research data. Over time, however, this information may be lost in the shuffle of old project records and in job transitioning among investigators and staff. The information loss resulting from reliance on memory, paper memos, and generally uneven documentation affects how accurately data sets can be combined for joint analysis, and how reliably data may be reused years after their creation [3].

If such "hidden" metadata can be identified, recorded, and transmitted alongside clinical research data sets in a well-structured format, it may reduce information loss and enhance the usability of research results. Extensible Markup Language (XML) is an established standard for this type of structured information exchange: its hierarchical text format is flexible, simple to implement, and easily interpreted by both humans and computers [4]. Web-based

applications and fields such as business, chemistry, and mathematics have employed XML extensively for information sharing, encoding semantic content, and data modeling [5-7]. In biomedicine, XML has been used to encode clinical guidelines, support electronic messaging between hospitals, and annotate nucleotide sequences [8-11]. The structure and syntax of such XML models is often described using XML Schema, the World Wide Web Consortium's recommended means of expressing XML schemata[12].

We report here the development of methods for capturing and storing clinical research metadata in an XML format. Preliminary work will focus on data collected by the International Epidemiologic Databases to Evaluate AIDS (IeDEA), an initiative of the U.S. National Institutes of Health to "establish international regional centers for the collection and harmonization" of HIV data and the management of region-wide analyses [13]. Participating sites include hospitals and clinics located throughout the Americas, Asia, Africa, and Oceania. The IeDEA collaboration will focus on regional analyses in its first three years, after which sites may choose to submit their data sets worldwide data pooling and analysis.

The majority of IeDEA member sites are based in developing countries, where the process of collaborative research differs significantly from that in developed nations [14]. Consequently, an appropriate data model must take into account the information exchange needs and capabilities of clinical research centers in the developing world. We assert that a simple, structured model designed to capture and preserve clinical research metadata can improve the quality of international data exchange and merged-data research, as well as preserve study metadata for future research projects.

Approach to model development

Six IeDEA-participating sites submitted descriptions of their pre-existing patient care and research databases. We reviewed these descriptions and developed a list of elements required to represent the characteristics of the content. Approximately thirty participants, including statisticians, HIV researchers, epidemiologists, database managers, and international clinical trials specialists, reviewed the proposed element list and suggested additional fields they might use to determine the compatibility and usefulness of different data sets. The resulting metadata fields were categorized and mapped to a collection of XML examples and a corresponding XML Schema.

The authors and collaborators identified the following design objectives for the metadata model, based on the input of the researchers and descriptions of locally-developed databases at IeDEA sites.

Design objectives for model

Flexibility: An ideal model for collaborative international data exchange must be adaptable to a range of data storage formats, from elaborate databases in fully-electronic clinical trials data management systems to the locally-developed systems of treatment centers in the developing world. While some sites may upload database files to their regional coordinating centers, others will struggle to deliver Microsoft Excel spreadsheets or plain text files. The internal structure of the data exchange format must accommodate all possibilities.

Document Independence: The metadata solution should not rely on any single data management system to repackage data and metadata into a cohesive unit. Instead, the metadata document should accompany data files originating from preexisting local databases – such as those maintained at IeDEA's participating sites – when they are submitted to collaborators or study coordinating centers.

Transparency: The syntax for recording research metadata should be easy to generate and interpret, for humans and computers alike. The information should be accessible to individuals without a specialized data viewer and it should be possible to automate part of the metadata documentation process.

Design objectives for model content

Data Descriptors: The model describing a clinical research data set should contain traditional data dictionary content, such as variable names, ranges, datatypes, and special values for missing or pending measurements. It is particularly important to capture this information for clinical databases in developing countries, where often no formalized data dictionary exists.

Measurement: The documentation for numeric values such as laboratory results should include the measurement unit and a description of the measurement method used.

Geography: The model should include some marker of location that indicates, geographically, the site(s) where the data have been collected and the location of the sponsoring or coordinating organization. This geographic code should be able to pinpoint the country, region, or city from which the data originated, but also allow fuzzy mappings to protect the privacy of individuals in small villages or accommodate the address-less locations of traveling or wilderness clinics.

Contacts: The name, contact information, and website of the organization (clinic, hospital, research center) responsible for producing and maintaining the clinical research data should be recorded in the metadata, along with a list of persons responsible for preparing the data or metadata.

Terms of Use: A section of text should clarify under what conditions the data set may be used.

Quality: The data exchange model should include a description of the quality assurance measures applied to each variable as it exists in the current data set. Examples of quality assurance tests include range and valid values checking (either manually or using an algorithm), double data entry with comparison, and source document verification.

Ontologies: The XML model should allow users to map locally defined variables to existing concepts in multiple controlled vocabularies and vocabulary versions. The

resulting structure should be able to represent compound concepts, also known as data tuples or post-coordinated terms.

Proposed model

The model overview contains XML samples and brief descriptions of the more complex elements in the document hierarchy. The full XML Schema is available online at http://ccasanet.vanderbilt.edu/xml.php. The website is hosted by CCASAnet, the Caribbean, Central, and South America Network for HIV research, which is one of seven IeDEA-funded regional coordinating centers and is located at Vanderbilt University [15].)

Model overview

The XML structure has a single root element, <metadata>, that encloses all other metadata-related elements. The first levels of the hierarchy are shown in Figure 1 - Root structure of the XML model for exchanging pre-existing international research data1:

```
1<metadata>
2   <about_dataset>
3       <db_software>...</db_software>
4       <owning_organization>
        See Figure 2 - XML description of
the organization that owns the data set2
5       </owning_organization>
6       <contact_people>
7           <contact_person>...
                </contact_person>
8       </contact_people>
9       <dataset_composition>
        See Figure 3 - XML structure for
storing information about the data set's
patient cohort3
10      </dataset_composition>
11  </about_dataset>
12  <dataset>
13      <datafile>
14          <variables>
15              <variable>
                    See Figure 4 - XML
structure for variables in a data set4
16              </variable>
17          </variables>
18      </datafile>
19      <datafile>...</datafile>
20  </dataset>
21</metadata>
```

Figure 1 - Root structure of the XML model for exchanging pre-existing international research data

As seen above, the XML elements are partitioned into two sections of meta-content: metadata related to the *data set as a unit* (<about_dataset>, line 2), and metadata related to the *observations contained in the data set* (<dataset>, line 12).

Metadata associated with the data set unit can answer questions such as "what software was used to maintain this database?," "who owns this data?," and "whom should I contact if I have questions related to this data set?" This information is captured between the <about_dataset> tags. The <dataset> tag, in contrast, encompasses the metadata related to the observations in the dataset.

The four fields within the <about_dataset> tag include <db_software>, which contains the product name and version of the database software used, and <owning_organization>, which contains the name, contact information, and geographic location of the hospital, clinic, or research institution that owns the data. The syntax of the <owning_organization> hierarchy is shown in Figure 2 - XML description of the organization that owns the data set2. Geographic location markers, or geocodes, are recorded in latitude and longitude, as these measures are easy to compute using publicly accessible web tools such as EarthTools or Getty Thesaurus of Geographic Names Online [16, 17]. Data owners may also stipulate a Terms of Use clause, <terms_of_use>, that restricts the applications of their data.

```
1<owning_organization>
2   <name></name>
3   <contactinfo>
4       <address></address>
5       <city></city>
6       <country></country>
7       <phone></phone>
8       <fax></fax>
9       <email></email>
10      <url></url>
11  </contactinfo>
12  <geocode>
13      <name></name>
14      <description>
            </description>
15      <latitude></latitude>
16      <longitude></longitude>
17      <coordinate_source>
            </coordinate_source>
18  </geocode>
19  <terms_of_use></terms_of_use>
20</owning_organization>
```

Figure 2 - XML description of the organization that owns the data set

A third element, <contact_people>, on line 6 of Figure 1, contains the names, titles, roles, and contact information for people associated with the creation, maintenance, or packaging of the data set. A high-level description of the data set, a description of the study cohort, and a set of geocodes demarking the data's source regions are part of the fourth and final descriptive metadata element, <dataset_composition>, shown below:

```
1<dataset_composition>
2   <description></description>
3   <study_cohort></study_cohort>
4   <data_region>
5       <geocodes>...</geocodes>
6   </data_region>
7</dataset_composition>
```

Figure 3 - XML structure for storing information about the data set's patient cohort

The second section of the data exchange model describes the data files that accompany the XML metadata docu-

ment. Each file is described within its own <datafile> tag (Figure 1, line 13), and the set of such tags constitutes the <dataset> structure of the model. These files may be database tables, spreadsheets, or even text files, but they contain variously formatted table information. All variables or equivalent table columns appearing in a data file are represented in the <variables> tag. The <variable> structure is detailed in Figure 4 - XML structure for variables in a data set4.

```
1<variable>
2    <name></name>
3    <description></description>
4    <datatype></datatype>
5    <length></length>
6    <count></count>
7    <measurement>...</measurement>
8    <quality>
9        <qa_measure>
10       <qa_type></qa_type>
11       <description>
         </description>
12       <date></date>
13       <result></result>
14       </qa_measure>
15   </quality>
16   <value_types>
17       <pending>...</pending>
18       <missing>...</missing>
19       <range>...</range>
20       <valid_values_sets>
21       <valid_values_set>
 See Figure 5 - XML structure for internal
and external valid values lists5
22       </valid_values_set>
23       </valid_values_sets>
24   </value_types>
25   <content_mappings>
26       <content_mapping>
         See Figure 6 - XML structure for
   ontology mappings
27       </content_mapping>
28   </content_mappings>
29</variable>
```

Figure 4 - XML structure for variables in a data set

The XML fields in lines 2 through 6 of Figure 4 include a variable name, description, datatype, type length, and count. The <measurement> field, shown on line 7, records the measurement unit, if applicable, and a description of the measurement method used to ascertain a numeric or subjective value. The subsequent <quality> tag (line 8) encloses a list of quality assurance measures applied to the variable, along with the quality test date and result. Special value codes for missing and pending data and counts of these code occurrences are recorded in the <value_types> field beginning on line 16, along with valid ranges for numeric variables and sets of valid values, discussed below.

```
1<valid_values_set
    location="inline|file|external">
2    <name></name>
3    <description></description>
4    <values_list>
5        <value></value>
6        <value></value>
7    </values_list>
8    <filename></filename>
9    <filetype></filetype>
10   <url></url>
11</valid_values_set>
```

Figure 5 - XML structure for internal and external valid values lists

The <valid_values_set> tag set, highlighted in Figure 5 - XML structure for internal and external valid values lists5, specifies a list of names, codes, or numbers considered to be valid content for a given variable. An *inline* list of valid values is often one developed locally, by the database designers. In this case, all valid codes are specified between the <values_list> tags, beginning on line 4 of Figure 5. A valid value set of type *file* indicates a value list included in a separate document transmitted along with the data set. A reference to *external* authorities suggests the valid values for this field are governed by controlled vocabularies such as SNOMED, ICD-9, or LOINC.

Each <variable> also contains a <content_mappings> structure that allows users to associate the variable semantics with one or multiple ontologies, specified by the vocabulary and version tags. The <content_mapping> structure is shown in Figure 6 - XML structure for ontology mappings6. Multiple <concept> tags within a single content mapping indicate a concept tuple, or post-coordinated term.

```
1<content_mapping>
2    <concept>
3        <vocabulary></vocabulary>
4        <version></version>
5        <term_name></term_name>
6        <content_id></content_id>
7    </concept>
8</content_mapping>
```

Figure 6 - XML structure for ontology mappings

The hierarchy of subsequent closing tags is specified in Figure 1 - Root structure of the XML model for exchanging pre-existing international research datagure 1. Many structures, such as <content_mappings>, are optional, and any text field can take a language attribute. Repeatable elements such as <contact_person> or <variable> are fully specified in the XML Schema available online.

Discussion

The XML model described in this paper meets the specified design objectives and constitutes a complete structure for international clinical research data exchange. The metadata fields include traditional data dictionary values, mappings to controlled vocabularies, and markers of geographic location, as well as specifics about local data entry and storage methods, definitions, measurements, quality control procedures, data ownership, and use restrictions. In addition, the model employs a flexible data exchange language (XML) that is system-independent and interpretable by both humans and computers.

By making this metadata uniformly available among IeDEA collaborators, researchers can analyze count and content metadata to determine in advance whether sufficient data are present to perform an analysis across global regions. Thus, the decision to make the data itself available for pooled analysis still remains with the owners of the data, while still allowing outside researchers to see what categories of information are available.

Although current database modeling and healthcare messaging standards exist, these exchange formats are either too rigid or too complex to use to describe data collected outside the controlled environment of a clinical trial. Data dictionaries are often focused on database structure and fail to capture information outside of the database syntax, while medical messaging and data exchange formats such as HL7 are designed for the exchange of patient care documents among electronically-enabled hospitals and care providers, rather than the merging of clinical research data.

We intend to validate the XML model presented here using real-world data sets contributed by IeDEA member sites. In areas where the model is shown to be not sufficiently flexible, we will update the schema to accommodate new categories of metadata. In addition, we intend to develop a tool to automatically populate most fields of the XML model by pre-processing the source database, and create a secure web interface for online maintenance of the metadata. This will reduce the burden of documentation placed on the data set owners. We will test the usefulness of the metadata and its XML model for the harmonization and meta-analysis of HIV electronic clinical research data from different countries, languages and cultural traditions.

The work described here represents a two-fold approach to improving the documentation that accompanies clinical research data in international studies. We have described seven categories of metadata that researchers find useful in determining the semantics and compatibility of pre-existing research data sets, and have formulated an XML model to document this metadata in a flexible, structured format. Future work will assess whether this innovation will improve the quality of international data exchange and merged-data analyses, thereby creating a more sustainable model for research collaborations worldwide.

Acknowledgments

Supported in part by NIH Cooperative Agreement 1 UO1 AI069923-01, International Epidemiologic Databases to Evaluate AIDS (IeDEA), Latin American Region. The authors wish to thank Bryan Shepherd, Paul Biondich, William Tierney, Claudio Faulhaber, and Bill Lober for their reviews of and comments on the proposed metadata model.

References

[1] Luukkonen T, Persson O, Sivertsen G. Understanding Patterns of International Scientific Collaboration. Science Technology & Human Values. 1992 Win;17(1):101-26.

[2] Ioannidis JP, Lau J. Uncontrolled pearls, controlled evidence, meta-analysis and the individual patient. Journal of clinical epidemiology. 1998 Aug;51(8):709-11.

[3] Detmer DE. Building the national health information infrastructure for personal health, health care services, public health, and research. BMC medical informatics and decision making. 2003 Jan 6;3:1.

[4] Wyke RA, Rehman S, Leupen B. XML Programming (Core Reference). Redmond, Washington: Microsoft Press 2002.

[5] Medjahed B, Benatallah B, Bouguettaya A, Ngu AHH, Elmagarmid AK. Business-to-business interactions: issues and enabling technologies. Vldb Journal. 2003 May;12(1):59-85.

[6] Miner R. The Importance of MathML to Mathematics Communication. *Notices of the American Mathematical Society* 2005:532-8.

[7] Murray-Rust P, Rzepa HS. Chemical markup, XML, and the Worldwide Web. 1. Basic principles. Journal of Chemical Information and Computer Sciences. 1999 Nov-Dec;39(6):928-42.

[8] Achard F, Vaysseix G, Barillot E. XML, bioinformatics and data integration. Bioinformatics. 2001 Feb;17(2):115-25.

[9] Araki K, Ohashi K, Yamazaki S, Hirose Y, Yamashita Y, Yamamoto R, Minagawa K, Sakamoto N, Yoshihara H. Medical markup language (MML) for XML-based hospital information interchange. Journal of medical systems. 2000 Jun;24(3):195-211.

[10] Dolin RH, Alschuler L, Behlen F, Biron PV, Boyer S, Essin D, Harding L, Lincoln T, Mattison JE, Rishel W, Sokolowski R, Spinosa J, Williams JP. HL7 document patient record architecture: an XML document architecture based on a shared information model. Proceedings / AMIA Annual Symposium. 1999:52-6.

[11] Shiffman RN, Karras BT, Agrawal A, Chen R, Marenco L, Nath S. GEM: a proposal for a more comprehensive guideline document model using XML. J Am Med Inform Assoc. 2000 Sep-Oct;7(5):488-98.

[12] Walmsley P, Fallside DC. XML Schema Part 0: Primer Second Edition. W3C Recommendation. 2004 [last accessed 2006 Dec 1]; Available from: http://www.w3.org/TR/xmlschema-0/

[13] IeDEA: Project Overview. 2006 [last accessed 2006 November 29]; Available from: http://www.iedea-hiv.org/about/

[14] Duque RB, Ynalvez M, Sooryamoorthy R, Mbatia P, Dzorgbo DBS, Shrum W. Collaboration paradox: Scientific productivity, the Internet, and problems of research in developing areas. Social Studies of Science. 2005 Oct;35(5):755-85.

[15] CCASAnet: Caribbean, Central and South America Network. 2006 [last accessed 2006 Nov 29]; Available from: http://ccasanet.vanderbilt.edu/index.php

[16] EarthTools. [last accessed 2006 Dec 3]; Available from: http://www.earthtools.org/

[17] Getty Thesaurus of Geographic Names Online. [last accessed 2006 Dec 3]; Available from: http://www.getty.edu/research/conducting_research/vocabularies/tgn/

Address for correspondence

Stephany Duda, 4th Floor Eskind Biomedical Library, 2209 Garland Ave, Nashville, TN 37212, USA; stephany.duda@vanderbilt.edu

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Framework for Clinical Data Standardization Based on Archetypes

Jose A. Maldonado[a], David Moner[a], Diego Tomás[a], Carlos Ángulo[a], Montserrat Robles[a], Jesualdo T. Fernández[b]

[a] *Biomedical Informatics Group, ITACA Institute, Technical University of Valencia, Spain*
[b] *Departamento de Informática y Sistemas, University of Murcia, Spain*

Abstract

Standardization of data is a prerequisite to achieve semantic interoperability in any domain. This is even more important in the healthcare sector where the need for exchanging health related data among professional and institutions is not an exception but the rule. Currently, there are several international organizations working on the definition of electronic health record architectures, some of them based on a dual-model approach.

We present both an archetype modeling framework and LinkEHR-ED, an archetype editor and mapping tool for transforming existing electronic healthcare data which do not conform to a particular electronic healthcare record architecture into compliant electronic health records extracts. In particular, archetypes in LinkEHR-ED are formal representations of clinical concepts built on a particular reference model but enriched with mapping information to data sources which define how to extract and transform existing data in order to generate standardized XML documents.

Keywords:

medical records, archetypes, standardization, data translation, information systems

Introduction

Health care is a sector where the need for sharing information is the norm rather than the exception. However, the health data of one patient is usually scattered among the different health facilities where she/he has been attended. This leads to distributed and heterogeneous data resources, all of them containing health data, making the exchange of data across systems and organizations very difficult. This situation has created a large gap between the potential and actual value of the information content of electronic health records (EHR). Closing this gap by making efficient use of the health data held by these systems, could improve significantly patient care, clinical efficiency and empower research activities.

Due to the special sensitivity of medical data and the wide range of ethical and legal constraints, data exchange must be done in a meaningful way, avoiding all possibility of misunderstanding or misinterpretation. The faithful communication in EHR crucially depends on the standardization of the EHR architecture (EHRA) used to communicate data. Currently there are several international organizations [1] working on the definition of an EHRA. Health Level 7 [2] supports two message protocols: HL7 Version 2 and HL7 Version 3. The technical committee 251[3] of the European Committee for Standardization is working on a new full European Standard and future ISO norm for the communication of the EHR called EN13606 [4][5]. The OpenEHR foundation [6] maintains an EHR architecture designed to support the constructions of distributed, patient-centered, life-long, shared care health records.

Due to the complexity and the continuous evolution of the health domain a new approach for the development of EHR systems has been proposed. The methodology is known as the dual model methodology which is based on a clear separation between information and knowledge. The former is described through a Reference Model (RM) that contains the basic entities for representing any entry in an EHR. The latter is based on archetypes, which are formal definitions of clinical concepts in the form of structured and constrained combinations of the entities of a RM. Examples of Dual Model EHRA are CEN/TC251 EN13606 and openEHR.

Currently, in most organizations there is a vast amount of health related data that does not conform to an EHRA that need to be converted into standardized EHR extracts in order to be exchanged with other organizations. In this paper we address the problem of how to use archetypes to make public these legacy data in the form of standardized EHR. We argue that archetypes are suitable for this purpose. First, archetypes allow the formal description of the semantics of legacy and un-standardized health data. On the other hand, by mapping the structure of archetype definitions to the elements of the data sources, it is possible to generate normalized EHR extracts compliant with the underlying RM.

Dual model approach

The dual model approach distinguishes a reference model and archetypes.

A reference model is an object oriented model that is used to represent the generic and stable properties of health record information. It comprises a small set of classes that define the generic building blocks to construct EHRs. It

specifies how health data should be aggregated to create more complex data structures and the context information that must accompany every piece of data in order to meet ethical and legal requirements.

An archetype is a formal definition of a distinct, domain-level concept in the form of structured and constrained combinations of the classes of the RM. What is important is that for each concept in the domain we want to use, a definition can be developed in terms of constraints on structure, types, values, and behaviors of RM classes. Basically, archetypes are means for providing semantics to data instances that conform to some reference model by assuring that data obey a particular structure (combination of classes of the reference model) and satisfy a set of semantic constraints. This is achieved by linking data structures and content to knowledge resources as terminologies and ontologies. Their principal purpose is to provide a powerful, reusable and interoperable way of managing the creation, description, validation and query of EHRs.

ADL (Archetype Definition Language) [7] is a formal language for expressing textually archetypes developed by OpenEHR that has also been adopted by CEN. ADL is a textual language for specifying constraints on data instances of an RM in a formal way. An archetype expressed in ADL is composed of four main parts: header, definition, ontology and revision history. The header section contains the archetype metadata. In the definition section is where the modeled clinical concept is represented in terms of a particular RM class. This description is built by constraining several properties of classes and attributes, such as existence, occurrences or cardinality or by constraining the domain of atomic attributes. It is important to notice that in this section only those entites that need to be constrained should appear. The ontology section is where the entities defined in the definition section are described and bound to terminologies. Finally the revision history section contains the audit of changes to the archetype.

Archetype modeling

The current ADL specification is not precise enough regarding archetype specialization; this hinders a precise understating of archetypes and their implementation. As a consequence our first work was to define a precise archetype modeling framework as a prerequisite for implementing tools providing enhanced support for archetypes. Since our main concern is the generation of standardized EHR extracts we have focused on the data definition facet of archetypes. We view the definition section of archetypes as a database schema which describes subset of instances of a class from a particular RM. In this section we present briefly the data model used to describe data instances (EHR extracts) and the schemas that formalize the definition section of archetypes, the composition and specialization of archetypes and the relationship between a business concept and its archetypes and between an archetype and its instances. An extended formal definition and additional examples can be found in [8], due to lack of space these are omitted here.

Data model

Archetypes impose a hierarchical structure to the EHR, therefore we have chosen a data model based on trees with labeled nodes to formalize their data instances. It is similar to the models presented in [9,10] but our data model supports both ordered and unordered nodes. Although archetypes do not impose an order on class attributes or attribute values, it is possible to define ordered multi-valued attributes such as attributes whose value is a list.

The representation of data instances (in our context EHR extracts compliant with a RM) is straightforward. Each object is described by a data tree. The root node is labeled with the class name and has one child for each attribute. The children are labeled with the attribute names and each of them has one child labeled with the corresponding type (class) name. This mechanism is repeated iteratively. Atomic values are represented by a leaf node labeled with a value.

Schema model

For the representation of the definition section of archetypes we have developed a type system that allows the specification of decidable sets of data trees, i.e., in our context set of instances of the RM. We assume the existence of a finite set of primitive types C, i.e., the set of primitive types defined in the reference model, an infinite set of type variables T disjoint with C and an infinite set P of label predicates.

Definition 1. A multiplicity list is a regular expression of the form:

$$t_1^{(l_1:u_1)} \ldots t_n^{(l_n:u_n)} \tag{1}$$

Where n≥1, t_iÎCÇT, l_i£u_i and $t_i^{(l_i:u_i)} = t_i^{l_i} \mid t_i^{l_i+1} \mid \ldots \mid t_i^{u_i}$

Definition 2. A constrained multiplicity list (CML) is language definition expression of the form:

$$\left(t_1^{(l_1:u_1)} \ldots t_n^{(l_n:u_n)}\right)^{|l:u|} \tag{2}$$

Where $t_1^{(l_1:u_1)} \ldots t_n^{(l_n:u_n)}$ is a multiplicity list, l≤u,

$$\sum_{i=1}^{n} l_i \geq u \quad \text{and} \quad \sum_{i=1}^{n} u_i \geq l \; .$$

Intuitively the language generated by a constrained multiplicity list is composed by all the word defined by the regular expression whose length (number of symbols) is between l and u inclusively. As an example consider the CML $\left(A^{(1:2)}B^{(1:3)}\right)^{|3:4|}$ which defines the language {ABB, ABBB, AAB, AABB}.

We formalize the definition section of archetypes as a set of type definitions (a schema).

Definition 3. A type definition has either the form $t:=l_t<r_t>$ or $t:=l_t\{r_t\}$. Where t is a type name, l_t is a label predicate and r_t is a CML over the set the type names and primitive types.

A type definition has two parts. The first is a label predicate that describes the valid labels of nodes. The second is a CML that is used to describe the sequence of children that a node may have. An expression of the form t:= l_t<r_t> defines a sets of data trees t whose root node is labeled with a label that satisfies l_t, its children are ordered and are describe by r_t. On the other hand, an expression of the form t := l_t{r_t} defines a set of data trees t whose root node is labeled with a label that satisfies l, its children are unordered and at least one permutation is described by r.

Definition 4. A schema is a set of type definitions, one of which must be declared to be the root type, i.e. whether roots of data trees can be assigned this type.

We need to define the semantics of a schema, i.e. the set of data tree that it models. Intuitively, a data tree D conforms to a schema S if it is possible to assign to every node d of D a type T_i from S, and d satisfies the label predicate and the CML of T_i.

Both RM and definition sections of archetypes can be modeled as a schema. Each type in the schema represents an RM/archetype entity, i.e. a class or an attribute, and the CML represent its structure. The label predicate describes the name of the entity or the valid domain of atomic attributes. As an example consider the class *PERSON* from Figure 1. It can be formalized by the type definition:

$$t_{\text{PERSON}} := \text{is\_}\ person(X)\left\{\left(t^{(1:1)}_{addresses}\, t^{(1:1)}_{name}\right)^{|2:2|}\right\}$$

where is_person(X) is a unary predicate that is true only when X is equal to the string "person" and $_{\text{t}}$addresses and t_{name} are the types that, respectively, model the attributes addresses and name.

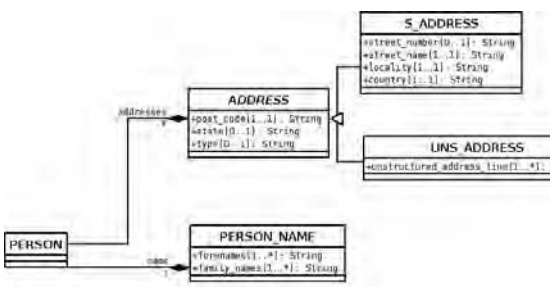


Figure 1 - Example of OO model

Let see how different types of archetype constraints are modeled with our type system. Existence constraints of attributes can be easily expressed by the CML attached to the type that models the attribute. Cardinality and occurrence constrains are mutually related. Actually, it only makes sense to constrain the occurrence of node objects that are inside a block introduced by a multi-valued attribute. Cardinality constraints can be modeled by the length constraint of CML while occurrence constraints by regular expressions. Alternative constraints for mono-valued attributes can be expressed by a CML whose length constraint is equal to |0..1| if the attribute is optional or to |1..1| otherwise, and the regular expression is of the form $t_1^{(0:1)}$.... $t_n^{(0:1)}$ where t_1,..t_n are the alternative types. Finally, domain constraints on primitive types are straightforwardly modeled as label predicates. As an example, let us consider the following ADL expression:

PERSON[at001] matches {
addresses cardinality matches {1..} matches {*
ADDRESS[at002] occurrences matches {0..1}
matches {...}
ADDRESS[at003] occurrences matches {0..}*
matches {...}
}

Which can be modeled by the following set of type definitions:

$$PERSON[at001] := is\_PERSON(X)\left\{\left(t^{(1:1)}_{addresses}\right)^{|1:1|}\right\}$$

$$t_{addresses} := is\_addresses(X)\left\{\left(t^{(0:1)}_{ADDRESS[at002]}.t^{(0:*)}_{ADDRESS[at003]}\right)^{|1:*|}\right\}$$

$$t_{ADDRESS[at002]} := is\_ADDRESS(X)\{...\}$$

$$t_{ADDRESS[at003]} := is\_ADDRESS(X)\{...\}$$

Archetype specialization

Archetypes can be defined by further constraining other archetype, i.e., by specialization, in order to obtain a more adequate or fine grained representation of a clinical concept. An archetype is specialized by providing narrower constraints on data. The overall idea is that all data instances that conform to the more specialized archetypes also conform to the more general but there can be data instances of the more general archetype which are not data instances of the specialized archetype. Multiple inheritance is not allowed. We formalize the inheritance relationship between archetypes by means of the subsumption relation [10], i.e., an archetype A is more general than archetype B if A subsumes B. The subsumption relation is also used to formalize the relationship between business concepts and archetypes, we say that an archetype A specialized a business concept B if B subsumes A. What makes the proposed subsumption relation interesting is that it not only captures the containment relationship between the set of data instances defined by two archetypes, but also captures some of the structural relationship between node objects from both archetypes.

Intuitively, subsumption is based on defining a mapping function () between types, on inclusion between both label predicates and languages defined by the CML of these types. All these can be translated to archetype specialization. Subsumption mappings specify specialization relationships between node objects and attribute object from parent and child archetypes. This is compatible with the syntactical rules used in ADL to specify the specialization of node object. In ADL specialization of coded concepts, e.g. PERSON[at0001], is indicated by using the same root, followed by an extension, e.g. PERSON[at0001.1]. Note that this defines partly the subsumption mapping, e.g. (PERSON[at0001.1]) = PERSON[at0001]. On the other hand,

inclusion of label predicates assures both that only class and attribute names from the reference model are used in archetypes, and the domain of atomic attributes in specialized archetype are a subset of the domain in the parent archetype. Finally, CML controls that super-types and subtypes have the same structure.

As stated before, RM and the definition section of archetypes can be modeled by the proposed type system. One interesting consequence is that "archetypable" classes of RMs can be considered as archetypes (from a data definition point of view). Therefore, the same logic can be applied both to the specialization of an exiting archetype or to the definition of a new one by constraining a RM class. This brings about the possibility of building flexible archetype editors capable of working with several RM or different version of the same reference model. This possibility has been explored in LinkEHR-Ed.

Mapping and data translation

Since the health data to be made public resides in the underlying data sources, it is necessary to define some kind of mapping information that links entities described in the archetype to data elements in data repositories (e.g. tables and attributes in the case of relational data sources). In the health care domain very few generic EHR data transformation efforts exist. Commercial tools with limited capabilities exist but they are mainly focused on the generation of HL7 v2.x or EDI messages and none of them supports archetypes. Furthermore, definition sections of archetypes can not be represented by XML schemas. Thus, current tools for data translation for XML schemas can not be used for this purpose.

In our scenario an archetype is considered to be a view that provides abstraction in interfacing between the data sources and the RM used to communicate the EHR extracts. Since EHR extracts have an inherent hierarchical structure, we have chosen XML as canonical data model, i.e. data sources are viewed a XML documents.

There exists two kinds of mappings: atomic attribute mappings and class mappings. Atomic attribute mappings define how to obtain a value for an atomic attribute of an archetype by using a set of values from the data sources. For this purpose rules relating a set of source paths to an archetype path identifying an atomic attribute are used. This kind of mappings preserves node paths from the root, i.e. the node contexts. It is possible for an archetype attribute to have more than one mapping and it also possible to utilize functions involving more than one source path (for instance the addition of the value of two source attributes). A rule may also contain a condition specifying the subset of values of the data source that can be used to compute values. Obviously, there must be at least one of this kind of mapping for each mandatory atomic attribute.

On the other hand, for each constrained class there exists a class mapping which contains both the query to be used to retrieve all the data necessary for generating data instances and the set of attributes that identify univocally the class instances. The combination of both components allows the conversion from source data to XML documents compliant with the RM. The query extracts the relevant information and for each different combination of values of the identification attributes a new instance of the class is generated.

Archetype designers are responsible of defining the atomic attribute mappings and the system tries to generate [11] from them a set of candidate class mappings by taking into account the structure of the RM entity being used, the archetype constraints and the integrity constraints of data sources. This approach alleviates the work of defining how to populate archetypes since it is easier for the designer to indicate which data elements of the data sources are relevant to a certain archetype attribute, rather than to specify the possible complex query required to extract and transform all the relevant information. As a result, an XQuery expression is generated, which transforms the XML view of the data source into an XML document that satisfies the constraints imposed by the archetype and at the same time is compliant with the RM.

Results

LinkEHR-Ed (http//pangea.upv.es/linkehr) is a visual tool implemented in Java under the Eclipse platform which allows the edition of archetypes that can be based on different RMs, mapping specification between archetype and data sources and the generation of data conversion scripts which generate XML documents compliant with the RM. Figure 2 describes the overall edition process. LinkEHR Ed is composed of four main components.

Reference model manager. In LinkEHR-Ed new reference models expressed as a W3C XML Schema can be imported at any time. Note that the XML Schema is supposed to describe the XML documents compliant with the RM. The import process generates archetype representations of the RM archetypable classes that will be immediately available as basis for the creation of archetypes by means of specialization. Currently, many of the characteristic of W3C XML schemas are supported, such as all the data types, name spaces, imports and includes (reference models can be defined by several files) and several structures such as complex and simple types, elements, attributes, inheritance by extension and restriction, sequence, choice, all, attributes, patterns and groups and their respective facets. Two reference models have been imported and used successfully: EN13606 and OpenEHR. The XML schema of EN13606 schema has been developed by us from the UML model due to the lack of an official one. For OpenEHR the official schema (actually a set of XML schemas) has been used [6]. This feature has been very useful in keeping in pace with their evolution without modifying a single line of code.

Semantic validation module. In LinkEHR-Ed only the logic that guides archetypes specialization, which is based on the subsumption relationship described before, is hard coded. This module, given an archetype, tests whether the archetype is valid with respect the RM entity that it constrains (i.e., it is subsumed by the RM entity) or in the case that there exists a parent archetype it test that its constraints are narrowed that those of the parent archetype (i.e., it is subsumed by the parent archetype).

Mapping module. It is in charge of managing data sources schemas, attribute mappings and the generation of candidate class mappings. Given a candidate class mapping, it generates an XQuery expression that outputs standardized XML EHR extracts satisfying all the constraints stated in the archetype.

Visual interface. LinkEHR-Ed provides two different interfaces, one for the health domain experts and one for the information technologies professionals. On the one side, the Health domain expert will be in charge of archetype definition and so they must have some knowledge about the RM they are working with. But the main idea for this perspective is to hide the underlying complexity of the system and the Dual Model architecture logic involved in the designing of an archetype as we can not presuppose any computer management skills for this expert. A set of available constraints and applicable restrictions is provided during design time. On the other side is the information technologies expert, who knows the structure of data sources of the organization and his role is to map the archetype definition tree nodes to them. A mapping definition interface fills nearly all this edition perspective. It is composed by a graphical representation of the archetype definition tree and a graphical representation of the diverse data sources available. Users can then add or modify mapping transformations between elements of both representations. In this case, LinKEHR-Ed can be seen as visual mapping and data translation tool.

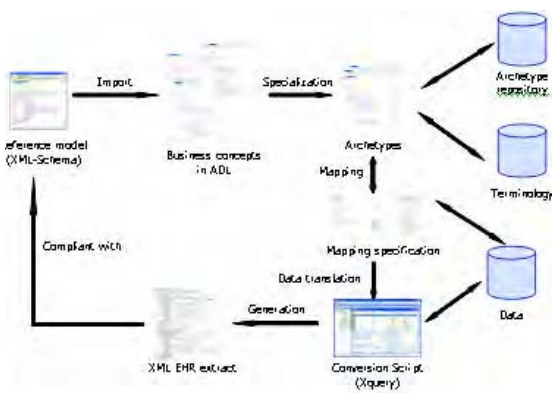


Figure 2 - Archetype edition and mapping process

Conclusion

In this paper, we have presented an archetype modelling framework and LinkEHR-Ed, a tool that allows the utilization of archetypes for upgrading already deployed systems in order to make them compatible with an EHR architecture standard. The overall objective is to maintain in-production systems and applications without any changes while providing a means for making public clinical information in the form of standardized EHR extracts, hiding technical details, location and heterogeneity of data repositories. Therefore, we use archetypes as a semantic layer over the underlying databases associating them with domain specific semantics. LinkEHR-Ed combines in an easy manner the formal representation of knowledge of a heath domain expert, represented by an archetype, with the mapping information to clinical data sources for semantic interoperability and standardization purposes.

Acknowledgments

This work was supported in part by the Spanish Ministry of Education and Science under grant TSI2004-06475-C02.

References

[1] Eichelberg M, Aden T, Riesmeier J, Dogac A and Leleci GB. A survey and analysis of electronic healthcare record standards. ACM computing Surveys 2005: 37(4): 277-315.
[2] Health Level 7: www.hl7.org
[3] European Committee for Standardization. Technical Committee on Health Informatics: www.centc251.org.
[4] European Committee for Standardization. Health informatics-Electronic health record communication- Part 1: Reference model. Draft European Standard for CEN Enquiry prEN13606-1, 2006.
[5] European Committee for Standardization. Health informatics-Electronic health record communication- Part 2: archetypes. Draft European Standard, prEN13606-2, 2006.
[6] OpenEHR foundation: http://www.openehr.org.
[7] Beale T, Heard S. Archetype Description Language 1.4 (ADL). The OpenEHR Foundation.
[8] Maldonado JA. Historia Clínica Electrónica Federada Basada en la Norma Europea CEN/TC251 EN13606 PhD. Dissertation. Technical University of Valencia, 2005.
[9] Beeri C, and Milo T. Schemas for integration and translation of structured and semi-structured data. Proceedings of the 7th International Conference on Database Theory (ICDT'99) , pp. 296-313, 1999
[10] Kuper GM, Simeon J. Subsumption for XML types. Proceedings of the 8th International Conference on Database Theory (ICDT'01), pp. 331-345, 2001.
[11] Popa L, Velegraki Y, Miller RJ, Hernández MA, Fagin R. Translating web data. Proceedings of the 28th VLDB Conference, pp. 598-609, 2002.

Address for correspondence

Jose Alberto Maldonado, Ph.D.
Biomedical Informatics Group. ITACA Institute.
Technical University of Valencia
Valencia 46022, Spain
jamaldo@upv.es

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Reaching Standards for Dissemination: A Case Study

Helen Christensen[a] **and Kathleen Griffiths**[b]

[a b] *Centre for Mental Health Research, The Australian National University, Canberra, Australia*

Abstract

At what stage are web applications ready for dissemination? A set of standards for efficacy, effectiveness and readiness for dissemination of (prevention) interventions has been developed by the Society for Prevention Research (SPR) (Flay et al., 2005)[1]. In this case study paper, we examine these standards criteria with reference to MoodGYM, an automated web application designed to reduce depression symptoms using cognitive behaviour therapy training. We examine evidence for its efficacy, its effectiveness in real world situations, and its capacity to meet recommended standards for dissemination. We conclude that MoodGYM has substantially met the standards of evidence required for dissemination. This successful application of the SPR standards to the evaluation of a web application, suggests that these prevention standards might be usefully applied to web/Internet context. Web applications might be assessed for dissemination by the International Society for Research into Internet Interventions (ISRII), or some other professional organization to assist policy makers in making decisions about the funding, adoption and promotion of applications.

Keywords:

internet, depression, efficacy, effectiveness, implementation

Introduction

Over the last 10 years, web applications have proliferated in the areas of mental health and substance abuse. This growth in the development and research of web applications is illustrated below using a review of the frequency of research papers identified using search terms for 'web therapy' trended for year from 1996 to 2003 (Figure 1)[2]. Web applications have been found to be both feasible and effective in reducing common mental health disorders such as anxiety and depression [3,4].

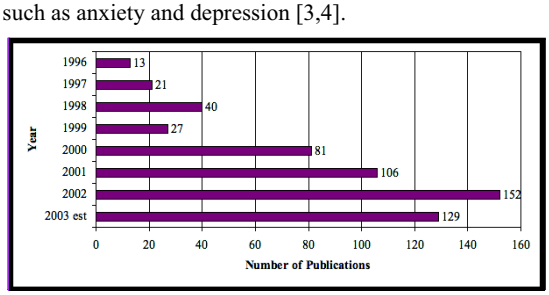


Figure 1 - Web therapy papers as a function of year

Policy makers and governments are now seriously interested in these applications, but they require guidelines for determining which applications are likely to yield positive public health outcomes, and which should be supported financially. One set of standards, which has been recently developed by the Society for Prevention Research (SPR), may be suitable, not only for prevention programs, but also for web applications designed as early intervention programs. Flay et al., 2005 [1] provide a set of 47 standards for efficacy, effectiveness and dissemination, which may prove to be useful in assessing web applications in terms of their success and their potential for implementation and wider dissemination. Currently, most web interventions are disseminated using marketing strategies, and their uptake is determined by market forces, consumer preferences and other factors. Little is known about the benefits or potential adverse effects of these interventions. Computer-based CBT programs have recently been endorsed by government bodies such as the UK-based National Institute for Health and Clinical Excellence, and it is likely that such bodies will similarly endorse Internet programs in the future. This endorsement will also be expected to drive demand.

In this case report paper we describe the research taken to establish the efficacy and effectiveness of a web intervention – MoodGYM –an application designed to reduce symptoms of depression in community users. We ask (i) whether it meets the SPR criteria for efficacy, effectiveness or dissemination, and, if not, (ii) what additional research might be required. We also (iii) comment on the potential usefulness of SPR standards in the evaluation of web applications.

In the SPR standards, the demonstration of *efficacy* requires the demonstration of positive treatment outcomes under optimal conditions: The following are required: clear description of the project, and clearly stated outcomes [decrease in depression symptoms, over a (minimum) six month period, using psychometrically sound measures, with one form of data collection not susceptible to demand characteristics]. The strongest form of control design should be used where possible (RCT), and the sample requires definition, [how it was obtained], the statistical analysis should be unbiased [intention to treat, adjustment for pretest differences, differential attrition adjustment, correction for experiment wise error]. At least two different high quality studies/replicates must exist. Desirable standards include process variable measure-

ment, and measures of side effects [See Flay for more details [1]].

Effectiveness studies focus on the importance of the project in real-world situations, implementation procedures, fidelity and adaptation (p. 153). Standards include the provision of manuals and technical support, the delivery of the project, as in the real world, with a clear statement for whom the intervention works. The level and integrity of the intervention, in addition to the level of engagement is needed, including measures of adherence, and or involvement in both the control group and the treatment group. RCT designs are required, although often more difficult to implement. The sample investigated should be the sample for which the application is targeted. A desirable standard is replication with different responses. The practical importance or public health impact of studies requires assessment. Examples include cost effectiveness, effect size measures and percent relative change. Replication is also required.

Programs worthy of *dissemination* must meet all requirements for efficacy and effectiveness, as well as additional standards for dissemination such as training of staff, and the provision of project support materials. The criteria include: the provision of provider materials, and evidence that the program can be implemented with fidelity, evidence for scalability, and exact estimates of cost. Monitoring and evaluation tools must be available to providers (p.167). A desirable standard is "a clear statement of the factors that are expected to assure the sustainability of the program or policy once implemented".

Materials and methods

Checklist and raters

A checklist of standards was constructed based on the Flay et al. criteria. We reduced the number of SPR criteria from 47 to 19 in order to summarize outcomes within the space of this report. To achieve this some sub-criteria, definitions and some outcome criteria were not included. All published or completed research reports of the MoodGYM website were collected. Two raters reported whether the MoodGYM site reached standards for efficacy, effectiveness and potential for dissemination. Any difference between raters was resolved by references to criteria and published reports.

Results

Efficacy Standards

All of the required standards for efficacy were met by the MoodGYM application, although one of two desirable standards was not met. The MoodGYM website provides a clear indication of its content and aims (http: moodgym.anu.edu.au). An efficacy trial published in 2004 compared the website to an attention placebo control condition (health coach) and to a psychoeducation site (BluePages) using participants from a community sample with elevated depression symptoms, recruited by survey using addresses from the Australian Electoral Roll [6,7]. An intention to treat analysis demonstrated that the effects of the website persisted at 12 months follow-up [8]. Figure 2 illustrates that both MoodGYM and the psychoeducation site were associated with better outcomes immediately, and at 6 and 12 months. A second study has now been undertaken of the use of the website in 29 schools around Australia, using a clustered randomized controlled trial design. Although not complete (two of the 29 schools currently in the intervention phase of the project) and the long term follow-up is yet to be collated, there is evidence of a small effect for the (Universal) intervention in students in Years 9 or 10 (mean age 14.3 yrs). An additional efficacy trial has been completed in Norway which found that MoodGYM (in combination with BluePages) reduced depression symptoms compared to a control condition [9]. A New Zealand trial is about to commence. The day-to-day implementation of these trials is under the control of researchers external to developers. The MoodGYM site measures a number of variables that might mediate the effect of the website on the reduction of depression symptoms. The major measure is the Warpy Thoughts Questionnaire. This questionnaire was developed by the Centre for Mental Health Research (CMHR), and has been validated against other scales of dysfunctional thinking. The users of the site record scores on this questionnaire three times over the course of the intervention. In the first instance, this permits the calculation of correlations between symptom change and thinking style [10]. A second desirable criterion for the efficacy of a trial is a measure of side effects. MoodGYM does not explicitly collect information on side effects. This represents a common omission in psychological therapy studies, but one that requires rectification.

Effectiveness standards

The 8 criteria for effectiveness are outlined in Table 1.

The MoodGYM application provides a user manual and a clinician manual which can be purchased from the CMHR website at http://www.anu.edu.au/cmhr/shop.php. The user manual provides basic information on access to the site and trouble shooting for technical problems. The clinician manual provides week-by-week instructions, which allow clinicians to implement the self-help program with clinical populations, and normative data on website users.

Table 1 - SPR Criteria

Simplified Standards Criteria (Society for Prevention Research)	
Criteria for Efficacy	**Rating**
Clear description of project (R)	**Yes**
One high quality RCTs (ITT, sound measures) (R)	**Yes**
Long term follow-up (at least 6 months) (R)	**Yes**
Additional long term high quality RCT (R)	**Yes**
Process variable measurement (D)	**Yes**
Measures of side effects (D)	**No**
Criteria for Effectiveness	**Rating**
Provision of manuals and technical support (R)	**Yes**

Simplified Standards Criteria (Society for Prevention Research)	
Evidence of delivery in the real world to target (R)	**Yes**
Level of exposure, engagement and adherence (R)	**Yes**
Two RCTs in identified settings (R)	**Current**
Replication with different samples (D)	**Current**
Cost effectiveness (D)	**Yes**
Dose analyses (D)	**Yes**
Demonstration of public health impact (R)	**No**
Criteria for Dissemination	**Rating**
Provision of provider support and fidelity (R)	**Yes**
Evidence for scalability (R)	**Yes**
Clear cost information available (R)	**No**
Monitoring and evaluation tools (R)	**Yes**
Factors to assure sustainability (D)	**No**

Note: R=Required, D=Desirable.

One of the advantages of the MoodGYM application is that it is both automated and self-directed. As such it requires less intensive training for those who implement it, especially in comparison to face-to-face interventions. Technical support is available via an email address posted on the site. One potential problem with the application when used outside a research trial that incorporates interviewer-instigated tracking is its failure to engage all registrants for the complete program of five modules. Studies of usage [11] and mental health change suggest that treatment benefits are obtained once two or more modules are accessed. As many spontaneous users engage with the site for more than this number of modules, we believe that adherence level is relatively acceptable for a proportion of users. Moreover, the interpretation of drop out from web applications, where there is no expectation held by most users that they are required to engage for hours of therapy online, is less straightforward than conventional drop out in non-Internet RCTs. Many users may well attain what they set out to achieve in two or three modules rather than the full six.

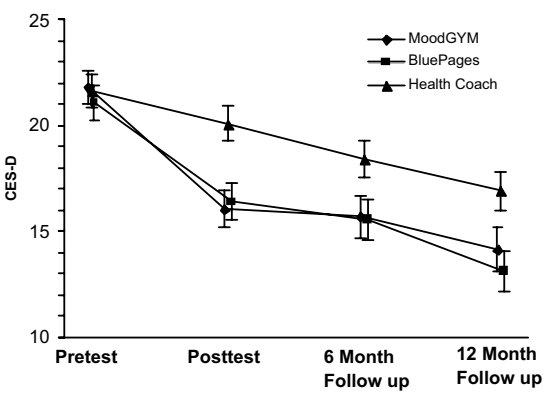


Figure 2 - Least squares means for CES-D estimated using time by condition model with unstructured covariance matrix. Note: CES_D: Higher scores indicate higher levels of depression symptoms [8]

MoodGYM has not yet met the standard of two randomized controlled effectiveness trials, although comparisons of spontaneous users and those in our first randomized controlled trial were positive [12] and the schools trial, mentioned earlier, is both an efficacy and effectiveness trial as it is implemented in real world conditions. An additional study is currently in progress, evaluating MoodGYM's effectiveness in a general practice setting. Results from these evaluations may well provide the necessary evidence of effectiveness.

A study of the cost effectiveness of MoodGYM has also been conducted. There are few cost-effectiveness studies of Internet programs, although one computerized therapy program used in general practice settings has been assessed to date [See Kaltenthaler et al, 2006] [13]. Cost effectiveness analysis is particularly interesting for Internet programs given the scalability of Internet applications and their very low marginal costs [14]. The cost effectiveness study indicated that MoodGYM (M-CBT) was dominant over conventional treatments consisting of a general practitioner-delivered anti-depressant medication (GP-DRUG) and psychiatrist led cognitive behaviour therapy (P-CBT). The above assumes a similar level of efficacy for therapist administered CBT, for anti-depressant medication and for Internet CBT. A separate randomized controlled trial of community users found that different components of the program were associated with improved outcomes [15], but, in particular, that extended CBT component of the program was associated with large effects.

A final effectiveness criterion is the demonstration of the public health benefit of the application. This has not yet been conducted.

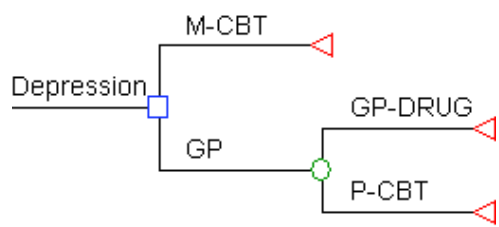


Figure 3 - Cost effectiveness comparisons. M-CBT=MoodGYM, P-CBT=Psychiatrist CBT and GP-DRUG=General Practitioner Anti-depressant medication[14]

Dissemination standards

Of the identified five standards in Table 1, MoodGYM meets three. These are fidelity, scalability, and monitoring tools. Fidelity (as in the provision of the exact user experience of the website) is guaranteed in automated web applications (although there is no guarantee that each user will access all pages in a similar way, or even receive the same pages if the application is personalized or tailored). The site is also scalable, with over 160,000 registrants, and up to 14,000 visitors a month. Moreover, organizations or researchers are able to download SPSS formatted data

from their specific trials or organisations, to allow monitoring and evaluation. MoodGYM is currently offered free of charge to users on the World Wide Web. The costs of the development and maintenance of the MoodGYM website are borne by our research group within a University. Clearly, this model of funding the site is not sustainable. The capacity to provide the service will cease when the research agenda has concluded or grants are completed. This as-yet-to-be developed model for sustainability represents the most serious threat to meeting the criteria for standards for dissemination.

Discussion

Usefulness of the efficacy, effectiveness and dissemination standards

There may be merit in examining web applications with reference to the standards developed by the Society for Prevention Research. In this study, the areas in need of further research and development to reach standards for the MoodGYM site were relatively easily identified. For example, we identified the need to investigate and include side effect assessment in our online MoodGYM satisfaction questionnaire (available from the menu), and to consider methods to estimate the public health impact of our intervention more systematically. The review of the web application against the standards criteria suggests possible areas for the refinement of the criteria. For example, it is sometimes difficult to distinguish for automated applications such as MoodGYM whether an intervention study should be classified as an efficacy study or an effectiveness study. Some aspects that differentiate an efficacy from an effectiveness trial of a psychological intervention may be less relevant in an automated application. For example, in contrast to human-delivered therapy, automated delivery of web therapy ensures fidelity of treatment content whether or not the intervention is delivered in a research or specialist environment. Another potential area in which the SPR might be profitably refined relates to the definition of an acceptable 'control' group. Although there is a large 'effectiveness' trial of the MoodGYM site which was conducted entirely online [14], it does meet the specified criteria for an effectiveness trial, as the criteria state that the comparison group must be one of waitlist, attention placebo or TAU. In this trial, which is offered directly to users, the control condition was a one-module version of the site. Perhaps effectiveness trials could be permitted to include a minimal intervention condition as a control.

Advantages of web applications

This exercise of comparing MoodGYM against the standards criteria reveals the strengths of web applications compared to non-Internet interventions. For example, because of the capacity for tracking, web applications should be particularly strong in comparison to face-to-face interventions in the following areas: determining dose responses, process evaluation and the collection of routine implementation and evaluation data. Demand characteristics might be reduced by self-report data submission. Websites can be readily disseminated, and dissemination will occur regardless of whether a website meets the SPR or other quality standards. This underscores the importance of researchers and policy makers introducing standards like these to rate the readiness of applications for dissemination.

Major problems for sustainability

A major problem we identified for the MoodGYM site was lack of sustainability. A number of methods to financially sustain the site might include a user-pays system, an employer-pays system for the workforces of large organizations, or a health company pays system delivered via health maintenance organizations. However, as developers of the site, we preferred to offer the site free of charge to individuals with mental health problems with the aim of maximising the potential public health impact of the application. To offer cost-free access, the site needs to gain support from government. The mechanisms within government to offer this support must be in place, and this itself may be a stumbling block. Because of the international nature of website applications, it is difficult to determine how users from different countries might contribute. For example, the MoodGYM site is regularly accessed by users of the UK National Health Service. There may be some capacity here for governments to agree to contribute to costs of approved sites. However, at this stage, in Australia, the sustainability for public health websites is yet to be convincingly demonstrated.

Conclusion

The standards of evidence for prevention research might be usefully applied to web/internet interventions. Web applications might be assessed for dissemination by the International Society for Research into Internet Interventions (ISRII), or some other professional organization to assist policy makers in making decisions about the funding, adoption and promotion of applications.

Acknowledgments

The authors are the developers of the MoodGYM site. The site is supported by e-hub at the Australian National University.

References

[1] Flay BR, Biglan A, Boruch RF, Castro FG, Gottfedson D, Kellam S, Moscicki E, Schinke S, Valentin, JC, and Ji, P. Standards of evidence: Criteria for efficacy, Effectiveness and Dissemination. : Prevention Science 2005;6 (3): 151-175.

[2] Wantland DJ, Protillo CJ, Holzemer WL, Slaugher R, McGhee EM. The effectiveness of web-based vs non-web-based interventions: A meta-analysis of behaviour change outcomes. JMIR 2004; 6(4):e40.

[3] Griffiths KM & Christensen H. Review of randomized controlled trials of Internet interventions for mental disorders and related conditions. Clinical Psychologist 2006; 10(1), 16-29.

[4] Griffiths KM, Farrer L, Christensen H. Clickety-click: e-mental health train on track. Australasian Journal of Psychiatry, in press.

[5] National Institute for Health and Clinical Excellence. Computerised cognitive behaviour therapy for depression and Anxiety. Review of Technology Appraisal 51.

February2006. http://guidance.nice.org.uk/TA97/guidance/word/English

[6] Christensen H, Griffiths KM, Jorm AF. Delivering depression interventions using the Internet: positive results from a large randomised controlled trial. British Medical Journal 2004, 328:265. Online 23-01-04.

[7] Griffiths KM, Christensen H, Jorm AF, Evans K,Groves C.. Effect of web-based depression literacy and cognitive-behavioural therapy interventions on stigmatising attitudes to depression: a randomised controlled trial. British Journal of Psychiatry 2004;185:342-349.

[8] Mackinnon AJ, Griffiths KM, Christensen H. Twelve month followup of the CBT site. Submitted for publication.

[9] Lintvedt OK Sorensen K Ostvik AR. Evaluating the effectiveness of an Internet based intervention preventing depression. A randomized controlled trial. 2006. Unpublished Thesis for Psychology Degree. Department of Psychology. University of Tromso, Tromso, Norway.

[10] Parslow R, Christensen H, Griffiths K, Groves C. The Warpy Thoughts Scale: a new 20 item instrument to measure dysfunctional attitudes: Factor analysis and associations with mental health messages. Cognitive Behaviour Therapy 2006;Vol 35(2) 106-116

[11] Christensen H, Griffiths KM, Korten AE. Web-based cognitive behaviour therapy (CBT): analysis of site usage and changes in depression and anxiety scores. Journal of Medical Internet Research 2002;4(1):e3.

[12] Christensen H, Griffiths K, Korten A, Brittliffe K, Groves C. A comparison of changes in anxiety and depression symptoms of spontaneous users and trial participants of a cognitive behavior therapy website. Journal of Medical Internet Research 2004;6 (4):e46..

[13] Kaltenthaler E, Brazier J, De Nigris E, Tumur I, Ferriter M, Beverley C, Parry G, Rooney G and Sutcliffe, P. Computerised cognitive behaviour therapy for depression and anxiety update: a systematic review and economic evaluation. Health Technology Assessment, 2006; 10 (33).

[14] Butler, J, Griffiths, KM, Bartlett, L and Christensen, H. Direct costs of Internet based cognitive behaviour therapy for depression compared to face to face treatment and medication. Submitted for publication.

[15] Christensen H, Griffiths KM, Mackinnon AJ, Brittliffe K. Online randomised controlled trial of brief and full cognitive behavior therapy for depression. Psychological Medicine 2006 Vol : 36 (12) 1737-1746

Address for correspondence

Centre for Mental Health Research, The Australian National University, Canberra, 0200, Helen.Christensen@anu.edu.ua

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

A Discrete Time-Space Geography for Epidemiology: From Mixing Groups to Pockets of Local Order in Pandemic Simulations

Einar Holm[a], Toomas Timpka[b,c]

[a] *Department of Social and Economic Geography, Umeå University, Sweden*
[b] *Department of Computer Science, Linköping University, Sweden*
[c] *Department of Social Medicine and Public Health, Linköping University, Sweden*

Abstract

The World Health Organization urges all nations to develop and maintain national influenza preparedness plans. Important components of such plans are forecasts of morbidity and mortality based on local social and geographic conditions. Most methodologies for simulations of epidemic outbreaks are implicitly based on the assumption that the frequency and duration of social contacts that lead to disease transmission is affected by geography, i.e. the spatial distribution of physical meeting places. In order to increase the effectiveness of the present methods for simulation of infectious disease outbreaks, the aim of this study is to examine two social geographic issues related to such models. We display how the social geographic characteristics of mixing networks, in particular when these significantly deviate from the random-mixing norm, can be represented in order to enhance the understanding and prediction of epidemic patterns in light of a possible future destructive influenza pandemic. We conclude that social geography, social networks and simulation models of directly transmitted infectious diseases are fundamentally linked.

Keywords

health informatics, methodology, social geography, epidemiological modeling, simulations

Introduction

In response to the threat of a destructive influenza pandemic similar to that of 1917-1918, the World Health Organization urges all nations to develop and maintain national influenza preparedness plans [1]. Important components of such plans are forecasts of morbidity and mortality based on local social and geographic conditions. Such forecasts require both surveillance of disease activity and simulations of disease dissemination. Epidemiological simulations of contagious diseases based on social network theory, where hosts and contacts are modeled as 'actors' and 'relations', have attracted considerable attention [2,3,4]. While an arbitrary level of complexity may be included in network and agent-based simulations of spatial epidemics, computational intensity and analytical intractability mean that such models often lack transparency into the determinants of epidemiological dynamics. Another approach to simulation of epidemics is based on stochastic models. These strategies have the advantages that they are computationally less complex and also can be validated using quantitative epidemiological data. Discrete-time, stochastic simulations have been used to study both the dissemination of infectious diseases and the outcome of interventions. To deal with the unrealistic assumption of homogenous contact patterns in large populations, individuals are usually allocated at each time step to discrete places, with specific localized mixing groups, where within social mixing is assumed to be more homogenous than between [5,6]. The computations in such a mixing group simulation can hereby be reduced to mixing people in the discrete spaces with homogenous mixing, while maintaining the dynamic state of infection of each sample person at each time step. Although numerous approaches have attempted to resolve the complexity-tractability/validation trade-off between social network-based and stochastic simulations, there is today no established solution to the problem.

A social geography of infectious disease

The available methodologies for simulations of epidemic outbreaks are implicitly based on the assumption that the frequency and duration of social contacts that lead to disease transmission is affected by geography, i.e. the spatial distribution of physical meeting places of different types. However, current socio-geographic concepts and empirical findings have seldom been adequately applied to specify these assumptions. It is for example only the neighborhood around the living place that increases contact probability only for spatial reasons, i.e. the closest neighbors within some 100 meters to 1 km range [7,8]. Beyond that, the organization of workplaces and other meeting places are primarily not spatially but socially stratified. If the work place, friends, relatives, or shopping mall, etc. are around the corner or 20 km away does not significantly influence behavior; these places are still visited in favor of random places that are closer. The next distance that matters somewhat is the dwelling area (cp. village), say within some 3 km range. In Sweden, one third of the working population have their place of work within that range, often also a local shop, a primary school, and a day-care center. Obviously, the likelihood of meeting others at such a place chosen at random is larger than beyond. But more impor-

tantly, already within this spatial scale, you will almost never visit the other majority of workplaces, inhabitants, etc., except for your own specific ones. Once selected, they remain fixed with respect both to location and members for the duration of a pandemic. Next level that might matter is municipality range, approx. within 30 km (cp. the municipality). Most of everyday work (two thirds) and services are situated there. The following boundary is an extended commuting distance, approx. within 100 km (cp. county). Although, it is a relatively small fraction of the population that interacts daily on that distance. For instance, the median commuting distance all over Sweden is approx. 5 km. Beyond that, we have low frequency meetings within and outside the country more or less regardless of how many there are living in between or the distance. This low frequency travel is, however, important for the initial transmission of disease between regions.

A realistic social geographic representation is not the same as a mapping of the socio-spatial trajectories of all individuals in a population in full detail. Contrarily, most of such intrinsic peculiarities probably have minor impact on general transmission of infectious diseases and should therefore rather be modeled by help of more abstract entities with random variations. This is also the most prominent property of the mixing group model for stochastic simulations of epidemics. The most important thing is instead to represent the relevant discrete spaces of close contact, i.e. the only places where the transmission of infectious agents occur. Second is the spatial configuration of these meeting places, with Hägerstrands terminology, "the pockets of local order" [9]. They are constructed by people in order to perform activities for certain projects that must be done at a particular place and to shield these from destructive outside influences [8]. The spatial configuration is mainly a constraint on the initial set up and choice of meeting places. It can in most cases be regarded as a constant for the duration of the pandemic, not changing the set of participants at the places.

Study aims

In order to increase the effectiveness of the present methods for simulation of infectious disease outbreaks, the aim of this study is to examine two social geographic issues related to such models:

1. Contact endurance at meeting places. Most mixing groups used in simulation models represent the mixing of more or less the same set of individuals day after day during the evolution of the pandemic. The degree to witch this is the case is one of the most important distinguishing features between mixing groups and should be explicitly considered for modeling.
2. Social vs. ecological fallacy. In a mixing group simulation, populations are divided into submodules based on, e.g. municipalities or counties. Such regionalization that is heterogeneous with respect to distances, from a few hundred meters to hundreds of kilometers between random inhabitants. Depending on context this distance may or may not matters for whom you will meet, but not if it is 200 or 20000 people in between. In either case, most of those latter individuals you will never meet.

This paper reports from a research program aimed at supporting pandemic response planning by simulations of contagious disease outbreaks in local public health contexts [10,11,12]. The methods section of the paper outlines the basic theoretical concepts necessary for application of social geography in the context of influenza pandemic simulations. The results section presents an application based on the Swedish ASTRID database. The final section summarizes the paper and draws conclusions.

Materials and methods

Mixing groups as social pockets of local order

Our approach to inform models for simulation of pandemics by social geography is to transfer Hägerstrand's concept of 'pockets of local order' to homogenous localized population mixing groups. A network structure is used to represent the mixing groups as nodes with contact frequencies between individuals within them as well as the other properties of social pockets. Instead of a fixed arbitrary hierarchy of household groups and regions, we suggest a system of meeting place centered, floating, overlapping reference areas. We propose the following structure for modeling of the time individuals spend at social pockets.

1. Close neighborhood (100-1000 m^2around the family residence)
2. Dwelling area (range 3 km around the residence)
3. Municipality range (30 km around the residence)
4. Extended commuting range (100 km around the residence)
5. Rest of nation and world.

The reference spatial entity is the family residence, where the infection enters and is distributed out with the family members. In principle, each family has its own unique set of ranges with a unique localization of the different types of meeting places for each individual family member. Their closest neighbor has almost the same range attributes but the set of meeting places within the ranges are very different for the individual members of each family. Information on the location of the meeting places is almost only necessary because of our interest in the spatial outcome of the disease diffusion. For global results, the individual flows between, and the characteristics of the meeting places would be enough.

A network representation of mixing groups

A central question for epidemic modeling is how to represent a differential description of who meets whom, how close and for how long time and how to represent the interaction (flows of people) between different geographical contexts. We know relatively well people's behavior with regard to their social contracts (employments, compulsory schools, etc.) and can model visits at corresponding social pockets accordingly. What needs to be dynamically modeled is the time spent in pockets that are visited during leisure time. Following pockets of social order with rele-

vance for transmission of infectious disease can be identified:

Reference pocket with fixed location and mainly fixed set of mixing subjects

1. Family residence (by type, size)

Pockets regularly visited by social contract with the same subjects

2. Place of work (by size, type of employees and social mixing patterns)
3. Day care center (specifically)
4. Primary school (specifically)
5. Other education

Pockets visited at random occasions with the same subjects

6. Relatives' homes (grandchildren, former wife etc.)
7. Friends' homes (a large part of social interaction goes on here)

Pockets visited at random with random new subjects each time

8. Public services and entertainment

Transition pockets (random or fixed occasions, random subjects)

9. Local daily travel with public means of transport (commuting, train, air transports - short duration but sometimes close contact with new people)

Transitions (movements between communities)

10. Short and long haul recreation and leisure travel
11. Business travel (rapidly growing, partly replacing commuting and migration)

Table 1

Social pocket	n	*p (c)*	*p* (t)	Contract
Family home	1-7	0.3-0.6	0.12	Yes/No*
Workplace	20	0.06	0.12	Yes
Day care				
Family day-care	3-6	0.35	0.12	Yes
Day-care center	10-20	0.15-	0.12	Yes
Schools				
Primary school	15-25	0.03-0.04	0.12	Yes
High school	30-60	0.03	0.12	Yes
Public services				
Shopping mall	25-150	0.01-0.001	0.12	No
Arena	200-800	0.001-0.002	0.12	No
Public transportation				
Short distance	40-80	0.03-0.04	0.12	No
Long distance	100-200	0.01-0.005	0.12	No

* Relatives and friends visiting a family home do not own or rent the facility.

Table 1 displays some characteristics of such social pocket types. The contact probability ***p*(c)** reflects the probability for any person visiting the social pocket to meet one single person in the pocket, while ***p*(t)** reflects the probability of transmission of the infectious agent at a given contact. Estimates are from Swedish data and the literature [5,6,13]. This representation is mainly consistent with the mixing group model. A spatially explicit mixing group simulation can thereafter be divided into two stages. At the first stage, an abstract representation of contacts based on social contracts is generated from spatially explicit population data using fixed types of social pockets and itinerary types specific for age-group and socioeconomic categories. Each person is initiated by allocating the person first to an age-specific itinerary type, thereafter to the relevant types of social pockets, and finally to instances of social pockets, , called local pockets (localized mixing group) from the database. That allocation is based on observed or estimated cross sectional frequencies for each of the social pocket types over each of the ranges for certain groups of individuals.

Table 2

Age	1km	3km	30km	100km
0-6	Own home *17.5/7*		Day-care center *3.5/5* or Family day-care *3.5/5*	
7-15	Own home *4/7*	Primary school 5/5		
16-19	Own home *14/7*		High school *5/5* Short-distance transports 0.5/8	
20-64	Own home *14/7*			Workplace 5/5 Short-distance transports 0.5/8
65-	Own home *17.5/7*			

Using empirical data, the patterns for visiting the social pockets can thus be represented as itinerary types for regularly visited social pockets and for arbitrarily visited pockets, respectively. While the pockets visited by social contract can be represented by one particular mixing group instance (school, workplace, etc.) during the instantiation, the arbitrarily visited social pockets are randomly instantiated (as one or several local pockets) within the corresponding geographical region determined from the individual's place of residence. Table 2 shows itinerary types for pockets visited by social contract displayed by concentric geographical regions determined from the individual's place of residence. Numbers in italics indicate 8-hour time units spent in the pocket per week and number of visits per week (8-hour units/visits).

Table 3 shows itinerary types for arbitrarily visited mixing groups displayed by concentric geographical regions determined from the individual's place of residence. The examples in the results sections are based on structures and data from the ASTRID database covering the Swedish population (9 million).

Table 3

Age	1km	3km	30km	100km	Nation
0-6	Homes	Homes Shopping malls	Homes Shopping malls	Homes	Homes
7-15	Homes	Homes Shopping malls	Homes Shopping malls Arenas	Homes Shopping malls Arenas	Homes
16-19	Homes	Homes Shopping malls	Homes Shopping malls Arenas	Homes Shopping malls Arenas Short-distance transports	Homes
20-64	Homes	Homes Shopping malls	Homes Shopping malls Arenas	Homes Shopping malls Arenas Short-distance transports	Homes Work places Long-distance transports
65-	Homes	Homes Shopping malls	Homes Shopping malls Arenas	Homes	Homes

Results

Mixing-group instantiation of social pockets

In our instantiation of the social geographic model for epidemiological simulations, each social pocket type I has type specific properties, such as the average disease passage between pairs of visitors (infectious or susceptible) per unit of time. In order to enable adjustments to different infectious agents, this risk is factorized into contact probability per time unit and transmission risk per contact. The specific instances i of I are given the same properties as the type, and provided a random variation reflecting observed empirical distributions. In addition each i has a location and size etc. The outcome probability for a specific infectious/susceptible pair is also affected by those persons' individual properties.

Parameterization of individuals

Parameters for each individual is derived from the ASTRID database. The basic object in this modeling approach is each individual person p in the country. All other local pockets are relational properties of the person but also when convenient objects by themselves. The approach lends from time geography, micro simulation and agent based simulation [14,15]. Each person p has a set of individual properties (age, sex, education, origin, income, mother, family etc.). Person p's family f has other properties common to its members (single, adults only, child family etc., size, location, housing type etc.) Person p also has individual relations r(I) to the social pockets types I visited by social contract (work, day care, school) and those visited less regularly but within a small set of persistent members (friends' and relatives' homes) and those visited irregularly with random members (public services, entertainment, long-distance travel, etc.). For the duration of a pandemic, p's properties, including location and social composition of family and meeting places, can be regarded as fixed, i.e. as an itinerary. For example, suppose that the partly overlapping ranges (family, 1 km^2, 3 km neighborhood, 30 km commuting distance, and 100 km regional range) are outlined for two (p and q) out of the nine million localized individuals in the database. Say that two example local pockets, a school and a workplace are in place. The school might be within reach only for q whereas both p and q might work at the work place. The floating ranges are only used as a tool for the initial allocation of persons to local pockets. When that is done, the resulting set of "arrows", the network of connections, is the only thing necessary to feed into the simulation.

Each person also is given an individual computed state with regard to the disease. The only property that has to be dynamic is the person's current infectious status. The set of people m(i) "mixing" at each specific social pocket i of type I, consists of persons being members of pocket i (p.r(i).member = true). That set is therefore also fixed for all persons and local pockets (meeting places) during simulation by the individual properties of the persons as derived from the database.

Simulation

When running the simulation, a population is first broken down in administrative-political segments (by residence) and then divided into standardized units. The division into standardized units is, among other reasons, based on the necessity to validate the simulations against historical data. For each time step (8 hours) each person in the population is exposed to a context, time and person specific computed probability to transmit and receive the disease at each mixing group and thereafter updated accordingly to the outcome of the lottery. The consequence events are then conventionally modeled without time delay either immediately after the transmission or for aggregates at the end of the time step.

After that the table for simulation is set and fixed, the main property changing value during simulation is the persons current infectious status. In a more elaborate model also other dynamic response actions can be introduced, like keeping children at home instead of at day care or school before they become infected or that hospital personnel obstruct by staying home from work in order to avoid personal risk. Such behavior responses can only be based on vague and anecdotic evidence but it is nevertheless crucial to enable such hypothetical experiments.

Discussion

Social geography, social networks and the epidemiology of directly transmitted infectious diseases are fundamentally linked. The foundations of epidemiology and early epidemiological models were based on population wide random-mixing, but in practice each individual has a finite set of contacts to whom they can pass infection; the ensemble of all such contacts forms a 'mixing network'. Knowledge of the structure of the network allows models to compute the epidemic dynamics at the population scale from the individual-level behaviour of infections [16]. We have in this study displayed how issues related to the social geographic characteristics of mixing networks, in particular how these deviate from the random-mixing norm, can be resolved in order to enhance the understanding and prediction of epidemic patterns in light of a possible future destructive influenza pandemic.

Acknowledgments

This work was supported by the Swedish Emergency Man-agement Agency (KBM) under contract 0700/2004. Henrik Eriksson, Britt Åkerlind and Kristina Kinnunen from Linköping University, Sweden, Johan Jenvald and Magnus Morin, VSL Research Labs, Sweden, Magnus Strömgren, Umeå University, Sweden, Elin Gursky, ANSER Analytic Services, Arlington, VA, USA, and James Nyce, Ball State University, USA have contributed valuable comments to the analyses presented in this report.

References

[1] World Health Organization (WHO). Responding to the avian influenza threat. Geneva: World Health Organization, 2005. Available at http://who.int.org/csr/resources/publications/influenza (accessed March 15th 2007).

[2] Eubank S, Guclu H, Kumar VS, Marathe MV, Srinivasan A, Toroczkai Z, and Wang N. Modelling disease outbreaks in realistic urban social networks. Nature 2004 May 13;429(6988):180-4.

[3] Browers L. MicroPox: a Large-scale and spatially explicit microsimulation model for smallpox transmission. http://www.sgem.se/about_sgem/researchers/associated_members.html

[4] Burke DS, Epstein JM, Cummings DA, Parker JI, Cline KC, Singa RM, and Chakravarty S. Individual-based computational modeling of smallpox epidemic control strategies. Acad Emerg Med. 2006 Nov;13(11):1142-9.

[5] Longini IM Jr, Halloran ME, Nizam A, and Yang Y. Containing pandemic influenza with antiviral agents. Am J Epidemiol. 2004 Apr 1;159(7):623-33.

[6] Germann TC, Kadau K, Longini IM Jr, and Macken CA. Mitigation strategies for pandemic influenza in the United States. Proc Natl Acad Sci U S A. 2006 Apr 11;103(15):5935-40. Epub 2006 Apr 3.

[7] Holm E, and Öberg S. Contagious social practice? Geografiska Annaler 2004;86B:297-314.

[8] Ellegård K, and Vilhelmson B. Home as a pocket of local order: Everyday activities and the friction of distance. Geografiska Annaler 2004;86 B(4):281-296.

[9] Hägerstrand T. Time-geography: focus on the corporeality of men, society and environment. In The Science and Praxis of Complexity. New York: United Nations University, 1985.

[10] Timpka T, Morin M, Jenvald J, Eriksson H, and Gursky E. Towards a simulation environment for modeling of local influenza outbreaks. AMIA Annu Symp Proc. 2005; 729–33.

[11] Timpka T, Morin M, Jenvald J, Gursky E, and Eriksson H. Dealing with ecological fallacy in preparations for influenza pandemics: Use of a flexible environment for adaptation of simulations to household structures in local contexts. Medinfo 2007;

[12] Eriksson H, Morin M, Jenvald J, Gursky E, Holm E, and Timpka T. Ontology-based modeling of pandemic simulation scenarios. MedInfo 2007;.

[13] Halloran M., Longini IM, Cowart DM, and Nizam A. Community trials of vaccination and the epidemic prevention potential. Vaccine 2002;20:3254–3262.

[14] Boman M, and Holm E. Multi-agent systems, time geography, and microsimulations. In Olsson MO, Sjöstedt G (Eds.). Systems approaches and their application: Examples from Sweden. Dordrecht: Kluwer Academic Publishers, 2004, 95-118.

[15] Holm E, Holme K, Mäkilä K, Mattsson-Kaupi M, and Mörtvik G. The SVERIGE spatial microsimulation model – Content, validation, and example applications. **Gerum Kulturgeografi 2002:4. Umeå: Umeå Universitet,** 2002.

[16] Keeling MJ, and Eames KT. Networks and epidemic models. J R Soc Interface. 2005 Sep 22;2(4):295-307.

Address for correspondence

Professor Einar Holm PhD
Department of Social and Economic Geography
Umeå University
SE-901 87 Umeå
Sweden
einar.holm@geography.umu.se

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

A National Study of eHealth Standardization in Finland - Goals and Recommendations

Juha Mykkänen[a], Maritta Korhonen[b], Jari Porrasmaa[a], Tuula Tuomainen[b], Antero Ensio[c]

[a] *HIS R&D Unit, IT Services Centre, University of Kuopio, Finland*
[b] *Business and Administration, Savonia University of Applied Sciences, Kuopio, Finland*
[c] *Ensitieto Oy, Varkaus, Finland*

Abstract

The role of standards is constantly increasing in health services, electronic health records, and eHealth applications. There are many areas of standardization which affect the healthcare work and health information systems. On a national level, the organization of the development and support for standardization should be a key priority. This paper summarizes a national study in Finland which reviewed the current status of eHealth standardization and made recommendations for the national standardization and the use of standards. The recommendations are related to the organization model and balanced participation, international and cross-domain collaboration and quality assurance of standards-related activities. In addition, education and support services and support for introductions and pilots are recommended to improve the know-how of standards in system acquisitions.

Keywords:

standardization, interoperability, eHealth, health information systems

Introduction

The role of standards in eHealth and national initiatives

The provision and use of health services are facing increasing change pressures. The aging population, recruitment problems, biomedical advances, diversity of treatments and examinations, and increasing need of services by the elderly and by those having multiple diseases are seen as international trends [1, 2]. On the other hand, the national and international health and economic policies affect the market of the eHealth solutions. For example, the transition from hospital-centric to patient-centric and local health service provision has been seen as a pan-European trend [3]. On the other hand, specialized services such as clinical laboratories are transferred to large specialized units, for example in Finland. Electronic health information is increasingly used to improve quality and to enable new care models and treatments [1]. Healthcare networks require connections between disparate healthcare units and professionals. These needs are reflected in the requirements of eHealth applications and information systems.

One of the central challenges for electronic health services is the *interoperability* of eHealth solutions [4]. Health information systems contain lots of diverse information which have to be preserved and used for a long time. Large amount of new information and knowledge is constantly emerging in healthcare [1]. In addition, information is often sensitive or has accountability requirements. It has been observed that in general, the interoperability of eHealth applications and health information systems is far from optimal. This is partially due to the fact that *standards* are missing, their existence is not known, or they have not been implemented to an adequate extent [2].

A central goal of the national health project in Finland is to produce interoperable national electronic healthcare records. This includes national solutions for the long-term storage of electronic patient records and a secure access to the patient records by health service providers, patients and other actors [5]. The national architecture requires that the electronic patient record systems correspond to the national specifications as well as the establishment of national healthcare IT services. These services include an electronic archive and related services for the registration of documents as well as solutions for messaging. The realization of the architecture requires steering by the

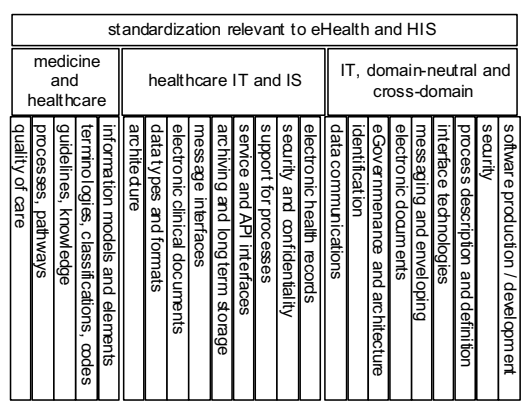


Figure 1 - Areas of standardization addressed in the study

authorities and the establishment of an actor that is responsible for the implementation of the national services. Shared standards are a key factor in pursuing these goals. In addition to Finland, similar large-scale projects are underway in many countries. One central component of these initiatives is the promotion of the standardization of the healthcare IT.

This paper presents the central results of a study which aimed to document the current status of the national healthcare IT standardization in Finland and to identify and suggest improvements in the organization and development of standardization and the use of standards.

Dynamics of standardization

To illustrate the context of this work, we briefly discuss some aspects of standardization as a basis for this study. A *standard* is a document approved by an accepted body which contains rules, guidelines or features for generic and repeated use in products, processes or services [6]. In general, standardization has many meanings and motivations [7]: the *uniformity* of production, the *compatibility* of technologies, the *objectivity* in measurement, the means for *justice* and a form of *hegemony*. Out of these goals, compatibility in the form of *interoperability* of software applications has been emphasized in recent years. Rapid change, global connectivity and the need for systems to interoperate are also increasingly evident in the future [8]. However, there is a wide range of interoperability standards available, both on healthcare-specific and domain-neutral levels. The lack of implemented standards hinders the development and use of healthcare knowledge, causes risks to the patients and sub-optimal use of resources [2].

Central tasks related to standards include the organization and steering of the standards work, the production and use of standards (experiments, introduction and use in products and procurements), education and support services, and the evaluation of standards and products conforming to them. The tasks are different in various phases of the lifecycle of standards, and they affect different participants. Standardization organizations are only responsible of some of the costs of these activities.

Most standardization organizations emphasize the creation of standards as a response to the *needs expressed by the companies and users*. On the other hand, standards are created as a result of complex *social negotiations* [7]. These viewpoints to the creation of standards (functionalistic and constructivistic) emphasize the quality and accuracy, and the acceptance and dissemination of standards, respectively.

In health information systems, standards support the *vision of open systems* which can be complemented with new products. The direct benefits of eHealth standards provided for healthcare organizations, professionals and patients include the reduction in costs and errors, and improvements in care quality, usability, work practices and the availability of information. In addition, the application developers must attain advantages from the use of standards, such as new markets, consistent and accurate requirements, improved integration with partners, specialization possibilities and reusable models for subsystems or interfaces. In addition, healthcare financiers, medical research and the government benefit from standardization [3,2].

However, several problems related to standardization have been identified. Standards and standard families are incompatible and overlapping. Despite many standards available nationally and internationally, the introduction and assignment of different standards has not been completed. As the participation in the standards work is voluntary, not all relevant areas of standardization have been considered. In addition, the participation has been based on personal interest and it has not been properly coordinated, organized or resourced.

Materials and methods

In November 2004, the challenges and problems of standardization were considered in a meeting by the representatives of the Finnish Ministry of Social affairs and Health, Ministry of Trade and Industry, the Association of Finnish Local and Regional Authorities, the Technical Research Centre of Finland (also representing HL7 Finland), the National Research and Development Centre for Welfare and Health, the Finnish Agency for Technology and Innovation TEKES, the University of Kuopio and the Savonia University of Applied Sciences. In conclusion, it was acknowledged that standardization is an important enabler of the health service provision, the development of the national economy, and the growth of the small and medium enterprises. The lack of sustained solutions for the development, utilization and participation in standardization was identified as a central challenge. The action was taken to produce a report which would review the previous studies and recommendations and combine them with a wider view of standardization. The result of the study would be recommendations for the improvements in the development of activities related to standardization on a national level.

The work was scheduled from December 2004 to March 2005 for two projects funded by the Finnish Agency of Technology and Innovation TEKES. A *literature survey* was first performed, and national and international recommendations for eHealth standardization were reviewed, e.g. [3,4,9,10,2]. The study also produced a description of central *international and national actors* in eHealth standardization and a description of the *different areas of standardization* related to eHealth and health information systems. In addition, a previously defined framework for the evaluation and selection of standards was updated. The standardization areas were intentionally described from a wider view than only health IT to identify relationships and improvements beyond healthcare-specific considerations (see Figure 1). The standardization areas were divided in three main classes: the standards requiring specific medical or healthcare knowledge, the standards requiring a combination of IT and healthcare skills, and the IT or domain-neutral standards relevant in healthcare. These classes have some correspondence with the specificity levels and perspectives of the health informatics

profiling framework by ISO TC 215 [10]. The notion of complementary standards to allow the modular construction of solutions [4] was one of the guiding principles in the identification and classification work.

After the identification of the main actors and areas, data collection was designed to survey the *current status* and the *target state* of the field. A recent web-based survey from the SerAPI project which included questions related to the responsibilities of integration and standardization was used as a basis. The survey was continued with an e-mail survey to named experts in different organizations (hospital districts, companies, organizations in the board of the project). This survey had four main categories and 60 detailed points. The main categories were the goals and policies, areas of standardization, activities related to standardization and the objectives for improvement. Each respondent received questions from two or three categories to lower the response threshold.

The questions of the e-mail survey were also discussed in several meetings with various experts. The number of responses in the web-based survey was 10 (of 18) and in the e-mail survey and meetings 13 (of 18), the total response rate being 66,7 %. In addition to the surveys, interviews were performed with several experts responsible of the different areas of official and industry standardization. Furthermore, the participants of the project board reviewed the work in two mid-project meetings. The results of the literature survey and the data collection were then used by the authors who wrote the descriptions of current and target state, and other parts of the report.

The main result of the study was the description of current status and issues and a set of recommendations related to the standardization of healthcare IT. For the recommendations, the material of the study was combined with the personal experience of the authors. The results of the study were published in a 92-page report in Finnish for further actions [11].

Even though the work covered many areas, the application of standards was delimited to consider mainly *health information systems*. This was due to the fact that the majority of challenges of the national project and of the hospital districts were related to the interoperability of these systems. Excluded were also the devices and device interfaces, for example the medical equipment. In addition, the scope of the work was limited to open standards and specifications, and it did not consider internal implementation technologies of applications or development models such as open source.

Results

The results of the surveys and interviews

The respondents of the surveys and interviews saw the quality assurance and applicability of standards as a key factor in relation to the *goals and policies* of standardization. In addition, the usability requirements and pragmatism of standards were emphasized. Fast introduction and implementation were highlighted by software companies, as well as consistent solutions across different business domains. However, the opportunities of international markets were not among the most central goals. There were conflicting views regarding the special consideration of the existing systems in relation to standards, as well as the need for accurate identification of specific fine-grained standards versus widely applicable frameworks.

According to the respondents, the most central *areas of standardization* were the consistency in the support and specification of healthcare processes, the structure of information, data types, semantic consistency and the storage of electronic patient documents. In addition, unified information models and desktop integration were among the most important areas in some responses. Terminologies, codes, knowledge, shared IT services and workflow support were advocated to some extent. There were different opinions about the necessity of the standardization of architectures, security solutions and many technical or domain-neutral aspects. The code sets and information models were seen as enablers for more advanced standards such as guidelines and processes[1]. According to the responses, eHealth solutions should be based on generic technologies and cross-domain standards, but the specific requirements of healthcare were identified in relation to security, safety and privacy, for example.

Many respondents saw a clear need for many different types of *participants* in the standardization. Healthcare organizations were emphasized in the standardization of healthcare-specific information, processes and guidelines. In the technology standardization (including healthcare IT), the role of companies and standardization organizations was highlighted. The role of authorities was also seen central, especially in relation to the healthcare-specific aspects and architectural guidance.

Goals, problems and challenges

Based on the literature survey and the results from the data collection, the national problems and challenges were identified and the target state was specified. In eHealth standardization, no organization can master all areas. This stresses the importance of coordination, the utilization of experts in different areas and the relationship between international and national standardization. In addition, clear relationships between healthcare information and processes, healthcare IT and technical standardization, and the requirement to base the solutions on the actual needs of the market are central. Some research results suggest that global standards succeed only if they can be adjusted to the local processes and activities, and they are applied or modified according to the local requirements [7, 12]. This observation was supported by our study.

The local or national *definition of responsibilities, guidelines and resources* is needed for the selection and production of standards and recommendations [9]. The selection and production of specifications, steering of these activities, support for their use in products and pro-

1 This statement has been emphasized also in other research [7].

curements, and education and conformance evaluation are central parts of these activities.

The *coordination and ownership* of standardization activities was identified as one of the key challenges in standardization. The various earlier recommendations for the organization models of standardization have not been realized. The collaboration and knowledge of different activities should be increased. Even though standardization requires participation from the industry and their customers, the government can maintain a working standards infrastructure. This support can be realized as clear recommendations about the selected standards, clarification of responsibilities and guidelines such as national architecture for electronic health records. The linkage between the standardization and system introductions, and the creation of conditions which balance the participation in standards activities can also be centrally supported [9].

The *quality assurance and the evolution of requirements* require special attention in standardization. There is a contradiction between the fast introduction of standards and their quality [9]. The subtle balance between accuracy and flexibility depends on the type of the specification. The conformance and certification require accurate specifications and dedicated services. Implementation examples, support for the projects which introduce standards, and guidelines for the procurement are useful means for the quality assurance.

The *relationship between official and industry standardization* has not been clearly defined. Industry standardization has been more agile in responding to the needs of the users. The free availability of key specifications promotes their use. In addition, the *commitment and participation* in standards-related work has been increasing very slowly. In particular, the companies and healthcare organizations have not participated actively in the international standardization. In general, successful standards require demand from the market, and standards can not be easily enforced by the officials.

New standards and specifications which are potentially relevant for healthcare, health IT or technology aspects, are continuously emerging. Constant *learning* by the users and developers is required. The careful selection of the specifications to be studied, and education about standards are necessary to keep the workload reasonable. The knowledge of key areas and activities of standardization should be included in the education of healthcare IT professionals.

Discussion

Main recommendations

Based on the surveys and previous parts of the study, the following eight main recommendations were made:

1. The standardization relationship between the healthcare IT and the domain-neutral IT must be intensified. *Shared national goals, policies and procedures for IT-related standardization* must be specified. National standardization steering groups are proposed for overall coordination and the healthcare-specific standards.
2. The *continuity* of domain-neutral and healthcare-specific IT standardization must be assured using permanent *funding* from various departments.
3. The primary preference for all domains, including healthcare, must be given to *cross-domain and generic standards*. Healthcare-specific standards should be developed and introduced cautiously and only on areas where they are essential.
4. The participation to the *international standardization* work and the observation of international key developments in standardization must be intensified and resourced. The goal is to identify mature standards which solve current local needs and to avoid contradictions with the international standardization.
5. The participation of healthcare *application vendors and health service providers* in the development, localization and introduction of standards must be supported by funding projects which aim at standards compliance and by developing models to support the balanced participation to the standardization.
6. The national standardization in healthcare IT must primarily support the *goals of the national health project*, especially the interoperable electronic health records. This requires a quick prioritization of the most immediate areas and goals, and the establishment of realistic funding and scheduling for them.
7. The *status, normativeness and mutual relationships* of healthcare IT standards, guidelines and recommendations must be defined unambiguously and accurately.
8. To ensure the interoperability of information systems and to support the introduction of standards, a *support and education network* must be created with adequate expertise to promote these recommendations and to support the steering groups.

In addition to these high-level main recommendations, the report identified 51 detailed goals and 127 recommended actions to support these goals. The *policy recommendations* were related to the organization of the standards-related activities, steering and funding. The *relationship recommendations* focused on improving the international and cross-sectoral linkages in standardization. For the *quality assurance* of standards and guidelines, 20 recommendations were made. Central recommendations were also related to the *improved know-how for the system acquisitions* and to the *establishment of education and support* related to standards. In addition, the support for the *introductions and pilots* involving standards, and *balance in the participation* of standards activities received recommendations. In addition, detailed recommendations were made for many *areas of standardization* and national policies. The most urgent recommendations were related to the core information sets, clinical documents, architecture and security of the national electronic health records.

Current trends of interoperability standardization

Some central trends of standardization are related to the evolution towards advanced eHealth interoperability.

Information systems are increasingly evolving towards networked and service-oriented solutions, which increases the role of standardization. In technical standardization, the aim is to *increase the flexibility* of systems to support different processes and adaptability. On the other hand, accurate *functional and semantic interoperability* is pursued, and *processes and workflows* themselves are increasingly standardized. In particular, *profiles* which constrain the standards are increasingly used to promote plug and play interoperability or automatic adaptability. Such profiles are emerging on the technology level [13], on the functional level [14], on the semantic level [4] and in the field of application of interoperability standards [15, 3]. In addition, emerging support services such as conformance testing and certification are closely related to standardization.

Conclusions

Achieving the benefits of standardization in the eHealth domain requires coordination, the selection and development of standards on many areas, the identification of central actors and close collaboration between them. The different areas require specific expertise and multidisciplinary collaboration. The success of standards is measured only through their utilization on the market. The support for the introductions of standards, the clear scope of standards, quality assurance, and the availability of standards and their support services create a vortex of successful standardization activities. The national participation in international standardization and the evaluation and selection of standards are necessities in this process.

The rapid evolution of IT and complementary standards offer many opportunities. However, generic standards require healthcare-specific profiling and evaluation, and healthcare also has specific standardization areas. In a wider context, standardization and the use of standards promotes competition and open market, and progresses economy as a whole.

The recommendations of this study were published and given to the project board after the report was finalized in 2005. Since then, the Social Insurance Institute of Finland has been selected as the national actor for the development of national healthcare IT services. HL7 CDA (Clinical Document Architecture) R2, DICOM and HL7 version 3 Medical Records have been recommended as some of the key standards of the national EHR in relation to clinical documents, medical imaging and messaging, respectively. In addition, some technical specifications such as X.509 for certificates, http(s) and SOAP for data communications and WS-Security for mediated secure messaging have been recommended. Furthermore, some of the key recommendations related to the steering model of standards-related work are being refined as of March 2007.

Acknowledgments

This work is part of the SerAPI, ZipIT-ojo and ZipII-ojo projects, funded by the Finnish Funding Agency for Technology and Innovation TEKES grants no. 40251/06, 644/06, 40353/05, 644/05, 40437/04 and 790/04 together with a consortium of companies and healthcare organizations.

References

[1] Haux R. Health information systems - past, present, future. *Int J Med Inf* 2006: 75 (3-4): 268-281.

[2] Klein G. Standardization of health informatics – results and challenges. Yearbook of Medical Informatics 2002, 103-114, IMIA, 2002.

[3] Current and future standardization issues in the e-Health domain: Achieving interoperability. Report from the CEN/ISSS e-Health Standardization Focus Group. CEN, 2005.

[4] Beale T. Health Information Standards Manifesto. Revision 2.5. Deepthought informatics, 2004. http://www.deepthought.com.au/health/HIS_manifesto/ Output/his_manifesto.html.

[5] Iivari A, and Ruotsalainen P. Principles for national data system architecture in the field of health care; from regional solutions to a national entity. Working Group Memorandum, Ministry of Social Affairs and Health, 2005.

[6] Guide to the Project Management Body of Knowledge (PMBOK Guide). Project Management Institute, 2000.

[7] Feng P. Studying Standardization: A Review of the Literature. In: Proceedings of the 3rd conference on Standardi-zation and Innovation in Information Technology (SIIT2003). IEEE, 2003, pp. 99-112.

[8] Boehm B. Some future trends and implications for systems and software engineering processes. Systems Engineering 2006: 9 (1): 1-19.

[9] Swann P. The Economics of standardization - Final Report for Standards and Technical Regulations Directorate. Department of Trade and Industry. Manchester Business School, University of Manchester, 2000.

[10] Health Informatics - Health Informatics Profiling Framework. ISO/TR 17119:2004(E), Technical report, ISO TC 215, 2004.

[11] Mykkänen J, Korhonen M, Porrasmaa J, Tuomainen T, Ensio A. Tietojärjestelmien standardointityön organisointi ja kehittäminen terveydenhuollossa: nykytila ja toimenpide-ehdotukset. Osaavien keskusten verkoston julkaisuja 3/2005. Helsinki: Stakes, 2005.

[12] Berg M, and Timmermans S. Standardization in action: Achieving local universality through medical protocols. Social Studies of Science 1997: 27(2): 273-305.

[13] Basic Profile Version 1.1. Web services interoperability organization, 2004. http://ws-i.org/Profiles/BasicProfile-1.1-2004-08-24.html

[14] Dickinson G, Fischetti L, and Heard S, eds. HL7 EHR System Functional Model - Draft Standard for Trial Use. Health Level Seven, Inc., July 2004.

[15] Vegoda P. Introducing the IHE (Integrating the Healthcare Enterprise) concept. *J Healthc Inf Manag* 2002: 16 (1): 22-24.

Address for correspondence

Juha Mykkänen
University of Kuopio, IT Services Centre, HIS R & D Unit
P.O.B. 1627, Fin-70211 Kuopio, Finland
juha.mykkanen@uku.fi

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Building a Womb-to-Tomb Health Record in Hong Kong – an Application of Information Architecture

Vicky Fung[a], N T Cheung[a], Eric Ho[b], Clara Cheung[b], Hudson Chan[b], Kitty Tsang[b], Joycelyne Cheung[b], William Ho[b], Maggie Lau[a], Veronica Hung[a], Austen Wong[a], Anna Tong[a], W N Wong[a], Antonio Sek[a]

[a]*Health Informatics Section, Hospital Authority, Hong Kong Special Administration Region*
[b]*Information Technology Department, Hospital Authority, Hong Kong Special Administration Region*

Abstract

The Hospital Authority developed the Information Architecture (IA) model in 2002 to support a fast, robust, flexible and accurate electronic patient record (ePR) to meet the high-tempo health care environment in Hong Kong.

With several successful applications in sharing data that were created for the same patients in various systems, the IA model was further developed to extend the longitudinal ePR to include one's fetal data as entered in the mother's record. This paper describes how various IA elements: Section, View, Form, Group, Entity, Content, Document supports the building of a true womb-to-tomb ePR for the HA patients.

The future focus of Information Architecture in the HA will include building a Information Architecture Management System and linking the ePR with other patient records in the community.

Keywords:

Information Architecture, electronic health record, terminology, standards, patient master index

Introduction

Healthcare is a person-centred process with multiple carers providing care to a person at different times at various locations. With the aim to improve a person's health, a nation's health and the world's health, developing a longitudinal electronic health record (EHR) is one of the greatest challenges to the health informatics field.

Electronic health record refers to 'a repository of information regarding the health of a subject of care in computer processable form' [1]. It is essential to ensure the longitudinal EHR is functionally and semantically interoperable such that the systems are able to exchange and further process the received information which is understood with formal defined domain concepts [1,2].

In recent years, various countries started their own initiatives to build health information infrastructure to facilitate a person to access one's health record regardless of where, when and by whom the record is created [3, 4, 5]. The keys to building a 'longitudinal' EHR include a universal unique identifier for the data subject and care provider, the carers and the facilities; terminology standards for recording, storing and reusing information and supporting decision support systems; messaging standards for exchanging information amongst systems; information model such that various standards are able to work together; and security standards to enable the information to be communicated within a consistent security framework [5, 6, 7, 8].

Some common terms to describe the 'longitudinal' record are 'cradle-to-grave', 'womb-to-tomb', or even 'sperm-to-worm' [1]. There is no question that one's health could be related to conditions one had during his/her fetal life. This relationship is so important that the World Health Organisation requests countries to report these prenatal conditions that affect one's health using International Classification of Diseases (ICD) [9].

Yet we find little discussion on how to electronically link the health record of one's fetal life to that created after one's birth. A powerful generic Information Architecture allows us to share data from the record of one's immediate family and utilize this data in subsequent health care. This paper describes how the Hospital Authority (HA) realized the development of a true 'womb-to-tomb' record.

Clinical information systems in the HA

The Hong Kong Patient Master Index (HKPMI)

At HA, all patients are uniquely identified with their Hong Kong Identity Number (HKID), a corporate based patient identifier. To date, the HKPMI has around 6 million unique HKIDs of living Hong Kong residents, which accounts for around 83% of the Hong Kong population. For each admission, outpatient visit or emergency attendance, a separate episode number is created. The episode number serves as a unique identifier to link all clinical information that created in that particular episode. The HKID links all clinical information in various episodes under the same patient.

The Clinical Management System & ePR

The Hospital Authority developed the Clinical Management System (CMS) from 1994. The CMS is an integrated

workstation for doctors to directly enter clinical orders and patient documentation. To create a record in the CMS, there must be an episode in the Hong Kong Patient Master Index (HKPMI).

From 2000, all essential clinical data from all HA hospitals and clinics were consolidated in the ePR repository, building up an enterprise-wide, longitudinal patient record [10]. Currently, the ePR houses 6.3Tb data (not including radiology images) with the records of 7.9 million patients, and garners 300,000 hits per day from clinical users.

The Obstetrics Clinical Information System

The Obstetrics Clinical Information System (ObsCIS) was first implemented in the HA hospital in 1994. The ObsCIS aims to support ongoing patient care and clinical audit. In 2005, out of 57,124 births in Hong Kong, 41,258 (72%) of them were born in HA hospitals, and the antenatal, delivery and postnatal information for all these mothers and newborns are recorded in the ObsCIS.

Information Architecture in the HA

In 2002, HA developed a simple yet generic model, HA's Information Architecture (IA), with the aim to build a fast, robust, flexible electronic patient record (ePR) from data being captured in various systems [11]. The IA framework was developed based on the following principles :

- Simple – easily understood, and implemented
- Generic – able to apply to data captured in any technical environment
- Flexible – support different presentations of the same data
- Concept oriented – ensure the consistency of the meaning of the information

More recently, the model has been refined with groupings of Entities into Forms and Groups to facilitate data capture and management. Under this model (figure 1), medical facts in ePR are described with :

- Sections – the macroscopic structure of ePR
- Views – dynamic presentations that fit according to various requirements
- Forms – source from which Entities are originated from
- Groups – logical grouping of Entities to facilitate data entry and subsequent data reporting and analysis
- Entities – the label that stores data
- Contents – the value that being stored, this could be a code, text data
- Documents – a special type of Content with human readable images

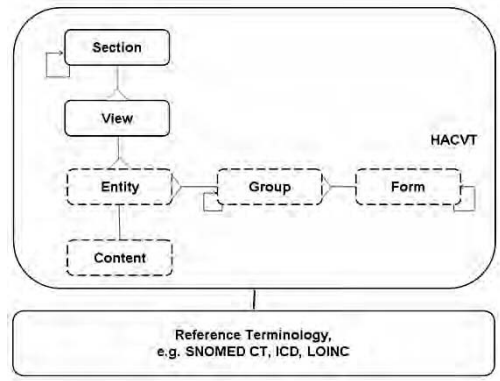


Figure 1 – The Information Architecture Model

Every medical fact has a concept

Central to the IA model is the "concept" attached to Entity, Document and Content. All Entity descriptions are added to the HA Clinical Vocabulary Table (HACVT) [12]. Each HACVT term is uniquely identified with a TermID. Where applicable, these terms will be mapped to reference terminologies, such as, SNOMED CT, or LOINC, ICD 9 CM, or ICD 10 to facilitate subsequent interpretation and data retrieval.

Contents may be codes, numbers, text or even images, and since coded Contents is referenced to the HACVT as well, mapping to the reference terminologies may be required.

Every medical fact has a context

Under the IA, all medical facts are stored in Entities. Entities are related through Groups and Forms. The IA defines relationships between these constructs, insuring that the context of data is preserved. Each Entity is identified with a unique EntityID. Each EntityID, apart from storing Entity description, also includes a set of attributes, such as data format or repetitive indicator.

Every medical fact has a presentation

Sections and Views define how information is organized and presented in ePR. To support flexible data retrieval, the same Group / Entity / Content / Document can be put under different ePR Sections. For example, Psychiatric Nursing Discharge Summary document can be placed under Nursing Section and also the Psychiatric Summary Section.

Although in 2002 other models were already in development, e.g. HL7's Clinical Document Architecture [13], ASTM's standard for the patient record [14], the Good Electronic Health Record [15], HA's IA was able to scale rapidly to the demand of data retrieval on an enterprise-wide scale in a busy clinical environment.

The IA model takes a comprehensive approach in managing information from its generation to how it is used, presented and aggregated. It allows semantic interoperability between the different modules and subsystems of the CMS. IA helps the HA to build a flexible, comprehensive ePR which meets the high volume of data retrieval with

subsecond response whilst still allowing advanced functionality.

The journey begins

The mother-baby relationship table

Since 2003, all newborns born in HA hospitals are admitted and a medical record is created. This allows clinicians to record the newborn's data, e.g. laboratory results, directly in the newborn's own CMS record instead of keeping them in the mother's CMS record as before.

In 2005, it was decided to share the mother's delivery data to the baby's record. The first step is to ensure there is a correct linkage between the mother and the baby in the HKPMI.

The mother's delivery episode number and the newborn's birth episode number are stored in the HA Mother-Baby Relationship Table (M-B table) which is incorporated in the HKPMI. This forms the basis for subsequent sharing of data from a mother's ObsCIS record to the newborn's record.

Data to be shared

The project is governed by the Perinatology Committee with representatives of senior clinicians from all Obstetrics and Paediatrics Department of the HA hospitals. The committee agreed that the following information is shared from the mother's ObsCIS:

- birth datetime
- maturity
- mode of delivery
- duration of membrane rupture
- birth weight
- Apgar score at 1, 5 & 10 minutes
- related diagnoses and procedures

For diagnoses and procedures, only those related to the fetus, the uterus, the membranes, liquor, and umbilical cord are passed to the baby's record. Examples include preterm labour, cord presentation, oligohydromnios, fetal distress, and fetal growth retardation.

Sharing mother's ObsCIS Data to baby's ePR & CMS

Each of the selected ObsCIS information to be shared is assigned with a unique EntityID and referenced to SNOMED CT where applicable. With the same ObsCIS diagnosis / procedure, separate HACVT terms are assigned for the mother's and the baby's records to ensure correct ICD codes are generated in the corresponding records.

The EntityID and TermID, plus the baby's episode number is passed to the ePR repository where the information is separately stored under both the mother's and the baby's episode number. Birth data and related diagnoses and procedures will be retrieved from the ePR repository and stored in the baby's CMS and ePR record.

Doctors can retrieve the baby's birth data while preparing the discharge summary (figure 2).

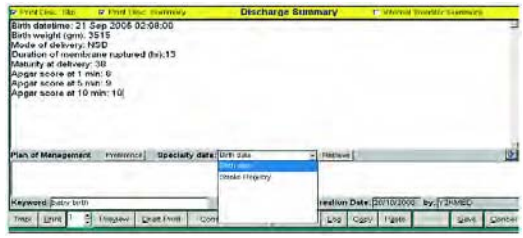


Figure 2 – Retrieving birth data for baby's discharge summary

A Section on Birth Record is created in the baby's ePR to display the baby's birth data (figure 3).

Figure 3 – Birth record in baby's ePR

Subsequent application

In 2006, a Neonatal Form was developed based on the IA and the same data sharing model. The ObsCIS data and mother's antenatal problems are shared to the Neonatal Form via the ePR repository.

The way forward

One of the greatest challenges in the use of the IA is to identify which Entities and which Contents should be referenced to terminologies like SNOMED CT. The CMS includes modules that are developed purely for patient care purpose but also some modules developed to facilitate auditing. Given the diversity, it is neither necessary nor worthwhile to reference all Entities and Contents to a Reference Terminology. Work started with defining Entities and Contents for ancillary services such as laboratory, and radiology – data will most likely contribute to the development of decision support systems – and has proceeded to other areas.

Despite SNOMED CT [16] being the single largest clinical terminology covering quite a number of domains, an internal evaluation of SNOMED CT indicated that it is better developed in areas like findings, but would need improvement in terminologies on others like procedures and allied health. More work, such as using post-coordination, is required to use SNOMED CT to represent concepts for the whole spectrum of healthcare domains. The requirement of post-coordinated terms also creates challenges for future data retrieval and analysis [17].

Given the massive number of Entities and Contents in the HA CMS, the need of a tool to manage these Entities and Content is inevitable. To this end, the Information Architecture Management System (IAMS) is being explored. IAMS aims to provide robust functionalities such as semantic locality and lexical matching to facilitate the management of terminologies. Functions supported

include mapping concepts to classification, managing the terminology lifecycle and managing reference information [18, 19]. In addition, IAMS will also manage relationship between individual IA elements.

Currently, the mother's ObsCIS data are shared to the baby's record at birth. For singleton, it is clear to whom the identified information belongs. More work would be required for linking correct fetal data to records for babies of multiple pregnancy.

Another challenge to be addressed is to protect the mother's privacy. The ObsCIS is adding a 'confidential flag' to indicate the mother particularly requested to keep the information confidential. This flag will also be passed to the baby's record to alert paediatricians to be more careful when discussing the baby's condition with a third party, e.g. baby's father, where applicable.

Conclusion

IA is a relatively simple but comprehensive concept when compared with other development in this area. Emphasizing on context, concept and presentation, the model can be flexibly applied to facilitate the building of a dynamic ePR while retaining the semantics of the data captured.

The HA will complete building a platform to share data with the Department of Health (DH) by March 2007. Discussion has started on sharing the newborn baby's discharge summary to the Maternal & Child Health Centre (MCHC) which is under the DH management. With nearly 90% of babies born in Hong Kong attending the MCHC, and 72% of babies born in HA hospitals, linking the HA and the MCHC baby's record marks another significant step in building a womb-to-tomb record for Hong Kong residents.

Acknowledgments

We would like to thank members of the Obstetric & Gynaecology Coordinating Committee with their contribution to the development of the Obstetrics Clinical Information System. We are grateful to the Perinatology Committee for their support in the design and implementation of linking mother's and the baby's record.

References

[1] ISO TC 215. 2005. ISO/TR 20514. Health informatics – electronic health record – definition, scope and context.
[2] Norgall T. Interoperability – a key infrastructure requirement for personalized health services. Studies in health technology and informatics. 2005;117:125-33.
[3] A strategy for building the National Health Information Infrastructure. National Committee on Vital and Health Services. Retrieved 4 May 2006 from website : http://aspe.hhs.gov/sp/NHII/Documents/NHIIReport2001/default.htm
[4] National Programme for IT, National Health Services. Retrieved 24 Oct 2006 from website : http://www.connectingforhealth.nhs.uk/
[5] Canada Health Infoway. 2004. Electronic Health Record (EHR) Standards Needs Analysis. Retrieved 20 Oct 2006 from website : http://www.infoway-inforoute.ca/en/WhatWeDo/Overview.aspx
[6] Connecting for Health. The Data Standards Working Group final report. Markle Foundations; 2003 Jun 5.
[7] Philip Aspden, Janet M. Corrigan, Julie Wolcott, Shari M. Erickson, Editors, Committee on Data Standards for Patient Safety. 2003. Patient safety: achieving a new standard for care. Institute of Medicine.
[8] Fung V. White paper on standardization of health data. Retrieved 1 Oct 2006 from Hong Kong Special Administrative Region, Office of Government Chief Information Officer website: http://www.ogcio.gov.hk/eng/infra/download/consortium_whitepaper_0609.pdf
[9] World Health Organisation. 1993. International statistical classification of diseases and related health problems, 10th revision.
[10] Cheung NT, Fung KW, Wong KC, Cheung A, Cheung J, Ho W, Cheung C, Shung E, Fung V, Fung H. Medical informatics – the state of the art in the Hospital Authority. International Journal of Medical Informatics. 2001 Jul;62 (2-3):113-9.
[11] Cheung NT, Fung V, Kong JH. The Hong Kong Hospital Authority's information architecture. Medinfo. 2004; 11 (Pt 2):1183-6.
[12] Fung V, Cheung NT, Szeto K, Ngai L, Lau M, Kong JHB. Hospital Authority Clinical Vocabulary Table – the past, the present, and the future. 2004 International Federation of Health Records Organization Congress Proceedings.
[13] Dolin RH, Alschuler L, Beebe C, Biron PV, Boyer SL, Essin D, Kimber E, Lincoln T, mattison JE. The HL7 Clinical Document Architecture, Journal of American Medical Informatics Association. 2001 Nov-Dec;8(6):552-69.
[14] ASTM. E1384-02a, Standard Practice for Content and Structure of the Electronic Health Record (EHR), 2006.
[15] Schloeffel P, Beale T, Heard S, Rowe D, Background and overview of the Good Electronic Health Record (GEHR), retrieved 15 Sept. 2003 from GEHR website: http://www.gehr.org/Documents/BackgroundOverview_
[16] SNOMED Clinical Terms, retrieved 20 Sept 2006 from SNOMED website : http://www.snomed.org/snomedct/documents/july_2006_release.pdf
[17] Dolin RH, Spackman KA, Markwell D. Selective retrieval of pre- and post-coordinated SNOMED concepts. Proceedings of the American Medical Informatics Association 2002 Annual Symposium.
[18] Chute CG, Elkin PL, Sheretz DD, Tuttle MS. Desiderata for a Clinical Terminology Server. Proceedings American Medical Informatics Association Symposium. 1999;:42-6.
[19] Rector AL, Solomon WD, Nowlan WA, Tush TW. A terminology server fro medical language and medical information systems. Methods of Information in Medicine. 1995 mar;34(1-2):147-57.

Address for correspondence

Vicky Fung
Rm 121N, Hospital Authority Building
147B Argyle Street
Kowloon
Hong Kong
Email: funghv@ha.org.hk

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Another HISA – The New Standard: Health Informatics – Service Architecture

Gunnar O. Klein[a], Pier Angelo Sottile[b], Frederik Endsleff[c]

[a] *Dept. of Medcine, Karolinska Institutet, and Cambio Healthcare Systems AB, Stockholm, Sweden*
[b] *GESI, Gestione Sistemi per l'Informatica, Rome, Italy*
[c] *Dept. of Informatics, H:S, Copenhagen, Denmark*

Abstract

In addition to the meaning as Health Informatics Society of Australia, HISA is the acronym used for the new European Standard: Health Informatics – Service Architecture.

This EN 12967 standard has been developed by CEN – the federation of 29 national standards bodies in Europe. This standard defines the essential elements of a Service Oriented Architecture and a methodology for localization particularly useful for large healthcare organizations.

It is based on the Open Distributed Processing (ODP) framework from ISO 10746 and contains the following parts:

Part 1: Enterprise viewpoint

Part 2: Information viewpoint

Part 3: Computational viewpoint

This standard is now also the starting point for the consideration for an International standard in ISO/TC 215.

The basic principles with a set of health specific middleware services as a common platform for various applications for regional health information systems, or large integrated hospital information systems, are well established following a previous prestandard. Examples of large scale deployments in Sweden, Denmark and Italy are described.

Keywords:

medical informatics, hospital information systems, standard, SOA, ODP, HISA, CEN, middleware, data-carrying integration platform

Introduction

Healthcare structure consists of networks of units over a territory characterised by a high degree of heterogeneity and diversity, from organisational, logistic, clinical, technological and cultural perspectives. The structure of individual centres such as hospitals of different sizes and outpatient clinics including primary care is evolving from a vertical, aggregated organisation towards the integration of a set of specialised functional areas. These need to share common information and to operate according to integrated workflows within and between centres of a larger enterprise e.g. a county or a large city.

On the one hand it is necessary to support the specific requirements of each unit or user in the most appropriate and cost-effective way whilst on the other hand it is vital to ensure the consistency and integration of the overall organisation, both at local and territorial level. This integration requirement is not only related to the need for improving clinical treatment of the subject of care but is also demanded by the necessity of to control and optimise the current level of expenditure for health.

It is today common with a large number of different databases and applications, isolated and incompatible, already operational in healthcare organisations to support specific needs of users. Even within the same centre, information systems are frequently fragmented across a number of applications, data and functionalities, isolated and scarcely consistent with each other.

A main need for care delivery organisations is to integrate and to make available the existing information assets, with the interoperability of existing applications, thereby protecting investments made. During integration activities, continuity of service needs to be achieved whilst gradual migration of existing proprietary, monolithic systems towards the new concepts of openness and modularity occurs. The cost-effectiveness of the solutions, especially when projected on the scale of the whole healthcare organisation, represents a crucial aspect to be evaluated carefully.

The goal can be achieved through a unified, open architecture based on a middleware of information services independent from specific applications and capable of integrating common data and business logic. These services shall be made available to diverse, multi-vendor applications through many types of implementations. All aspects (i.e. clinical, organizational and managerial) of the healthcare structure must be supported by the architecture. This must be able therefore to comprise all relevant information and all business workflows, structuring them according to criteria and paradigms independent from specific subdomain aspects, temporary requirements or vendor specific technological solutions.

Standards and technological solutions already exist and will continue being defined for supporting specific requirements, both in terms of *in situ* user operations and with respect to movement of information. The architecture must be able to accommodate such requirements by allowing the specific models to be integrated with the complete information assets of the healthcare organisation and the

communication messages to be "services" extracting or importing data from/to the common information as shown in Figure 1.

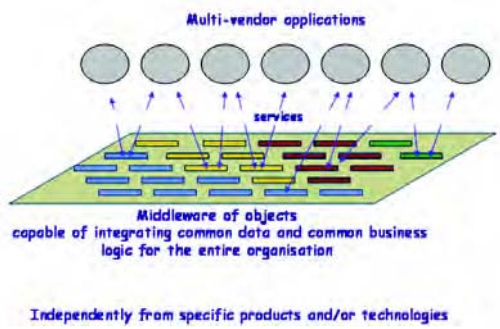


Figure 1 - A Service Oriented Architecture

The purpose of this standard is twofold:

- to identify a methodology to describe healthcare information systems through a language, notation and paradigms suitable to facilitate the planning, design and comparison of systems;
- to identify the fundamental architectural aspects enabling the openness, integration and interoperability of healthcare information systems.

The architecture is therefore intended as a basis both for working with existing systems as well as for the planning and construction of new systems.

Materials and methods

The standards body and process

This standard was developed within CEN/TC 251 (www.centc251.org) which is the technical committee for Health Informatics within the federation of 29 European national standard bodies.

The work in health informatics has been mandated by the European Union and the European Standard is published as a national standard in all of the member countries following the approval based on a weighted vote. In this case only one country was opposing the approval.

This work was developed within Working Group I: Information models by the Task Force HISA led by Gunnar Klein, who was at the time also chairman of CEN/TC 251 and with the two other authors as member of the core group with Frederik Endsleff having a particular responsibility for part 1 and Pier Angelo Sottile for parts 2 and 3.

Experts from the following countries also contributed actively to the development of this standard: Belgium, France, Germany, Hungary, Italy, Norway, Sweden and UK but also as non-European contributors Australia and the USA.

During the extensive formal review process in several stages a total of 112 written submissions where adding to the final version.

Open distributed processing

The standard framework for Open Distributed Processing was first developed by the Object Management Group (OMG) and later approved by ISO/IEC as the International Standard ISO 10746 from 1996.

This standard contains a rich set of specification elements and recommendations for the development of open distributed systems. It has successfully been applied in a large number of different industry sectors including, telecom and banking.

At the time of issuing the first ODP standard, there was no available standard for information modeling. Since then UML (the Unified Modeling Language) has been developed and since it has been adopted as the method of choice for health information modelling by both CEN and ISO and many other organizations, UML was selected as the information modelling language for the new HISA standard.

The ODP framework contains five viewpoints. In the HISA standard we have provided health specific advice and definitions for the three upper levels. The two lower viewpoints, Technology Viewpoint and Engineering Viewpoint are applicable for a concrete development and implementation project but there is no point in providing health specific restrictions in a standard.

Results

Part 1: Enterprise viewpoint

The Enterprise Viewpoint specifies a set of fundamental common requirements at the enterprise level with respect to the organisational purposes, scopes and policies that must be supported by the information and functionalities of the middleware. It also provides guidance on how one individual enterprise (e.g. a regional healthcare authority, a large hospital or any other where this model is applicable) may specify and document additional specific business requirements. This should aim at gradually achieving a complete specification, adequate for all the characteristics of that enterprise.

The strategic paradigm

The specification of the architecture shall start with a very concise, managerial-oriented document (the "Strategic Paradigm") that identifies (at a high level of abstraction) the overall requirements and strategic objectives of the envisaged system. It describes, in natural language:

- the rationale and the scope of the IT system with respect to the overall enterprise;
- the fundamental organisational processes (as defined under terms and definitions) that can be identified in the enterprise and that are relevant for the envisaged system;
- the fundamental constraints and objectives to be satisfied.

The HISA standard gives important guidance for an Enterprise on how to model its detailed requirements. The importance of starting from a model of the care processes

is emphasized and the following model is offered to clarify the different processes.

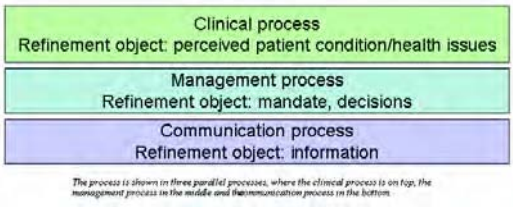


Figure 2 - Subdivision of the care process

The Enterprise Viewpoint deals with the requirement specification and contains *use-cases*, *process descriptions* and models leading to the identification of the overall Service Architecture and the basic clusters of objects. The following are identified:

- Subject of Care Workflow
- Activities management workflow
- Clinical Information Workflow
- Management of authorizations
- Management of resources
- Management of dictionaries and coding
- Interactions with other systems

In the standard, there are detailed use case descriptions specifying common requirements on information management.

It should be emphasized that this standard does not claim to have the description of all possible requirements, only the commonly shared processes of healthcare enterprises.

Part 2: Information viewpoint

The Enterprise Viewpoint of the HISA standard has identified certain processes. The Information Viewpoint is detailing the information structures using UML class diagrams. The information model is specified without any – explicit or implicit- assumption on the physical technologies, tools or solutions to be adopted for its physical implementation in the various target scenarios. The specification is nevertheless formal, complete and non-ambiguous enough to allow implementers to derive an efficient design of the system in the specific technological environment that will be selected for the physical implementation.

This specification does not aim at representing a fixed, complete, specification of all possible data that may be necessary for any requirement of any healthcare enterprise. The information model of the standard only specifies the set of characteristics identified as fundamental and common to all healthcare organisations.

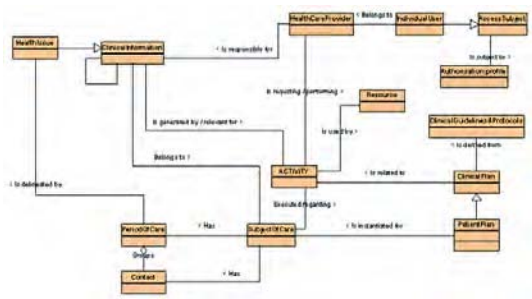


Figure 3 – High level Information Model

HISA Information Objects in each package shall be classified as *operational* or *descriptive*:

- "**Operational**", usually representing the actual (clinical, Organizational, etc.) objects that are continuously generated during (and for) the daily activities. These include the personal and healthcare treatment information on patients, the individual resources used for carrying out the actual activities, etc..
- The operational information objects model the entities involved in the daily activities of the healthcare enterprise in the treatment of subjects of care and in the functioning of the enterprise itself.
- "**Descriptive**", usually Organization-related, specifying the criteria according to which the Organization works and is organized. It includes general classifications of clinical concepts, rules according to which the activities are performed, and more (e.g. the types of activities which are carried out in the radiology department, the diagnostic classification in use in the clinical setting, etc.).

The descriptive information objects model the entities required for the overall knowledge base that is required by the healthcare enterprises to carry out daily activities related to the treatment of subjects of care and in the functioning of the enterprise itself.

For each "operational" information object, therefore, the model foresees one "descriptive" information object, containing the main classification data, the properties, the rules and the default values that are necessary for the management of the live data instantiated in the "operational" object.

The HISA information model is where relevant using existing standards. One example is the European EN 14822: Health Informatics - General Purpose Information Components which is based on the HL7 Reference Information Model.

Part 3: Computational viewpoint

This part of the standard specifies the fundamental characteristics of the computational model to be implemented by a specific architectural layer of the information system (i.e. the middleware) to provide a comprehensive and integrated interface to the common enterprise information and

to support the fundamental business processes of the healthcare organization.

The computational model provides the basis for ensuring consistency between different engineering and technology specifications (including programming languages and communication mechanisms) since they must be consistent with the same computational object model. This consistency allows open inter-working and portability of components in the resulting implementation.

The basic computational objects, corresponding to the information objects, will be equipped with *standard* lower-level **basic** interfaces having the scope of *adding, updating and deleting –in short maintaining-*, *listing*, and *getting one instance* of the main classes described in the information viewpoint. These basic methods allow the access to and the manipulation of each element of the underlying model and secure the openness of the system.

The higher-level computational objects implement more complex business transactions on the objects of the information model, simplifying and ensuring consistency of developments for common fundamental procedures of the organisation.

Examples are:

- Patient/person area, including registering a person, Patient Administration (ADT), merging patient identifiers, period of care, etc.
- Activity management and life cycle, including requests, planning, booking, etc.
- Clinical and EHC record, including terminologies, classifications, problem-orientation, etc.
- Resource management, including standard usages

Examples of use of HISA based architectures

In Europe there are a number of large healthcare organizations that have based their strategic planning on the HISA middleware principles and various technical solutions exist from different vendors and technical generations. Here are a few examples:

UppsalaCounty Council, Sweden

This is one of seven healthcare regions in Sweden that have decided to use the Cambio Spider middleware product. It is used for all its three hospitals and 35 primary care centres in one installation. This HISA implementation is based on a modern Java (J2EE) architecture with choice of Application Server and SQL database.

This system handles healthcare for all the 300 000 inhabitants plus as a regional highly specialised care for a million people. Today there are around 10 000 daily users of the system that includes, the following applications on top of the HISA platform:

- Care Administration
- Resource Planning
- Electronic Health Record
- Medication management including electronic transfer of prescriptions
- Order management for lab, imaging and consultancy services

The same system is also used in the counties of Kronoberg, Jönköping, Östergötland, Västmanland, Kalmar and Värmland in Sweden and also in Odense in Denmark and in the Faroe Islands.

The "Policlinico A.- Gemelli" in Rome "Università Cattolica del Sacro Cuore"

The UCSC information system consists of several applications based on a common architectural approach suitable to ensure the integrity and consistency of the information assets of the organisation, from the clinical, organisational and managerial point of view.

It relies on a healthcare-specific middleware product from GESI (the DHE®, Distributed Healthcare Environment) of services allowing different applications to access the common information heritage and to perform common business processes through a set of services. On top of the DHE, several applications provide specific support to the user activities.

The following figure 4 shows at a high level of abstraction, the overall heterogeneous structure of the healthcare information system of the hospital and the centrality of the HISA-based DHE middleware.

Figure 4 - The Gemelli system in Rome

Copenhagen Hospital Corporation (H:S), Denmark

H:S selected in 2002 the DHE product to implement its HISA strategy for 6 hospitals, in all comprising 4000 beds. The DHE serves as the joint data-carrying Integration Platform, forming the basis and common information heritage for a number of clinical and managerial applications.

Among the applications are the medication module and the master patient index with over 10.000 users, fully rolled out and in integral daily running operation throughout the organisation.

H:S is from 1 January 2007 being merged with a.o. Frederiksborg County and Copenhagen County, into the Capital Region, responsible for regional healthcare of all of the region, including 10 hospitals and 1.6 million inhabitants. The master patient index and the medication module is for this purpose currently undergoing rollout in the Frederiksborg County.

Several further DHE based applications are under current advanced development, such as Reporting, Accounting and the Patient portal, utilising the common information heritage.

Information is exchanged not only with the direct HISA interfaces (provided by the DHE) but also through messaging with other standards (utilising the underlying HISA interfaces) e.g.:

- The Danish national implementations of European standards in Edifact for e.g. laboratory communication and ePrescribing (MedCom).
- Information exchange with other messaging standards such as HL7, XML and Diagnostic Equipment such as for ECG and Imaging

Discussion

HISA is a new standard defining a "Service Architecture" identifying the general principles to secure openness and vendor-independence:

a) information must be separated from specific applications and accessible through services

b) service logic must be independent from technological issues (i.e. multiple technologies and mechanisms must be allowed for accessing the same services)

HISA is also identifying the fundamental elements of a comprehensive information model capable of supporting the whole healthcare organisation and finally the fundamental characteristics of a set of services for managing common information and for performing common business logic

This standard is not an alternative but a complement to other standards for health informatics such as specific messages developed by e.g. HL7 or the general EHR communication standards from CEN and openEHR [4].

Conclusion

After many years of research and standardization activities, there is now a formal standard for a Service Oriented Architeture (SOA) specific for the requirements of health care enterprises based on the general principles of Open Distributing Processing. It is now a European Standard but it has been submitted to ISO and is formally under consideration to become an International Standard.

Acknowledgements

This work has been made possible through a large number of expert contributions from many countries in different phases from 1993-2006. In the first phase during the development of the European prestandard, the work was financially supported by the European Commission.

References

[1] EN 12967-1: Health informatics – Service Architecture – Part 1: Enterprise Viewpoint.
[2] EN 12967-2:.Health informatics – Service Architecture – Part 2: Information Viewpoint.
[3] EN 12967-3: Health informatics – Service Architecture – Part 3: Computational Viewpoint.
[4] Blobel BG. Advanced EHR architectures-promises or reality. Methods Inf Med 2006; 45(1): 95-101.

Address for correspondence

Gunnar O Klein,
can be contacted at Karolinska Institutet,
SE 171 77 Stockholm, Sweden
E-mail: gunnar.klein@ki.se

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Protecting Privacy while Sharing Medical Data Between Regional Healthcare Entities

Tyrone Grandison[a], Srivatsava Ranjit Ganta[b], Uri Braun[c], James Kaufman[a]

[a] *IBM Almaden Research, 650 Harry Road, San Jose, California 95120*
[b] *Pennslyvania State University, University Park, PA 16802*
[c] *Harvard University, Cambridge, MA 02138*

Abstract

Economies of scale, corporate partnerships and a need to increase the efficiency of Information Technology in the Healthcare sector are leading to the construction of Regional Health Information Organizations (RHIOs) across the United States. RHIOs are normally aligned by service provision given by particular healthcare payers (e.g. Blue Cross-Blue Shield, PacifiCare etc.) in particular geographies. Globalization has created a transient workforce that may require their healthcare provider access their patient data across several sovereign RHIOs. The barrier to enabling RHIO to RHIO collaboration lies in the need to respect the data disclosure policy of each RHIO, to adhere to the geography-specific healthcare legislation and also to not violate the express privacy wishes of the patient(s) involved. In this paper, we propose a data-level control called Sticky Policy Enforcement which allows sharing to occur across RHIOs, while adhering to the concerns mentioned.

Keywords:

privacy, healthcare systems, collaboration

Introduction

Beyond the move to transform the American healthcare landscape by leveraging Information Technology to deliver better care, to reduce medical errors and to improve the quality of life, the existing healthcare topology dictates that computer networks be built that preserve the established business alliances that exist between payers and providers. This motivates the formation of connected information centers called Regional Health Information Organizations (RHIOs). The mandate to create a National Healthcare Information Network (NHIN) in the United States is based on the emerging existence of these RHIOs [1, 2]. A NHIN is a realization of a collaborative network of Healthcare Information Systems.

Formally, a RHIO is an independent regional collective that facilitates the development, implementation and application of secure health information exchange among participating care providers. Each RHIO has independent policies regarding the privacy of health records stored within the RHIO. Figure 1 presents the typical RHIO environment. A limited number of RHIOs exist today in the US and they vary in the ways they approach data sharing. The formation of a RHIO is based on the understanding that all the stakeholders agree to follow a specific set of guidelines.

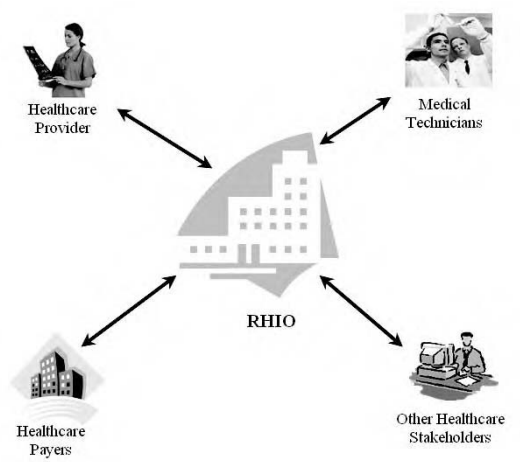


Figure 1 - Regional Health Information Organization (RHIO)

These guidelines detail policies regarding access, dissemination and processing of the patient data in the RHIO. Currently, the mechanism for cross-RHIO information sharing differ so greatly that collaboration across RHIOs to deliver care is still a daunting problem, but critical for the successful adoption of RHIOs.

One of the key challenges is the protection of patient privacy when clinical documents are shared. This challenge is compounded by the facts that there can be no assumption of a central authority; that enforcement may involve multiple privacy policies based on source, destination and the documents involved in the transfer and the fact that data can be forwarded to an entity with additional rights, such as remote update rights.

This paper presents a first step towards addressing these concerns and marches healthcare systems towards achieving inter-RHIO collaboration while adhering to data disclosure constraints (e.g. privacy and security concerns).

The technology is called *Sticky Policy Enforcement* and provides a way to ensure that policy constraints are enforced wherever patient data travels.

In this paper, we will provide a discussion on the related work in the field, highlight the *Sticky Policy* architecture, demonstrate the technology by walking through the enforcement steps and provide concluding remarks.

Related work

The first mention of *Sticky Policies* appeared in computer science literature at the start of the 21st century [3]. It emerged from IBMs work on Enterprise Privacy Authorization Language and was recognized as a concept that is important for privacy preservation in distributed computer systems. The underlying notion behind *Sticky Policy Enforcement* is that the policy applicable to a piece of data travels with it and is enforceable at each point it is used. Though identified over a half a decade ago as a critical problem, application-independent solutions that were technically feasible and scaleable were not realized.

Rivest and Lampsons work [4] embodies the earlier efforts in this space. Their focus was on the establishment of trust for a single disclosure object with a single policy. A data recipient is either granted access to the entire document, or must request authorization from the source. In healthcare environments, this is not sufficient. *Sticky policy* functionality should handle data disclosure to a party with well-defined constraints that allow data release to less privileged parties without requiring the originators involvement. This avoids the potential pitfall of having to contact a (potentially) large number of third parties before making a decision to disclose a specific piece of information.

The work by the Trusted Computing Group (TCG) consortium [5] represents another, more popular, approach to establishing the trust in single object, single policy environments. The general concern with the traditional approaches is that they require targeted application development and are not application and data agnostic, which is a mandatory requirement for situations with a complex web of pre-existing infrastructure, which are likely to be from differing vendors and running different, even proprietary, systems. Additionally, an ideal approach to *Sticky Policy Enforcement* should account for the fact that data changes occur frequently. It is not clear how approaches like that of the TCG would handle this without incurring a severe performance penalty.

System description

The goal of our system is to enable distributed privacy policy enforcement. The difficulty in achieving this lies in the fact that there is no single entity with *a priori* access to all the policy constraints applicable to a document in a given state of the system. Our solution to this problem involves identifying the applicable privacy policy constraints for a document(s) to be shared and sticking them together, forming a single entity of transfer.

In taking the approach of packaging policy with data, we maintain centralized decision making in a distributed enforcement. As only policy constraints that apply to the disclosed data are transferred, the communication impact is relatively small and the system does not require prior agreement among all medical organizations, states and patients.

Hippocratic database technology

Our solution approach to the distributed privacy policy enforcement problem leverages the Hippocratic Database (HDB) Active Enforcement (AE) technology [6], which provides cell-level, policy-based disclosure management functionality, such that databases only return data that is compliant with company policies, applicable legislation, and customer preferences. The AE component ensures that enterprise applications accessing a database adhere to fine-grained data disclosure policies. These policies, which may be security policies, privacy policies or data management policies, are distilled from the companys own policy, legal and regulatory requirements, customer preferences, and opt-in and opt-out choices. The component automatically rewrites user requests (i.e. queries) and returns only data that is consistent with these policies, allowing applications to enforce disclosure policies on arbitrary data elements at query execution time. A detailed description of HDB AE can be found in [7]. A quick overview of its operation is that, for a centralized or federated data system, HDB AE allows the definition of fine-grained disclosure policy, the creation of user preferences, the resolution of conflicts between preferences and policy and the enforcement of all the applicable constraints in an architecture that is application and database agnostic.

HDB Active Enforcement technology was chosen as our platform because: (i) it offers a general platform for handling and codifying privacy policy and preference information; (ii) its enforcement mechanism is transparent to enterprise applications (integration currently assumes a database interface such as ODBC or JDBC); (iii) it is agnostic to underlying database technology; (iv) it allows policy changes without any modifications to the applications in use; and (v) in the typical case, it improves query processing speed.

Sticky privacy policy

The format of our *Sticky Policy* package (Figure 2) consists of three parts: Data, Policy and Audit information. The Data segment contains the health documents requested. The Policy segment embodies the policy constraints applicable to the documents to be made available. The semantics of the policy entries are:

1. Requestor: The entity requesting access to part(s) of clinical document(s) from the source. The values for this entry could be taken from the roles mentioned in CDA standard.
2. Recipient: The entity that will be the final consumer of the data. The domain of possible values is similar to the set used for a requestor.
3. Purpose: The purpose for which the document(s) is being requested.
4. Retention: Time period until which access to the data is allowed. This could be computed based on various organizational policies.

5. Copy-forward: The condition specifying whether the recipient is entitled to forward the requested document(s) to a third party after copying. The set of possible values are:

Copy forward	
Yes w/notification	*May copy and forward the data with a notification to the sender*
Yes w/o notification	*May copy and forward the data without any notification*
No	*May not forward at all*
Ask	*Must ask the sender on forward*

6. Append/Modify: The Boolean condition specifying whether the recipient can append/modify the document.

Append/Modify	
Yes w/notification	*May append/modify with a notification*
Yes w/o notification	*May append/modify without any notification*
No	*May not append/modify at all*
Ask	*Must ask on append/ modify attempt*

The Audit section of the sticky policy consists of information including the source, requestor, a timestamp and digital signature to verify the authenticity of the sticky policy.

Figure 2 - Format of sticky privacy policy

Data	
×	Clinical Document 1
×	..
×	Clinical Document n
Policy	
×	Requestor
×	Recipient
×	Purpose
×	Retention
×	Copy forward
×	Append Notification
Audit	
×	From
×	To
×	Timestamp
×	Verifiable Signature

The source and requestor information is used by the auditor to track the data while the timestamp is used to determine the causal ordering. The digital signature serves two purposes: 1) to maintain the integrity of the healthcare documents, i.e. guaranteeing that the document(s) are not tampered with, and, 2) to ensure the non-repudiation of the sticky policy.

Architecture & model

In solving the distributed policy enforcement problem, we assume a trust model where authenticated users are honest-but-curious. Thus, our attack model focuses on the user who inspects the data they receive and attempts to gather data that they are not entitled to, and does not address malicious users who attempt to gain access to data they have not received even if it violates the policies.

Figure 3 illustrates the conceptual architecture of our system and showcases the creation and management of sticky policies.

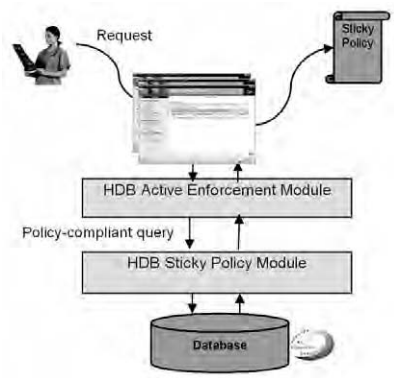


Figure 3 - Sticky policy enforcement architecture

Our enforcement model consists of two approaches:

Proactive enforcement involves prevention of unauthorized disclosure before it occurs by blocking operations or suppressing results that may lead to a privacy violation.

Reactive enforcement involves detection of the violation through audits. This is based on an optimistic assumption that the environment is non-malicious.

These approaches differ in their points of execution. Proactive enforcement eliminates violations before they occur, which has limited use in scenarios when *a priori* knowledge of all the possible access situations is not possible. For example, in an emergency situation, violations should be allowed as long as they are auditable and necessary for the delivery of care. Reactive enforcement achieves this by tracking all the access information and assuming the existence of a trusted auditor system. The auditor must be able to access data from the source and recipient including any intermediaries between those parties. For the US healthcare environment, this could be the responsibility of the Department of Health and Human Services or its delegates. In the case of a violation, the system presumes the node to be guilty and demands records to certify innocence.

For distributed environments, proactive and reactive enforcement can be achieved through either a centralized or a federated approach leveraging a set of cooperating enforcers.

In our design, we perform proactive enforcement through HDB Active Enforcement technology. In the case of reactive enforcement, a centralized approach employs a single auditor which is trusted and authorized to investigate all aspects of a suspected policy violation. In a decentralized model, a set of cooperating auditors can be employed with each team responsible for a specific set of nodes, data or both.

Beyond the issues of proactive and reactive enforcement, there is a question of where the enforcement occurs. We consider two possible locations: at the source and at the recipient. Enforcement at the source is simpler because it relies on the sources controls. Enforcement at the recipient places trust in all recipients. However, some enforcement can only occur at the recipient. For example, restricting the recipient from forwarding the result on to others is not something the source can enforce. Our approach is to attempt to perform all enforcement at the source and only rely on enforcement at the recipient when no alternative exists.

Scenario

Initially, a request for clinical documents is placed through an application interface, which issues a query or a set of queries. On receiving the query, the proactive enforcement module rewrites the query to account for all applicable disclosure policies. In a traditional HDB-enabled system, the rewritten query would continue directly to the data systems native query execution engine. However, for our system, the rewritten query is redirected to the sticky policy module where the query is further modified to create a sticky policy result set. A digital signature is computed on the source using a User-Defined Function (UDF) and included in the sticky policy. The final sticky policy is then transferred to the recipient in its entirety. Figure 4 provides a sample healthcare sticky policy.

For this prototype, we chose to use XML as the data format for representing the sticky policy for multiple reasons. The interoperability among various EMR, HER and PHR systems used in the healthcare industry has been limited based on the proprietary interfaces and standards [8]. Although, XML processing as in the case of any text processing involves a lot of overhead, it offers the much needed features of platform independence and simplicity demanded by the healthcare industry. When the XML processing overhead becomes unbearable, specialized hardware and software can be used to mitigate this concern.

For the purposes of this work, we assumed an agreement on vocabulary among interoperating parties. We also assume that proactive enforcement is achieved by leveraging HDB Active Enforcement. On the recipient, the AE component accepts the received sticky policy, assigns a unique id to the policy and stores an unaltered copy of policy for the purposes of auditing. The policy elements are then de-coupled and the corresponding data and policy constraints are extracted. The policy rules are then entered into the HDB metadata tables. The document(s) or their part(s) are then stored in a database and links are created from the entries in the HDB metadata

Reactive enforcement is achieved by traversing the sticky policy audit logs to find a violation or to prove innocence. An audit begins with the suspicion of a privacy violation. We presume that any party with access to the data but without a sticky policy is guilty. In essence, the sticky policy is a certificate of innocence. This is similar to not having a license demonstrating legal ownership of a software product.

The auditor searches the database for the data item for which enforcement is presumed as violated. Once identified, the auditor checks the HDB metadata, and identifies the relevant policy and archive entries. If the ability to identify these entities does not exist (i.e. there is no policy or the archive entry is missing), then a violation has occurred. The auditor then tries to verify the signature on the sticky policy stored in the archive table. Again if the signature is not valid, a violation has occurred.

Figure 4 - Sample sticky policy for healthcare RHIO sharing

<ClinicalDocument 1 ...>

<patient><name><given>James</given><family>Beach</family><suffix>PhD.</suffix></name></assignedPerson>... <section><title>History of Present Illness</title><text>James Woods has suffered from calcific bursitis.</text>... <title>Vital Signs</title><text>The patient's height, weight, and body mass index were measured to be 2.29489 meters, 400.05671738536824 pounds, and 34.45 kilograms per meter squared, respectively.</text> ...

<Clinicaldocument2 ...> ...

Requestor - Alice/ Admin Role/ ArizonaCare

Recipient - Bob/ Staff Physician/ ArizonaCare

Purpose - Emergency Case

Retention - 30 days

Copy-Forward - Yes With Notification

Append/Modify - No

From - Trina/ Admin Role/ CalShield

To - Alice/ Admin Role/ Arizona Care

Timestamp - Nov 19 2006

Signature - ...

If everything has been successful so far, the auditor checks to make sure the sticky policy and data content agree. The auditor compares the policy in the policy table to its counterpart in the sticky policy and similarly compares the data in the sticky policy with that in the database. The auditor also verifies that the HDB metadata tables cover all the data included in the sticky policy. Even if everything checked so far is OK, the audit is not stopped.

It is possible for an enforcement breach to have occurred before the current node even received the sticky policy. The auditor then traverses through the sticky policy to identify the node that forwarded the sticky policy. The auditor continues the audit up the chain to the originators

until he reaches the bounds of his jurisdiction or the first sender.

Performance discussion

We ran experiments to evaluate the overhead cost of sticky policy generation and sticky policy consumption. In the interest of terseness, we will just provide the results here. Our experimental platform used a synthetically generated dataset based on the Clinical Document Architecture [9]. All experiments were run using IBM DB2 UDB 8.2. The operating system was Microsoft Windows XP with Service Pack 2. The hardware consisted of a PC with Pentium-4 2.4GHz processor and a 60GB disk. The buffer pool was set to 1 MB. All other DB2 default settings were used.

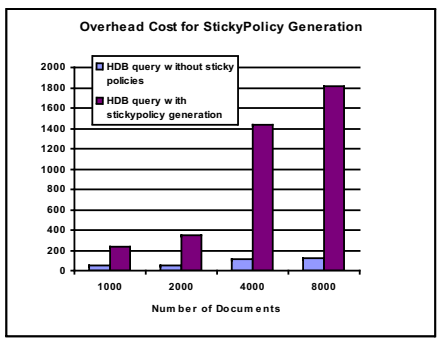


Figure 5 - Overhead cost for sticky policy generation

We examined document sharing for sets of 1000, 2000, 4000 and 8000 documents, which is well over the current limits for healthcare document sharing. It was observed that the overall cost introduced by sticky policy generation over the privacy preserving query processing in HDB is acceptable (Figure 5) considering that the generation is done using XML.

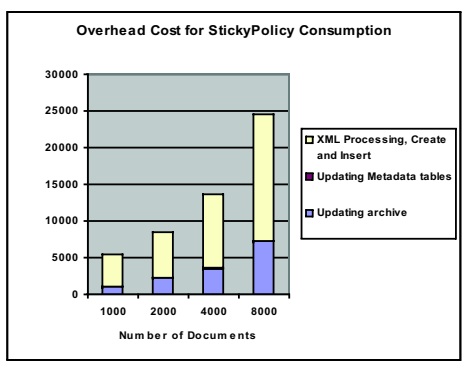


Figure 6 - Overhead Cost for Sticky Policy Consumption

For policy consumption, the time elapsed in updating the metadata tables and the archive is less than 30% of the overall policy consumption cost (Figure 6). As pointed earlier, XML hardware appliances may be used to reduce the overall consumption cost.

Future work

Our future efforts will focus on 1) further deployment and enhancement of the current *Sticky Policy* functionality, and 2) innovating new approaches to enable *Sticky Policy Enforcement* in distributed systems with a central repository. On the first task, we plan to start by including technology that increases the transmission security strength. i.e. removing the assumption that the channel is inherently secure and encrypting the shared documents before transfer. Then we will remove the assumption of an honest-but-curious user and consider hostile environments; touching on the difficulties in considering Byzantine failures or collusion among several participants. For the second task, we will assume a system of shared policies and construct mechanisms to provide privacy guarantees when data is transferred.

Conclusions

The construction of RHIOs and the sharing of information between them is an important pre-requisite for the successful creation and deployment of a National Healthcare Information Network (NHIN). Very little attention has been placed on technology to enable this RHIO to RHIO collaboration in a privacy preserving manner, till now. In this paper, we present *Sticky Policy Enforcement* technology, which provides mechanisms to perform proactive and reactive enforcement when sharing clinical documents between RHIOs.

References

[1] The Goals of Strategic Network. http://www.hhs.gov/healthit/goals.html.

[2] The California Regional Health Information Organization. http://www.calrhio.org/.

[3] Stufflebeam, William, Antn, Annie I., He, Qingfeng, and Neha Jain. Specifying Privacy Policies with P3P and EPAL: Lessons Learned. Proceedings of the 2004 ACM workshop on Privacy, Washington DC, USA (2004).

[4] Rivest, Ronald L., and Lampson, Butler. SDSI A Simple Distributed Security Infrastructure. Working Document v1.1

[5] TCG Specification Architecture Overview. Revision 1.2 Retrieved 2006 from https://www.trustedcomputinggroup.org/specs/IWG/TCG_1_0_Architecture_Overview.pdf

[6] K. Lefevre, R. Agrawal, V. Ercegovac, R. Ramakrishnan, Y. Xu, D. DeWitt. "Limiting Disclosure in Hippocratic Databases". Proc. of the 30th Int'l Conf. on Very Large Databases (VLDB 2004), Toronto, Canada, August 2004

[7] IBM Almaden Research. Hippocratic Database Active Enforcement: User's Guide. Retrieved Nov. 28, 2006 from http://www.almaden.ibm.com/cs/projects/iis/hdb/Publications/papers/HDBEnforcementUserGuide.pdf

[8] Srivatsava Ranjit Ganta, Eishay Smith, Sarah Knoop, Sondra Renly, James Kaufman. The Eclipse Open Health Framework, 5th International Conference on Healthcare Technology and Management (HCTM 2006), Chicago, August 2006.

[9] Dolin RH, Alschuler L, Boyer S, Beebe C, Behlen FM, Biron PV, Shabo Shvo A, HL7 Clinical Document Architecture, Release 2, Journal of American Medical Informatics Association (JAMIA), 2006

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Pulling Back the Covers: Technical Lessons of a Real-World Health Information Exchange

Atif Zafar, M.D.[a,b], Brian E. Dixon, M.P.A.[a]

[a]*Regenstrief Institute, Inc., Indiana*
[b]*School of Medicine, Indiana University, Indiana*

Abstract

Several nations and local communities are striving to achieve widespread, secure exchange of clinical data between various health care providers and public health organizations. Most of the literature on health information exchange focuses on the financial, political, and privacy aspects of these initiatives. Perhaps just as important are the technical and organizational factors that have influenced development of data exchange methods and results.

One mature network in the Midwestern United States has had success in establishing consistent, secure exchange of clinical data for more than ten years. Presented here are the technical lessons learned and design decisions made from this initiative with the hope that they can be used by others striving to connect disparate clinical information systems for the improvement of health care quality and safety.

Keywords:

computerized medical records systems,
computer communication networks, equipment reuse

Introduction

Everything in health care ultimately revolves around the accessibility and effective use of clinical data. When physicians, nurses, and other health care professionals have the information they need when they need it, they serve patients better, in terms of both the quality and the safety of the care they provide. Making these data available electronically, then, appears to make good sense.

Electronic clinical data abound. The problem is that they are often inaccessible to providers, because health care organizations tend to house their clinical data in distinct, isolated repositories. Many providers and policy makers now recognize that the sharing of data among hospitals, doctors, and other health care organizations in a given city, state, or region often referred to as health information exchange (HIE) can make health care safer, more efficient, and more effective [1].

The Indiana Network for Patient Care (INPC)

Indianapolis has pioneered an extremely successful HIE initiative, the Indiana Network for Patient Care (INPC), launched in 1993 under the leadership of the Regenstrief Institute. Initially, the INPC provided data from one hospital to providers in emergency departments at three other hospitals. By 2005, a more mature network, with a membership comprising 95 percent of all hospital and emergency care in Indianapolis, expanded to include providers in other parts of the state. By the end of 2006, the INPC contained more than six million distinct patient registration records, 850 million discrete observations, 17 million text reports, 50 million radiology images, and 40 million orders.

A number of factors have contributed to the INPCs success, including political and legal dynamics, which have been addressed in other publications [2][3][4]. Here we outline the reasons for the technical design decisions and functionality of the INPC, highlighting the technological and organizational factors that have contributed to the networks growth, ease of use, and sustainability.

Technological factors

Regenstrief has examined, deployed, refined, and evaluated a variety of operating systems, programming languages, software applications, and database management systems over its thirty-year history. The philosophical approach has been to *select a technology, stick to it, and make it work*. This means we rarely make radical changes (e.g., redesign a program written in one language using another language) unless there is a clear need (e.g., the new language is far superior to the old one). For example, early use of the Web (prior to 1995) for results aggregation brought many challenges. To enable asynchronous communication between clients and servers, we developed customized tools that possessed modern asynchronous JavaScript and XML (AJAX) functionality. Only recently have we begun to redesign our tools to utilize current AJAX frameworks.

Although sometimes slow to change, *our organization is not afraid to experiment*. We have worked with state-of-the-art image-compression technologies, such as JPEG

2000 and Wavelets, and motion video (MPEG) for display of radiology images and cardiac echo movies, respectively, in Web browsers. We have also experimented with voice-recognition and voice-annotation for clinical notes and display of laboratory results on mobile devices in wireless settings. Currently, we are experimenting with nomadic computing technologies so that clinicians can have an access-anywhere system for clinical data.

The INPC is composed of many moving parts, some of which are legacies while others are more modern. It is not the formula of a certain operating system with a specific database management system that has produced success for the INPC. Rather, the technical success of the INPC may be attributed to its adherence to these philosophies and the principle that, *when possible, one should build upon existing infrastructures rather than inventing or implementing new ones*. Below we discuss this principle in the context of the INPCs security, speed, flexibility, and reusability.

Security

Secure exchange of information between the INPC and participants operates using point-to-point connections. In the past, the INPC has employed T1 lines (data pipes as weve referred to them in other publications). We are now phasing out T1 lines in favor of virtual private networks (VPNs).

INPC security policy dictates the use of up-to-date encryption methods and good password hygiene and RSA keys. Currently the INPC relies on 128-bit SSL encryption to protect data on the rare occasion we use the public Internet to exchange information. Passwords must be changed at specified intervals, require a certain combination of letters and numbers, have a minimum length, and cannot be reused by the same individual.

The INPC also requires users to sign a confidentiality agreement and devices to be equipped with time-limit controls to prohibit unauthorized access. However, users do not log in directly to the INPC. User authentication is done via providers. First, users login to a provider portal or local area network (LAN), then they access INPC applications through a gateway created between the provider and the INPC network. The INPC relies on providers to implement access and time-limit controls on devices and ensure that users have signed a confidentiality agreement, which is required of them anyway to access the providers electronic resources.

Federated data sharing model

At its core, the INPC is a series of federated vaults, sometimes referred to as edge proxies or silos, storing data from the various participating institutions. Each institution has its own privileged silo where only data from that institution resides. The architecture of each silo closely resembles that of the Regenstrief Medical Record System (RMRS), a well known electronic medical record system [5]. A simplified data model of the RMRS is presented in Figure 1.

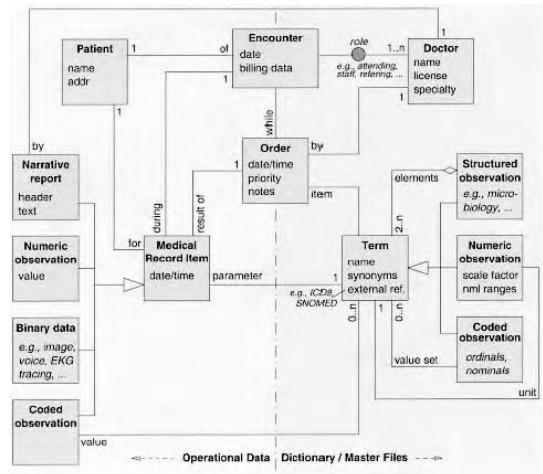


Fig. 1. Simplified data model for Regenstrief EMR.

Figure 1 - Simplified data model for the Regenstrief Medical Record System (RMRS)

Each silo represents mirrored data from one participating institution. Patient registry data, such as name, medical record number, and date of birth, and clinical data, like laboratory results, immunizations, and free-text notes recorded by the doctor during an encounter (e.g., clinical visit), are stored in the silos.

Silos can be created technically using a variety of methods. Hierarchical databases using large flat files running on clustered VMS nodes have been employed in the past. More modern relational databases can also be used to develop silos. Each silo can function as a separate database, or indices can partition data stored in a table on separate physical disks.

The specific technologies employed to create silos are not as important as the concept. Creating federated vaults gives participants peace of mind that their data will be segregated and secure. Yet data in federated vaults can exist within a single network access storage (NAS) unit, reducing latency when retrieving data during a clinical encounter. To date, the centralized, federated model developed by Regenstrief has yielded better performance (speed) when compared to decentralized federated networks used by other HIE initiatives.

A centralized, federated model also simplifies the process of data standardization ensuring identical blood urea nitrogen (BUN) results from various laboratories are interpreted the same way. The burden of reconciling various tests is shifted from individual provider organizations to Regenstrief. We can employ a single data model and dictionary and resolve errors as they arrive at the central hub for processing. Troubleshooting and mapping data elements to the standardized model requires overhead approximately 1-3 FTEs for the volume of messages we process. However, the costs and occasional painstaking mapping efforts are justified, because *we value the quality of the data stored in the INPC*. Our methods ensure that the

data retrieved from the INPC is reliable and standardized, which builds trust amongst network users and makes retrieval and delivery methods efficient.

Standards

Since its inception, the INPC has strived to provide optimal access to relevant clinical data at the point of care. To that end, the INPC has invested significant time and resources into the development and use of health information technology (health IT) standards (e.g., HL7, LOINC, CPT, etc.). These standards permit disparate systems to share data among one another, making them *interoperable*. They also permit the INPC to quickly add new types of data by reducing the time required to create customized interfaces for information delivery to providers.

Standards also enable data reusability, the ability to store a single concept and use it multiple times in a variety of applications. For example, a physician may order an HIV 1 AB (LOINC #7917-8) to indicate the presence of HIV in a patient. Once the result is reported to the INPC from the lab, three separate actions can be taken using the same HL7 message and LOINC code. First, the result would be stored in the institutional silo corresponding to the provider identified in the message. Second, the result could be delivered electronically to the physician using a clinical messaging application. Finally, the result could be reported to one or more public health agencies.

The example demonstrates that a single element, a standardized clinical message, can be used by three very different components of the INPC to store and exchange clinical information. Standards are employed so that the provider, physician, and health department interpret the result in the same way (e.g., all three receive a message indicating a positive value for LOINC #7917-8). This reuse of the same data is efficient, flexible, and cost-effective.

At the time of inception for the INPC, standards were *immature* and *limited*. We had to *invent* standardized methods for transmitting and mapping data between networked provider organizations. One such invention, the LOINC standard, was created because SNOMED and other existing terminologies lacked breadth for laboratory and some clinical concepts. We hope that other organizations can benefit from early experiments by us and other organizations. We believe that *field tested standards*, such as HL7 2.x and LOINC, can help others create interconnected systems in less than half the time it has taken us to develop the INPC.

Although we are experimenting with HL7 Version 3, current INPC members continue to transmit data using HL7 2.x. We encourage continued development and refinement of standards, and we will support them as they mature and become adopted by INPC participants.

Applications

An important lesson learned from building an **aggregated**, **standardized** data repository is that *data can then be reused for many applications*. For example, the same dataset that is reported out to clinicians using a clinical messaging application, we call ours DOCS4DOCS, can be sent to the State Health Department for communicable disease reporting. Similarly, data received from private practices could be aggregated and presented to an ER physician for delivering emergency care. These compelling applications allow INPC *stakeholders* to get some value-added by joining the collaborative, with an understanding that *their data* will only be used according to the agreed upon terms within the data-sharing contract.

Component-based architecture

The INPC employs a number of component technologies to process much of the data that travel across the network. Technologies like *interface engines*, *message processors*, and a *global patient and provider index* perform specific tasks that are generic enough to be re-used from application to application. Thus each component can be optimized for its task and easily modified to include a newly interfaced system. This creates a network in which components are not only interoperable but also reusable.

The idea of developing and reusing components is not unique to Regenstrief or the INPC. The object-oriented paradigm has influenced software development practices, with conventional modular techniques abandoned in favor of component-based approaches [6]. This is especially true in the open source software movement [7].

Regenstrief has embodied the philosophy of component-based development into the INPC, which has enabled the network to remain flexible. Expansion over the last thirteen years has involved the addition of new participants (e.g., hospitals, laboratories), new applications that use the data for a variety of tasks, and new forms of data (e.g., we added pathology reports in 2003 and dictated notes in 2005). With each new addition, the network has required slight modification. Development time is shorter, because components can be reconfigured and redeployed faster than monolithic programs. New components can be developed more quickly, because insertion into the network does not require recompilation or reconfiguration of other components.

Organizational factors

Designing, constructing, and operation of a working technical infrastructure for interoperable exchange do not guarantee success. In addition to its technical infrastructure successes, the INPC has also benefited from a number of organizational factors that have shaped its development over the last 13 years.

Incremental evolution

Incremental change has played a significant role in the INPCs long-term success. What began as an experiment to connect emergency rooms together slowly evolved into a large network that provides clinical information to emergency rooms, hospital staff in other departments, and ambulatory providers. This growth was guided by steady leadership that focused consistently on the INPC vision rather than on trends in the budding HIE industry.

Some HIE projects can and will evolve more quickly than the INPC. However, leaders of such projects should temper expansion with a clear vision for their network and agreement from all their partners.

Human resources

Technology is not the single most critical factor for successful HIE. To succeed in the development of a broad clinical data exchange, the INPC employed many capable people to manage and support the technology of the network. A knowledgeable staff is necessary on both ends of the network, at each participating organization as well as at the data exchange entity. Technical difficulties and bugs are inevitable, so capable humans are needed to troubleshoot errors, resolve data issues, and continue to move the vision of the exchange forward. And as the INPC has grown, so too has its need for more staff members to effectively monitor all of its members relations and data connections.

For example, we recently had a lab send us an HL7 message using unexpected units (up/mL). The INPC exception processor detected the anomaly (unrecognized units), which resulted in 26,000 records being dumped into an exception queue for analysis. Turns out the lab system had an embedded typo (the units should have been ug/mL), and the problem was resolved after a phone call and a few emails.

A more common problem we face is reporting of units in any other field, usually the notes field, except the appropriate HL7 units field (OBX-6). This is a problem common to all the labs from which we receive data, and it is a recurring problem for newly created tests.

Exceptions require human intervention as subtle differences between common clinical concepts are difficult for computers to resolve, despite several attempts in the past [8][9].Given the need for regular human intervention, we employ 2-3 FTEs to constantly monitor and troubleshoot the more than 150 message streams from the major hospital systems in Indianapolis, regional referral laboratories, specialty providers, several rural providers throughout the state of Indiana. We predict the need to add trained personnel in the future as the INPC continues to grow.

Our human resources also drive innovation. For years we have benefited from the talents of National Library of Medicine (NLM) informatics fellows typically post-doctoral physicians. These individuals have been key players in designing, creating, testing, and evaluating innovative components of the INPC infrastructure, including add-on programs such as CHICA™ and PHESS™ that extend the INPC beyond clinical messaging. Many of these fellows have stayed on as faculty at Regenstrief and the Indiana University School of Medicine, continuing to enhance the INPC and mentoring new fellows.

Sustainability

The ability of senior leadership to repeatedly make a clear, evidence-based business case for the INPC has contributed significantly to its sustainability. Initial funding for the INPC came from the NLMs high-performance computing and communication initiative. Subsequent funding has come from the NLM, the Agency for Healthcare Research and Quality (AHRQ), the Health Resources and Services Administration (HRSA), the National Cancer Institute (NCI), the Indiana Genomics Initiative, and the Indiana Twenty-First-Century Fund [2]. Each grant supporting a portion of the INPCs development enabled Regenstrief to measure clinical, financial, and community outcomes. These data provided support to the INPCs business case, which allowed the network to secure additional funding for expansion of existing services and development of new ones.

In addition, the networks interoperable, flexible design supports a variety of clinical and research activities. Applications such as CareWeb™, DOCS4DOCS, the Shared Pathology Information Network (SPIN) anonymous query tool, and the Public Health Emergency Surveillance System (PHESS™) for syndromic surveillance build on the INPCs core infrastructure. However, each creates a unique service for all or specific network members. An innovative, legally separate organization, the Indiana Health Information Exchange (IHIE), has also capitalized on the INPC infrastructure, creating a highly reliable, customer-oriented organization to support care delivery organizations using Regenstrief technology. By vertically expanding the INPC in this way, senior leadership has successfully created new resource opportunities for the INPC to grow and improve.

Conclusion

Reliable, up-to-date clinical data at the point of care remain the key to improving both the quality and safety of health care. Successful exchange of clinical data occurs only when all participating providers and organizations share not only data, but an understanding of what those data mean. Incremental change and growth are key to the success of data exchange networks. Over time, effective networks tend to expand the types of data they carry, as well as the applications for those data. Standards and reusable components help HIEs to maximize their efficiency through shorter development time and lower costs creating opportunities for integration with new systems and organizations. If other exchanges are as successful as the INPC, valuable improvements in care will be achieved in many communities.

Acknowledgments

This work was produced under Agency for Healthcare Research and Quality contract no. 290-04-0016. The authors thank staff at Regenstrief Institute, Inc. and the Indiana Health Information Exchange (IHIE) for their contributions to this work and tireless pursuit of health information exchange to improve the quality, safety, and efficiency of health care.

References

[1] Walker J, Pan E, Johnston D, Adler-Milstein J, Bates DW, and Middleton B. The value of health care information exchange and interoperability. *Health Affairs*. 2005

[cited 2006 May 1]. Available from: http://content.healthaffairs.org/cgi/content/abstract/hthaff.w5.10

[2] [2] Biondich PG and Grannis SJ. The Indiana network for patient care: An integrated clinical information system informed by over thirty years of experience. *J Public Health Mgmt*. 2004 Nov; 10 Suppl 1:S81-6.

[3] McDonald CM, Overhage JM, Barnes M, Schadow G, Blevins L, Dexter PR, Mamlin B, and the INPC Management Committee. The Indiana network for patient care: A working local health information infrastructure. *Health Affairs*. 2005; 24(5): 1214-1220.

[4] Sears C, Prescott VM, and McDonald CJ. The Indiana network for patient care: A case study of a successful healthcare data sharing agreement. *ABA Health eSource* [serial on the Internet]. 2005 [cited 2006 Nov 29]; 2(1). Available from: http://www.abanet.org/health/esource/vol2no1/Sears.pdf.

[5] McDonald CJ, Overhage JM, Tierney WM, Dexter PR, Martin DK, Suico JG, Zafar A, Schadow G, Blevins L, Glazener T, Meeks-Johnson J, Lemmon L, Warvel J, Porterfield B, Warvel J, Cassidy P, Lindbergh D, Belsito A, Tucker M, Williams B, Wodniak C. The Regenstrief Medical Record System: a quarter century experience. *Int J Med Inform*. 1999 Jun; 54(3):225-53.

[6] Waguespack L and Schiano WT. Component-based IS architecture. *Information Systems Management*. 2004; 21(3): 53-60.

[7] Madanmohan TR and Rahul D. Open source reuse in commercial firms. *IEEE Software*. 2004 Nov-Dec; 21(6): 62-69.

[8] Schadow G, McDonald CJ, Suico JG, Fohring U, and Tolxdorff T. Units of measure in clinical information systems. *J Am Med Inform Assoc*. 1999 Mar-Apr; 6(2): 151-62.

[9] Friedman C and Johnson SB. Natural language and text processing in biomedicine. In: Shortliffe EH and Cimino JJ, eds. Biomedical Informatics. 3rd ed. New York: Springer, 2006.

Address for correspondence

Brian E. Dixon, M.P.A.
Regenstrief Institute, Inc.
410 West 10th Street, Suite 2000
Indianapolis, IN 46202
U.S.A.
(317) 423-5582
bdixon@regenstrief.org

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Geographically Distributed Complementary Content-Based Image Retrieval Systems for Biomedical Image Informatics

Sameer K. Antani[a], Thomas M. Deserno[a,b], L. Rodney Long[a], George R. Thoma[a]

[a] *National Library of Medicine, National Institutes of Health, Bethesda, MD, USA*
[b] *Department of Medical Informatics, Aachen University of Technology (RWTH), Aachen, Germany*

Abstract

There is a significant increase in the use of medical images in clinical medicine, disease research, and education. While the literature lists several successful systems for content-based image retrieval and image management methods, they have been unable to make significant inroads in routine medical informatics. This can be attributed to the following: (i) the challenging nature of medical images, (ii) need for specialized methods specific to each image type and detail, (iii) lack of advances in image indexing methods, and (iv) lack of a uniform data and resource exchange framework between complementary systems. Most systems tend to focus on varying degrees of the first two items, making them very versatile in a small sampling of the variety of medical images but unable to share their strengths. This paper proposes to overcome these shortcomings by defining a data and resource exchange framework using open standards and software to develop geographically distributed toolkits. As proof-of-concept, we describe the coupling of two complementary geographically separated systems: the IRMA system at Aachen University of Technology in Germany, and the SPIRS system at the U. S. National Library of Medicine in the United States of America.

Keywords:

medical informatics applications, image information storage and retrieval, Internet services

Introduction

There has been an explosive growth in the acquisition of medical images for clinical diagnosis, and use in medical research and education [1]. Hospitals have been adopting technology such as Picture Archiving and Communication Systems (PACS) and Hospital Information Systems (HIS) to assist in the digital collection, organization, and storage of patient data. The goal of these systems is to make patient data more accessible; in reality, the amount of data that is entered and stored in these systems have created new challenges in effective information indexing and retrieval. Retrieval of image information from these systems is done using limited text keywords in special fields (e.g., unique patient identifier, fields in the image header). These keywords, however, do not capture the richness of features depicted in the image itself. It would be beneficial if the images could be retrieved by their visual content to help improve research, education, or medical practice. Content-Based Image Retrieval (CBIR) has received significant attention in the literature as a promising technique to ease the management of large image collections in a variety of domains [2-4]. Recently there has been an increasing interest in applying it to medical image repositories [5]. Rather than limiting queries to textual keywords, users can also query by example image or image feature (e.g., color, texture, or shape computed from a region of interest) to find similar images of the same modality, anatomical region, and disease along with the matching associated text records.

While the literature lists several successful systems for content-based image retrieval and image management methods [2, 5], they have not made significant inroads in routine medical informatics. In addition, although many large imaging databases exist, such as the National Cancer Imaging Archive (NCIA) or the Lung Imaging Database Consortium (LIDC) created under the aegis of the Cancer Imaging Program[1] at the U.S. National Cancer Institute (NCI), these efforts have concentrated on data collection and transmission but have left development of applications to the research community. Lack of CBIR adoption has been attributed partly to the difficulty of integrating current implementations with existing healthcare systems [6]. The following reasons may further explain this anomaly: (i) challenging nature of medical images, (ii) need for specialized methods specific to each image type and detail, (iii) lack of effective image indexing methods, and (iv) lack of a uniform data and resource exchange framework between complementary systems. Most systems tend to focus on varying degrees of first two items, making them very versatile in a small sampling of the variety of medical images, but unable to share their resources or strengths. This requires each project to redevelop what may exist as an advanced implementation, but inaccessible. The lack of suitable image indexing methods is a problem for large image collections. Image comparisons performed linearly are inefficient and too slow for practical use. This paper proposes to overcome these shortcomings by defining a data and resource exchange framework using open stan-

1 http://imaging.cancer.gov (Last accessed: March 27, 2007)

dards and software to enable such specialized systems to act as geographically distributed toolkits. The approach enables communication between two or more geographically separated complementary systems with possibly different architectures and developed on different platforms, and specialized for different image modalities and characteristics. The resulting system provides the user with a rich functionality operating within a familiar interface on the Web-browser, making it portable and independent of location and underlying user operating systems. Figure 1 illustrates this concept.

In this figure, each circle represents a CBIR system specializing in particular image types, pathologies, or CBIR techniques. Using a standard open protocol any system could act as a client using the services available elsewhere to provide the user with a rich medical image informatics resource.

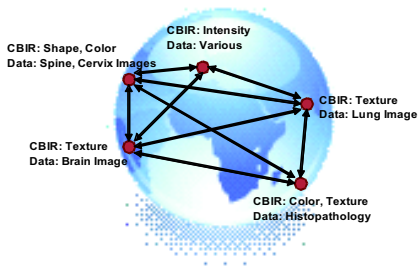


Figure 1 - Concept: Geographically distributed framework of complementary CBIR systems

As proof-of-concept, we describe the coupling of two leading, complementary geographically separated image informatics CBIR systems: The Image Retrieval in Medical Applications (IRMA) project at the Aachen University of Technology (RWTH) [7, 8] and the Spine Pathology & Image Retrieval System (SPIRS) project at the U.S. National Library of Medicine (NLM) [9-11]. IRMA and SPIRS retrieve images based on different approaches: Traditionally, IRMA has focused on image retrieval by computing overall (or global) image characteristics such as color, intensity, and texture. In particular, the image distortion model was developed and proven as a robust distance measure for differences on smaller sub-regions in the image [12]. Such an approach permits queries on a varied image collection and helps identify similar images, e.g., all chest X-rays in the A-P view. The IRMA system lacks the ability to find particular pathology that may be localized in specific regions within the image. In contrast, SPIRS can retrieve images that exhibit pathology that may be localized to a particular region under the assumption that the query to a large image collection containing images of only one type, e.g., vertebral pathology expressed in spine x-ray images in the sagittal plane. SPIRS lacks the ability to select pertinent images from a large varied image collection typical in a hospital PACS system, for example. We believe that combining the strengths of these two complementary technologies of whole image and local feature-based retrieval is unique and valuable for research into the retrieval of images in large repositories that are similar in type as well as pathology.

Background

IRMA project

The IRMA project[2] aims to develop and implement high-level methods for CBIR with prototypical application to medico-diagnostic tasks on radiological image archives. Stated goals include support for semantic and formalized queries to the medical image database with support for inter- and intra-individual variance and diseases. Example tasks are the staging of a patient's therapy or the retrieval of images with similar diagnostic findings in large electronic archives. Formal content-based queries also take into account the technical conditions of the examination and the image acquisition modalities. The system classifies radiological images in a general way without restriction to a certain diagnostic problem or question. Pattern recognition and structural analysis methods describe the image content in a feature-based, formal and generalized way. The *mean* image description enables a fast and reliable image comparison and retrieval. The project also includes an automatic classification and indexing process for insertion of new data into the system without manual interaction.

The IRMA project has several interfaces and can be characterized by its features: (i) Automated classification of radiographs based on global features with respect to imaging modality, direction, body region examined and biological system under investigation; (ii) Identification of image features that are relevant for medical diagnosis; these features are derived from a priori classified and registered images; and (iii) Image retrieval on similarity to an a priori selected feature set based on the visual similarity of certain image structures. Current image data consists of radiographs, with future plans to include medical images from other modalities. An IRMA retrieval interface supporting query refinement is shown in Figure 2.

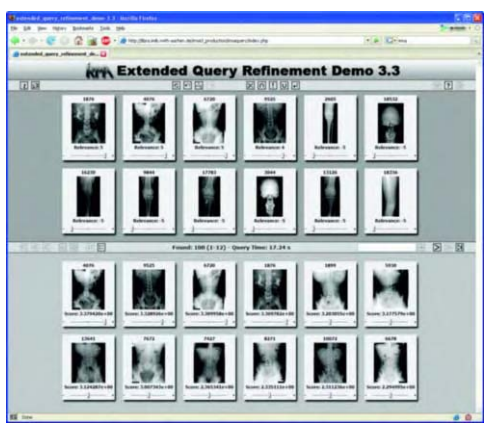


Figure 2 - IRMA retrieval interface with query refinement.

2 http://irma-project.org (Last accessed: March 27, 2007)

SPIRS project

SPIRS[3] provides a Web-based interface for performing image retrieval from a database of digitized spine x-rays using the morphological shape of the vertebral body. Its framework enables interaction with and retrieval of relevant information from large databases of image and patient data using rich hybrid image and text query methods.

A query editor enables users to pose queries by sketching a unique shape, or selecting or modifying an existing shape from the database. It aims to capture query semantics through support of advanced mechanisms like multiple partial shape matching. Additional text fields enable users to supplement visual queries with other relevant data (e.g., anthropometric data, quantitative imaging parameters, patient demographics). These hybrid text-image queries may be annotated with pertinent pathologies by selecting and weighting local features to indicate importance. Query results appear in a customizable window that displays the top matching results and related patient data. SPIRS provides a working proof-of-concept that is capable of accommodating large amounts of imaging data expected in the near future.

Figure 3 - SPIRS interface with (a) crop and (b) detail view

At NLM, the focus of CBIR research has been to develop systems capable of performing a range of queries on large medical multimedia databases comprising various biomedical images and patient health data. Such a database in current use contains digitized spine x-rays and associated person metadata that comes from a large nationwide survey, the Second National Health and Nutrition Examination Survey (NHANES II). NHANES is conducted regularly by the National Center for Health Statistics in the United States. The goals of NHANES include estimating prevalence of selected diseases, monitoring disease trends, and studying the relationship between nutrition and health. The NHANES II collection is considered very valuable to radiologists, bone morphometrists, researchers in osteo-arthritis, and medical educators. Domain experts reviewed a sample of the data and identified 23 key biomedical features that were exhibited in the x-rays. Of these, anterior osteophytes, subluxation/spondylolisthesis, and disc space narrowing were determined to be frequently occurring and reliably detectable. Each of these key features may be identified by examining the boundary shape of the vertebra. SPIRS may be used to determine what features (e.g., protrusion on the anterior edge of the cervical vertebra) are consistently associated with a certain symptom (e.g., neck discomfort) or whether a certain feature is a precursor to more serious illnesses (e.g., arthritis). Pathology or medical condition may also be documented in the text of the patient record as a survey response or in the medical diagnoses. The SPIRS Web-interface is shown in Figure 3. In addition to the Web-interface, SPIRS also offers a service to its core shape similarity algorithms and data. Although SPIRS focuses on shape-based queries its framework is extensible to adopt features particular to other biomedical image and data collections, e.g., its core architecture is being extended to include color, texture, and spatial location in uterine cervix images from the National Cancer Institute at the NIH [11]

SPIRS-IRMA multilevel distributed CBIR

Medical CBIR systems can be classified by the type of numerical features used to characterize the images: (i) **global** approaches extract a single feature vector from the entire image, (ii) **local** approaches assign the feature vector to distinct regions of interest, and (iii) **structural** approaches additionally cope with spatial and temporal relations between the image objects of interest. In this context the IRMA system is a global feature system with particular structural aspects of some image regions that are also computed. In contrast, SPIRS takes a more local approach to image similarity with some structural knowledge available from the vertebral image labels. In this sense, SPIRS-IRMA jointly may be considered a multilevel distributed CBIR system [11, 13].

In this joint system, IRMA serves as the front-end for the end user, while SPIRS provides specific shape similarity algorithms and supplies formatted responses to structured queries. In this proof-of-concept system, extensive data interchange was minimized through data mirroring. It is conceivable, however, that the data could also be securely shared in a purely service oriented setup. Screenshots from the combined SPIRS-IRMA system are shown in Figure 4.

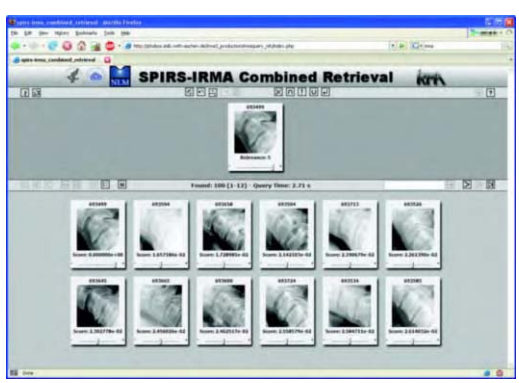


Figure 4 - SPIRS-IRMA interface

3 http://archive.nlm.nih.gov/spirs (Last accessed: March 27, 2007)

The system has plans for development in several phases. SPIRS-IRMA is utilizing only limited services provided by SPIRS with near term goals to use all available features. Future phases will also include permit users to upload their own images, as in the current IRMA system, and segmented or sketched boundary outlines of interest. Next steps include multi-resolution shape queries and support for query logging and relevance feedback.

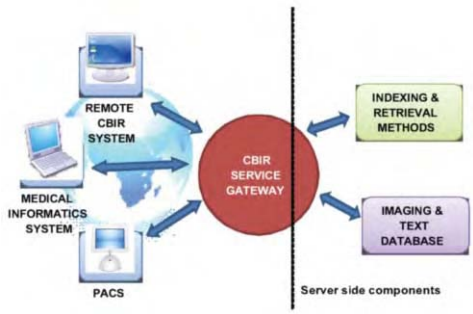


Figure 5 - The distributed architecture of SPIRS.

Data and resource exchange framework

CBIR systems sharing their data and computational resources need to be developed with a distributed architecture as shown in Figure 5. Each CBIR system would require the following components: (i) a gateway that acts as a mediator between client and server-side components, (ii) the indexing and retrieval server, which performs the feature representation and similarity matching, (iii) the databases containing images and associated text data, and (iv) an open communication standard for sharing of data and commands between the systems. Key components are described using the protocol developed for the SPIRS-IRMA system.

Gateway

The gateway is the entry point for all service requests. It can be implemented in any standard Web interface. In case of the SPIRS-IRMA collaboration, SPIRS has the service gateway implemented as a Java servlet. The gateway acts as a mediator between client requests and server-side components. It manages multiple simultaneous connections (users) as separate sessions and queues requests to the core CBIR engine. It translates query components that require information from the MySQL text database into SQL queries. The gateway is also responsible for formatting responses from server-side components and sending them to the client.

Server-side components

The core CBIR algorithms that operate on image feature data and text databases comprise the server-side components. While these algorithms rarely need any modification, their implementation is of particular importance. Typically, monolithic CBIR systems have indexing and similarity retrieval components closely wedded to the query and result visualization user interfaces. It is necessary to decouple these core algorithms from their user interfaces and implement them such that the local interfaces also use the gateway for all system communication. SPIRS indexing and retrieval algorithms use a variety of shape similarity metrics embedded within feature indexing techniques for efficient retrieval essential for Web use [11]. All text data is stored in a MySQL database and is accessible via both the gateway and the CBIR algorithms. In this initial prototype both systems hold all images and shape contours to minimize data exchange. Future plans include allowance for secure image and feature data exchange.

Communication protocol and data exchange format

A distributed computational framework relies extensively on a robust communications protocol and data exchange format. In case of SPIRS-IRMA, the systems are loosely-coupled over the Internet making it possible to use open communication standards. XML documents styled to a developed DTD are used for data exchange. The entire transaction is divided into three primary events, viz., *querystatus*, *query*, and *queryresult*. Each element in the XML file is designed for a particular event. For example, the <*querystatus*> element is used to determine if a desired service is available and to obtain a list of currently available services.

The <*query*> element is used to make shape queries where each query contains: (i) the query vertebra contour (boundary data points), (ii) partial shape indices and weights (if any) (iii) requested matching method (iv) maximum number of responses requested, and (v) range of similarity result scores. SPIRS responds with the <*queryresult*> element populated with: (i) matching image - vertebra tuples, and (ii) similarity scores. The design of the SPIRS-IRMA collaboration, the DTD, and sample XML data, are discussed in [13].

Discussion

The field of medical informatics has been unable to take advantage of image retrieval methods and systems in spite of their acknowledged importance in the face of growing use of image data in medical practice, research, and education. Image management and pathologically sensitive image content-based retrieval systems are thus increasingly necessary. The challenging nature of images and lack of comprehensive systems have hindered their acceptance. While it is difficult to develop a single comprehensive system, it may be possible to take advantage of the growing research interest and several successful systems with techniques developed for specific image collections. In addition to supporting rich segmentation, validation, indexing, query, retrieval, and visualization methods, inclusion of open interfaces using standard communication protocol, such as that described above, to the projects can enrich individual systems. The proposed approach has the following advantages:

- Simplicity: The simple communication interface allows for rapid development of methods for individual systems by expanding the available resources reachable through open communication standards.

- Extensibility and Flexibility: By separating the user interface from the core image informatics algorithms, the proposed approach allows systems offering services to continue development of other techniques and add them as they mature. It is no longer necessary to rebuild entire applications. In unusual cases, the protocol allows removal of some services advertising their unavailability.
- Security: Separating the core algorithms enables selective additional data security, user authentication, and encryption of the communication component (the gateway), where appropriate. The flexibility also allows sharing of just the methods.

We have demonstrated some of these features through the SPIRS-IRMA collaboration. In the IRMA system, it is possible for a user to find images from their database similar to an uploaded image. The query is limited to this extent, however. It requires manual viewing of each resulting image to identify those with pathology similar to that in the query example. While the SPIRS system allows shape queries with multiple parts highlighted, which indicates their relative importance, and includes text fields for further refined responses, it is limited to the spine x-ray database. When completed, users familiar with the IRMA interface can extend their searches beyond finding similar x-rays to include localized searches. Goals include enabling image segmentation services for user-supplied images. The resulting system is a multi-level, pathologically sensitive, geographically distributed system.

Conclusions and future work

In addition to formalizing the implementation of all aspects of the SPIRS-IRMA system, we plan to develop and publish a formal specification to enable sharing of image informatics resources among various systems. As a test, we will expand this framework to the CBIR system under development for uterine cervix images that uses color, texture, and spatial location information in generating an image description.

In this article, we have proposed a resource sharing strategy for increasing the impact of traditionally limited medical CBIR systems on medical informatics, and possible applications to clinical medicine, medical research, and education. We propose the use of open standards and a distributed framework that is simple to implement, extensible, flexible, and secure for the development of CBIR systems. We demonstrate its impact through a prototype SPIRS-IRMA combined retrieval system.

Acknowledgements

The authors would like to thank Leif Neve at NLM, William Hsu at University of California at Los Angeles, and Mark Oliver Güld at Aachen University of Technology for their assistance.

This research was supported by the Intramural Research Program of the National Institutes of Health (NIH), National Library of Medicine (NLM), and Lister Hill National Center for Biomedical Communications (LHNCBC).

References

[1] Andriole KP. Addressing the Coming Radiology Crisis: The Society for Computer Applications in Radiology Transforming the Radiological Interpretation Process (TRIP) Inititative. White Paper (http://www.siimweb.org). November 2005. Last Accessed: March 26, 2007.

[2] Smeulders A, Worring M, Santini S, Gupta A, Jain R. Content-based image retrieval at the end of the early years. IEEE Trans Pattern Rec. & Machine Intel. 2000; 22(12):1349-80.

[3] Antani S, Kasturi R, Jain R. A survey on the use of pattern recognition methods for abstraction, indexing and retrieval of images and video. Pattern Rec. 2002;35(4):945-65.

[4] Catarci T, Costabile MF, Levialdi S, Batini C. Visual query systems for databases: a survey. J. Visual Lang. & Comp. 1997;8(2):215-60.

[5] Muller H, Michoux N, Bandon D, Geissbuhler A. A review of content-based image retrieval systems in medical applications—clinical benefits and future directions. Int. J. Med. Info. 2004; 73(1):1-23.

[6] Shyu CR, Brodley CE, Kak AC, Kosaka A, Aisen AM, Broderick LS. ASSERT: A physician-in-the-loop content-based retrieval system for HCRT image databases. Computer Vision & Image Understanding 1999;75(12): 111-32.

[7] Lehmann TM, Güld MO, Thies C et al. Content-based image retrieval in medical applications. Methods of Information in Medicine 2004; 43(4): 354-61.

[8] Thies C, Gueld MO, Fischer B, Lehmann TM. Content-based queries on the CasImage database within the IRMA framework. Lec. Notes Computer Sci 2005;3491:781-92.

[9] Antani S, Long LR, Thoma G. Content-based image retrieval for large biomedical image archives. Proc. 11th World Cong. on Med Info (MEDINFO) 2004, pp. 829-33.

[10] Long LR, Antani SK, Thoma GR. Image informatics at a national research center. Computerized Med Imaging & Graphics. 2005;29:171-93.

[11] Thoma GR, Long LR, Antani S. Biomedical imaging research and development: knowledge from images in the medical enterprise. Lister Hill National Center for Biomedical Communications, NLM, NIH; Sep. 2006. Report No.: LHNCBC-TR-2006-002.

[12] Keysers D, Dahmen J, Ney H, Wein BB, Lehmann TM. A statistical framework for model-based image retrieval in medical applications. J Electronic Imaging 2003; 12(1): 59-68.

[13] Antani S, Lehmann TM, Long LR, Güld MO, Neve L, Thoma G. Interfacing Global and Local CBIR Systems for Medical Image Retrieval. Accepted by Workshop on Image Research for Medicine (Bildverarbeitung für die Medizin). March 2007, Munich, Germany.

Address for correspondence

Sameer Antani, PhD
National Library of Medicine, NIH
8600 Rockville Pike, MS 3824
Bethesda, MD 20894, USA
Email: santani@mail.nih.gov

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Utilizing SELinux to Mandate Ultra-secure Access Control of Medical Records

Peter R Croll, Matt Henricksen, Bill Caelli and Vicky Liu

Information Security Institute, Queensland University of Technology, Brisbane, Australia

Abstract

Ongoing concerns have been raised over the effectiveness of information technology products and systems in maintaining privacy protection for sensitive data. The aim is to ensure that sensitive health information can be adequately protected yet still be accessible only to those that "need-to-know". To achieve this and ensure sustainability over the longer term, it is advocated that an alternative, stable and secure system architecture is required. This paper considers the adoption of a model targeted at health information that provides much higher degrees of protection. A purpose built demonstrator that was developed based on enterprise-level systems software products is detailed. The long term aim is to provide a viable solution by utilizing contemporary, commercially supported operating system and allied software. The advantages and limitations in its application with a medical database are discussed. The future needs in terms of research, software development and changes in organizational policy for healthcare providers, are outlined.

Keywords:
information security, health information systems, operating systems, access control

Introduction

Advances in storage and communication technologies have made large repositories of data available even when they are maintained on separate systems and geographically distributed. The access to such data sets is often subject to varying degrees of legal, social and ethical constraints. Further, the data sets may not be available for open scientific research due to the sensitive and private nature of the information they contain. For example, individual personal health records offer significant value for medical and related research but such information cannot be readily accessed in a manner suitable for such research. The reasons include the need to abide by national privacy legislation, the reluctance to change to electronic record format due to security fears and the requirement to maintain end-user trust in the overall healthcare information system. Even when access rights have been granted - for example the data has been sanitized by having all personal identification removed - there are still legitimate concerns over the degree of privacy that contemporary IT applications will provide. The present reality is that IT applications operate on commodity level computer operating systems that do not - and cannot - provide the required level of assurance needed for scientific research to be undertaken on sensitive data. They were never designed for this purpose.

Information security mechanisms do exist to ensure sensitive information is protected and only accessible on a "need-to-know" and approval basis. It is imperative that both adversaries, such as external system "hackers" and technical/operations personnel with in-house knowledge be denied inappropriate access. The security mechanism known as Mandatory Access Control (MAC) is described below and can be used, as adapted, to enforce the necessary security and privacy processes required for handling sensitive health data. Such mechanisms have been studied and understood for over 30 years, mainly in defence related systems. However, they have not been evident in contemporary commodity level operating system or allied application level environments.

In particular, the ICT industry's move towards development of application systems based on so-called "Service Oriented Architectures (SOA)" and "Web Services" software environments presents new security challenges in the healthcare environment. Security standards, even for these high level service structures are often based on "interpreter" sub-systems such as XML, etc., and Internet/World-Wide-Web "browser" packages; they are complex and are only in the early stages of development and deployment. Moreover, all of these structures critically depend upon the overall security and dependability of the underlying "middleware", operating system and computer hardware systems. An application cannot be any more secure than the underlying systems upon which it depends. The same holds true for contemporary computer grid technologies, e.g., the Globus [1]. In other words, trying to adequately secure the shared "virtual machine" environments that grid technologies currently exploit is next to impossible (i.e., far too challenging for the foreseeable future).

The primary aim of this project is to build a "Concept Technology Demonstrator", based upon advanced cryptographic and information research and technologies (e.g., Cryptocards[1] and Starlight InfoSec technology[2]) in order

1 "Cryptographic plug-in cards", Eracom (Safenet) Australia.
2 "Starlight InfoSec Technology", Defence Science and Technology Organisation (DSTO), Australia.

to provide ultra-secure and sanitized access to the protected data sets.

The non-sustainability of current approaches

There is a strong vested business interest by mainstream suppliers of computer operating systems and similar middleware products to perpetuate the belief that computer applications can be *"made secure from within"*, irrespective of other software or even hardware components. In other words, the correct use of their technology will ensure a sufficiently secure operating environment upon which application programs can run safely. Unfortunately, any such assumption is flawed since in reality they represent a "fortress built upon sand" [2].

Current base operating systems in the commercial arena are based on what is designated as "Discretionary Access Control" or DAC. Essentially this design allows the "owner" of any object in a system to, at their discretion (from where the term DAC comes) pass on their access rights to any other person or entity in the overall information system's environment. In particular, DAC really does not acknowledge the essential difference between a computer "user", a person, and the individual "processes" in the system that act on his/her behalf. Moreover, the DAC structure assumes that the user is completely familiar with and trusts any program or related software system which he chooses to execute in the computer. In the current environment these assumptions, while possibly valid in the era of large mainframe computer systems with large "in-house" software development and support organisations prior to the development of global "packaged" software industry, are simply no longer true. The DAC approach, moreover, assumes that users formulate their own "discretionary" policy, a situation that is no longer valid as overall information systems become subject to overriding legal, societal and enterprise policy requirements.

The applications will not be secure unless the underlying operating system and hardware have been specifically developed with security in mind. A system that is built to utilize a MAC mechanism will provide levels of security relating to all aspects of the computing system, i.e., *"enforce an administrative set policy over all subjects and objects in a system, basing decisions on labels containing a variety of security-relevant information"* [3].

Practical implementations of MAC

There are many implementations of MAC based systems, but one of the most popular is called SELinux (Security-Enhanced Linux) [4]. It was designed and engineered by the National Security Agency, an intelligence-gathering organization belonging to the government of the United States. Released in 2000 as a patch to the Linux operating system, SELinux quickly gathered popularity within the Linux community due to its structural simplicity and the impeccable credentials of its designers. It now exists as an open-source module transparently integrated into the Linux OS kernel. Optimistically, the casual user may receive security support from SELinux without even noticing that the module is active, due to generic security configurations created by some of the recently arisen SELinux support groups.

SELinux shares two fundamental properties with many MAC systems. Firstly, the super-user concept of DAC systems (i.e., "root" or "Administrator") is banished, so that all users of the system are controlled by the same configuration policy. The policy is written by a security administrator. If an attacker can acquire the privileges of the security administrator, then the policy can be changed to suit his or her ends. But unlike the super-user, who controls any aspect of the system, the security administrator exists only to secure the machine, and not to make use of it.

Secondly, like some other MAC systems, SELinux is based upon the concept of "type enforcement", in which all the objects in an operating system (be they files, network sockets or processes) are labelled and classified as "domains" or "types". The system configuration determines how domains with label *x* are able to access types with label *y*. Typically the access rights will be described, in a broad sense, as "domain *x* can read type *y*", "domain *x* can write to type *y*", "domain *x* can execute type *y*", combinations of these, or "domain *x* cannot interact with type *y* in any way". This last option is the implicit default, so only positive relationships between domains and types need to be configured.

Nevertheless, the flexibility provided by SELinux tends to be its undoing, at least from the perspective of the casual user. Because there are many domains and types within even a single-user system, and because each possible positive interaction needs to be considered, configuration of access rights is a laborious and error-prone process. Add to this the fact that each system is guaranteed to be different to the one on which the prepackaged configuration was prepared, and a nightmare scenario in which SELinux denies essential accesses - such as allowing the system's Graphical User Interface to start - becomes a common one [5].

There are well-known strategies that can help to reduce configuration complexities. One of the most popular of these is Role-Based Access Control (RBAC), which intersperses a "role *r*" into the relationship between domain and type, such that, for example, if "role *r* can read type *y*" and "if domain *x* is a member of role *r*", then "x can read y". Since there are a small number of roles relative to the number of domains and types, then the number of rules relating roles to types and domains to roles should be much fewer than the number of rules that relate domains directly to types. SELinux supports a primitive version of RBAC, yet a typical SELinux configuration file still runs to about 50,000 lines.

The RedHat company sells RedHat Linux Enterprise and sponsors the "Fedora Core" open-source software activity, both of which sport an extension of SELinux that includes "strict mode" and "targeted mode" structures. Strict mode is no different to the "vanilla" version of SELinux, but targeted mode protects only a subset of domains and types, usually those which have interaction with the external world via network sockets, etc. (that is, those objects which are most likely to be attacked by hackers). The

remaining objects within the system are labelled "unconfined" and can "run amok" with only the discretionary access controls regulating their behaviour. RedHat ships SELinux in the default mode of "targeted", so that basic protection is afforded to the system without the mechanism becoming invasive, in turn preventing the user from being productive, and swamping RedHat support with basic administrative support requests. The flipside to this is that RedHat does not offer support to issues arising from the strict mode of SELinux. As will be seen later, this has a substantial impact on the use of SELinux to protect medical or other application data.

Protecting medical application data

The primary intention of SELinux is to protect objects embedded within the operating system, with security of application data being an afterthought. SELinux in effect partitions the operating system space into a set of "sandboxes", protected areas between which communication is tightly regulated. The mechanism is generic, and consequently, the security administrator can create an additional series of sandboxes at the application level to protect medical and other kinds of data. For example, the administrator may configure a web-browsing sandbox that permits a web browser such as Internet Explorer or Mozilla Firefox to access the internet. In addition, the administrator may also configure a medical-related sandbox in which a medical application is permitted to access medical records. However, unless explicitly permitted, the web browser does not have access to the medical records. Neither does the medical application have the same level of exposure outside the network as the web browser. The security administrator can create arbitrary levels of complexity in the application layer by constructing sandboxes for different applications, yet the enforcement mechanism of SELinux treats them all equally and prevents unauthorized accesses. Whereas if a hacker attacked a DAC system through the network interface, and managed to acquire super-user permissions, in an SELinux scenario, the hacker would control only a single sandbox, and would need to launch additional exploits, each of which became increasingly infeasible with distance from the network interface.

An important caveat is that the "targeted" mode of Red Hat Enterprise Linux and Fedora Core does not permit application-level sandboxes, because all application process run in the unconfined domain. Any system supporting application level security is compelled to run in strict mode, which in turn means that it is likely not to be fully supported by its commercial vendor.

Building an SELinux proxy

Application data tends to be much more dynamic and flexible than operating-system level data. There may be many users of an application level database, whereas the number of owners of operating system processes tends to be very small. By default, SELinux is configured for four users, including system, staff, sys-admin and ordinary users. Adding new users involves recompiling and reloading the configuration policy, as does adding new rules for interactions between domains and types. As operating-system level relationships tend to be very static, for example, changing only when new software is installed, this is not especially disadvantageous for the normal use cases for SELinux but is not well suited for creating rapidly changing sandboxes.

Our solution to this problem, which also avoids the problem of creating additional complex interactions between application and operating system level objects, is to create a proxy. The proxy runs at the application level and is secured in its own sandbox by SELinux, preventing unwanted interactions with other processes. The proxy regulates access by application-level process to protected data, using its own set of configuration files. In one sense, this solution can be viewed as nested SELinux, whereby the proxy represents a micro-instance of SELinux that deals only with application data. Operating system level processes see only a monolithic object (the proxy) representing application processes, meaning that the number of configuration rules between the two layers is linear rather than multiplicative.

The proxy deals with the added levels of interaction complexity at the application layer by using an enhanced version of RBAC, in which role permissions are inherited throughout a hierarchy. By collating roles into hierarchy, and associating the lowest member of each hierarchy with each type, this obviates the need to associate every role with every type. As an example, a vertical slice of a role hierarchy may consist of "Doctor is a subset of role Clinician" and "Surgeon is a subset of role Doctor". Configuring the policy with "any user in the role of Clinician has access to type *y*" automatically covers the rules for "any user in the role of Doctor has access to type *y*" and "any user in the role of Surgeon has access to type *y*" by virtue of their membership of the family. Portions of the hierarchy can be overridden: configuring "any user in the role of Surgeon does not have access to type *y"* does not cause a contradiction but allows only Clinicians and Doctors access to type *y.*

An option in this research was to build the extended RBAC functionality natively into SELinux for which the source code is freely available. However, the benefits to operating-system level objects, which are not ordered hierarchically, are unlikely to outweigh the disadvantage in branching the SELinux source code, consequently reducing the successful uptake of this solution.

The mechanism by which the proxy works is very simple, and abstractly mirrors the SELinux mechanism. A client interacts with the proxy via a pair of Client and Server messages. For each client message received, the proxy sends exactly one server message.

The client authenticates itself to the proxy using a client message with type CREDENTIALS and with a payload containing the user, role and password that describe the client. Until the next such message is received, the proxy caches the credentials. This mimics the SELinux mechanism, which authenticates a user via a password before transitioning the user into the requested role. The proxy

generally responds to credential messages by sending a dummy OK response.

The credentials are evaluated whenever the client requests access, either a read or a write, to a record in the proxy database. The proxy passes the credentials, along with the record identifier and the policy to the security filter. The security filter assesses the credentials, decides whether the record can be accessed in the way intended and passes this decision to the proxy. In the case of a read request, the proxy relays the appropriate record back to the client. If the client has requested a write, then the material passed in the payload of the REQUEST_WRITE_FILE message is appended or overwritten to the record.

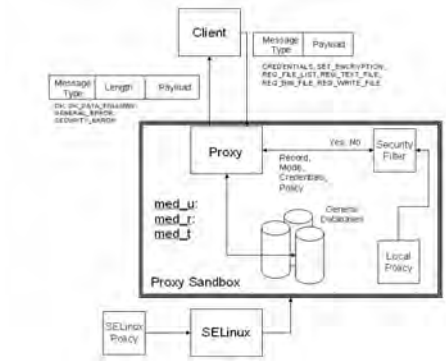


Figure 1 - Architecture of the SELinux proxy

Whereas SELinux can protect data to the granularity of the file, the proxy has arbitrary granularity, as determined by tags exchanged between the proxy and its client. The client may wish to retrieve a single word from a database, or an entire collection of files. Our mechanism allows this with as little as a single configuration, although for more complex cases, the number of configuration rules will increase linearly in the number of database items.

There are some cases when records must be accessible even in the absence of legitimate credentials. For example, if the authorized viewer of a patient's case file is not present, but the patient requires emergency treatment, then the availability of the information is more important than its privacy. So the proxy is programmed to respond to a special role of "Emergency", in which case it moves into auditing mode, until a new set of credentials with a differing role is provided. In auditing mode, all records can be retrieved and modified, but each action is recorded and flagged for review by the security administrator. Appropriate punishment for abusing this mode can be metered out at a social level. Our prototype does not handle differential records, whereby the deltas between subsequent versions of records are stored, although this would be advantageous for malicious or accidental modification of records in auditing mode.

It is not essential for the proxy and the client to maintain an encrypted channel, since access control on the channel can be maintained by SELinux. For ease of configuration, all communication can be encrypted using commonly available algorithms such as the Advanced Encryption Standard. Our research did not consider key management issues between the client and the proxy, although the usual public key establishment protocols, such as Diffie-Hellman can be used.

To prove the effectiveness of the proxy, we developed a simple prototype of the proxy and a client, as shown in Figure 2. Auditing data for the client is shown in Figure 3. We used the proxy and client to demonstrate the security advantages of SELinux over DAC-based systems such as Windows XP. In DAC-based systems, it was relatively easy to use hacking tools such as rainbow tables [6] to break weak Windows system administrator passwords, and modify the proxy and client code to allow unauthorized and unaudited accesses. As the proxy was housed in its own sandbox under SELinux, traditional hacking tools did not provide an avenue for breaking into or changing the proxy. The issue remains that this security is present only in the unsupported "strict" mode of SELinux which is still too complex to deploy in commercial situations.

Although the proxy significantly simplifies configuration of application data, it does not address problems at the operating-system level that need to be resolved. Further research in this area needs to focus on simplifying generic SELinux configuration, to allow realistic deployment of "strict" SELinux, which supports protection of application data. This is indeed happening, as witnessed by the development of modular policy logic in Fedora Core 5, which allows the configuration to be developed and loaded in blocks relating to the processes or daemons being protected. The efficacy of this strategy has yet to be solidly determined.

Figure 2 – The Proxy Client

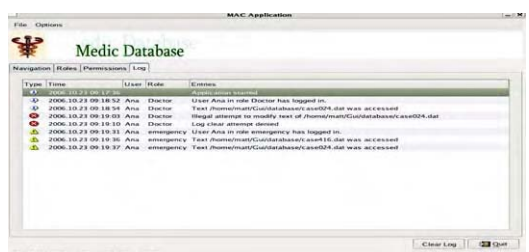


Figure 3 – Auditing data for the proxy client

Conclusions

Sufficient evidence is emerging that the security requirements and obligations for the protection of sensitive health data cannot be sustained using contemporary data access control and protection mechanisms in current commercial, commodity computer systems. *"Mandatory Access Control"* or MAC, incorporated into basic operating systems and allied supporting software structures, provides an alternative, strict, security policy driven approach far superior to industry standard DAC mechanisms. MAC can strengthen protection from unauthorised access to sensitive health related information from both outside and inside an organization. This provides enhanced privacy protection from staff, including knowledgeable ICT professional staff members, gaining access to such sensitive data for which they are not authorised (i.e., view, modify, copy, transmit, delete, etc.). It further provides enhanced 'boundary' security from outside intrusion whereby adversaries, such as hackers and spyware operatives, are unable to gain full control of an information system. In the MAC case it can be demonstrated that damage can be limited to violation of an individual user's account [7].

This research has found that a MAC based medical data system, although viable, still presents some key research and practical deployment challenges. In particular, the "strict" operational mode offered by SELinux may be seen as being too rigid for deploying Role-Base Access Control or RBAC structures with the required levels of flexibility needed in practical healthcare situations. Without this flexibility, system reconfiguration may be required each time a user is added or removed. This is infeasible in practice and is already the subject of a number of active research projects. It was shown with the demonstrator described in this paper that a compromise can be derived that provides an application level proxy to facilitate a secure, role-based access interface. A balance has to be struck between strict access control security and the degree of flexibility for dynamic modification of any system in the "real world". Any approach taken should be determined from a privacy impact oriented risk assessment process. For example, such an assessment might readily determine a need for emergency over-ride capability to enable at least wide read-only access to medical/health data. Such a facility would, however, have to be subjected to new audit and control requirements as well as to limitations potentially related to time periods and the location of users. In this regard an implementation that can support dynamic reconfiguration in a manageable and understandable manner may be essential. Earlier MAC systems were simply not designed for this environment where the security policy may need to be dynamic, not just in detail but also in structure.

Future needs in terms of research also involve a better understanding of the complementarity of SELinux's concept of "type enforcement" versus more traditional security structures based on hierarchical "Multi-level Secure" or MLS schemes. In the health information area it needs to be determined whether or not such hierarchical security schemes have a place or not and, if so, to what level are modifications of the basic concepts involved necessary. Likewise, the concept of "compartmentalization", reflected in the SELinux type enforcement system, needs to be assessed in relation to its suitability for all levels of information services needed in a nationwide health information structure. At the same time, application software development needs to become aware of the new parameters afforded by the MAC facilities and determine to what level such applications may or may not make use of the security mechanisms and services offered, i.e., to determine the distinction between what may be labelled as "security aware" versus "security ignorant" applications. Moreover, the integration of existing software systems into this environment must be understood, requiring further research into appropriate techniques for system integration in higher security environments. In turn, this places new demands on education and training as ICT professionals need to develop the skills needed to understand, utilise and manage this new environment. This indicates that necessary or desirable changes in organizational policy and management structures for healthcare providers may be also needed. At present, full guidance to policy makers and operational management in relation to deployment of newer MAC based overall information systems do not appear to exist. This leads to the need for further research and experimental system development in the area to enable study of the economic, cultural, social and legal responses required.

Acknowledgments

This research was supported under the Australian Research Council's Special Initiative on e-Research funding scheme (project number SR0567386). The authors would like thank RedHat (Asia Pacific) Ltd. and the E-Health Research Centre (a joint CSIRO and Queensland Health initiative) who have been highly supportive.

References

[1] Foster I, Kesselman C. The Globus Project: A Status Report. Proceedings of the Seventh Heterogeneous Computing Workshop, 1998, ISBN: 0-8186-8365-1.

[2] Loscocco PA, Smalley SD, et al. The Inevitability of Failure: The Flawed Assumption of Security in Modern Computing Environments. Proc. 21st National Information Systems Security Conference, Crystal City, VA, 1998.

[3] Loscocco PA, Smalley SD. Meeting Critical Security Objectives with Security-Enhanced Linux. Proceedings of the 2001, Linux Symposium, Ottawa, 2001.

[4] National Security Agency. Security Enhanced Linux homepage. Available at http://www.nsa.gov/selinux, 2000.

[5] Thompson, K. Does SELinux support X-Windows? http://www.crypt.gen.nz/selinux/faq.html#CP.7 Nov. 2006.

[6] Oeschlin P. Making a Faster Cryptanalytic Time-Memory Trade-Off**,** Proceedings of Crypto 2003, pp 617-630, 2003.

[7] Henricksen M, Caelli W, Croll PR. Securing Grid Data Using Mandatory Access Controls. to appear 5th Australasian Symposium on Grid Computing and e-Research (AusGrid 2007) Ballarat, Australia, Feb. 2007.

Address for correspondence

Prof. P.R Croll, Information Security Institute, QUT, 126 Margaret Street, Brisbane 4001, Australia. Email: croll@qut.com

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Proposal of a French Health Identification Number Interoperable at the European Level

Catherine Quantin[a], François-André Allaert[b], Paul Avillach[c,d], Benoît Riandey[e] Marius Fieschi[c], Maniane Fassa[a], Olivier Cohen[f]

a. Service de Biostatistique et Informatique Médicale, CHU de Dijon, BP 77908, 21079 Dijon Cedex, INSERM EMI 0106, France
b. Department of Epidemiology & Biostatistics, Mc Gill University, Montreal, Canada
c. LERTIM Faculté de Médecine Université de la Méditerranée 27 Bd Jean Moulin 13385 Marseille cedex 05, France
d. ISPED Université Victor Segalen Bordeaux II, 146 rue Léo Saignat, 33076 Bordeaux cedex, France
e. Institut National d'Etudes Démographiques (INED), 133 Bd Davout, F-75980 Paris Cedex 20, France
f. Laboratory TIMC-IMAG UMR 5525, CNRS University of Joseph Fourier Medical School of Grenoble, France

Abstract

The French ministry of Health is setting up the Personal Medical Record (PMR). This innovative tool has long been expected by French Health Authorities, Associations of Patients, other Health's associations, those defending Individual Liberties and the French National Data Protection Authority. The PMR will lead to improvements in many areas such as Diagnosis (Research and monitoring) Healthcare (Management of emergencies, urgent situations, Temporal health monitoring and evaluation), Therapy (Cohorts of patients for Clinical trials and epidemiological studies). The PMR will foster safe healthcare management, clinical research and epidemiological studies. Nevertheless, it raises many important questions regarding duplicates and the quality, precision and coherence of the linkage with other health data coming from different sources. The currently planned identifying process raises many questions with regard to its ability to deal with potential duplicates and to perform data linkage with other health data sources. Through this article, using the electronic health records, we develop and propose an identification process to improve the French PMR. Our proposed unique patient identifier will guarantee the security, confidentiality and privacy of the personal data, and will prove to be particularly useful for health planning, health policies and research as well as clinical and epidemiological studies. Finally, it will certainly be interoperable with other European health information systems. We propose here an alternative identification procedure that would allow France to broaden the scope of its PMR project by making it possible to contribute to public health research and policy while increasing interoperability with European health information systems and preserving the confidentiality of the data.

Keywords:

electronic health records, unique patient identifier, security

Introduction

In the majority of industrialized countries, at the heart of many of the concerns relating to electronic processing of health information lies the problem of patient identification. In August 2004, the French government decided by law to initiate a national project concerning Electronic Health Records called the "Dossier Médical Personnel", the Personal Medical Record (PMR) [1]. It intends to promote health care coordination, enhance the communication of health information and reduce iatrogenic accidents. The most important aim of this article is to show and present the principles of the French unique patient identifier created relative to the PMR. However, it will also explain and show its major disadvantage, which is its interoperability characteristic. For example, this current patient identifier seems to be incompatible with the identifier of the European health card [2]. This paper will thus demonstrate, in its first section that the French health identification number does not ensure interoperability at the European level. The second section proposes an alternative identification process that would allow France to improve the quality, security, precision and coherence of its PMR project. The developed and proposed health identifier will be extremely useful for health planners, those dealing with health policies, public health research and clinical and epidemiological studies, and this, at national, regional and international level.

Materials and methods

The issues and criteria of a unique patient identifier

The identification of the patient in a health-care structure and particularly in the framework of the electronic healthcare record of a care network is a major issue [3]:

- Patient care continuity requires secure, precise, coherent and reliable patient identification through all of the health information systems.

- The reduction and management of identification-related error is one of the major constraints necessary to maintain and improve the quality of care;
- A reliable unique format for patient identification is required in order to design and implement an interoperable health information system focused on the patient.

In the context of the "Principle and Process of Patient Identification" project, the French Group for the Modernization of Hospital Information Systems (GMSIH) [4] inventoried the principles and the architecture of identification systems by specifying two aspects of the problem:

1. "Identification" will specify which information will be used to uniquely identify a patient. Multiple identifiers can be proposed for the same patient in different applications.
2. "Merging" identification areas, requiring the implementation of specific methods (intra-health structures and among health structures).

This group also inventoried the international patient identification experience of ten countries (Germany, Australia, Canada, Denmark, Finland, Luxembourg, the United States, New Zealand, The Netherlands and the United Kingdom) considered by the GMSIH as representative of best practices in the areas of socialized health care, electronic health care information systems or patient identification.

The American Society for Testing and Materials (ASTM), a standards development organization accredited by the American National Standards Institute, identified 30 criteria that were published in the *Standard Guide for Properties of a Universal Healthcare Identifier* [5]. The most recent standard on Health Care Client Identification we know is from Australian Standard® AS 5017-2006 published in June 2006 [6]. This standard includes data elements that jointly comprise a unique identifier for health care clients. It is explained that the combination of the client identifier and the health care establishment that assigned the identifier is one way to indicate unique identification. The logical structure for these data elements has shown that, there may be multiple identifiers collected for any one individual. Together, the data elements that makeup one complete Health Care Client Identifier are

a) Health Care Client Identifier Designation b) Health Care Client Identifier Geographic Area c) Health Care Client Identifier Issuer d) Health Care Client Identifier Type.

Most health care clients have more than one identifier. The most commonly used identifier should be collected as the Primary Client Identifier (the first listed client identifier). This identifier is generally that assigned by the organization as the means of uniquely identifying the client. Often, there are other identifiers also related to the client, and these should be collected and recorded as Other Client Identifiers (listed second or subsequently in the identifier list). There can only be one Primary Client Identifier, but there may be many Other Client Identifiers collected for any one application. Examples of Health Care Client Identifiers include 1)Person Identifier 2) Medical Record Number (MRN) 3) Local Client Identifier 4) Health (care client) Identification Number 5)Unit Record (UR) Number 6) Enterprise Identifier 7) Area Identifier 8) State/Territory Health Identifier 9) Unique Identifier (UID) 10) Unique Health Identifier (UHID) 11) National Health Identifier (NHI).

The five most important criteria and characteristics (Table 1) of the Health Care Client Identifier (HCCI) Designation are adapted from the AS5017-2006 and provided here as a guide to assigning a unique identifier [6].

Table 1 - criteria and characteristics of the Health Care Client Identifier (HCCI)

Atomic (the HCCI Designation should be a single data item. It should not contain sub-elements that have meaning outside the context of the entire HCCI Designation. Nor should the HCCI Designation consist of multiple items that must be taken together to constitute an identifier).
Content-free (the HCCI Designation should not depend on possibly changing or possibly unknown information pertaining to the health care client. Including content in the HCCI Designation will make it impossible to assign the 'correct' identifier if that information is not known. It also leads to invalid situations if the information changes: for example, what happens to an identifier based on sex if the health care client has a sex change procedure).
Longevity (an HCCI system should be designed to function for the foreseeable future. It should not contain known limitations that will force the system to be restructured or revised radically).
Permanent (once assigned, an HCCI Designation should remain with the health care client. It should never be reassigned to another client, even after the health care client's death).
Unambiguous (whether represented in automated or handwritten form, an HCCI Designation should minimize the risk of misinterpretation. Where using alphanumeric identifiers, be aware of possible confusion with the number '0' with the letter 'O' and the number '1' with the letter 'I')
Unique (a valid HCCI Designation should identify one and only one health care client. A health care client should have only one primary HCCI Designation)

These criteria are designed to support four basic functions of a universal health care identifier:

1. Positive identification of patients when clinical care is rendered,
2. Automated linkage of various computer-based records on the same patient for the creation of lifelong electronic health care files,
3. Provision of a mechanism to support data security for the protection of privileged clinical information (does not attempt to address all safety concerns, however)
4. Use of technology for patient record handling to keep health care operating costs at a minimum.

The founding principles of the French Health Identification Number

In September 2005, the French government defined the founding principles of the Health Identification Number (INS in France) of the personal medical file (PMR). This number must be:

- **unique**: to ensure that two people cannot have the same INS, thus avoiding erroneous attribution of a diagnosis. An identifier is therefore associated with a person whose identity can be validated.
- **content-free:** the number reveals no information about the holder (sex, age, place of birth…).
- **public:** the number can be legally used by all health care professionals authorized by the patient. It can be stored in the health care professional's information system. The number is therefore not secret.
- **permanent:** will remain the same for the life of the patient and, possibly, beyond.
- **irreversible:** impossible to determine the identifier by calculating backwards from the social security number. However, if lost, the same number can be recreated.

First proposal for the creation of the INS in the PMR

To meet the "uniqueness" condition, the French patient identification working group recommended that the identifier be associated with the social security number. The solution proposed by this working group involves the creation of a Health Identification Number (INS in France) for each patient managed at the patient's request by an independent organization known as a "trusted authority". The role of the "trusted authority", which may not be a host, is to guarantee that the identifier is not and will not be duplicated. The trusted authority will also be in charge of the secure access, the respect of confidentiality and the integrity of all the identification workflows.

To preserve the anonymity of the identifier, the trusted authority creates the INS through the following process:

- The patient chooses an approved host and advises the PMR office using a request form and the host using a membership form.
- By contacting the National Health Insurance organization the PMR office confirms the requester's affiliation and that it is a first-time request
- The trusted authority creates the INS using the request form number that it transmits along with patient attributes to the host.
- The host then informs the trusted authority that the patient's INS is operational

The use of the request form number and not the social security number to generate the INS is designed to guarantee that the number is content-free and the INS creation process mathematically irreversible, i.e. no mathematical operation would permit backward calculation from the INS to lead to the social security number. The guarantee of INS anonymity would rely on the absence of a table of correspondence *directly* linking INS and the social security number. Detection of collisions and duplicates would occur by way of the request form number. Nonetheless, the relationship between the social security number and INS would be indirectly maintained at both sides. One and only one request form number would be associated with each social security number and one and only one INS would be associated with each request form number. The trusted authority would not require knowledge of the social security number to prevent collisions; they would need only to ensure that the INS had been assigned to no other request form number. Similarly, by checking that each form number is given to only one INS, they would prevent duplicates provided that no duplicate form number existed further up the chain.

The trusted authority would send the INS to the host along with other patient identifiers, such as first and last name and date of birth, but never the social security number. The host would thus possess only those identifiers that could not be used to reconstitute the social security number.

Criticism of this method of generation

The principal criticism [7] addresses the second main function, as defined in the ASTM *Standard Guide for Properties of a Universal Healthcare Identifier* [5]: Automated linkage of various computer-based records on the same patient for the creation of lifelong electronic health care files.

Thus, the fact that the proposed system leads to an institutional disconnect between the INS as a central health system identifier and the social security number. The main reason cited to justify this disconnect is the desire to prevent the creation of a population table of correspondence between the social security number and the INS. This argument is, however, open to criticism insofar as all health professionals and in particular large institutions whose patient databases cover a large proportion of the population of a region must have access to the correspondence between the social security number and the INS.

If the link between the Social Security Number (SSN) and the INS is broken, relating PMR data with other patient data for research purposes becomes impossible. For example, relating PMR with other sources is necessary not only to complete their data but also to ensure multi-source validation in conformity with the recommendations of the report of the academy of sciences regarding epidemiology. If research using PMR data is not possible in the short term because of the time necessary to create databases that are sufficiently structured, exhaustive and validated, it can be expected that the data will be of use in less than a decade, if an acceptable identifier is adopted. That's why, like our British colleagues who complain that overzealous interpretation of UL laws stifles epidemiological research [8-11], we think that the current French identifier may cause particular difficulties regarding epidemiological and clinical research.

Moreover, it must be possible to reconstitute the INS, which represents the true signature of a patient, whatever the nationality or the country of origin starting from features of identification that are always available. These methodological principles must guarantee the secure and perennial identification of all the patients. These principles must apply whatever the future use of the information. Indeed, the technical and methodological choices must be able to follow evolutions in the legislation, which may be modified according to the evolution of mentalities and cultures.

Our proposal

It would be perfectly possible to preserve the confidentiality due to the patient and desired by patient associations by setting up anonymous procedures [12-14] such as those adopted by the Institut de Veille Sanitaire (Health Surveillance Institute) on the recommendation of the French National Commission for Data protection and the Liberties (CNIL), in the context of the follow-up procedures for the 30 diseases subject to mandatory reporting (including AIDS).

Unlike encryption methods that must be reversible to allow the legitimate recipient to decode the message, unidirectional hashing techniques are irreversible. Hashing produces a perfectly anonymous code (it is not possible to retrace the patient's identity) that is always the same for a given individual so that patient data can be linked. There are many medical applications, which include the creation of national databases (such as those relating to the national follow-up of infected subjects - approximately 100,000 patients – an excellent example of what can be put to the service of epidemiological research, with complete patient approval) as well as regional and inter-regional databases in many areas (cancer, perinatality, genetic diseases). This system has also allowed (based on the hash-coding of the Social Security Number (SSN), the gender and the date of birth), standardized hospital discharge abstracts to be linked, classified into French Diagnosis Related Groups at the French national level and linked to the data of the national medical insurance information system. An anonymous procedure based on hash coding is also used for chaining patient files in Switzerland [15].

In the case of the PMR the situation is not, in fact, more complex because several requirements must similarly be met in a similar way:

- demands of the CNIL, patient and healthcare professional associations regarding confidentiality of personal information contained in the PMR, and respect of the law of August 13, 2004 concerning the use of data,
- needs in public health or for individuals to have access to these data, particularly when the patient has given express consent.

Ideally, hashing the social security number would help meeting these requirements (Figure 1).

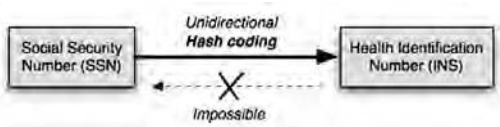


Figure 1 - Hash coding

Regarding confidentiality, insofar as the social security number could not be reconstituted using the INS, the link between them would be broken. Another advantage of using hash-coding is to meet the criterion of being focused (created and maintained solely for supporting health care cf. Table 1). As a consequence, using different keys for hash-coding will allow the creation of distinct identifiers according to different purposes (health administration, diagnosis and healthcare, epidemiological research). The same solutions also derived through an irreversible encryption of the unique Social Security Number (SSN) have been proposed in Belgium [16] and New Zealand.

Regarding data access for public health research, the use of the Social Security Number (SSN), after hash-coding, would thereby allow the linkage of the main national databases. However this solution raises many questions. Firstly, in France, foreigners do not a have a complete social security number. For instance, the last characters, corresponding to the country of birth, are not systematically filled. Moreover, as the date of birth in some foreign countries is not known precisely, two different patients may have the same date and the same country of birth, resulting in collisions, and linkage errors. Secondly, as the Social Security Number (SSN), has a different structure according the country, it can not be a solution for a unique patient identifier, interoperable at the European level.

One solution would be to add personal patient characteristics such as family name, first name, date of birth (separately hashed) to the social security number, which would help to conform to the recommendations of the International Association of Medical Information Technology and where possible to ensure interoperability of this identifier with a European identifier [17]. In fact, in the national medical insurance information system like in the standardized discharge abstracts, gender and date of birth have already been associated with the social security number, in the same identification field, before hash-coding, resulting in a single code. We could thus propose to build the INS through the simultaneous hash-coding of gender and date of birth associated with the social security number.

However, due to a simple error in the field for gender, linkage of data from the same patient can be definitively refused. It is important to note that this field is highly unreliable (numerous data entry errors have been observed in hospital settings). In addition, with regard to the quality of the linkage this variable is undiscriminating insofar as the probability of two people being the same gender is close to 50% and it does not discriminate for twins. It would therefore be preferable to replace this variable with one that is more sensitive, such as the first and last name of the beneficiary. As a consequence, our first proposal is to separately hash the social security number (removing the last digits), the date of birth and the last and first names, then merge this into a single signature of patient identity (Figure 2).

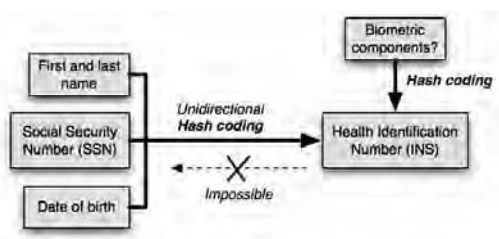


Figure 2 - Extended identifier for European interoperability

We could also propose, if possible to add biometric components (separately hashed). Biometric technologies are sometimes proposed in solving the problem of associating patients with their medical data, as they do not require the patient to bring any documents or remember information. Though this technology represents a real progress both in the identification and in the authentication of the patient, it raises many questions [18]. First, the accuracy and reliability of each biometric technology (for instance: finger print, iris scan, retinal scan and DNA) are not optimal in all circumstances. For example, fingers are frequently injured. Though retina scans and DNA analyses are the most accurate (with the exception of twins for DNA) they are also quite intrusive, which render even more difficult the possibility of combining these technologies, as proposed in some commercialized biometric systems. Moreover, the costs of the biometric solutions have to be considered. But the main problem lies in their acceptation by ethical organizations such as patients associations, national committees on ethics, human rights associations and national committees for data protection. In order to render this solution compatible with the current national procedures, it would be preferable to add the biometric component (separately hashed) to the hashed social security number, in order to ensure (at the European level) the interoperability of this identifier with national ones.

Conclusion

Our proposal for a French Health Identification Number will make it possible to uniquely identify and link a patient to his specific medical data. By hashing the social security number it will be possible to link the information of the personal medical file to other national health information sources, with the aim to complete or validate Personal Medical Data (PMR) or conduct epidemiological research. Adding personal patient characteristics such as first and second names and date of birth and/or biometric identifiers (separately hashed, then merged) to the hashed social security number would also contribute to the establishment of European public health statistics by matching healthcare data of the patients' records with other administrative data (mortality, social information …). This data linkage would thus meet the requirements of anonymous data of the European directive on data protection. This proposal could also be used in the discussions about the creation of European health care identifier. Of course, even if anonymised, the use of biometric components would require, the authorization of the National commissions for data protection.

Acknowledgements

This research was supported by The French National Agency for Research (ANR).

References

[1] Dossier Médical Personnel. [cited 2006 December 4th]; Available from: http://www.d-m-p.org

[2] Pre-commercial procurement of innovation - A missing link in the European innovation cycle - March 2006. [cited 2006 December 4th]; Available from: http://europa.eu.int/information_society/research/key_docs/documents/procurement.pdf

[3] Fieschi M, Les données du patient partagé: la culture du partage et de la qualité des informations pour améliorer la qualité des soins. Rapport au ministre de la santé. janvier 2003.

[4] Group for the Modernization of Hospital Information Systems [cited 2006 December 4th]; Available from: http://www.gmsih.fr

[5] American Standards for Testing and Materials (ASTM). (Standard Guide for Properties of a Universal Healthcare Identifier (UHID), Designation: E1714-95, Volume 14.01, Issued June 2000

[6] Australian Standard® AS 5017-2006 on June 2006 www.standards.org.au

[7] Quantin C, Guinot C, Tursz A, Salomez JL, Rogier C, Salamon R. Should epidemiological data be extracted from personal medical files? Revue d'Epidémiologie et de Santé Publique, 2006;54:177-84

[8] Walley T. Using personal health information in medical research. British Medical Journal 2006; 332; 130-1.

[9] Ward H, Cousens S, Smith-Bathgate B, et al. Obstacles to conducting epidemiological research in the UK general population. British Medical Journal 2004; 329:277-9.

[10] Verity C, Nicholl A. Consent, confidentially and the threat to public health surveillance. British Medical Journal 2002; 324:1210-3.

[11] Iversen A, Liddell K, Fear N, Hotopf M, Wessely S. Consent, confidentiality and the Data Protection Act. British Medical Journal 2006; 332:165-9.

[12] Quantin C, Bouzelat H, Allaert FA, Benhamiche AM, Faivre J, Dusserre L. Automatic record hash coding and linkage for epidemiological follow-up data confidentiality. Methods of Information in Medicine 1998; 37:271-7.

[13] Quantin C, Bouzelat H, Allaert FA, Benhamiche AM, Faivre J, Dusserre L. How to ensure data security of an epidemiological follow-up: quality assessment of an anonymous record linkage procedure, International Journal of Medical Informatics 1998; 49:117-22.

[14] ISO TS 25237 "Health informatics - Pseudonymisation practices for the protection of personal health information and health related services"

[15] Borst F, Allaert FA, Quantin C. The swiss solution for anonymously chaining patient files. MEDINFO 2001;10 (Pt 1):1239-41.

[16] Roger France FH, De Clercq E, Bangels S M. Purposes of Health Identification Cards in Belgium – EFMI, European Federation for Medical Informatics, IOS Press, 2005, Connecting Medical Informatics and Bio-Informatics

[17] Quantin C, Allaert FA, Gouyon B, Cohen O. Proposal for the creation of a European healthcare identifier. Stud Health Technol Inform 2005; 116: 949-54.

[18] Vaclav M, Zdenek R. Biometric authentication systems. A technical report. Retrieved April 15, 2006 from ecom-monitor.com web site: http://www.ecom-monitor.com/papers/biometricsTR2000.pdf

Address for correspondence

Professeur Catherine QUANTIN
Service de Biostatistique et Informatique Médicale
Centre Hospitalier Universitaire - BP 77908
21079 DIJON CEDEX - Tel: 33 3 80 29 36 29
Fax: 33 3 80 29 39 73
Email: catherine.quantin@chu-dijon.fr

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Clearinghouse: A Teleradiology Platform Emphasizing Security of Data and Communication

Michael Spitzer, Lars Brinkmann, Frank Ueckert

Department of Medical Informatics and Biomathematics, University Hospital, University of Muenster, Germany

Abstract

The Clearinghouse application platform is a web based solution for secure digital exchange of radiological images and other clinical documents among authorized researchers and physicians. It implements a sophisticated security and role model to protect privacy and to minimize the risk of eavesdropping of patient data. The Clearinghouse serves as a centralized platform for distributed, distantly located medical research and health care. It is based on Open-Source software, thus ensuring continued support, maintenance, security and last but not least continuity of the platform. The use of the Clearinghouse minimizes turnaround times by superseding comparably slow and insecure conventional communication methods otherwise used for the exchange of radiological images and clinical documents, such as standard mail and courier services. Furthermore, it alleviates the integration of distantly located expert knowledge into diagnostic routines, culminating in an increased health care quality regardless of location of patients or physicians.

Keywords:

telemedicine, teleradiology, radiology information systems, PACS, remote consultation, computer security, information storage and retrieval.

Introduction

In medicine and radiology applications the efficiency and contribution to health care quality of collaborations of distantly located participants is primarily dictated by the efficiency of the communication methods used. Travel times of several days for international mail are not unusual, and may be too lengthy to deliver advantages for a patient's treatment. Furthermore, important original documents may be lost irrecoverably during delivery. In case of e.g. rare diseases the probability not to find expert knowledge on-site is comparably high, and local physicians without a strong background in such diseases are forced to *(i)* send these patients to specialists or *(ii)* to rely on conventional mail to communicate diagnostic findings and subsequent treatment with colleagues.

The Clearinghouse application as an approach for a teleradiology platform aims to improve the speed and latency in communication of medical imaging and other clinical documents, while at the same time providing a maximum of flexibility, data security and privacy.

The integration of the Clearinghouse platform into the daily radiological routine enables effortless communication and collaboration with national and international experts. In reverse, patients in rural regions without ready access to specialists may benefit from expert opinions, improving the quality of their therapy. Due to the nature of e.g. rare diseases, and the fact that diagnosis, research and treatment are geographically separated in most of these cases, the Clearinghouse is especially suited to centrally organize and distribute relevant data. Exchange of data and knowledge in internationally relevant studies is intensified, while communication costs and turn-around times are lowered.

Materials and methods

Traditional workflow

The traditional workflow for exchanging clinical documents in general and radiological images in particular, involves mainly delivery by either patients themselves or by standard mail and courier services. Patients may be sent by their physicians to radiologists for taking radiological images, which in turn are sent back either via mail or are given to the patients to hand them over personally to their physician. If no other parties are involved in diagnosis and treatment based on such documents, this practice may be feasible, depending on the geographic distance of physician and radiologist.

Once multiple, probably distantly located, parties are involved in diagnosis and thus require access to existing or newly acquired documents of a patient, these documents have to be sent out (preferably as copies) to all participants by mail, courier services or as fax. Taken that none of these services are trusted with original documents the risk of irrecoverable loss of data is minimized, however the risk of eavesdropping of private and possibly compromising data still remains. This risk may be diminished by using e.g. sealed envelopes which in turn increase cost and expenditure, yet this measure does not eliminate disclosure of private data by accidental loss.

This traditional, decentralized workflow of exchanging patient data and documents bears several unratable risks of

data loss or delay in schedule which may negatively influence a patient's diagnosis and treatment.

The Clearinghouse platform

The Clearinghouse platform addresses these risks and shortcomings by unifying and optimizing access to a patient's data, documents and clinical history. It poses a centralized web application, amalgamating the contributions of geographically separated institutions to a patient's file and subsequent treatment. It features a sophisticated security concept and role model, both of which are described in the following paragraphs. On the client's side, no additional software needs to be installed apart from a web browser and Java, all other components (cf. section *Clearinghouse features*) are provided online and instantaneously by the server.

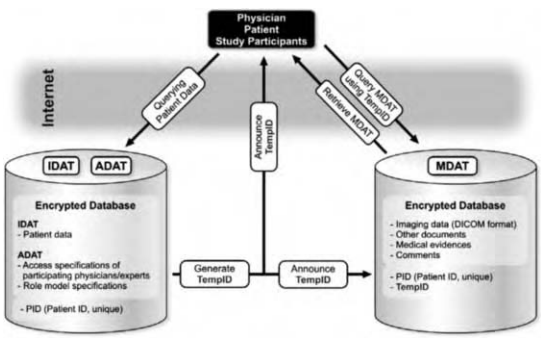


Figure 1 - Abstraction of the Clearinghouse security concept and communication flow

Data Security Concept

The security concept of the Clearinghouse platform (cf. Figure 1) is derived from a generic data security concept developed by the Telematic Platform for Medical Research Networks (TMF), Germany [1], and has been specifically tailored to meet the requirements of a web-based platform for digital image exchange. It implements a strict physical and logical separation of patient identification data (IDAT) such as name, gender and age, and medical treatment data (MDAT) such as digital images, clinical documents, medical evidences and diagnoses. In the case of DICOM objects, the header data is made pseudonymous by replacing data identifying patients in corresponding header fields with a so-called picture ID. It is generated by the MDAT component of the system and will not be propagated to other system components, except to clients (i.e. treating physicians) who are legally entitled to obtain knowledge of both data classes (IDAT and MDAT).

Each data class is stored on a separate database server featuring encrypted storage space. Each server is maintained by a separate system administrator and is located in the demilitarized zone of separate data centers, behind a firewall. Thus no persons except legitimate treating physicians (or in general, users with appropriate data access rights) are able to merge patient and treatment data. Furthermore, should one server be compromised, no association to the other data class is possible, thus protecting privacy of patients ([2], [3]). All internal and external communication is encrypted to prevent eavesdropping and possible "man in the middle" attacks (cf. Figure 2). Both data classes (IDAT and MDAT) are transmitted separately to the client using encryption and merged in the client's web browser only.

Queries on treatment data of specific patients are internally carried out using temporary IDs (TempID, cf. Figure 1). The TempID is generated by the IDAT server and propagated to the client *and* the MDAT server. The client then automatically queries the MDAT server for the corresponding patient's data using the TempID. The latter is invalidated upon completion of the transaction or after a specific idle period has passed. This procedure ensures that no transaction may be carried out twice by e.g. interception by unauthorized parties, and eliminates the possibility of malicious transactions not explicitly requested by authorized persons.

The Clearinghouse platform is implemented using inexpensive Open-Source software and is strongly based on the LAMP concept (**L**inux, **A**pache, **M**ySQL, **P**HP).

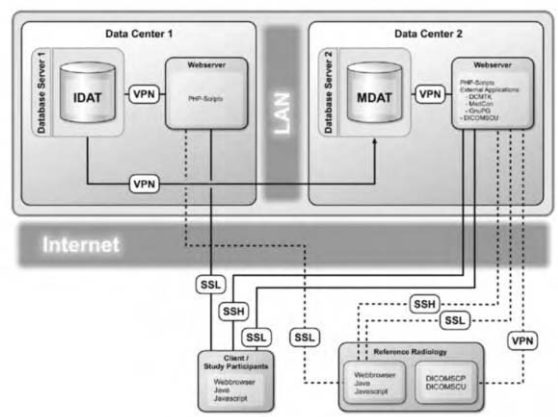


Figure 2 - Scheme of the Clearinghouse IT infrastructure

Infrastructure

The infrastructure of the central Clearinghouse teleradiology system (cf. Figure 2) resembles the theoretical data security concept as described in the previous section. The strict physical separation of the IDAT and MDAT data classes is realized by housing corresponding components in separate data centers, including system administration tasks carried out by different persons. Two web servers are used to deliver encrypted IDAT or MDAT content to the client's web browser. Communication to and among system components (i.e. database servers and web servers) is carried out exclusively via secured lines, e.g. using a virtual private net (VPN) for the Clearinghouse system components or SSL encryption for communication with clients. Reference centers, e.g. reference radiologies, may be integrated via VPN or using SSL/TLS encryption of the DICOM network protocol ([4], [5]).

Workflow

The Clearinghouse platform transforms the traditional decentralized workflow into a centralized one. The web application forms a central venue for all participants and data corresponding to clinical documents. On the one hand, the platform is suitable for mutual exchange of documents between several physicians and experts, to confer clinical diagnoses with high efficiency and low latency. On the other hand, it is also suitable as a central platform for patient and data acquisition within the scope of a clinical trial or study. Here, participating physicians, institutions and clinics contribute patients and corresponding relevant documents to a specific study, with the study center having access (probably pseudonymous only, depending on study treaties) to all contributed data.

Multiple studies may coexist within the same platform instance, without the need to have separate servers for each particular study. The security concept, including the role model, ensures that no data is exchanged between studies. Access to a study and patients or documents therein is granted by either the study manager and/or the person who contributes a specific patient. Read and modification rights may be granted to other users (e.g. of the corresponding study), while the right to delete objects, documents or patients is exclusively granted to the person who contributed the specific patient to the Clearinghouse platform. The study center always has read access to all contributed data and documents, as described above.

The Clearinghouse allows dispatching DICOM objects to previously specified, trusted Reference Radiologies by implementation of the DICOM STORESCP protocol. This way, reference diagnoses, which are then included in the corresponding patient's file, may be obtained when needed.

Figure 3 - The Java upload applet of the Clearinghouse teleradiology platform for upload of complete DICOM CDs

Clearinghouse features

The Clearinghouse teleradiology platform is specifically suited to handle, besides generic bitmap graphics formats, digital medical images encapsulated in the DICOM format. The DICOM format has become the most important format for the generation, storage and communication of digital medical images. This standard includes a comprehensive description of data organization within the DICOM container, as well as protocols for the networked communication of DICOM objects (e.g. STORESCP). The Clearinghouse supports the whole DICOM workflow by integration of Open-Source software and libraries (Offis DICOM toolkit [6], XMedCon [7]).

Apart from DICOM objects containing a single image the Clearinghouse is also able to handle multislice DICOM objects as generated by e.g. systems employed in cardiology diagnoses and treatment (e.g. heart catheter movies). Conversion of image objects into other formats (e.g. MPEG movies for multislice DICOM objects) is possible in almost any arbitrary way. Furthermore, up- and download of virtually any generic non-graphic file formats like PDF, Microsoft Office or Open Office documents etc. is possible.

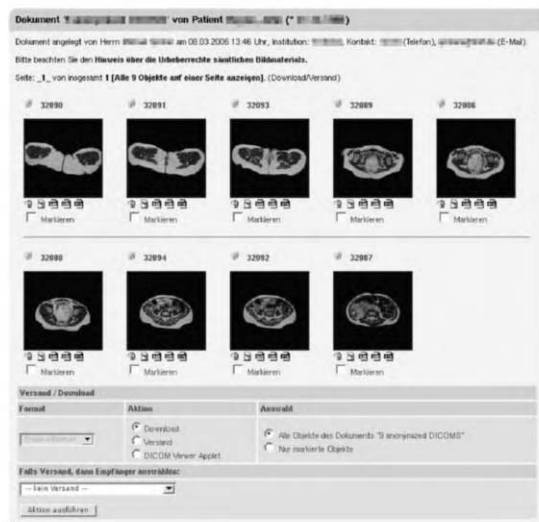


Figure 4 - Thumbnail overview of image objects contained in a specific document (some data garbled for privacy reasons)

The Clearinghouse enables straightforward upload of complete DICOM CD-ROMs by a sophisticated, operating system independent Java applet (cf. Figure 3). The applet is executed within the web browser of the user and automatically scans CD-ROMs for DICOM data. Once identified, these data are compressed, then transmitted to the platform via encrypted communication and stored within a previously defined document of the corresponding patient's record.

For each patient a virtually unlimited number of documents may be created, and each document may comprise any number of objects. Primarily, a document is presented as a thumbnail overview of all image objects therein (cf. Figure 4), automatically sorted according to the header data (e.g. based on study ID and/or slice number). By clicking on an object a detail view is entered, with (in case of a DICOM object) a listing of all available header data (cf. Figure 5).

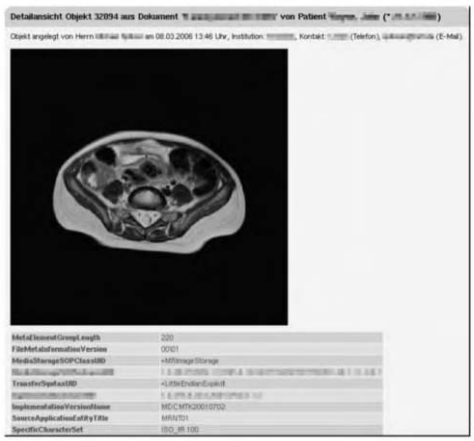


Figure 5 - Detail view of a DICOM image object including header data (some data garbled for privacy reasons)

To allow users the quick visualization of selected DICOM objects, the Radscaper [8] DICOM viewer (cf. Figure 6) was integrated into the Clearinghouse. As for the DICOM upload applet, the Radscaper viewer is written in Java and thus is operating system independent. This built-in DICOM viewer functionality relieves users to install standalone DICOM viewers on the client's workstations, and allows for portable and more detailed analyses of DICOM objects. The Radscaper applet features several tools e.g. for adjusting window center and width, zoom, image rotation and measurement functionalities. A cinematic mode allows advancing through multiple slices with adjustable speed.

Results

Advantages

By focusing on the exchange and annotation of digital image data the Clearinghouse platform improves the otherwise slow traditional transfer of medical images via standard mail. Both the risk of data loss and the response time are significantly decreased since

- geographical distances are transformed into virtual "technical" distances,
- the centralized structure guarantees for immediate access to any document as well as for reliable backup,
- IDAT and MDAT are securely stored in physically separate databases (cf. section *Data Security concept*),
- the sophisticated role-based user model enables to define access to documents with fine granularity.

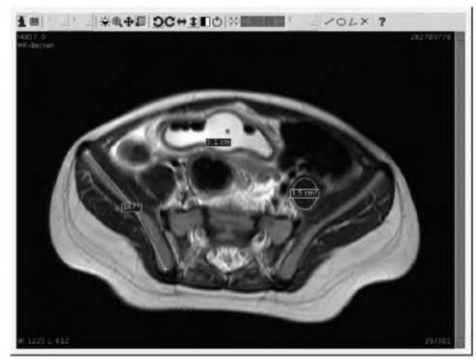


Figure 6 - The Radscaper Java DICOM viewer applet of the Clearinghouse teleradiology platform

In summary, the Clearinghouse telemedicine approach has the potential to improve diagnosis and treatment while data security and privacy are improved at the same time.

The cost-saving employment of the Clearinghouse platform is one of its major advantages. Due to consistent use of Open Source software the costs for installation, maintenance and support are comparably low. Furthermore, the platform constitutes an easily extensible and maintainable basis for future research projects.

Comparison to other teleradiology solutions

There are several solutions for teleradiological platforms, some following a similar architecture with a central web server communicated with by clients using a web browser [16], other implement a special Java-based solution for execution at the client's workstation [17] or use the DICOM-e-mail facilities of the DICOM standard [18]. The latter approach allows for simple participation in such teleradiology networks by use of a standard Email application. However, collaborative features such as instant access for multiple users are not available except by sending mails to several recipients simultaneously. Furthermore, utilization as an electronic patient record is not feasible since other document types would have to be encapsulated in a DICOM container first. The Java-based solution described in [17] features a security concept comparable to the Clearinghouse in terms of encrypted communication, yet supposedly without the strict separation of IDAT and MDAT data classes, missing in [16] as well. Additionally, the solution in [17] features teleconferencing abilities and an interface to existing PACS applications, a topic that has not been tackled yet in practice by the Clearinghouse system. Main advantages of the Clearinghouse platform over the aforementioned solutions are:

- no installation of specific standalone client software necessary, thus updates of the central system software are immediately available and usable by all clients,
- inexpensive setup and administration due to the use of Open-Source software and a centralized infrastructure,
- integration of reference centers by DICOMSCU/-SCP,

- immediate availability of uploaded documents and digital images for all legitimate users,
- a security concept based on the generic concept of the Telematic Platform for Medical Research Networks (TMF), Germany. The concept is commonly accepted among local institutions and fully complies with at least German legal requirements.

Current employment

The Clearinghouse platform has already been proven to be beneficial for e.g. studies and participating institutions by facilitating secure, instantaneous and geographically independent access to data and documents via the Internet. It is already successfully tested and used routinely within the scope of the Ewing study [9], the Nephroblastome and AML study [10], the competence network of pediatric oncology and hematology ([11], [12]), the Network for Ichthyosis and Related Keratinization disorders ([13], [14]) as well as the network for Epidermolysis bullosa [15].

Future development

The current development and research is based on already implemented functionalities of the Clearinghouse. It concentrates on the further development and full implementation of the data security concept as described in this document. Further projects involve the extension of DICOM integration and functionalities and embedding of a WYSIWYG editor for elementary online editing of documents, as well as internationalization and translation of the platform.

Acknowledgments

The project is partially funded by the German Federal Ministry of Education and Research within the scope of the Competence Network for Paediatric Haematology and Oncology GPOH/KPOH (code 01GI0413), the Network for Ichthyoses and Related Keratinization disorders NIRK (code 01GM0606), and the German beneficence for child cancer (Deutsche Kinderkrebsstiftung).

References

[1] http://www.tmf-ev.de/ [last access: 4th December 2006]
[2] Stanberry B. Legal and ethical aspects of telemedicine. J Telemed Telecare 2006: 12 (4): 166-175
[3] White P. Legal issues in teleradiology – distant thoughts! Br J Radiol 2002: 75 (891): 201-206
[4] Digital Imaging and Communications in Medicine: http://medical.nema.org [last access: 4th December 2006]
[5] Hussein R, Engelmann U, Schroeter A, and Meinzer HP. DICOM structured reporting: Part 1. Overview and characteristics. Radiographics 2004: 24 (3): 891-896
[6] http://dicom.offis.de/dcmtk/ [last access: 4th December 2006]
[7] Nolf E. XMedCon - An open-source medical image conversion toolkit. European Journal of Nuclear Medicine 2003: 30 (2):246.
[8] http://www.divinev.com/radscaper/ [last access: 4th December 2006]
[9] http://www.kinderkrebsinfo.de/e1664/e1676/e1758/e5468/index_ger.html [last access: 4th December 2006]
[10] http://www.kompetenznetz-leukaemie.de/ [last access: 4th December 2006]
[11] http://www.kinderkrebsinfo.de/kpoh/ [last access: 4th December 2006]
[12] Creutzig U, Jurgens H, Herold R, Gobel U, and Henze G. Concepts of the Society of Paediatric Oncology and Haematology (GPOH) and the German Competence Network in Paediatric Oncology and Haematology for the quality controlled development in paediatric oncology. Klin Padiatr 2004: 216 (6): 379-383 [Article in German]
[13] http://www.netzwerk-ichthyose.de/ [last access: 4th December 2006]
[14] Traupe H. Ichthyoses and related keratinization disorders. Management, clinical features and genetics. Hautarzt 2004: 55 (10): 931-941 [Article in German]
[15] http://www.netzwerk-eb.de/ [last access: 4th December 2006]
[16] Corella F, Lewison KP, Talukder M. A Teleradiology Architecture Featuring Security and High Performance. J Digit Imaging 2002: 15 Suppl 1: 214-215
[17] Munch H, Engelmann U, Schroter A, Meinzer HP. The integration of medical images with the electronic patient record and their web-based distribution. Acad Radiol 2004: 11(6): 661-668
[18] Weisser G, Engelmann U, Ruggiero S, Runa A, Schroter A, Baur S, Walz M. Teleradiology applications with DICOM-e-mail. Eur Radiol 2006 [Epub ahead of print]

Address for correspondence

Michael Spitzer
Department of Medical Informatics and Biomathematics
University Hospital, University of Muenster
Domagkstr. 11
48149 Muenster, Germany
Email: spitzer@imfl.de

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

K-Box: Automatic Structuring and Exchange of Medical Documents Based on the Clinical Documentation Architecture (CDA)

Minh H. Doan, Paul-Ludwig Lott, Marek Václavík, Prof. Dr. Frank Ueckert

Department of Medical Informatics and Biomathematics, University Hospital, University of Muenster, Germany

Abstract

Sustainable health systems should understand information as a resource which also has to be managed efficiently. Today's electronic documentation is only the mandatory first step to automatic information exchange between hospitals, physician offices, pharmacies and other participants. The K-Box is a modular system which allows different monolithic information systems to be connected and integrated by using the clinical document architecture (CDA) as the standard document exchange format. A prototype K-Box has been implemented and a trial run is scheduled. The workflow selected for the trial produces discharge summaries in CDA format.

Keywords:

information systems, systems integration,
public health informatics

Introduction

Defining historical medical attendance records as an instance of *information* in the context of health system, today's hospitals and other healthcare organizations should understand *information* as a resource beside other resources. Furthermore, we assume that within this context information will become one of the most expensive resources in the future. Missing information would mean additional or double tests in some cases and cause extra costs as the consequence. Therefore, the development of sustainable health systems also has to pay attention to *save* the resource information by endeavoring after its *reuse*.

Today, in German hospitals medical data are processed and handled by hospital information systems (HIS') which also include electronic patient records (EPRs) [1]. Also, most of the medical offices in Germany usual use computer-aided patient management systems to save medical data related to constituent patients. Regarding the German health system we can notice, therefore, that the evolutional step from paper supported documentation to electronic documentation is mostly done.

Even though the documentation is done electronically, there is a big gap in electronic *communication* between German healthcare organizations because the electronic documentation was not forced to follow a nationwide or international standard. As a result, there is no integration or direct communication between different information systems in different organizations possible [2]. Some organizations joined together and implemented electronic communication between their systems. But the majority is still exporting their medical information by using word processing software to write a report. The importing party then has to copy the data from the written down report and paste it into their information system manually.

Regarding building sustainable health systems, we are sure that the ability to automatic information exchange between the system elements would enhance the property "sustainability" of the system itself. The *K-Box* described in this paper is such a software system which allows different medical information systems to be connected and exchange structured information.

The idea behind the K-Box is to have handy servers ("boxes") at hospitals, physician offices, pharmacies and other participants, which are connected to local information systems to provide them a unique communication and data exchange capability. At this phase of the project we focus on clinical document exchange between physicians such as exchange of discharge summaries. The "K" in the name of the project is representing "Kommunikation" as the German word of communication. But actually, beside the communication subsystem the K-Box also comprise another subsystem which we call the *structuring* subsystem. The structuring subsystem is able to convert every document it receives in a CDA (Clinical Document Architecture) document by using machine-learning supported methods. Based on the generated CDA particular information entities then can be extracted and imported directly into foreign systems.

The K-Box architecture is modular. Each K-Box installation can be extended with the support of new document formats or communication interfaces by adding pluggable components ("Plugins") to the running system.

Materials and methods

Component-Based Software Engineering

The K-Box development follows the paradigm of Component-Based Software Engineering (CBSE). That means that we focus on software components in all phases of development. Furthermore, all components have their own

development cycle, and are released independently from each other.

The term of *software component* is not new to the software engineering domain. It was first used in 1968 by Doug McIlroy at the NATO software engineering conference in Garmisch where the discipline Software Engineering itself was born. In his article McIlroy's described his vision - based on the model of mass production in the industry - about mass development of software by combining *reusable* components instead of programming blocks of similar functionality from zero every time [3]. The term reusability is often used in the context of Object-oriented Programming (OOP). From this background CBSE is often confused with OOP. In fact, reusability is the main motivation of both paradigms. However, we distinguish between white-box reusability and black-box reusability: White-box reusability means that we reuse written source code and add modifications to it to solve a similar but not the same problem. On the other side black-box reusability means that we reuse independent software components in binary format to solve, intra-domain, independent, and repeating sub-problems [4]. It is the art of CBSE to design components that are not too specific to a context to make them widely reusable but also not too abstract. In other words, OOP makes use of white-box reusability and CBSE makes use of black-box reusability.

OOP and CBSE are no competitive paradigms. On the opposite, it is wise to combine both techniques in an advantageous way. In our project, we use OOP to develop the components themselves and follow CBSE to design the component hosting core-system. The OOP language used at the K-Box project is Java.

Information extraction

To process incoming messages, a typical commercial EAI system relies on explicit structure definitions. It expects a description written in a vendor specific meta-language. This approach reaches its limits when the explicit message structure is unknown (e.g. undocumented legacy systems) or too complex to be expressed this way. Integration engineers usually apply workarounds to solve this situation: they "reconstruct" the message structure by generalizing available examples or they content themselves with a functional definition: "How do I extract a certain data element?" Learning the structure "by example" can also be performed automatically or semi-automatically, by means of information extraction (IE) and machine learning. IE with text data (*text extraction*) has been widely applied in the medicine domain for biomedical data mining. (e.g. [5])

The task of text extraction comprises populating data slots in a given target format with information pieces from the input text. *Text classification* means analysing a text part and assigning it with a discriminating attribute (*tagging*), like "relevant" versus "irrelevant" or "regular mail" versus "spam". The choice of the appropriate text extraction method depends on the *structuredness* of the text. Extracting information from an *unstructured* text (words, sentences, paragraphs) requires linguistic analysis (*natural language processing, NLP*) to maintain a certain level of *text reasoning*. A s*emi-structured* text contains additional information (e. g. presentation mark-up), enabling non-NLP techniques to reach comparable or even better results. Non-NLP methods are typically based on sets of extraction rules (*wrappers*), often modelled as finite-state automata [6]. Both approaches can utilize machine learning, i.e. to adapt the extraction process according to the received feedback.

We apply adaptive text extraction to discharge summaries, trying to turn them from an unstructured form (binary Word-File) to a structured XML-document conforming to Clinical Document Architecture (CDA). CDA is a document standard based on the HL7 v3 Reference Information Model. In Germany, the SCIPHOX-specification (CDA Release 1) developed into the "Electronic Medical Report" (eArztbrief), derived from "CDA Release 2".

Results

Architecture of the K-Box

The K-Box architecture conforms to the layered model for message-based integration systems, as described by Keller [7]. Four functional layers can be distinguished: An adaptor layer which supports various communication protocols, an internal transport layer which is responsible for intra-component-communication, a format transformation layer which converts a document to a new format, and a workflow layer which routes documents though the K-Box.

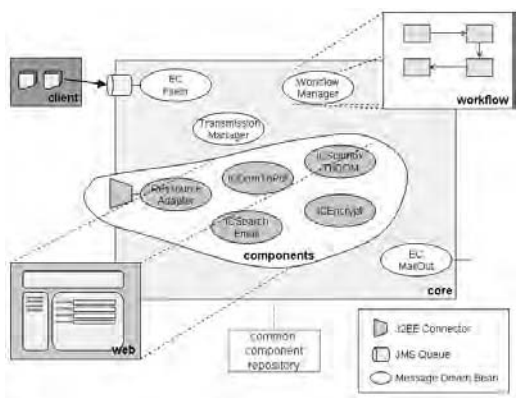


Figure 1 - Modules of the K-Box architecture

The entry point for a document into the K-Box is an inbound adaptor. Inbound adapters are components of the communication subsystem. They are responsible for the receipt of documents through different communication protocols, e. g. FTP, SCP, or SMB/NetBIOS. After a document is received the inbound adapter creates a new transaction within the K-Box. Each transaction is related to an internal document envelope (IDE) which contains the document itself as an attachment and further meta information in the header part of the IDE. The body part of the IDE is reserved for the CDA. Subsequently, the document is routed through the K-Box following the information

from a predefined workflow. After the document is finally converted the result will be sent by an outbound adaptor.

From the software design point of view, K-Box has been conceived as a component-based software system, which adds new functionality dynamically. The development of the component infrastructure takes place in the project module we call the *core*. Beside the core we declared four further project modules, namely: *components, web, workflow* and *client* (figure 1).

kbox.core is a J2EE application containing the essential *SessionBeans*, a *WorkflowManager* and a *TransmissionManager* for tracking and logging. It also includes the generic *AbstractIC* and *AbstractEC* classes, designed as abstract *MessageDrivenBeans* (MDB). Each MDB is bound to a JMS-Queue (*Java Message Service*). The queues enable asynchronous message exchange between the components. The internal message type used in the K-Box is a SOAP message with attachments, a protocol related to the Web Services technology.

Within the core we develop **kbox.components** for external connectivity (resource adaptors - *J2EE Connector Architecture*) and message transformation.

The **kbox.web** GUI allows K-Box administrators to manage and configure the K-Box core. It also provides a uniform view of the components and their properties.

kbox.workflow is a stand-alone Java application for graphical design of workflow objects. The constructed workflow object can be saved locally or uploaded to the connected K-Box core by employing the remote WorkflowManager.

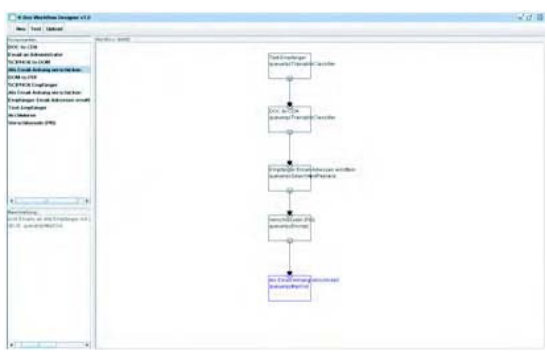


Figure 2 - Screenshot of the graphical workflow designer

kbox.client serves as an alternative K-Box adaptor, meant for external systems without a J2EE resource adaptor. kbox.client is a generic, extendable JMS client, capable of sending byte stream messages to a specific EC in a configured K-Box.

Information extraction from unstructured and semi-structured documents

Discharge summaries in an unstructured form are unsuitable for further automatic processing. Provided there are enough document examples available, machine learning can be applied. One of K-Box's transformation components extracts CDA body information using text classification. CDA section captions (Level 2) such as "Diagnosis" or "Therapy" are being assigned to a particular text-block. For higher domain independence the method considers only unstructured (free)

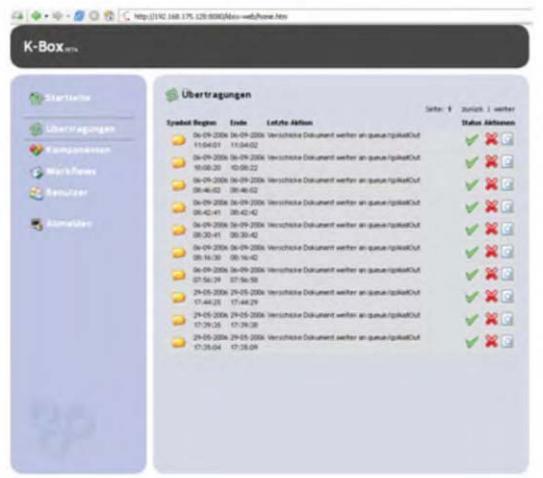


Figure 3 - Screenshot of kbox.web

formatting information. For this method to be effective, the system has to be trained with typical data. During the training process the user marks up a few documents manually and tells the system to which categories the document contents should be mapped.

Document transformation consists of three steps: plain text extraction, text-block classification and classified block processing. In the first step control characters, tags and proprietary mark-up are removed from the document by a learning algorithm comparable to those used in today's email-spam-filters. The normalized document is then passed to the second step where it is divided into distinct text blocks. Each block is then automatically assigned with a certain content-category (e.g. „Diagnosis"), thus adding semantic information to the document. After classifying all blocks, the document can be transformed to the destination format, using the semantic information from the previous step. Further, some more tightly focussed post-processing can be applied according to the content category. For example, a text-block classified as "Diagnosis" is likely to contain ICD-codes. These can be extracted automatically and converted to a highly structured, CDA Level 3 compliant document. The classification is based on NGRAM-Analysis, an adaptive classification algorithm which already proved to be effective in other areas of information extraction. [8]

Discussion

The K-Box is an integration tool with the focus on automatic structuring of unstructured medical documents. While commercial EAI-products are naturally developing into powerful and universal integration suites, a simple interface engine presents in many use cases a more transparent and manageable solution. Compared to other

branches, healthcare seems to adapt new IT trends rather reluctantly [9]. Replacing "interface engines" with more complex, business-process oriented EAI-tools is still at the beginning [10]. Straightforward and lightweight solutions might be particularly appreciated in smaller enterprises, with minimalist equipment and limited budget, such as medical practices. This is one of our target groups.

The component-based design should increase acceptance of K-Box among potential developers. The final component specification must find a reasonable trade-off between flexibility and complexity. A component repository should serve as a platform for component exchange inside the user community. The basic prerequisite is a complete development framework which is the objective of our current development.

Information extraction from websites [11], often wrapper-oriented and combined with machine learning [6], has been widely explored and exploited. However, there have been significantly fewer attempts to utilize these methods in the Enterprise Application Integration [12] or specifically in messaging systems. The goal of our approach is to minimize the need for explicit message structure descriptions or formal transformation-rules. At present we are exploring methods for precise and domain-independent extraction of CDA header data. Incorporating domain specific information might increase the quality of the extraction, but has to be added in a flexible, replaceable way.

The K-Box supports protocols and data formats that are usually used in healthcare. Such an emerging standard for inter-sectoral communication is the CDA. This format has been chosen for the nationwide health telematics platform in Germany. By means of the freetext-to-CDA transformation the K-Box provides the link between an isolated standard-unaware information system and the centralized infrastructure.

To receive a field test experience with the architecture, a prototype of the K-Box has been deployed in one of the departments at the University Hospital of Muenster. The implemented workflow transmits CDA discharge summaries to external recipients via encrypted emails. A trial run is the next step.

References

[1] Uslu A, Stausberg J. Nutzen und Kosten der Elektronischen Patientenakte. In: Jäckel, eds. Telemedizinführer Deutschland. Bad Nauheim: 2006; pp. 151-5.

[2] Kaiser RH. Elektronische Patientenakten aus aerztlicher Sicht. In: Jäckel, eds. Telemedizinführer Deutschland. Bad Nauheim: 2006; pp. 184-91.

[3] McIlroy D. "Mass Produced" Software Components. In: Bauer FL, Bolliet L, Helms HJ. Software Engineering. Report on a conference sponsored by the NATO Science Committee; 1968 October 7-11; Garmisch, Germany: pp. 138-57.

[4] Balzert H. Lehrbuch der Software-Technik – Software-Entwicklung. 2nd ed. Spektrum - Akademischer Verlag. 2000.

[5] Gaizauskas R, Demetriou G, Artymiuk PJ, Willett P. Protein Structures and Information Extraction from Biological Texts: The PASTA System. Bioinformatics. 2003; 19 (1): pp. 135–43.

[6] Kushmerick N. Finite-state approaches to Web information extraction. In: Pazienza MT, eds. Information Extraction in the Web Era: Natural Language Communication for Knowledge Acquisition and Intelligent Information Agents; Springer-Verlag; 2005: pp. 77-91. ISBN 3-540-40579-8.

[7] Keller W. Enterprise Application Integration - Erfahrungen aus der Praxis; dpunkt verlag; 2002. ISBN 3-89864-186-4.

[8] Cavnar WB, Trenkle JM. NGram Based Text Categorization. Proceedings of the Third Annual Symposium on Document Analysis and Information Retrieval; 11-13: April 1994; pp. 161-9.

[9] Laroia A. Healthcare's Digital Heartbeat. eAI journal. August 2002; pp. 44-7;

[10] Goth G. Integration Engines - A future of opportunities. Healthcare Informatics. 2005/08; pp. 14-8. Available from: http://www.healthcare-informatics.com/issues/2005/08_05/trends.htm [cited Jan 30, 2006].

[11] Henzinger MR, Motwani R, Silverstein C. Challenges in Web Search Engines; Proc. of the 18th International Joint Conference on Artificial Intelligence. 2003; pp. 1573-9.

[12] Kuhlins S, Korthaus A. A Multithreaded Java Framework for Information Extraction in the Context of Enterprise Application Integration. Proceedings of the 1st international symposium on Information and communication technologies; 2003 September 24-26; Dublin, Irland. Dublin: Computer Science Press Trinity College; 2003. pp. 535-40.

Address for correspondence

Prof. Dr. med. Frank Ueckert
Phone: +49 251 83-52773
Email: ueckert@imfl.de
Minh H. Doan
Phone: +49 251 83-58215
Email: doan@imfl.de
Institut fuer Medizinische Informatik und Biomathematik
Domagkstraße 11
48161 Muenster
Germany

Chapter 4.

Medical Knowledge, Ontologies and Terminologies

This page intentionally left blank

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Using Distributional Analysis to Semantically Classify UMLS Concepts

Jung-Wei Fan[a], Hua Xu[a], Carol Friedman[a]

[a] *Department of Biomedical Informatics, Columbia University, USA*

Abstract

The UMLS is a widely used and comprehensive knowledge source in the biomedical domain. It specifies biomedical concepts and their semantic categories, and therefore is valuable for Natural Language Processing (NLP) and other knowledge-based systems. However, the UMLS semantic classification is not always accurate, which adversely affects performance of these systems. Therefore, it is desirable to automatically validate, or, when necessary, to semantically reclassify UMLS concepts. We applied a distributional similarity method based on syntactic dependencies and -skew divergence to classify concepts in the T033 Finding class in order to determine which ones were biologic functions or disorders. A gold standard of 100 randomly sampled concepts was created that was based on a majority annotation of three experts. Precision of 0.54 and recall of 0.654 was achieved by the top prediction; precision of 0.64 and recall of 0.769 was achieved by the top 2 predictions. Error analysis revealed problems in the current method, and provided insight into future improvements.

Keywords:

semantic classification, UMLS, distributional similarity, natural language processing

Introduction

Ontological knowledge bases are important for developing Natural Language Processing (NLP) and other knowledge-based applications because they provide well-defined specifications of the concepts, semantic classification, conceptual normalization, and indicate relevant semantic relations between the concepts. The specifications help define the meaning and semantic category of each concept unambiguously whereas normalization helps unify synonymous surface terms to enhance retrieval. For example, blood platelets and thrombocytes should be normalized into a single concept. The relations defined in the ontology can be used in semantic parsing or pattern recognition as well in logical inference. For example, a semantic parsing rule that uses a broad semantic class, would be satisfied for concepts that are children of the broad class. The parent/child relations can also be useful in question answering systems by providing knowledge of hierarchical relation and use of inheritance. More specifically, information about malignant melanoma can be implied from texts containing its descendents, such as necrotic melanoma or choroidal melanoma, diffuse.

There are many such structured knowledge bases in the biomedical domain, e.g. the SNOMED-CT [1], MeSH [2], and Gene Ontology [3]. The most comprehensive integration of these resources is the Unified Medical Language System (UMLS) [4], which includes 139 source vocabularies in the current release (2007AA). Each concept in the UMLS has a unique identifier, called CUI, which is assigned one or more semantic types in the Semantic Network (SN) [5]. The SN also contains relations defined between the semantic types, and at the concept level, there are also various relations inherited from the source vocabularies. The semantic classification of a CUI, when appropriate, is very valuable for NLP applications. Textual patterns can be recognized on the basis of UMLS relations and the semantic types of UMLS concepts identified in text, or developers can define their own application-dependent relations based on UMLS semantic types. For example, the template [(*Food*) **interacts with** (*Pharmacologic Substance*)] can be used to screen potential food-drug interactions in text. The SN has been used to determine relations between the concepts extracted through specified semantic patterns [6] and used as semantic filters in literature-based knowledge discovery [7]. In this paper we focus on semantic classification within the UMLS because it is a comprehensive system in the biomedical domain with continuous maintenance and a large user population.

Although the SN semantic types are a valuable resource, there are issues about their granularity and correctness. For example, the current release assigns C0021494 Injections, Intravenous to T169 Functional Concept, but it seems that it would be more appropriate if it were assigned T061 Thera peutic or Preventive Procedure or a broader class *procedure* for general NLP purposes. Research has been performed to coarsen the granularity of the SN, which involves combining SN types [8] [9] [10], and to audit SN errors [11] [12]. However, the above coarsening approaches only aggregate existing types into broader classes without considering the validity of SN assignments themselves, and the auditing approaches work by finding contradictions derived from the existing semantic type assignments. The drawback of relying on existing assignments is that inappropriate assignments are retained along with the high-level reorganization. For example, through inheritance Injections, Intravenous would still not be con-

sidered a procedure-related class by the above methods. On the other hand, we cannot assume that all concepts in the Functional Concept class belong to *procedure*, since the SN type also includes concepts such as Solar system, Poisonous effects, and Hairy.

We have observed that some SN types are more problematic than others because they are high level, vague, or heterogeneous. We consider that a SN class is vague when it includes many CUIs that are very semantically heterogeneous or inappropriate for that class. For example, T033 Finding contains many concepts of the disorder sense (e.g. Nasal cartilage loss and Hemianopsia), but also contains many different types of concepts that are not clinically relevant, such as Spiritual state alteration, or that are not well-defined, such as Wanted, which requires additional context to determine the correct meaning. For example, the online Free Dictionary has over nine definitions for wanted (http://www.thefreedic tionary.com/wanted) with examples including context for clarity. This causes some applications, such as [6], to avoid using the concepts in these classes, but then relevant concepts are also dropped. For example, Progressive renal failure and Acute inflammation are in T033, and would be missed by applications using only well-defined classes, such as T047 Disease or Syndrome. Therefore, it is important to develop an automated method for reclassifying UMLS concepts, particularly those that are assigned vague SN types, into more appropriate semantic classes.

In the biomedical domain, Sibanda et al. showed that syntactic features are especially useful for determining the semantic categories of terms and clauses in discharge summaries, using Support Vector Machine classifiers [13]. Pedersen et al. used window-based contextual distributions from the Mayo Clinic Corpus of Clinic Notes to measure the semantic similarity between SNOMED-CT concepts, and reported high correlation with human experts [14]. Weeds et al. used distributional similarity based on syntactic dependencies and a nearest neighbor voting process to classify terms into semantic types of the GENIA ontology and achieved good accuracy [15]. In a previous paper [16] we showed that it was feasible to apply a distributional similarity approach to classify UMLS concepts. The distributional approach is based on Harris sublanguage theory: terms are syntactically dependent on other terms with unequal likelihoods especially in specialized domains [17], so that terms can be characterized and classified through the distribution of their syntactic dependencies. For example, the adjective psychogenic is more likely to be an adjective of a noun that is a disorder and unlikely to modify a noun that is a microorganism. We used Lees -skew divergence [18] as the measure of distributional similarity:

$$S(\mathrm{P}(C|t), \mathrm{P}(C|t_2))$$

$$= D(\mathrm{P}(C|t_2) \| \quad \mathrm{P}(C|t_1) + (1\text{-}) \quad \mathrm{P}(C|t_2)) \qquad (1)$$

where 0 1, t_1, t_2 are two terms, C is the union of the syntactic dependencies of t_1, t_2 , $\mathrm{P}(C|t_1)$, $\mathrm{P}(C|t_2)$ are the corresponding distributions, and D is KL divergence. We generalized the approach to one that measures similarity of concepts instead of individual terms. This is accomplished by forming each concepts distribution with the syntactic dependencies of all its synonymous terms in a training corpus. When reclassifying concepts of well-defined SN types into seven broader classes, we achieved a precision of 0.802 for the top prediction and 0.884 when considering the top two predictions. An advantage of the corpus-based approach is that it is automated and therefore high-throughput. In addition, it is based on real language usage and thus tends to be more reproducible than experts using their own judgments, which may vary from person to person and thus consistency would be more difficult to achieve. Consistency using multiple experts could be obtained but would necessitate a costly process involving consensus.

In this paper we focus on evaluating the classification of CUIs for the SN type T033 Finding because it includes many important concepts associated with disorders and functional information, e.g. some disorder-related adjectives like Hypertensive are only covered by T033. Our approach differs from the related work in that we used a typical distributional similarity measure, -skew divergence, based on more abundant types of syntactic dependencies obtained from a huge shallow-parsed corpus to reclassify/audit UMLS concepts. It also differs from other UMLS concept classification methods because our approach automatically proposes the most appropriate semantic class for a CUI, regardless of the SN type(s) that were originally assigned.

Materials and methods

The distributional classification method can be outlined as follows, and is elaborated on below: 1) determining clinically relevant semantic classes for building distributional profiles, 2) obtaining a training corpus, 3) extracting syntactic dependencies from the corpus for CUIs in the relevant classes, 4) using the syntactic dependencies to build distributional profiles for the classes, 5) using syntactic dependencies to build distributional profiles for the test CUIs, and 6) using -skew divergence to compute the distributional similarities between the CUI and class profiles for classification.

Based on previous work [16] we grouped subsets of the SN types (2006AC) into seven clinically relevant broad classes: *biologic function*, *anatomy* (above the molecular level), *disorder*, *gene or protein*, *microorganism*, *procedure*, and *substance*. For example, Laboratory Procedure, Diagnostic Procedure, and Therapeutic or Preventive Procedure were grouped to form the *procedure* class. Vague types, such as Finding, Functional Concepts, and Health Care Activity were excluded from forming the seven classes. We obtained a corpus of 199K MetaMap-processed abstracts fr[1]om the 2005 MEDLINE/PubMed Baseline Repository (MBR) database [20]. Perl scripts

1 MetaMap is a program that performs part of speech tagging, shallow parsing, and statistical methods to map terms in free text to the UMLS concepts. See [19].

were created to extract the part-of-speech (POS) tags, identify the phrase types, and trace the mapped CUIs to the original extractions from the machine-readable format of the requested corpus. For example, the machine-readable output of a sentence containing the concept Hemianopsia would be processed as shown in Figure 1 (simplified for clarity). Then we used a set of context-searching rules to extract syntactic dependencies for each concept. For example, in Figure 1 the syntactic dependency {associated noun: notch} can be obtained for Hemianopsia from the sentence The macular notch in hemianopsia was studied because the attached prepositional phrase in hemianopsia modifies notch. We aggregated the syntactic dependencies of all the CUIs in each class to build the distributional profile for that class. For a CUI to be classified, the syntactic dependencies of that CUI were used to build its distributional profile. Classification was performed by computing the distributional similarities between the test CUI and each of the seven classes, and the class with highest similarity score was selected.

S: The macular notch in hemianopsia was studies.
P: The macular notch NP
W: The det
W: macular adj
W: *notch noun
P: in hemianopsia PP
W: in prep
W: *hemianopsia noun
E: hemianopsia noun
P: was AUX
W: was aux
P: studied. VP
W: studied verb

Figure 1 - A sentence with POS tags, phrase types, and CUI mappings reconstructed from MetaMap output. Notations: S- sentence, P- phrase and phrase type, W- word and POS, asterisk marks the head noun, E- mapped term, concept term, and CUI

We semantically classified CUIs in T033 to determine the concepts in that type that were *biologic function* or *disorder*, but not other types because those two classes constituted the two most clinically important concepts in that class. There are a total of 55,445 concepts in T033 (2006AC), and 848 of them occurred in the 199K training corpus. We tested our method under feature-sufficient condition (10 syntactic dependencies) to ensure that we would have the appropriate number of features, since it would be possible to obtain more syntactic dependencies by expanding the training corpus. As we were also interested in studying if concepts that rarely occur in a large body of literature were of less value for clinical applications, we qualitatively evaluated some T033 concepts of varying frequencies in the training corpus.

To evaluate recall of our method in classifying T033 concepts into *biologic function* or *disorder*, we randomly sampled 50 CUIs from the set of CUIs that had 10 syntactic dependencies. For evaluating the precision, we randomly sampled 50 CUIs from the set that were classified as *biologic function* or *disorder*. Recall and precision by the top prediction and the top 2 predictions were both calculated, where the latter means the correct class was within the highest 2 classes in terms of the similarities ranking. The gold standard was generated by taking the majority judgments of three medical experts. They were given both the recall and precision test sets. For each set, the testing CUI and corresponding strings in the UMLS were displayed, and the experts were asked to annotate each concept as *biologic function*, *disorder*, or *neither of the two classes*. Kappa statistics [21] for the inter-annotator agreement were calculated. We also performed an error analysis on the recall and precision set respectively.

Results

Exemplary concepts of varying frequencies in the corpus are displayed in Table 1; we found that frequency does not reflect the usefulness of concepts, because concepts that seemed useful or not useful were distributed similarly over different frequencies. There were 365 CUIs (for evaluating recall) with 10 syntactic dependencies extracted from the corpus, and 238 of them (for evaluating precision) were classified as *biologic function* or *disorder*. There were 26 of the 50 sampled CUIs in the recall set annotated as *biologic function or disorder*, and they served as the denominator in evaluating recall. The recall and precision of our classifier is shown in Table 2. The Kappa statistics of all the three annotators on the recall set was 0.421, and was 0.462 on the precision set.

Table 1 - Concepts with different frequencies in the corpus

Frequency	Examples
0 or 1	Abdominal swelling, Gingival bleeding, Partial deafness, Abnormal or prolonged prothrombin time, Birth Place, Vegetarian
5 and < 10	Panting, Hypoesthesia, Hyperlactatemia, Diaphragmatic paralysis, Vegan, Vagabond
10	Albuminuria, Sudden death, Fever, Cardiomegaly, Hematuria, HIV positive, Nervousness, Problem, Unemployment, Divorced

Table 2 - Recall and Precision of the classifier

	Recall	Precision
By top prediction	0.654	0.54
By top 2 predictions	0.769	0.64

We performed an error analysis based on the results by the top predictions. The errors in the precision set could be

grouped into two types: 1) irrelevant concepts were classified as *biologic function* or *disorder*, 2) *biologic function* and *disorder* concepts were confused with each other. The first group had errors for concepts such as Wanted, Optimistic, and Spells. Some examples of errors in the second group are: Low birth weights as *biologic function*, Normal vision as *disorder*, and Hyperalgia as *biologic function*. Errors in the recall set could also be grouped into two types: 1) errors due to confusing *biologic function* with *disorder*, 2) concepts belonging to the two classes but classified otherwise. In the first category, Stress was classified as *biologic function*, as were Low birth weights and Unresponsiveness. In the second group Does move and Assisted were both classified as *procedure*, while the gold standard had them as *biologic function* and *disorder* respectively.

Discussion

From Table 1 we determined that some frequent concepts may be too general to be informative (e.g. Problem), suggesting the need to filter out some very general and uninformative concepts even before performing classification. In contrast, lengthy terms like Abnormal or prolonged prothrombin time are less likely to occur, which is consistent with the estimation by [22] that only about 10% of the UMLS strings could be found in MEDLINE. The Kappa statistics over the three annotators indicated only moderate degree of agreement, which manifests the difficulty of the task. However, noticing that one annotator often checked *neither of the two classes* when the other two agreed on *biologic function or disorder*, we considered the gold standard generated by voting to be adequate overall but future evaluations should be performed using more experts. The recall and precision by the top 2 predictions show that the similarity function did rank the correct class higher out of the seven, but the performance was not satisfactory. We hypothesize that the T033 set, with more heterogeneous, high-level, and ambiguous concepts, poses a harder task for automatic classification than the other well-defined semantic types we tested (e.g. Disease or Syndrome and Organism Function).

We summarize and provide examples of the main classification errors in Table 3. The first category contained concepts that are very general English terms, and that do not have a well-defined meaning without additional context. For example, Wanted was classified by our method as a *biologic function*; we believe this term alone is much too vague to be assigned as a concept. Another example was Spells, which is also very ambiguous, and thus, all three annotators marked it as being *neither of the two classes*. When examining the corpus, we found it was classified by our method as *disorder* because frequently the term Spells occurred in contexts such as apneic spells, cyanotic spells, or spells of respiratory distress. It is likely that the annotators would agree on the *disorder* sense, if given the corresponding contexts. Literally, Spells could also mean the *biologic function* of being able to spell. The second category characterized behavioral/affective observations and also included many general English terms. Some examples of differences between our method and the gold standard are Apprehension and Patient noncompliance, which were classified as *disorder*, whereas Optimistic and Well-being were classified as *biological function*; however, in the gold standard these are *neither of the two classes*. The above examples show that our method is very sensitive, but not specific enough to rule out imprecise classifications. The third error category occurred when there was a combination of a modifier (usually a qualitative adjective, such as *low or normal*) and a clinical attribute, such as *weight* or *blood pressure*. These errors could result from the limited coverage of our classes. For example, there was not a class for findings specified as normal. It is also possible that the distributional approach *per se* is not adequate to differentiate that type of nuance, and lexical features from the concept strings may need to be introduced. Concepts in the fourth error category were misclassified based on contextual distributions but could have been classified more appropriately using morphologic information. For example, the suffixes *-algia* and *-opsia* frequently occur with disorders. Thus, from the results, it appears that, other features in addition to contexts, should be used for semantic classification because syntactic dependencies are not adequate in certain situations.

There were also borderline cases where one of the annotators agreed with our classification. For example, we classified Resting state as *biologic function* and Atypia as *disorder*, and each agreed with one annotator. There were some ambiguous cases, which stemmed from the strings associated with the concept. For example, the concept Sweating, included the two strings Sweats, function and Sweating symptom NOS from the source vocabularies. The UMLS did not sharply differentiate the two different senses. However, our method obtained the latter sense, while the annotators agreed on the former.

Table 3 - Categories of the classification errors

Category	Examples
1 Common English term	Does move, Spells, Wanted
2 Behavioral/ affective observation	Apprehension, Patient noncompliance, Optimistic, Well-being
3 Modifier + clinical attribute	Low birth weights, Blood pressure normal, Normal vision
4 Disorder by morphology	Hyperalgia, Hyperosmolarity, Hemianopsia

In summary, we believe that most of the errors resulted from three main limitations of our current approach: 1) the specificity issue associated with the first two categories of error indicates the lack of a global cutoff value and/or a

negative class distributional profile, thus resulting in many false positives, 2) the third error category suggests that the coverage of our existing classes is not adequate to appropriately accommodate some concepts such as neutral clinical findings, 3) errors in the third and fourth error categories show that using only the contextual syntactic dependencies could miss morphological and lexical cues that are essential to precise semantic classification, especially for concepts that are equivocal for multiple classes when given only the contexts. For example, in previous work we observed that *biologic function* and *disorder* were easily confused by our method, contributing to one third of the misclassifications because many pathologic functions were classified as normal biologic function. We believe this issue was exacerbated when the method was applied to T033 because it contained many poorly defined concepts.

Although the performance of our automated method was lower than expected, we believe automatic classification is possible and highly desirable, and will explore adding additional features to our method as well as developing complementary methods and performing more extensive evaluation. More specifically, future work will focus on 1) filtering out concepts that are too general to be useful before and/or during the classification, 2) reconsidering the formation of the classes and possibly incorporating more classes, and 3) reducing the classification errors that occur between *biologic function* and *disorder*. After the implementation refinements, we would also like to reclassify concepts under other types such as T169 Functional Concept. For example, we have observed that some T169 concepts such as intramuscular injection were appropriately classified as *procedure* by our method.

Conclusion

We used a distributional similarity approach to classify the concepts of the UMLS T033 Finding class into *biologic function* or *disorder*. The precision and recall were 0.54 and 0.654 by the top prediction, and 0.64 and 0.769 by the top 2 predictions. The error analysis indicated this task was intrinsically difficult when using only a distributional similarity measure and also helped identify the main factors hampering performance. We conclude that there is room to improve the method and it is definitely worth exploring further.

Acknowledgements

We thank Drs. Amy Chused, Daniel Stein, and Peter Hung for annotating the gold standard. This work was supported by Grants R01 LM7659 and R01 LM8635 from the National Library of Medicine.

References

[1] College of American Pathologists. SNOMED Clinical Terms: http://www.snomed.org/snomedct/index.html

[2] National Library of Medicine. Medical Subject Headings: http://www.nlm.nih.gov/mesh/

[3] The Gene Ontology Consortium. Gene Ontology: tool for the unification of biology. Nature Genet. 2000;25:25-9.

[4] Lindberg DA, Humphreys BL, McCray AT. The Unified Medical Language System. Methods Inf Med. 1993;32(4):281-91.

[5] McCray AT. An upper level ontology for the biomedical domain. Comp Funct Genom 2003;4:80-84.

[6] Rindflesch TC, Fiszman M. The interaction of domain knowledge and linguistic structure in natural language processing: interpreting hypernymic propositions in biomedical text. J Biomed Inform. 2003;36(6):462-77.

[7] Weeber M, Klein H, Aronson AR, Mork JG, de Jong-van den Berg LT, Vos R. Text-based discovery in biomedicine: the architecture of the DAD-system. Proc AMIA Symp. 2000;903-7.

[8] McCray AT, Burgun A, Bodenreider O. Aggregating UMLS semantic types for reducing conceptural complexity. Medinfo. 2001;10(Pt 1):216-20.

[9] Chen Z, Perl Y, Halper M, Geller J, Gu H. Partitioning the UMLS semantic network. IEEE Trans Inf Technol Biomed. 2002;6(2):102-8.

[10] Zhang L, Perl Y, Halper M, Geller J, Hripcsak G. A lexical metaschema for the UMLS semantic network. Artif Intell Med. 2005;33(1):41-59.

[11] Gu HH, Perl Y, Elhanan G, Min H, Zhang L, Perg Y. Auditing concept categorizations in the UMLS. Artif Intell Med. 2004;31:29-44.

[12] Cimino JJ. Auditing the Unified Medical Language System with semantic methods. J Am Med Inform Assoc. 1998;5(1):41-51.

[13] Sibanda T, He T, Szolovits P, Uzuner O. Syntactically-informed semantic category recognizer for discharge summaries. Proc AMIA Symp. 2006;714-8.

[14] Pedersen T, Pakhomov SV, Patwardhan S, Chute CG. Measures of semantic similarity and relatedness in the biomedical domain. J Biomed Inform (e-pub ahead of print). Jun 10, 2006. DOI 10.1016/j.jbi.2006.06.004.

[15] Weeds J, Dowdall J, Schneider G, Keller B, Weir D. Using distributional similarity to organize biomedical terminology. Terminology. 2005;11(1):107-41.

[16] Fan JW, Friedman C. Semantic classification of biomedical concepts using distributional similarity. J Am Med Inform Assoc. (in press)

[17] Harris ZS. A theory of language and information: a mathematical approach. New York: Oxford University Press, 1991.

[18] Lee L. Measures of distributional similarity. Proc Annu Meet Assoc Comput Linguist. 1999;25-32.

[19] Aronson AR. Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program. Proc AMIA Symp. 2001:17-21.

[20] The MEDLINE/PubMed Baseline Repository (MBR): http://mbr.nlm.nih.gov/

[21] Green AM. Kappa statistics for multiple raters using categorical classifications. Proc 22nd Annu Conf SAS Users Group. 1997;1110-5.

[22] McCray AT, Bondenreider O, Malley JD, Browne AC. Evaluating UMLS strings for natural language processing. Proc AMIA Symp. 2001;448-52.

Address for correspondence

Jung-Wei Fan, M.S.,
email: fan@dbmi.columbia.edu

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

A Reappraisal of Sentence and Token Splitting for Life Sciences Documents

Katrin Tomanek, Joachim Wermter, Udo Hahn

Jena University Language and Information Engineering (JULIE) Lab, Friedrich-Schiller-Universität Jena, Germany

Abstract

Natural language processing of real-world documents requires several low-level tasks such as splitting a piece of text into its constituent sentences, and splitting each sentence into its constituent tokens to be performed by some preprocessor (prior to linguistic analysis). While this task is often considered as unsophisticated clerical work, in the life sciences domain it poses enormous problems due to complex naming conventions. In this paper, we first introduce an annotation framework for sentence and token splitting underlying a newly constructed sentence- and token-tagged biomedical text corpus. This corpus serves as a training environment and test bed for machine-learning based sentence and token splitters using Conditional Random Fields (CRFs). Our evaluation experiments reveal that CRFs with a rich feature set substantially increase sentence and token detection performance.

Keywords:

biomedical text mining, natural language processing, linguistics, text processing.

Introduction

Natural language processing systems expect their input to be properly delimited into consecutive sentences and sentences to be properly segmented into their basic tokens. Real-world documents lack such fine structure and, typically, some simple heuristic pre-processor performs these analysis steps. In the newspaper domain, manually defined patterns for sentence splitting rely on the fact that, e.g., a period is a sentence boundary, if it is followed by an uppercase letter and not part of a known abbreviation ("e.g.", "Dr.", etc.). Tokenization is often performed by splitting at a closed set of special characters (especially punctuation symbols, quotation marks, parentheses, brackets, etc.). While such rules may be adequate for the newspaper domain, they are underspecified for the biomedical domain, in which complex naming conventions lead, in particular, to quite unsatisfactory tokenization results when few and simple hand-written rules are applied. Here, both sentence and token boundary symbols are much more ambiguous because they often appear within entity names (organism, protein, cell names, etc.) and their abbreviations, as well as within (chemical) formulae, bibliographic references. If these challenges are not met, in an analysis pipeline for text mining each error occurring in the course of these preprocessing steps will unavoidably be propagated upwards in the pipeline. As a consequence, text mining modules, such as (named) entity recognition and relation detection, fed by erroneous segmentation results will inevitably suffer in terms of performance.

Rather than adding further complexity to manually maintained rule sets, supervised machine learning (ML) is becoming more and more the method of choice for many demanding natural language processing (NLP) tasks. From a given set of training examples a statistical model is learned automatically which is then used to assign labels to unseen data. Those approaches are to a large extent data-driven, i.e., by exchanging the training material they can be ported to other domains and languages, possibly without further changes. This constitutes a clear advantage over rule-based approaches which require a lot of manual tweaking and tuning, especially when more complex rules have to be supplied [1]. Furthermore, supervised ML approaches have been shown to outperform rule-based ones on several NLP tasks, including sentence boundary detection and tokenization, both in performance and breadth of coverage [2]. Therefore, we redesigned the entire pre-processing cycle and developed a new approach to sentence and token splitting based on Conditional Random Fields (CRFs), a sequential machine learning technique.

For sentence and token boundary detection, two annotated biomedical text corpora are available which may hold adequate training data for a supervised ML approach, *viz.* the GENIA corpus [3] and the PennBioIE corpus [4]. Whereas determining the sentence boundaries in a text corpus may be seen as a comparatively easy task, deciding on token boundaries is not as straightforward as it might appear at first sight. In particular, in biomedical language such as found in PubMed abstracts (or in full articles), crucial semantic units, such as entity names or even references to biological processes, can be contained within larger string units and hence are not simply delimited by white spaces. Thus, a closer look at both corpora reveals that only PennBioIE addresses the problem of semantically motivated word token boundary annotation at all, whereas GENIA annotates tokens around the same closed set of special characters used for English newspaper language [5].

In this paper, we report first on the compilation and annotation of the JULIE corpus which provides sentence and word token boundary information in a semantically moti-

vated and linguistically feasible way. The JULIE corpus is composed of documents from a large variety of biomedical subdomains and entity types with critical word token issues. We then report on an ML-based tool suite for sentence and word token boundary detection, the JULIE Tools. We evaluate both the corpus and the tools against GENIA and PennBioIE and against another ML-based tool suite, the OpenNLP Tools.[1]

Background

To train ML-based tools for sentence and/or token boundary detection, high-quality annotated text data resources are needed. In particular, several ambiguous non-alphanumeric character symbols may denote sentence or word token boundaries (or not) and, furthermore, these symbols may (or may not) be part of names for biomedical entities, such as protein, cell, or organism names etc. In the following, we outline the annotation guidelines considered for the compilation of the JULIE corpus which we annotated with sentence and token boundary information.

Sentence boundary annotation

For sentence boundary annotation, it is first necessary to determine potential sentence boundary symbols (*SBS*). For biomedical language texts, such as those from the PubMed literature database,[2] we defined the "classical" sentence boundary symbols (".", "!", "?", ":") and also two PubMed-specific ones (")", "]"). In particular, for periods (".") and colons (":"), we encountered many cases where they did not denote an SBS:

- General abbreviations ("e.g.", "i.e.", "et al.", "ref.", "*viz.*", "Dr.", "vs.", etc.);
- Numbers (0.05, 1.2, .4);
- Entity names, e.g., organism names ("E. coli", "F. oxysporum", "f. sp. Lycopersicim", "P. decumbens"). Classifying these symbols as SBS would break up organism names which are essential for disambiguating protein names and mapping them into their database entry (e.g., UniProt). Obviously, simply stating a rule that marks an SBS after every period if the following word starts with a capital letter would break up the organism entity "f. sp. Lycopersicim". Moreover, it would also fail to recognize an SBS in cases where the beginning of the following sentence starts with a lower-case letter, as is the case with many protein names (e.g. "p53", "tac", etc.);
- Author and journal names in literature citations which are contained in many Pubmed abstracts (e.g., "Am. J. Physiol.", "J. Biol. Chem.", "L. Hoffmann", "Schindler L.");
- Other alphanumeric strings, such as EC numbers, chromosome locations, database identifiers ("EC1.7.3.3", "LEN.PK113-78");
- Colons followed by enumerations ("Several cytokines interact with each other: IL-2, IL-5, and IL-18.").

Token boundary annotation

For the annotation of word tokens in biomedical text (or any other domain), it is essential to determine which word tokens denote semantic units (i.e., entities) of interest and thus should be recognized as such (e.g., by named entity recognizers). In biomedical text, there are various symbols which may (or may not) denote word token boundary symbols, such as "-", "+", "/", " ' ", "=", "%" "(", ")", etc., and thus are very ambiguous with respect to their status as token boundary symbols (*TBS*). The most important cases are:

- Parentheses must usually be split from regular words, but, e.g., in chemical terminology they are part of the name, such as in "Ca(2+)" or "(S,S)-Tartate". The same holds for enumeration list items such as "1)", "(2)", "a)", "b)", etc.;
- Plus ("+") symbols and hyphens may denote relevant semantic information, such as indicating the presentation (+) of an antigen on a cell or the absence (-) thereof: "CD34(+) T-cells", "CD83+ dendritic cells", "CD11c(++) B-lymphocytes", "CD8alpha(-) DCs". Here, it can be seen that biomedical language expresses complex biological processes by means of a single character symbol. In such cases, of course, these symbols should be tokens on their own;
- Hyphens often concatenate entity names (such as protein names) with other words ("IL-2-specific",[3] "CD28-dependent", etc.) or even with other entity names, such as cell names, as in "CD43-DC". Lacking recognition of these entities as word tokens would prevent entity recognizers from detecting them at all;
- Similar observations can be made with respect to slashes ("/") which often separate two (or more) entity references ("IL-2/CD34", "HA-1/2", etc.);
- Hyphens (and slashes) may also denote the knock-out status of a certain gene with respect to an organism, such as in "flt3L-/- mice".

One could argue that such problems might be overcome by simply splitting (i.e., marking a sentence or token boundary symbol) at every potential split symbol (i.e., at every parenthesis, hyphen, period, colon, etc.). This strategy would split a protein name such as "IL-2" into [IL], [–], and [2]. However, modules further up in a typical text mining pipeline (part-of-speech taggers, phrase chunkers, syntactic parsers, etc.) would not be able to perform adequately on such broken data because their linguistic representations (either rule-based or derived from training data) could not deal with such fragments.

Materials and methods

JULIE Corpus

Currently, only one biomedical text corpus, *viz.* PennBioIE [4], addresses both sentence boundary and tokenization annotation issues at all, although insuffi-

1 http://opennlp.sourceforge.net

2 http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?DB=pubmed

3 Here, only the second hyphen should be split because the first one is actually part of the protein name "IL-2".

ciently. Furthermore, it is limited to two highly specialized biomedical subdomains, *viz*. the CYP450 enzyme and oncology, focusing on few organism types (human and mouse). No manual token annotation was done on the second well-known annotated biomedical corpus, GENIA [3], whose domain scope is also rather limited (transcription factors in human blood cells). Hence, by using different sets of MeSH terms,[4] we assembled a PubMed text corpus both for sentence and for word token annotation which covered a more varied set of biological subdomains, including gene expression and regulation, stem cell biology/transplanta-tion and immunology. We refer to this as the *Subdomain Corpus*. Additionally, by using the MeSH thesaurus, we also assembled subcorpora with respect to biomedical entity types which are known to pose severe problems for sentence and token boundary detection, *viz*. organisms, chemicals, cell types and cell components. We refer to this as the *Entity Corpus*.

The Subdomain and Entity Corpora for sentence boundary information were automatically preprocessed with the OpenNLP Sentence Splitter trained on the PennBioIE corpus [6]. Then the annotations were manually inspected and, if necessary, corrected by a computational linguist (the second author of this paper) and a biologist. In addition, we have added the sentence annotations from GENIA and PennBioIE. Altogether, the **JULIE Sentence Corpus** contains 62,400 sentences. A similar procedure was performed for token annotation. Due to the more complex task, only the Subdomain Corpus was annotated with word token boundary information. Furthermore, we corrected the token annotation in the PennBioIE corpus when it did not conform to our guidelines. The **JULIE Token Corpus** contains both the annotated Subdomain Corpus and the corrected PennBioIE data, summing up to around 35,900 sentences. Both corpora (sentence and token) are here referred to as the **JULIE Corpus**.

ML-based JULIE tools

Our tools for sentence splitting and tokenization are based on supervised machine learning, i.e., from a given set of training examples a statistical model is learned. Such a model can be used to predict labels for unseen data. Here, we employ Conditional Random Fields (CRF), a *sequential* learning approach which assigns a label sequence to an observation sequence [7]. This approach fits well the inherent sequential structure of natural language text. As an implementation of CRFs, we employed the machine learning toolkit MALLET.[5] In the following, the sentence and the token splitter (referred to as **JULIE Tools**[6]) are explained in detail. As a general rule, we tried to optimize our tools not only in terms of accuracy, but also with respect to the kind of errors made, *viz*. false positive (FP) and false negative (FN) errors. As for the sentence splitter, we prefer FNs over FPs. As many consecutive NLP components work on the sentence-level the costs of erroneously splitting a sentence is higher than not splitting. For the tokenizer we favor FPs over FNs because in many consecutive processing steps tokens are considered as atomic units; and if these are too coarse information is lost.

JULIE Sentence Boundary Detector (JSBD)

The input text is broken down into a sequence of (observation) units by splitting at all white space positions. For each such unit our sentence splitter has to decide whether it is at the end of the sentence or not (binary classification). Thus, the following piece of text from a PubMed abstract would be split into the following units (with the actual SBS after "T-cells"):

> … [on] [IL-2-activated] [CD34(+)] [cytotoxic] [T-cells.] [p3hr-1,] [the] [Burkitt's] [lymphoma] [cell] [line,] [was]

Each unit is represented by the following features:

- the unit itself (lexical feature) and its size in characters;
- sentence boundary symbols (SBS):[7] whether unit ends with SBS, whether unit contains SBS;
- brief word class: capital letters are replaced by "A", lowercase letters by "a" etc. and then identical consecutive letters are collapsed (e.g. *IL2* → *AA0* → *A0)*;
- orthographical features: based on regular expressions (e.g. *HasDash, AllCaps, InitalCap, hasParenthesis*, ...);
- abbreviations: whether the unit is contained in the list of known abbreviations, whether the unit conforms to abbreviation classes (*"[A-Z].", "([A-Za-z].)+*", "*[a-z]+.*");
- local context: features of neighboring units in window [-1,1] copied.

JSBD also employs a rule-based post-processing routine to avoid that a sentence is split within opened parentheses or brackets. This is a useful extension, as scientific papers and abstracts often contain complex bibliographic references within parenthetical or bracket-like expressions. Such a reference should be considered as one sentence. However, within these parenthesized sentences there are often many SBSs, especially periods, which could also be considered as sentence boundaries. Preliminary experiments showed that this processing routine, though it does not improve the overall performance, shifts the FP/FN ratio favorably, i.e., false positives are avoided (clearly at the cost of some more false negatives).

JULIE Token Boundary Detector (JTBD)

For the tokenizer we have defined a set of critical token boundary symbols[8] (TBS) where it should check for possible token boundaries. It should be noted that periods, exclamation and question marks are not considered as TBS because these SBS symbols have been disambiguated by the sentence splitter before. This is in line with the paradigm of running NLP tools in a (sequential) pipeline, where each tool is designated to one NLP task and there is a (natural) order to the tools due to input/output dependen-

4 http://www.nlm.nih.gov/mesh/meshhome.html
5 http://mallet.cs.umass.edu/
6 The tools can be downloaded from http://www.julielab.de
7 Sentence boundary symbols are, e.g., period, question and exclamation mark, etc.
8 Token boundary symbols: { } , + – ( ) [ ] ; = / < > % &

cies. Thus, tokenization is typically performed after sentence splitting.

At all white space positions and at each TBS we split the sentence into single units, the TBS itself is treated as a separate unit. Our example piece of text from above would thus contain the following units (for each sentence fragment):

[on] [IL] [–] [2] [–] [activated] [CD34] [(] [+] [)] [cytotoxic] [T] [–] [cells] SBS [p3hr] [–] [,] [the] [Burkitt] [’] [s] [lymphoma] [cell] [line] [,] [was]

For each such unit, JTBD decides whether or not this unit is the end of a token. A token thus consists of a sequence of *n* units where the last unit was labeled as token end but the other *n-1* units not.

Furthermore, we assign each unit its so-called super-unit: therefore we also split the sentence into larger strings at each white space as we did for sentence splitting. Such a string is considered a super-unit for a unit if it covers this unit in the sentence. In our example sentence fragments, we would have the following super-units:

[on] [IL-2-activated] [CD34(+)] [cytotoxic] [T-cells] SBS [p3hr-1] [,] [the] [Burkitt’s] [lymphoma] [cell] [line,] [was]

The super-units are needed as context information when building the features. Each unit is then represented by the following features:

- the unit itself, the super-unit itself (lexical feature);
- whether the unit had white space to the right in original sentence;
- whether the unit is a TBS;
- features equivalent to the sentence splitter: size, brief word class, abbreviation class, local context, rich orthographical features of the unit;
- bracket information: whether the super-unit contains brackets (*hasOpeningBracketOnly, isInBrackets, hasClosingBracketOnly,...*);
- whether the super-unit is an enumeration (“(1)”), whether the super-unit has genitive (“enzyme's”), whether the super-unit has plural in brackets (“enzyme(s)”);
- other orthographical features of the super-unit with focus on hyphens, arrows, +/– symbols which are often contained in biomedical texts, and structure of chemical names

Experimental results and discussion

Sentence splitting and tokenization performance is typically evaluated in terms of the accuracy (A), i.e., the number of correct decisions divided by the total number of decisions being made. Here, the total number of decisions equals the number of units. In our evaluation, we focus on two question, *viz.* (1) how well are the different corpora suited for training and (2) whether the JULIE Tools perform better than another well-known ML-based NLP tool suite, the OpenNLP Tools, whose general applicability to the biomedical domain has already been shown [6]. Among other NLP tools, OpenNLP also provides a sentence splitter and a tokenizer, both based on conditional maximum entropy models [8], also known as logistic regression. Compared to the rich feature sets of the JULIE Tools, the OpenNLP counterparts have substantially fewer features (mainly lexical ones) coupled with a non-sequential learning algorithm.

Sentence splitter evaluation

To address the first question, we trained the JULIE sentence boundary detector on both the GENIA (18,529 sentences with approximately 486,000 word tokens) and the original PennBioIE (23,277 sentences with approximately 590,000 word tokens) corpora. The models learned from this training material were then evaluated against the JULIE Subdomain and Entity Corpora. We did not evaluate against the complete JULIE Sentence Corpus because it comprises both the GENIA and the PennBioIE corpus. The results are depicted in Table 1. There are only very small differences in the accuracy of the two models (accuracy about A=99.6). In addition, we also trained and cross-validated JSBD on the complete JULIE Sentence Corpus which yields an accuracy of A=99.8. Here, we encounter a small improvement over the last three experiments. Table 1 also shows the performance of OpenNLP's sentence splitter on the complete JULIE Sentence Corpus (10-fold cross-validation). In this setting, JSBD performed notably[9] better than OpenNLP's sentence splitter (A=99.8 vs. A=98.7).

Table 1 - Sentence splitting performance: evaluated on different tools and different corpora

tool	training material	evaluation material	accuracy
JSBD	GENIA	JULIE Subdomain + Entity Corpus	99.58 (FP=30%)
JSBD	PennBioIE	JULIE Subdomain + Entity Corpus	99.62 (FP=27%)
JSBD	10-fold cross validation on JULIE Sentence Corpus		99.8 (FP=30%)
OpenNLP	10-fold cross validation on JULIE Sentence Corpus		98.7 (FP=48%)

This can be explained by the rich feature set in our sentence splitter and the fact that we consider sentence splitting as a sequential learning problem. Also, our sentence splitter produces fewer false positives (FP) than OpenNLP's sentence splitter (FP=30% vs. 48%), which is more favorable for this task.

Tokenizer evaluation

We trained JTBD on GENIA and PennBioIE, and evaluated the models on the JULIE Token Corpus to determine how well these corpora are suited for training (see Table 2). As GENIA is only tokenized using newspaper language patterns, it is not so well suited for training; only an accu-

9 On such a level of accuracy, a 1-percentage-point difference is notable because this affects mostly critical cases (organism names etc.).

racy of about A=71.5% is reached. PENNBIOIE is more apt because its word token annotation is more semantically motivated. A 10-fold cross-validation of JTBD on the complete JULIE Token Corpus showed that the performance is thus improved by approximately 1 percentage point. Table 2 also indicates that in comparison with the 10-fold cross-validation of OpenNLP's tokenizer on the complete JULIE Token Corpus (A=95.0), both JTBD's machine learning algorithm (sequential learning) and its rich linguistic feature representation (super-units/units) are superior (A=96.7). Looking at the tokenization decisions, the OpenNLP tokenizer runs into particular problems with hyphens which either get split not at all or in a rather inconsistent way. Thus, an expression like "IL-2-activated" is sometimes tokenized as [IL-2][–activated], [IL][–2][–activated], or not at all. Similar errors occur with expressions such as "CD34(+)".

From the above experiments we conclude, that for sentence splitting the respective corpus used for training is not that critical because sentence annotations are not very controversial within different biomedical subdomains. Tokenization, however, is a much more complex task with respect to the relevant biomedical semantic units to be annotated. Still, performance of tokenization can be improved significantly by employing a tool with a linguistically adequate representation and a rich feature set. In both cases, the extension of an annotation corpus by a set of critical (and rare) subdomain and entity cases, as is done in the complete JULIE Corpus, boosts performance.

Table 2 - Tokenizer performance: evaluated on different tools and different corpora

tool	training material	evaluation material	accuracy
JTBD	GENIA	JULIE Token Corpus	71.5 (FP=3%)
JTBD	PENNBIOIE	JULIE Token Corpus	95.9 (FP=25%)
JTBD	10-fold crossvalidation on JULIE Token Corpus		96.7 (FP=45%)
Open NLP	10-fold crossvalidation on JULIE Token Corpus		95.0 (FP=47%)

Conclusion

We introduced a sentence splitter and a tokenizer based on sequential machine learning methods and rich feature sets as well as novel corpora for the biomedical domain to train these tools. These corpora are an extension and a correction of the PENNBIOIE corpus: an extension because we added training material so that this corpus covers a more complete section of the biomedical domain and is not specialized to a subdomain; a correction because we removed some inconsistencies, annotation errors, and extended the semantic motivation of word token annotation. Our evaluation experiments run on these corpora indicate that forboth the sentence splitting and the tokenization task, a substantial improvement in performance could be achieved. Compared to a maximum entropy approach with poor feature sets (OpenNLP), these results suggest a ML approach based on CRFs for such NLP preprocessing tasks. In the future, we will optimize our tools in terms of feature representation as well as extend the JULIE Token Corpus to include more critical entity types.

Acknowledgements

This research was funded by the Germany Ministry of Education and Research (BMBF) within the StemNet project (funding code: 01DS001A to 1C), and by the EU's 6th Framework Programm (the BOOTStrep project under grant FP6-28099).

References

[1] Ngai G, Yarowsky D. Rule Writing or Annotation: Cost-efficient Resource Usage for Base Noun Phrase Chunking. In: Proceedings of ACL-2000.

[2] Palmer DD, Hearst MA. Adaptive Multilingual Sentence Boundary Disambiguation. In: Computational Linguistics. Computational Linguistics 1997: 23(2): 241-67.

[3] Ohta T, Tateisi Y, Kim JD. The GENIA corpus: An annotated research abstract corpus in the molecular biology domain. In: Proceedings of the HLT 2002.

[4] Bies A, Kulick S, Mandel M. Parallel Entity and Treebank Annotation. In: Proceedings of ACL workshop on "Frontiers in Corpus Annotation II: Pie in the Sky", 2005.

[5] Marcus MP, Santorini B, Marcinkiewicz MA Building a large annotated corpus of English: the Penn Treebank. Computational Linguistics 1993: 19(2): 313-30.

[6] Buyko E, Wermter J, Poprat M, Hahn U. Automatically Adapting an NLP Core Engine to the Biology Domain. Proceedings of the BioLink & Bio-Ontologies SIG Meeting 2006, Brazil.

[7] Lafferty J, McCallum A, Pereira F. Conditional Random Fields: Probabilistic models for segmenting and labeling sequence data. In: Proceedings of the 18th International Conference on Machine Learning, 2001.

[8] Ratnaparkhi A. Maximum Entropy Models for Natural Language Ambiguity Resolution. Ph.D. thesis, University of Pennsylvania, 1998

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Corpus-based Error Detection in a Multilingual Medical Thesaurus

Roosewelt L. Andrade[a,b], Edson Pacheco[a,b], Pindaro S. Cancian[a,b], Percy Nohama[a,b], Stefan Schulz[b,c]

[a] *Paraná University of Technology (UTFPR), Curitiba, Brazil*
[b]*Pontifical Catholic University of Paraná (PUCPR), Curitiba, Brazil*
[c]*Department of Medical Informatics, University Hospital, Freiburg, Germany*

Abstract

Cross-language document retrieval systems require support by some kind of multilingual thesaurus for semantically indexing documents in different languages. The peculiarities of the medical sublanguage, together with the subjectivism of lexicographers' choices, complicates the thesaurus construction process. It furthermore requires a high degree of communication and interaction between the lexicographers involved. In order to detect errors, a systematic procedure is therefore necessary. We here describe a method which supports the maintenance of the multilingual medical subword repository of the MorphoSaurus system which assigns language-independent semantic identifiers to medical texts. Based on the assumption that the distribution of these semantic identifiers should be similar whenever comparing closely related texts in different languages, our approach identifies those semantic identifiers that vary most in distribution comparing language pairs. The revision of these identifiers and the lexical items related to them revealed multiple errors which were subsequently classified and fixed by the lexicographers. The overall quality improvement of the thesaurus was finally measured using the OHSUMED IR benchmark, resulting in a significant improvement of the retrieval quality for one of the languages tested.

Keywords:

controlled vocabulary, information storage and retrieval, quality control.

Introduction

The medical language presents several challenges to information engineering due to its specialized terminology and the large amount of texts found in literature databases, on the Web, and in medical record systems. Furthermore there exists a considerable mismatch between the language of medical science, the jargon used by health professionals, and the language used by laypersons such as patients and their relatives. Multilingualism is another important issue here, because even though the global tendency is to use English as the primary language of research, the local idioms are still used for patient-related everyday documentation and communication. Furthermore, medical language is extremely dynamic with novel terms, names, and acronyms being constantly created and English terminology increasingly permeating non-English medical documents.

In light of this scenario, document retrieval systems necessitate domain-specific thesauri [1]. We understand by multilingual thesaurus some organized repository of linguistic symbols that are mapped to language independent, concept-like descriptors tailored to the needs of a certain domain. Furthermore, a thesaurus usually provides additional semantic relations between these descriptors [2]. According to [3], the main rationale for the use of a thesaurus is the supply of a controlled reference vocabulary to represent the contents of documents and thus abstracting away from linguistic variation, for the sake of ameliorating and simplifying document retrieval.

In contrast to automatically generated word indices, such as the ones maintained by current Web search engines, thesauri are generally constructed manually, in a labor intensive process that requires the grouping of synonyms and translations into semantic classes, the addition of semantic relations between classes and the resolution of lexical ambiguities. It is therefore self-evident that such a scenario, which normally involves a group of domain experts, requires some measures for error detection and quality assessment.

The objective of this work is to demonstrate how a quality control mechanism can be implemented for the maintenance of the *MorphoSaurus* dictionary, a multilingual thesaurus tailored to support document retrieval in clinical medicine [4]. To this end, we exploit test samples of multilingual corpora exhibiting a high degree of content similarity, in order to discover weaknesses in the thesaurus content by comparing the distribution patterns of MorphoSaurus descriptors in these samples.

Materials and methods

Subwords and semantic indexing paradigm

The construction of a thesaurus starts with the identification of the most informative terms (i.e., words and phrases) for the domain under scrutiny [5]. The main difference between MorphoSaurus and virtually any other thesaurus is that the inclusion of lexical entries into the repository is

guided, as much as possible, by criteria of semantic atomicity. The main corollary of this decision has been the definition of a new kind of lexical item, the so-called *subword* entry. Subwords are defined as self-contained, semantically minimal units [6], since we assume that neither fully inflected nor automatically stemmed words constitute the appropriate granularity level for lexicalized content description. In the medical sublanguage we observe particularly complex words such as *"pseudo|hypo|para|thyroid|ism"*, *"append|ectomy"*, or *"tooth|ache"*. Such words are constituted, in our model, by two or more subwords (in here, e.g., *"pseudo-", "-ectomy", "tooth"*). In the MorphoSaurus system, subwords are listed as lexicon entries together with their attributes such as language and subword type (i.e., stem, prefix, suffix, invariant). Each entry is assigned to exactly one identifier representing a group of synonyms and translations. These concept-like semantic descriptors are named *MorphoSaurus identifiers* (*MIDs*).

The MorphoSaurus engine indexes texts by mapping every word to a subword sequence, which is then checked for morphological plausibility, using a finite-state automaton. Thus, invalid segmentations such as ones without stems or ones beginning with a suffix are rejected. Finally, each meaning-bearing subword is replaced by its corresponding semantic identifier (MID). This in turn constitutes the interlingual output representation of the system. For example, *#physioterap* is the MID of the meaning of the subwords *krankengymnast*^{GE}, *physiotherap*^{EN}, *fisioterap*^{PT} or *fysioterap*^{SW}, extracted from text words such as "krankengymnastisch", "physiotherapist", or "fisioterapia".

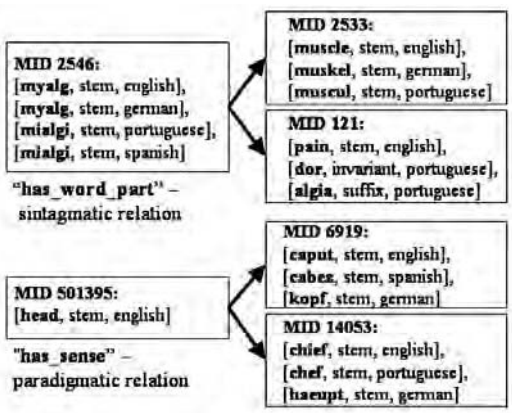


Figure 1 - Paradigmatic and syntagmatic relations: "has_sense" and "has_word_part" respectively

MorphoSaurus provides two different kinds of semantic relations between MIDs (cf. Figure 1):

- *has_sense*: This paradigmatic relation links ambiguous MIDs to their respective senses, e.g., the MID *#head* is linked to the MIDs *#caput* and *#boss* by *has_sense*.
- *has_word_part*: This syntagmatic relation links an MID having a composed meaning to its parts, e.g., the MID *#myalg={"myalg-", "mialg-",...}* to both *#muscle={myo-, muscle, mio, muscul, ...}* and *#pain={pain,-algy, -algi, -algia, dor, schmerz,... }*. The reason for this is the need to deal with composed meanings even in cases where a compound word cannot be properly dissected.

There are principally three approaches to treat ambiguous MIDs:

i. Substitute each ambiguous MID by the sequence of the MIDs representing its non-ambiguous senses.
ii. Substitute the ambiguous MIDs by exactly one MID which represents its most likely sense, computed from the MID frequency distribution.
iii. Substitute the ambiguous MIDs by exactly one MID representing its most likely sense, according to its context, using a corpus-derived co-occurrence matrix.

Whereas (iii.) has already experimentally tested [7], it has not yet been integrated into the MorphoSaurus system. For information retrieval purposes we use solution (i.), aware of the decrease in precision it may bring about.

The MorphoSaurus subword lexicon currently contains 90,550 entries, with 22,561 for English, 23,976 for German, 14,984 for Portuguese, 10,936 for Spanish, 7,812 for French, and 10,281 for Swedish. All of these entries are related in the thesaurus by 21,432 equivalence classes Fig. 2 depicts the process of semantic indexing, performed by the MorphoSaurus system.

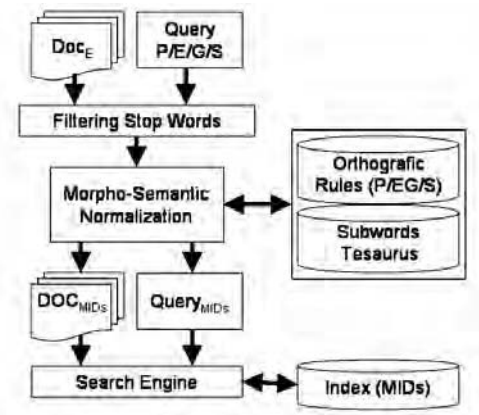


Figure 2 - Morpho-semantic normalization scheme

Pragmatics of thesaurus maintenance

An important advantage of constraining a dictionary to subword entries is the achievement of a high coverage with much less lexicon entries than in conventional lexicons. Nevertheless, the principal problems in the management of such a resource, as mentioned above, also exist with MorphoSaurus and require a principled solution. Although the lexicographers' work is based on written guidelines, many situations still require common-sense decisions which tend to produce more or less arbitrary results. Here, a particular difficulty of a subword thesaurus turns out to be that not only decisions about the semantic

relations between entries but also the proper delimitation of subwords have to be decided by the lexicographers.

Corpus-based error detection approach

We developed the following methodology to support the detection of weaknesses in the thesaurus. According to the current workflow, the lexicographers are already using a moderated mailing list in order to facilitate the communication of errors and to support consensus decisions of difficult modeling issues. However, we observed that this process is guided rather by serendipity than by systematic contemplations, although good responses can be found to any sort of posting. Our objective has been to improve this process by an automated detection of errors in the content of the thesaurus.

Our methodology is based on the hypothesis that in closely related corpora [8, 9] (i.e., texts that deal with the same subject-matter in different languages), the statistical distribution of semantic identifiers exhibits a high degree of similarity. In consequence, any exception to this expected conformity is likely to indicate some kind of fault in the indexing process, or a weakness in the semantic relations between entries in the lexicon.

General proposal

Our proposal aims at triggering the lexicographic activities with a ranked list of potential problems generated by the comparison of the "semantic extract" of comparable corpora, i.e., the MIDs extracted from them. In addition, we monitor this work via the repeated execution of a summative quality metric which uses an information retrieval benchmark that has already been applied in previous studies [10]. This benchmark measures the general appropriateness of a thesaurus to support medical text retrieval.

Multilingual related medical corpora

In order to create frequency distributions between the MIDs, generated from related multilingual corpora from different languages, we used the Merck Sharp & Dohme (MSD) manual of clinical medicine, a reference handbook of clinical medicine, available for the languages English (EN), Spanish (SP), Portuguese (PT), and German (GE), and freely available from the website http://www.merck.com. These corpora were submitted to the MorphoSaurus indexer and a frequency table was generated for each language.

Scoring of descriptors

For each MID in each language pair, an S value is calculated based on MID frequency distributions, using both relevance (S_a) and imbalance (S_d) measures:

$$S = \frac{2S_d - s_a}{3} \quad (1)$$

$$S_d = \left| \frac{f_1 - f_2}{f_1 + f_2} \right| \quad (2)$$

$$S_a = \frac{f_1 + f_2}{(f_{x1} + f_2)\max} \quad (3)$$

with f_1 being the MID frequencies in a normalized corpus of one language and f_2 the frequencies in another language; S_d scores the degree of imbalance between the MID occurrences between either language, and S_a relates the frequency of the MID under scrutiny with the frequency of the MID with the highest frequency in both corpora. Thus the S values range between zero and one. The overall score is therefore predominantly influenced by the degree of imbalance, but also gives an additional boost to highly frequent MIDs.

Redefinition of workflow

Guided by the sequence of problematic MIDs in the generated frequency lists, the lexicographers then started revising the thesaurus and the modifications were put down in a computer-based journal containing the following information in a semi-structured form: MID, problem description, problem class, solution, and rationale for modification.

Progress assessment

There are two different ways to assess the progress achieved in the thesaurus cleansing:

- Formative Evaluation: MID frequency lists are periodically generated, expecting a decrease of indices;
- Summative Evaluation: performance of a multilingual document retrieval system is measured, using the MorphoSaurus approach for indexing both documents and queries.

We subscribed to the second approach which is expected to yield stronger conclusions for the usefulness of the proposed error detection methodology.

According to [11], we use precision and recall as performance parameters in an IR system. Precision is defined as the proportion of relevant documents among all retrieved documents and recall as the rate of all relevant documents which could be retrieved. In IR systems which return all documents and produce a ranked output it is possible to measure precision at different recall points, thus yielding a precision / recall diagram. By interpolation it is then possible to compute precision values at defined recall points. As an overall assessment parameter, we computed the *eleven point average* value (*AvgP11*), defined as the arithmetic mean of the precision values at eleven recall points 0.0, 0.1, …, 0.9, 1.0.

For the precision / recall benchmark we used the OHSUMED collection, a subset of Medline abstracts, manually classified with regard to their relevance to a given set of authentic user queries [12]. In order to use this resource for benchmarking in a cross-language retrieval system, all queries had previously been translated to Portuguese, Spanish, German, and Swedish.

Table 1 - A sample of MID frequencies and related parameters. Top of the list for the language pair English / Portuguese

MID	Meaning	EqClass	f_1	F_2	S
#people	People	500783	6352	0	0.7155
#from	From	060077	4676	0	0.7026
#icas	icas	023555	0	3022	0.6899

During the correction period (three months), ten thesaurus backups were produced. Each of these backups was used for a complete IR experiment with the OHSUMED corpus. Each of these experiments in turn produced an AvgP11 benchmark value for each of the four languages.

To get a better visibility of the performance, we plot the average of Avg11P of all experiments in Fig. 2.

Results

Experiences with the correction process

During the process of problem analysis and correction it became clear that most of the highly scored MIDs spotted real problems which could be solved. Table 1 shows three rather extreme examples of imbalance between Portuguese and English, due to missing MIDs in one of the two languages. For example, the preposition *"from"* belongs to a MID which is marked for indexing, but its Portuguese analogue *"de"* is marked as a stop word and is therefore ignored for indexing.

Table 2 - Problems identified during the MID corrections

Reason for MID high score	Frequency Portuguese / English	Frequency German / English	Frequency Spanish / English
Ambiguities	0.23	0.38	0.14
Missing or dispensable MID	0.49	0.18	0.53
Same Sense in Different MIDs	0.06	0.12	0.19
One MID with Different Senses	0.04	0.05	0.06
No problem	0.11	0.10	0.04
Unclassified	0.07	0.17	0.04

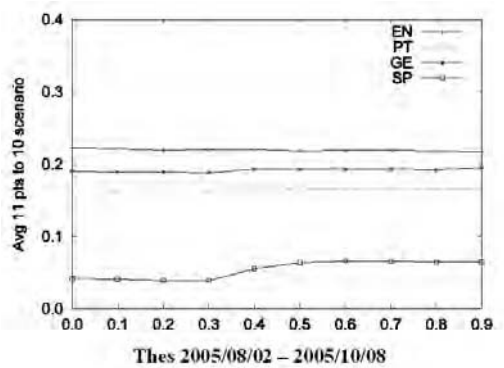
Table 2 depicts the most frequent problems:

- The ambiguity is mainly due to ambiguous lexemes (and, the according MIDs) in one language but not in another. In some cases the ambiguous MID was found not to be mapped to the unambiguous ones and was therefore used for indexing, since the normal procedure consisting of the substitution of an ambiguous MID by the MIDs that represent its non-ambiguous senses, did not take place. This is a problem that can easily be corrected by including the missing *has_sense* links.
- Missing or dispensable MIDs were common in borderline cases where a lexical entry had a very unspecific sense resulting in the fact that an entry was given a semantic identifier in one language but not in another one. An example is the preposition *"from"* (see discussion above and Table 1). The solution is to create a consensus about what should be considered as stop (sub)words, i.e., lexicon entries excluded from indexing.
- The same sense was found in different MIDs (which generally did not contain lexemes of all languages). This problem could be solved by merging different MIDs.
- Different senses were found in the same MID, and at least one of them was also present in another MID (generally with a focus on a different language). The solution consisted in splitting the non-uniform MID and redistributing its entries.

Surprisingly, other problems occurred only with quite low frequency. For instance, problems of string delimitation rarely had an impact on the MID distribution. Instead they seem to raise the overall noise level. So we conclude that other approaches may be better suited for detecting this kind of problem.

Summative evaluation measures

Fig. 2 demonstrates the evolution of the changes undertaken, using the IR benchmark described above. The AvgP11 values were calculated at 10 points within an evaluation period of nine weeks. During this time, about hundred hours were invested by experienced lexicographers who had already been working in the MorphoSaurus project before.



Thes 2005/08/02 – 2005/10/08

Figure 2 - Average of eleven point average value (AvgP11) evolution

For none of the languages under scrutiny there was a monotonous increase in performance. Comparing the first with the last AvgP11 value, there is a relatively insignificant growth of the values for Portuguese and German, namely by 1.8% and 2.6% respectively. This improvement seemed to be principally due to the addition of relations between MIDs and the rearrangement of MIDs. We even found an IR performance decrease of 1.9% in the case of English. We could argue that this value, just as the increase for German and Portuguese, lies within the range of normal variation, especially considering that the benchmark does not measure the whole information space but exactly the IR performance of a sample of 106 queries. Certainly, the more consolidated a resource, the less likely simple modifications will make it better. In contrast, the increment in performance of the Spanish benchmark amounted to a factor of 53% which cannot be attributed to chance. We interpret this finding as a good support for the hypothesis that the right problem selection – as done by our error detection approach – increases the performance in very short time.

The different degrees of maturity between the language-specific subsets of the thesaurus also became obvious when comparing the values from Table 2. The main difference is the relatively low rate of missing or unnecessary MIDs for German / English, a fact which may be derived from the maturity of the German part of the thesaurus on the one hand, but also from a more concordant treatment of stop words in this language pair.

Another interesting fact was that 10% of the MID disparities could not be attached to any thesaurus error. We interpret this as a consequence of lexical ambiguities occurring in one language but not in the other. So it frequently happens that one ambiguous MID (representing ambiguous terms in the same language) expands to one MID that is very common in the domain, and another one that is uncommon. As we disambiguate trough employing the expected frequency (cf. ii. in the second section) one or more readings will be ignored. For instance, if the English noun *"head"* corresponds a bit more likely to the sense of *"caput"* than to the one of *"boss"*, the latter sense will simply be ignored, which in turn may cause a disparity of the MID corresponding to the sense of *"boss"*, compared to other languages.

Conclusion

In this paper we presented an approach for data-driven error detection in the maintenance process of a multilingual thesaurus in the medical domain. The usefulness of the method could be demonstrated by the fact that most problems detected corresponded to real errors which could be fixed by the lexicographers. This showed the heuristic value of this approach in the process of continuous quality assurance and formative evaluation of the resource.

For the summative evaluation of our error-detection approach we used an IR benchmark during a nine-week period of thesaurus maintenance work. Whereas no significant increase in the benchmark parameter could be observed for those three languages which already exhibited a good quality after years of maintenance (i.e., German, English, Portuguese), there was a considerable increase for a fourth language, Spanish. This language, which had been added only much later, had therefore not received such a high level of attention due to the lack of Spanish language skills among the lexicographers.

Hence we can recommend the methodology especially for streamlining the lexicographers' efforts, particularly in the case of yet unconsolidated portions of a thesaurus. In addition to what we applied during this experiment we suggest the following elements to be included into a workflow for thesaurus quality assurance:

- For each language pair MID frequency, lists ordered by imbalance should be generated periodically, with a period covering one day to one week, according to the intensity of lexicon maintenance effort;
- An overall indicator for MID imbalance for each language pair should be generated and recorded over time;
- For each MID edited, the imbalance score S should be recorded over time;
- Alerts should be generated for every MID which exhibits an increase in imbalance above a given tolerance interval;
- IR benchmarks should be generated on a weekly basis and AvgP11 values monitored over time. Alerts could be generated in case of strong decrease in value, or cumulated decrease, considering both sequences of measurement for one language, or single-point measurement for all languages.

Acknowledgements

This work was supported by the CNPq, Brazil, (project number 550830/05-7), and by German-Brazilian cooperation project DLR/PUCPR (BRA 03/013). We are further indebted to Kornél Markó, Philipp Daumke, and Jan Paetzold for the development and maintenance of the MorphoSaurus engine.

References

[1] Schulz S, Hahn U. Morpheme-based, cross-lingual indexing for medical document retrieval. in: International Journal of Medical Informatics. 2000; 58-59: 87-99.

[2] Huge G. Natural Language Information Retrieval. Combining corpus linguistics and human memory models for automatic term association (1999). 75-98.

[3] Foskett, DJ. Thesaurus. Reading in Information Retrieval. Morgan Kaufmann, 1997; pp. 111-34.

[4] Markó K, Schulz S, Hahn U: MorphoSaurus - Design and evaluation of an interlingua-based, cross-language document retrieval engine for the medical domain. Methods of Information in Medicine, 4/2005(44): 537-45.

[5] Ribeira-Neto B, Baeza-Yates R. Information Retrieval - Data Structures & Algorithms, 1st. ed. New Jersey, EUA: Prentice Hall, 1992.

[6] Daumke P, Schulz S, Markó S. Subword approach for acquiring and cross-linking multilingual specialized lexicons. LREC 2006 Workshop: Acquiring and Representing Multilingual, Specialized Lexicons: the Case of Biomedicine. 2006; Genova, Italy.

[7] Markó K, Schulz S, Hahn U (2005); Unsupervised multilingual word sense disambiguation via an interlingua.

Proceedings of the 20th National Conference on Artificial Intelligence (AAAI '05): 1075-1080.
[8] Fung P. A statistical view of bilingual lexicon extraction: From parallel corpora to non-parallel corpora. In: Véronis J (Ed.). Parallel Text Processing; 2000
[9] Rapp R. Identifying word translations in nonparallel texts. Proceedings of the Meeting of the Association for Computational Linguistics; 1995; p. 320-322.
[10] Hahn U, Markó K, Poprat M, Schulz S, Wermter J, Nohama P. Interlingual indexing across different languages. In: Proceedings of RIAO'04, 7th International Conference "Recherche d'Information Assisté par Ordinateur"; (RIAO'04): Avignon, France. 2004 p. 100-115.
[11] Ribeira-Neto B, Baeza-Yates R. Modern Information Retrieval. 1st. ed. Boston: Addison-Wesley Longman Publishing Co.; 1999
[12] Hersh WR. Information Retrieval – A Health Care Perspective. New York: Springer; 1996.

Address for correspondence

Roosewelt Leite de Andrade,
LER, Pontifical Catholic University of Paraná.
Rua Imaculada Conceição,
1155 – CEP 80215-901,
Curitiba, Brazil
Email: roosewelt.andrade@gmail.com

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Defining Medical Words: Transposing Morphosemantic Analysis from French to English

Louise Deléger[a], Fiammetta Namer[b], Pierre Zweigenbaum[c]

[a] *INSERM, UMR _S 872, Éq. 20, Les Cordeliers, Paris, F-75006 France; Université Pierre et Marie Curie-Paris 6, UMR_ S 872, Paris, F-75006 France; Université Paris Descartes, UMR _S 872, Paris, F-75006 France*
[b] *ATILF and Université Nancy 2, CLSH, Nancy, F-54015 France*
[c] *LIMSI, CNRS UPR3251, Orsay, F-91403 France; INALCO, CRIM, Paris, F-75007 France*

Abstract

Medical language, as many technical languages, is rich with morphologically complex words, many of which take their roots in Greek and Latin—in which case they are called neoclassical compounds. Morphosemantic analysis can help generate definitions of such words. This paper reports work on the adaptation of a morphosemantic analyzer dedicated to French (DériF) to analyze English medical neoclassical compounds. It presents the principles of this transposition and its current performance. The analyzer was tested on a set of 1,299 compounds extracted from the WHO-ART terminology. 859 could be decomposed and defined, 675 of which successfully. An advantage of this process is that complex linguistic analyses designed for French could be successfully transferred to the analysis of English medical neoclassical compounds. Moreover, the resulting system can produce more complete analyses of English medical compounds than existing ones, including a hierarchical decomposition and semantic gloss of each word.

Keywords:

natural language processing, morphosemantic analysis, word definition, neoclassical compounds

Introduction

Medical language, as many technical languages, is rich with morphologically complex words, many of which take their roots in Greek and Latin. These so-called neoclassical compounds are present in many areas of the medical vocabulary, including anatomy (*gastrointestinal*), diseases (*encephalitis, cardiomyopathy*), and procedures (*gastrectomy*). Segmenting morphologically complex words into their components is the task of morphological analysis. When this analysis is complemented by semantic interpretation, the process is called morphosemantic analysis. Complex words are often "compositional," in the sense that the meaning of a complex word is often a combination of that of its parts.

Morphosemantic analysis can therefore help processes interested in semantics, such as the detection of similar terms or the generation of definitions. This was for instance the aim of [1], where Morfessor, a tool for unsupervised learning of morphological segmentation [2], was used to contribute mappings between WHO-ART and SNOMED terms. Early work on medical morphosemantic analysis focused on specific components such as *–itis* [3] or *–osis* [4], then on larger sets of neoclassical compounds [5]. Lovis [6] introduced the notion of morphosemantemes, *i.e.*, units that cannot be further decomposed without losing their original meanings. The Morphosaurus system [7,8] segments complex words using a similar notion called subword. The UMLS Specialist Lexicon [9], with its "Lexical tools," handles derived words, *i.e.*, complex words built through the addition of prefixes or suffixes. It provides tables of neoclassical roots, but no analyzer to automatically decompose compound words. DériF [10] morphosemantically analyses complex words. In contrast to [8] or [6], it computes a hierarchical decomposition of complex words. Moreover, it produces a semantic definition of these words, which it can link to other words through a set of semantic relations including synonymy and hyponymy. In contrast to the Specialist tools or to [11], DériF handles both derived and compound words. Designed initially for French complex words, then extended to the medical domain, its potential for cross-linguistic application was showed in [12]. Its transposition to English would fill a gap in the set of tools currently available to process complex English medical words.

This paper reports work on the adaptation of DériF to English medical complex words. It focuses on neoclassical compounds and is based on the hypothesis that this type of words is similarly formed in related European languages (here French and English). Our goal is to have DériF analyse English words and present its results in English.

We first describe the morphosemantic analyzer and our test set of words. We explain the modifications performed on this tool and the evaluation conducted. We then expose the results, discuss the method and conclude with some perspectives.

Selected for best paper award.

Material and methods

The principle on which this work is based is morphosemantic analysis, that is morphological analysis associated to a semantic interpretation of words. In other words, we want to obtain a decomposition and a description of the meaning of a complex word based on the meanings of its parts. A complex word may be formed through any combination of the following word formation rules:

- derivation, which adds an affix (prefix or suffix) to a base word, *e.g., pain/painful*;
- compounding, which joins two (or more) components together, those components being either neo-classical roots called Combining Forms (CFs) or modern-language words, *e.g., thermoregulation, arthritis*.

For this work we chose to analyze neo-classical compounds (formed from CFs). However, a compound may also undergo derivation, so that mixed-formation words must also be addressed. Therefore we included not only "pure" compounds, but also those neo-classical compounds that were prefixed or suffixed (*e.g., haemorrhagic*).

Our main hypothesis for the transposition from French to English is that a same linguistic analysis can be applied to neo-classical compounds of several languages. We assume that they are formed in a similar way and that the components involved are the same, the major differences being orthographic (such as *-algia/-algie*).

Material

We started from the French version of the DériF ("Derivation in French") morphosemantic analyzer. DériF was designed both for general language and more specialized vocabularies such as medical language. It performs an analysis purely based on linguistic methods and implements a number of decomposition rules and semantic interpretation templates. It also uses resources which include a lexicon of word lemmas tagged with their parts-of-speech and a table of CFs. The system goes further than simple decomposition and interpretation steps by predicting lexically related words. Another of its distinctive features is that it yields a structured decomposition of words and not simply a linear segmentation, so that we know which part is the head of the word.

As input the system expects a list of words tagged with their parts-of-speech and lemmatized (in their base form & no plural). It outputs the following elements:

1. a structured decomposition of the word into its meaningful parts;
2. a definition ("gloss") of the word in natural language;
3. a semantic category, inspired by the main MeSH tree descriptors (anatomy, organism, disease, etc.);
4. a set of potentially lexically related words. The relation identified can be an equivalence relation (*eql*), a hyponymy relation (*isa*) or a see-also relation (*see*).

For instance, the French word *acrodynie* (English *acrodynia*) is analyzed in the following way (*N* stands for *noun*, *N** is assigned to a noun CF):

acrodynie/N ==>
(1) [[acr N*] [odyn N*] ie N]
(2) douleur (de—lié(e) à) articulation *(pain of—linked to joint)*
(3) maladie *(disease)*
(4) eql:acr/algie, eql:apex/algie, see:acr/ite, see:apex/ite

It has been pointed out [12] that the method could be extended to other languages. Indeed the rules for generating lexically related words (item 4.) do not rely on any language-specific features so that this part of the system is fully language-independent. The morphosemantic parser however needs to be adapted to the language.

To test the transposition of DériF to English, we prepared a list of test words. These words were taken from the WHO-ART terminology, which describes Adverse Drug Reactions (ADRs), since one of the intended applications of this work is to contribute to the pharmacovigilance domain by grouping terms describing similar ADRs. We selected the English terms of this terminology; since DériF works on single words and not on multi-word units, we split them into single words; and since we adapted DériF to analyze neo-classical compounds, we only retained those types of words. The selection was done both automatically by removing all words of 4 characters or less (these words are practically never morphologically complex), and manually by reviewing the list to look for neoclassical compounds (the work was done by a language engineer, LD). This gave us a list of 1,299 words to be decomposed out of a total of 3,476 words. These words were lemmatized and tagged with their parts-of-speech using the Treetagger[1] part-of-speech tagger. We used a lexicon of tagged words from the UMLS Specialist lexicon[2] to help TreeTagger deal with unknown words.

Methods

Adapting the Morphosemantic Analyzer

The method of morphological analysis that we want to transpose to English is schematized in figure 1.

1 http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger/DecisionTreeTagger.html (last access 26/03/07)
2 http://www.nlm.nih.gov/pubs/factsheets/umlslex.html (last access 26/03/07)

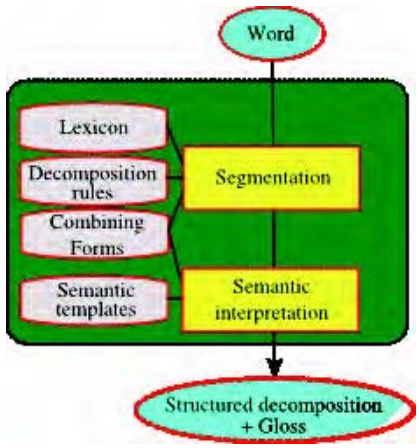


Figure 1 - Morphosemantic parsing

Language-specific material is located in:

1. *the lexicon* of tagged word lemmas which is used by the system to test whether a component exists and to retrieve its part-of-speech;
2. *the table of Combining Forms(CFs)*: each CF is associated with a modern-language word which describes its meaning, a semantic type, a part-of-speech, and a set of CFs related through relations *eql, isa, see*;
3. *the decomposition rules*: these rules are triggered in a certain order according to the part-of-speech of the word and to the affix identified (if any). They identify the head of the word and its other components, relying on the table of CFs and lexicon. Each rule may have a set of exceptions. Orthographic normalization is also performed on the components;
4. *the semantic templates*: they provide template glosses of a complex word based on the relation between its head and its other components.

We address each of these in turn.

(1) We replaced the French lexicon by an English lexicon derived from the UMLS Specialist lexicon.

(2) Our hypothesis is that CFs are mainly the same in English and French save for a few minor orthographic differences. We handled these by making small modifications to the French CFs (*e.g.*, removing accents as in *blépharo – blepharo*) to obtain the English equivalents. Semantic types assigned to the CFs might have been left with French labels since they are conceptual labels, but we considered that English names would be more consistent and translated them; this was done very easily since they were few. In the same line, we provided modern-English names for the CFs. They were obtained from two lists of CFs, one taken from the UMLS Specialist lexicon, the other one extracted from the Dorlands medical dictionary, where CFs were paired with modern-English words.[3] We automatically matched them to our English CFs and reviewed the results. Those roots that could not be matched were dealt with manually. The set of related CFs was also replaced by their English equivalents (obtained from the orthographic modifications mentioned above). The parts-of-speech were kept as is. An extract of the resulting table can be seen in table 1. The lexical relations between the CFs are labeled as follows: <- for an hyponymy relation, ~ for a see-also relation and no sign for an equivalence relation.

(3) Intervention at the rule level remained limited since we assumed that neo-classical compounds were formed in a similar way in French and English. So this mainly involved adapting the exceptions, the orthographic normalizations performed and the affixes (*e.g.,* French suffix *–ique* becomes *–ic*).

(4) Finally, we translated the semantic templates so that they could generate English glosses. For instance, the following template is now associated with the *–ia* suffix, the *hyper-* prefix and Y/X as nouns:
Affection of **X** *linked to the excess of* **Y**
where **X** and **Y** will be substituted with modern-language equivalents of CFs or with simplex words from the lexicon, as in *hypercalcemia*, which is analyzed into *Affection of blood linked to the excess of calcium*.

Table 1 - Table of Combining Forms (excerpt)

CF	English	Type	Relations
algia	pain	disease	odyn, algo, ~itis
blephar	eyelid	anatomy	palpebr, <- ocul, ~coro
ectomy	surgical excision	act	~tomy, ~stomy
gastr	stomach	anatomy	stomac, gaster, ~hepat, ~entero, <-abdomin, ~pancreat

Evaluation

We ran the updated DériF on the 1,299 complex words extracted from WHO-ART. The expected output for each word is a hierarchical decomposition into the word components, a definition in natural language, a semantic type and a set of related words, which we evaluated for coverage and validity. We define coverage as the proportion of words the system was able to analyze. An analysis was considered valid if its decomposition and definition were correct; in this evaluation, we did not take into account the semantic type nor the set of potentially related words. At this point we computed two standard measures of evaluation in natural language processing: precision and recall. Precision is the ratio of correct analyses over the total number of analyses. Recall is the ratio of correct analyses

3 http://www.merckmedicus.com/pp/us/hcp/thcp_dorlands_content.jsp?pg=/ppdocs/us/common/dorlands/dorland/dmd-a-b-000.htm (last access 26/03/07)

over the total number of analyses that should have been produced. The validity of the answers was determined by a language engineer when reviewing manually the output.

Results

In the present state of its transposition, DériF was able to analyze 859 out of the 1,299 words in our list (see table 2), thus obtaining 66% coverage. Example word analyses are given in table 3. They illustrate that both pure neoclassical compounds (*[[arthr N*] [algia N*] N]*) and derived words built on a classical base (*[a+ [dactyl N*] +y N]*) are analyzed.

Table 2 - Coverage

Total nb of words	Decomposed words	Coverage
1,299	859	66%

We identified several causes of non-decomposition:

- Certain suffixation rules are not currently implemented in the system (in both French and English versions). This is the case for *–ation* and *–ism* suffixes so that words like *lacrimation* and *hermaphroditism* are not decomposed;
- Some components are listed neither in the table of CFs nor in the lexicon; e.g. as *campt-* is not in the table of CFs, the word *camptodactyly* has not been decomposed;
- Errors at the preprocessing level (mistagged words), *e.g., corporal/N* was tagged as a noun while being an adjective in our context.

Table 4 - Precision/recall figures

Nb of correct results	Precision	Recall
675	78.5%	52%

We measured a precision of 78.5% (see table 4) which is fairly good. Combined to a moderate coverage, this yields a 52% recall. Incorrect results were mainly due to:

- Wrong structuring of the decomposition. An example can be seen in table 3 with the word *meningoencephalitis*. Its correct decomposition should be:
  [**[** [mening N*] [encephal N*] **]** [itis N*] N]
  -itis should be the head of the conjunction of *mening-* and *encephal-,* which would give a definition such as "*inflammation related to head and meninges.*"
- Unsatisfactory definition (often due to the fact that the meaning of the word was not sufficiently reflected by the meaning of its parts). See for instance the analysis of the word *acanthosis* in table 3. The decomposition is right but its meaning has evolved too much and cannot be derived from that of its parts. This word should not be considered a complex one.
- Mistagged words that could not be correctly analyzed. This is the case of *alveolar* (last row of table 3), which was treated as a noun derived from an adjective (*alveolar* as an adjective is correctly analyzed by DériF).

Discussion

The precision of the adapted system is good, especially for a first implementation. Recall is lower but should rapidly grow when increasing the size of the table of CFs and the lexicon.

The present state of the system already shows that this language-dependent system could be adapted to another, related language for the task of analyzing medical compounds. The advantage of doing so is that we did not have to start from scratch and implement a new system. Indeed, a certain amount of manual work is still necessary, such as preparing the table of CFs and adapting the semantic templates; but there is a stable basis on which to work, so that we believe that this solution is overall less time-consuming. An additional advantage is that in the process, we can

Table 3 - Example word analyses generated by the updated DériF

Word/POS	Decomposition	Definition	Type	Related words
Correct analyses				
arthralgia/N	[[arthr N*] [algia N*] N]	pain (of -- linked to) joint	disease	eql:arthr/algesia see:arthr/itis
adactyly/N	[a+ [dactyl N*] +y N]	Affection characterized by the absence of digit	disease	_
gastroesophageal /ADJ	[[gastr N*] [oesophag N*] al ADJ]	Related to oesophagus and stomach	anatomy	eql:stomac/oesophag isa:abdomin/oesophag
Errors				
meningoencephalitis/N	[[mening N*] **[** [encephal N*] [itis N*] N**]** N]	(Part of -- Specific type of) encephalitis related to meninges	_	_
acanthosis/N	[[acanth N*] [osis N*] N]	(Part of -- Specific type of) disease related to prickle	disease	_
alveolar/N	[[alveolar A] N]	Entity being alveolar	_	_

transfer to English linguistic analyses which were initially designed for French. For instance, Namer [13] proposed an analysis of pathology nouns of the form Pref-(Y)X-ie, for instance *hypercalciurie*, which she showed could also apply to German, Spanish, Italian, and English. In the present work, the implementation of this analysis in the French DériF is directly transferred to its English version.

Using a linguistics-based morphosemantic analyzer such as DériF has a number of advantages. The system performs both morphological decomposition and semantic interpretation while other methods remain at the level of morphological segmentation [8] or add semantics after using a tool for decomposition [1]. We also provide a hierarchical decomposition as opposed to a linear one [1,6,8]. In [1] the method was also applied to the WHO-ART terminology, so that we could compare results. The same word list was submitted to the Morfessor segmenter in exactly the same conditions as in [1], and obtained a coverage of 93.7%, but a precision of 53.2% (25% less than DériF) and a recall of 49.9% (slightly less than DériF). This means that about the same number of words was correctly analyzed (recall), but DériF was much more to the point by proposing significantly less incorrect analyses. By relying on a statistical segmenter, the method in [1] has the advantage of being language-independent; however, its implementation also uses a table of morphemes, so transposition to another language is not immediate either. Besides, statistical segmentation also brings certain types of errors that could be avoided with linguistic rules as implemented in DériF, *e.g.*, *chemosis* segmented as *c+hem+osis*. Moreover, as pointed out above, DériF outputs more complete information than Morfessor (simple linear segmentation).

This work also suggests the perspective of a system that could work with several languages. We transposed the system from French to English, but a possible next step would be to be able to use it for translation: i.e. producing an English definition of a French word (or vice-versa). This would require a multilingual table of CFs (as suggested in [12] and prepared here for French+English), multilingual semantic templates (as obtained from the present work for these two languages), and translations of simplex words involved in compounds. Such a system could contribute for instance to cross-language information retrieval, with the same principles as [14]. Another potential next step would be to test the adaptation method on other related languages.

Conclusion

In this work, we successfully transposed a linguistics-based morphosemantic analyzer from French to English in order to provide definitions for English neoclassical medical words. This verifies our hypothesis that neoclassical compounds from different languages can be analyzed in a similar way, as can be expected in the medical vocabularies of languages in the Romance (*e.g.*, French) and Germanic (*e.g.*, English) families. This can be seen as a first step towards a multilingual system.

Future work includes improvement of the system as well as using the results for a specific application such as pharmacovigilance. The decompositions and definitions generated can be used to measure proximity between WHO-ART terms and group similar terms, but without necessarily relying on SNOMED as done in [1]. Grouping similar terms describing Adverse Drug Reactions would allow better signal detection in pharmacovigilance.

References

[1] Iavindrasana J, Bousquet C and Jaulent MC. Knowledge acquisition for computation of semantic distance between WHO-ART terms. Stud Health Technol Inform. 2006; 124:839-44.

[2] Creutz M, Lagus K, Linden K, and Virpioja S. Unsupervised morpheme segmentation and morphology induction from text corpora using Morfessor 1.0. Technical report, Publications in Computer and Information Science, Helsinky University of Technology, 2005.

[3] Pacak MG, Norton LM, and Dunham GS. Morphosemantic analysis of -ITIS forms in medical language. Methods of Information in Medicine 1980;19:99-105.

[4] Dujols P, Aubas P, Baylon C, and Grémy F. Morphosemantic analysis and translation of medical compound terms. Methods of Information in Medicine 1991;30:30-5.

[5] Wolff S. The use of morphosemantic regularities in the medical vocabulary for automatic lexical coding. Methods of Information in Medicine 1984;4(23):195-203.

[6] Lovis C, Michel PA, Baud R, and Scherrer JR. Word segmentation processing: a way to exponentially extend medical dictionaries. Methods of Information in Medicine 1995:28-32.

[7] Schulz S, Romacker M, Franz P, et al. Towards a multilingual morpheme thesaurus for medical free-text retrieval. In: Proceedings of MIE 99, Ljubliana, Slovenia. IOS Press, 1999:891-894.

[8] Markó K, Schulz S, and Hahn U. Morphosaurus - design and evaluation of an interlingua-based, cross-language document retrieval engine for the medical domain. Methods of Information in Medicine 2005;(44):537-45.

[9] McCray AT, Srinivasan S and Brown AC. Lexical methods for managing variation in biomedical terminologies. Proc Annu Symp Comput Appl Med Care. 1994:235-9.

[10] Namer F and Zweigenbaum P. Acquiring meaning for French medical terminology: contribution of morphosemantics. In Marius Fieschi, Enrico Coiera, and Jack Li, eds. Proc 10th World Congress on Medical Informatics, San Francisco, Ca, 2004;11(Pt 1):535-9.

[11] Baud R, Rassinoux AM, Ruch P, and Lovis C. The power and limits of a rule-based morphosemantic parser. In: Proceedings of AMIA 1999 Annual Symposium, 1999: 22-6.

[12] Namer F and Baud R. Predicting lexical relations between biomedical terms: towards a multilingual morphosemantics-based system. Stud Health Technol Inform. 2005;116:793-8.

[13] Namer F. Guessing the meaning of neoclassical compounds within LG: the case of pathology nouns. In: Proceedings of Generative Approaches to the Lexicon, Geneva, 2005:175-84.

[14] Hahn U, Honeck M, Piotrowski M, and Schulz S. Subword segmentation: Leveling out morphological variations for medical document retrieval. 2001:229-33.

Address for correspondence

Louise Deléger, email:- louise.deleger@spim.jussieu.fr

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Finding Malignant Findings from Radiological Reports using Medical Attributes and Syntactic Information

Takeshi Imai [a], Eiji Aramaki [a], Masayuki Kajino [b], Kengo Miyo [a], Yuzo Onogi [c], Kazuhiko Ohe [a]

[a] *The University of Tokyo Hospital, Japan* [b] *Japan Research Group for Medical Ontology, Japan*
[c] *Clinical Bioinformatics Research Unit, Graduate School of Medicine, The University of Tokyo, Japan*

Abstract

Radiology reports are written primarily in natural language. Automated extraction of malignant findings from narrative reports is an important technique for clinical support or alert generation for physicians. This paper proposes a method for automatically extracting malignant findings from narrative radiological reports written in Japanese. First, sentences are parsed and a medical attribute of each phrase is determined. Next, sub-trees related to radiological findings are extracted from a dependency tree using medical attributes. Finally, the malignant findings in each sub tree are extracted with their positive or negative assertions, each of which is determined by the multiplication of pos/neg signs along a path in a sub-tree. The recall and precision for the extraction of malignant findings with their positive or negative assertions were 76% and 91% respectively. The experimental results showed the validity of the proposed method for extracting malignant findings with correct assertions.

Keywords:

information extraction, radiology report, radiological findings, syntactic analysis, natural language processing

Introduction

Radiological reports contain a great deal of information about a patient's medical condition, and malignant findings such as "*carcinoma*" are especially important. As such, for a physician who orders an examination, automated extraction of such malignant findings is an important technique for clinical support or for alert generation from the viewpoint of medical safety control.

However, simple methods such as extracting only malignant keywords from each sentence do not work well. For example, if a malignant keyword "*cancer*" is found in a sentence in a patient's record (e.g., post operation), it may not be relevant to a current malignant finding and should not be extracted. Furthermore, if the phrase "*suspect for metastasis*" is found in a sentence, "*metastasis*" should **not** be extracted as "positive study" for cases in which the entire sentence is: "*There was* ***no nodular lesion suspect for metastasis."***

Note that an approach that attaches greater importance to recall than to precision and that depends on later screening burdens the users and is impractical. Thus, an effective alert system requires high precision.

Several studies have examined information extraction (IE) from narrative reports [1][2][3][4][5][6]. However, few studies have examined IE from narrative reports written in Japanese [7][8]. The previous approach for extracting findings from Japanese radiological reports [7] is based on pattern matching using assertions at the end of sentences. However, this approach does not consider dependency relationships between non-contiguous terms so that plural findings cannot be extracted from one sentence with their correct assertions. As a result, the previous approach achieved very high precision but the recall was no more than approximately 50%.

In the present paper, we propose an improved method for extracting malignant findings from narrative reports written in Japanese, using both medical attributes and syntactic information (in other words, "dependency structure"). The Japanese dependency structure is usually defined in terms of the relationship between phrasal units called *bunsetsu* segments (hereafter "phrase"), each of which consists of a content word and a function word. In addition, each content word or function word consists of morphemes and the process of recognizing plural morphemes as one phrase is called "chunking".

The dependency structure is necessary not only for extracting plural findings from one sentence, but also for determining positive or negative attributes of long and complex assertions (e.g., double negation). However, as is often mentioned in the previous studies [4], there are many difficulties in parsing radiology reports because of ungrammatical writing. The existing Japanese syntactic parser based on newspapers does not work well. Most errors occur in excessive "chunking" caused by shorthand and telegraphic writing style. Therefore, in the proposed method, the system performs the correction of excessive chunking of parser output. The system then determines the positive or negative attribute of each malignant finding using [+]/[-] signs in a sub-tree of the entire dependency tree. An outline of the proposed system is shown in Figure 1.

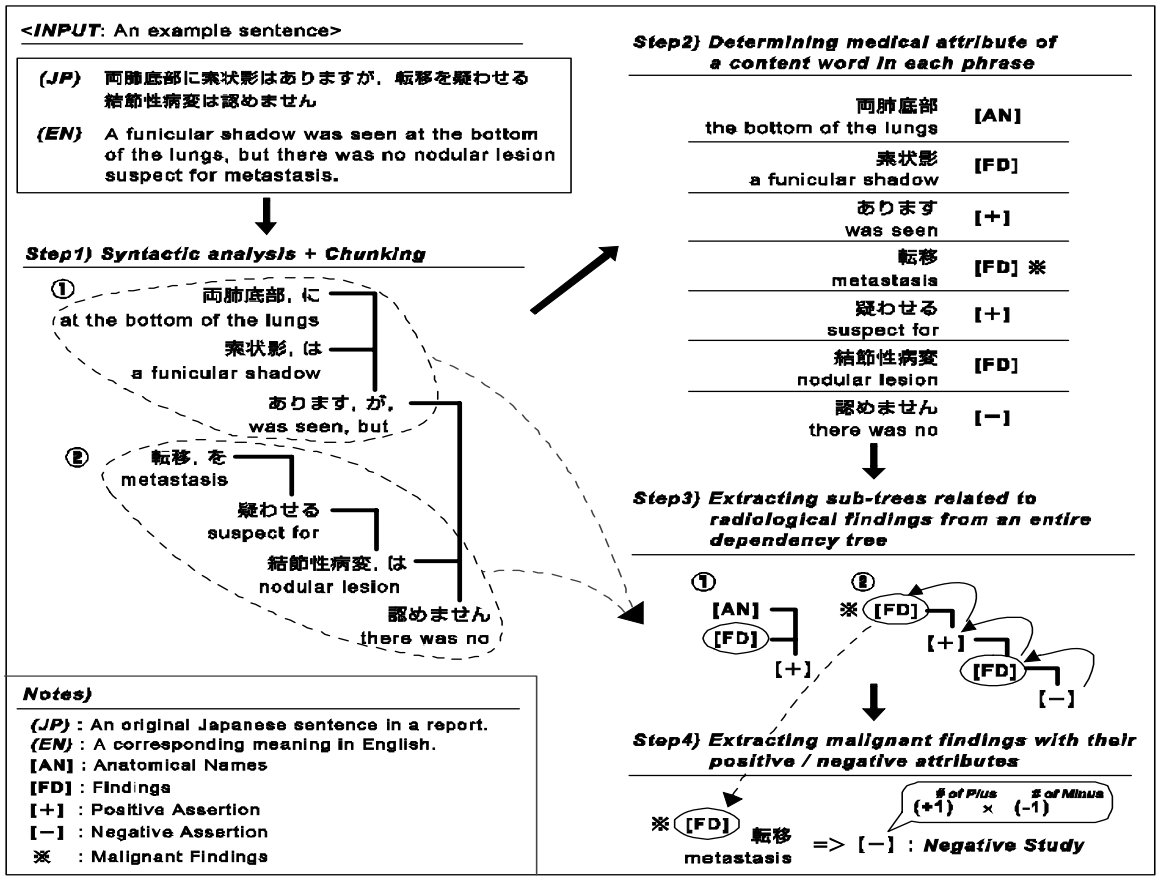


Figure 1 - Overview of the proposed method

Materials and methods

Materials

We used 22,496 CT and MRI reports stored in the Department of Radiology of the University of Tokyo Hospital from February 1, 2005 through August 31, 2005.

Radiology reports usually consist of various sections such as “Exam”, “Comparative Study”, “Patient Status”, “Recommendation”, “Findings”, and “Additional Information”. First, the sentence classification process was performed using a rule-based classifier described in a previous paper [7] (Recall = 99.1%, Precision = 98.8%). Then, sentences that include “keywords expressing malignant findings (Table1)” were randomly selected from the “Findings” category. We prepared 10,000 sentences for use in developing the rules of the proposed system and another 100 sentences for the evaluation set.

These typical malignant keywords (Table1) were selected by medical specialists for trial, and their future extension is discussed later herein.

Medical attributes

In order to determine a medical attribute of a content word in each phrase (in Step 2), a medical dictionary was constructed using four resources, as shown in Table 2.

We unified these four resources, while omitting some duplicate records, and then defined their medical attributes as shown in Table 3. Terms that have “Finding” attributes are (1) entries in ICD-10, where the first character of the ICD-10 code is ***‘R’*** (i.e., “Finding” category) in resource (D1) and (2) findings defined in resource (D4).

Table 1 - Keywords expressing malignant findings

[Straightforward words expressing malignant findings]
malignant, malignancy, cancer, carcinoma, recurrence, recurrent, melanoma, metastasis, glioma, lymphoma, leukemia, sarcoma, blastoma, cytoma, (and their corresponding words in Japanese)
[Abbreviations of malignant findings]
HCC, AML, CML

Table 2 – Unified medical dictionaries to determine medical attributes

Dic.	Resource	# of words
(D1)	'Standard Disease Names in Japan corresponding to ICD-10 for electronic medical record'*	34,634
(D2)	'Thesaurus for Medical and Health related Terms version 5'†	172,782
(D3)	Japanese Medical Terms in 'UMLS2005AA'‡	54,274
(D4)	Terms for radiology region collected from past radiological reports	2,773

* http://www.dis.h.u-tokyo.ac.jp/byomei/
† http://www.jamas.gr.jp/thesaurus.htm
‡ http://www.nlm.gov/research/umls/

Table 3 – Medical attributes

Attributes	# of Subtypes	# of Words	Examples
Anatomical Name	2	9,043	'liver', 'lung',..
Disease	1	75,051	'osteosarcoma',
Finding	3	4,994	'carcinoma',..
Examination	3	19,426	'CT', 'MRI',..
Modifier	8	777	'right','upper',.
*General	5	106,140	'protein',..

[Notes]:

Each attribute may have subtypes.
(ex) Modifier – 'for Temporal Coordinates'
Modifier – 'for Anatomical Names'

*"General" attributes were not used for selecting subtrees related to radiological findings.

Methods

Step 1) Syntactic analysis and chunking

First, syntactic analysis and chunking are performed. We used the general Japanese parser KNP [9] and added only 137 ***Sahen-nouns*** and 55 ***Na-adjectives*** into the KNP dictionary, both of which function not only as nouns or adjectives but also as predicates, and are especially important in avoiding parsing errors in Japanese syntactic analysis. KNP performs a chunking process using case particles, so that compound nouns or phrases that are unique to medical terminology can be detected as one chunk with high accuracy, even if these terms are not included in the dictionary of the parser.

Step 2) Determining the medical attribute of the content words in each phrase

Next, a content word in each phrase is detected, and morphological analysis is performed for each content word, using Japanese Tokenizer, JUMAN [10] in which the unified medical dictionary mentioned earlier is added to the default dictionary. As in the previous morphology-based approach like [11], we used the attributes of the morphemes in a content word in order to determine the medical attribute of the content word. This is basically performed by the Head-Final rule (i.e., the last morpheme decides the attribute of the content word), but the attributes of the previous morphemes or the sub-types of the attributes are used when the last is a continuation of a modifier (see Figure 2).

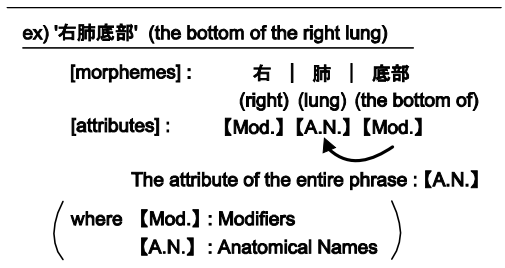


Figure 2 – Determining medical attributes

In addition, the system also detects positive or negative assertions about the existence of findings. We manually composed a list of 258 words (Table 4) with declined and conjugated endings, which should be recognized as positive or negative assertions, respectively, based on the frequency in the rule set.

Table 4 – Positive forms of negative assertions

Attributes	# of Terms	Examples (in English)
[+] (pos.)	165	'is seen', 'is found',…
[-] (neg.)	93	'is not seen', 'is notfound',…

The system then corrects excessive chunking caused by a lack of case particles. If the positive or negative assertions in the above list are at the end of the phrase (such assertion should be separated as another phrase), then the phrase with excessive chunking is separated into two phrases.

Step 3) Extracting sub-trees relating to radiological findings from the entire dependency tree

Next, searching from the end of the dependency tree, all subtrees that begin with the [+] or [-] assertion are recursively extracted (see Figure 1) by the following rules:

1. Each sub-tree must consist of only [+], [-], [Disease], [Finding], [Anatomical Name], and [Examination].
2. If a dependency tree is terminated with [Disease] or [Finding] (i.e., a shortage of an assertion), then [+] node is added as the top node of the sub-tree.
3. An edge between two [+/-] signs (direct dependency between two assertions) is separated.

For example, in Figure 1, the system extracts two sub-trees from the input sentence.

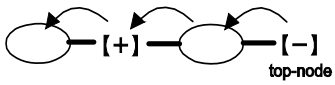
Step 4) Extracting malignant findings using their positive/negative attributes

In the last step, for each phrase that includes a malignant keyword in a sub-tree, the system determines whether it should be extracted. We consider only [Disease] or [Finding] phrases that include malignant keywords as candidates, so that malignant keywords in [Examination]

or [Patient Status] phrases are excluded. The positive or negative attribute of the existence of each malignant finding is then determined by multiplication of positive or negative signs along the path from the top node to each node (phrase) in the sub-tree, as defined below.

$$\text{Pos/Neg Sign} = (+1)^{\#\text{ of Plus}} \times (-1)^{\#\text{ of Minus}} \quad (1)$$

# of Plus or **# of Minus** is counted along the path from the top-node to each node (phrase) in the subtree (see Figure1).



For example, in Figure 1, "metastasis" should be extracted with "Negative attribute" because there are [-] and [+] in the path from the top-node to "metastasis" in the sub-tree.

Experiments

Experimental set-up

In order to evaluate the performance of the proposed method, we used the evaluation set (100 sentences) mentioned earlier. All phrases that include a malignant keyword were manually annotated with 'Positive study (+)', 'Negative study (-)', or 'Not an object to be extracted (N)' (gold standard). For example, if a phrase including a malignant keyword is not related to current malignant findings (e.g., 'past patient status' or 'examination'), then the phrase is annotated with '(N)'.

Our intent is to extract malignant findings that have (+) or (-) attributes. Therefore, we label such findings ***"(OBJ)".***

***"(OBJ*) = A phrase of [Disease Name] or [Findings]**

that includes a malignant keyword"

The system output was judged as to whether ***(OBJ)s*** were properly extracted using their positive or negative attributes. The evaluation set includes 97 pairs of ***(OBJ)*** and ***(Positive/Negative Sign)*** as the gold standard. In the evaluation, we used (BASE1) and (BASE2) as baselines and compared them with the proposed method (PROP).

Table 5 – Comparison of methods

(BASE1)	Extracting ***(OBJ)*** as ***Positive***
(BASE2)	***(OBJ)* + *Pos/Neg sign*** of the phrase upon which the ***OBJ*** is directly dependent
(PROP)	***(OBJ)* + *Pos/Neg sign*** determined by multiplication of [+][-] signs along the path in a sub-tree (defined in Equation (1))

Results

The results are shown in Table 6. The recall and precision of (PROP) are much better than the baselines. The precision of the proposed method is greater than 90%, whereas the recall rate is not so high. Some pairs could not be extracted because the system cannot detect the beginning of a sub-tree owing to the insufficiency of rules to detect certain assertions, such as ***"doutei dekinai*** (cannot identify)". On the other hand, badly extracted pairs were caused by errors in determining the **[+] / [-]** sign. For example, from sentences such as ***"Syuyou ha / ohkisa ni / tyohen / arimasen.*** (The tumor did not change in size.)", the system extracted "***syuyou*** (A tumor)" with the **[-]** sign, which was caused by the word ***"arimsen*** **[-]** (not)" (i.e., ***"syuyou"*** directly depends on ***"arimasen"***). In this case, the system recognized the negation as being in regard to the existence of a tumor, whereas, in fact, the negation was regarding a "change in size".

Table 6 - Results

Methods	Extracted Pairs	Correctly Extracted	Recall	Precision
(BAS1)	97	47	48.5%	48.5%
(BAS2)	74	58	60.0%	78.3%
(PRO)	81	74	76.2%	91.4%

Discussion

Both (PROP) and (BASE2) use dependency information. However, there is significant improvement in both recall and precision. Improvement in recall shows that the proposed method can extract a finding that does not directly depend on assertion, while an improvement in precision shows that the proposed method works well for a finding that changes its pos/neg attribute through plural assertions, such as "metastasis" in Figure 1.

In order to achieve a higher recall rate, further development of the rules to detect various patterns of assertions is necessary. In particular, it is important to deal with assertions that consist of more than one phrase, such as ***"~ wo / hiteisuru koto ha / konnan de aru*** (It can not be denied that ~)". In addition, we should construct a system to process English sentences (which occasionally appear in Japanese radiological reports) so as to deal with both Japanese and English sentences. Although the present study focuses on Japanese sentences, the proposed method for determining pos/neg signs along a path in a dependency tree is essentially language independent.

There is no tagged corpus of Japanese radiological reports, so that we cannot perform machine learning methods at the present time. We are currently constructing a large tagged corpus to carry out an evaluation on a larger scale and to make comparison with machine-learning methods in the near future. The proposed method will also be effective for supporting the construction of such a tagged corpus.

In the present study, we used approximately 30 keywords expressing malignant findings for the trial. Those keywords are not sufficient, to be sure. However, in the future, determining the malignancy of each [Disease Name] or [Findings] should be performed using hierarchical information of controlled vocabulary, such as ICD-10 or SNOMED-CT, that is, malignancy can be determined using attributes of hypernyms or top-node in the hierarchy. To extend malignant keywords by constructing such a module is an important future task, but the essentials of the proposed method will not be affected by the changes.

Related work

Many studies, such as MedLEE [1][3] and so on [2][4][5][6], are being conducted to extract findings from narrative reports, however there have been few studies concerning the extraction of information from Japanese radiological reports [7][8]. Above all, previous studies about Japanese narrative reports did not use syntactic information efficiently because of ungrammatical writing of narrative reports, which differs from 'standard' sentences such as newspapers.

The proposed method is novel because it (1) enables the efficient use of syntactical information by correcting excessive chunking caused by lack of case particles and (2) calculates the sign (positive or negative) of each malignant finding, which is determined based on medical attributes, by focusing on a path in a sub-tree.

Conclusion

In the present study, we proposed a new method for extracting malignant findings using their positive or negative attributes, through the multiplication of [+][-] signs in sub-trees related to radiological findings. Experimental results showed the validity of the proposed method for correctly determining the positive or negative attribute of each malignant finding.

Acknowledgments

The authors would like to thank Dr. Tomohiko Masumoto in the University of Tokyo Hospital for providing the corpus of radiological reports.

References

[1] Friedman C, Alderson P, Austin J, Cimino J, Johnson S. A General Natural Language Text Processor for Clinical Radiology. J Am Med Inform Assoc. 1994; 1(2): 161-174.

[2] Haug PJ, Ranum DL, Frederick PR. Computerized Extraction of Coded Findings from Free-text Radiologic Report. Radiology. 1990; 174: 543-48.

[3] Hripcsak G, Kuperman GJ, Friedman C. Extracting Findings from Narrative Reports: Software Transferability and Sources of Physician Disagreement. Methods Inf. Med. 1998; Jan 37(1): 1-7.

[4] Taira RK, Soderland SG, Jakobovits RM. Automatic Structuring of Radiology Free-text Reports. Radiographics. 2001; 21(1): 237-245.

[5] Hahn U, Romacker M, Schulz S. MEDSYNDIKATE – a Natural Language System for the Extraction of Medical Information from Findings Reports. Int. J. Med. Inf. 2002; 67:63-74.

[6] Sibanda T, He T, Szolovits P. Syntactically-Informed Semantic Category Recognizer for Discharge Summaries. Proc. AMIA 2006; 714-718.

[7] Imai T, Onogi Y. Extracting Findings from Japanese Radiological Reports based on Sentence Pattern Classification using Case Frame (in Japanese). JCMI 2004; 24(Suppl.): 800-1

[8] Imai T, Onogi Y. Extracting Numerical Measurement and Temporal Coordinates from Japanese Radiological Reports. Proc. SPIE 2004; 5371:268-276.

[9] Kurohashi S, Nagao M. KN Parser: Japanese Dependency / Case Structure Analyzer. In Proceedings of the International Workshop on Sharable Natural Language Resources. 1994; 48-55.

[10] Kurohashi S, Nakamura T, Matsumoto Y, Nagao M. Improvements of Japanese Morphological Analyzer JUMAN. In Proceedings of The International Workshop on Sharable Natural Language Resources. 1994; 22-28.

[11] Marko K, Schulz S, Hahn U. MorphoSaurus -Design and Evaluation of an Interlingua-based, Cross-language Document Retrieval Engine for the Medical Domain. Methods Inf. Med. 2005; 44(4): 537-545.

Address for correspondence

Takeshi Imai, The University of Tokyo Hospital;
7-3-1 Hongo, Bunkyo, Tokyo 113-8655, Japan
+81-3-5802-7115; ken@hcc.h.u-tokyo.ac.jp

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

A Normalized Lexical Lookup Approach to Identifying UMLS Concepts in Free Text

Vijayaraghavan Bashyam[a], Guy Divita[b], David B Bennett[a], Allen C Browne[b], Ricky K Taira[a]

[a] *Medical Imaging Informatics Group, University of California, Los Angeles, United States*
[b] *National Library of Medicine, National Institutes of Health, United States*

Abstract

The National Library of Medicine has developed a tool to identify medical concepts from the Unified Medical Language System in free text. This tool – MetaMap (and its java version MMTx) has been used extensively for biomedical text mining applications. We have developed a module for MetaMap which has a high performance in terms of processing speed. We evaluated our module independently against MetaMap for the task of identifying UMLS concepts in free text clinical radiology reports. A set of 1000 sentences from neuro-radiology reports were collected and processed using our technique and the MMTx Program. An evaluation showed that our technique was able to identify 91% of the concepts found by MMTx in 14% of the time taken by MMTx. An error analysis showed that the missing concepts were largely those which were not direct lexical matches but inferential matches of multiple concepts. Our method also identified multi-phrase concepts which MMTx failed to identify. We suggest that this module be implemented as an option in MMTx for real-time text mining applications where single concepts found in the UMLS need to be identified.

Keywords:

MetaMap, natural language processing, text mining, UMLS

Introduction

The recent past has seen a phenomenal growth in the amount of textual information generated in various areas of biomedicine such as clinical reports associated with electronic medical records and scholarly publications [1],[2]. There have been repeated calls for the development of text mining and natural language processing tools to extract information from the biomedical literature [3]. To cater to such needs, the National Library of Medicine (NLM) has developed several open-source text processing tools for the biomedical domain. The *MetaMap Program* is a tool developed at the NLM to map free text to concepts in the Unified Medical Language System Metathesaurus (UMLS) [4]. *MetaMap* and its java version *MMTx*, henceforth referred synonymously as MM, have been extensively used for a wide variety of biomedical text mining applications. We have developed a technique for fast and accurate mining of whole UMLS concepts. This is one of the applications for which MM has been used extensively, but for which MM is considered to be slow [5]. We report the details of this system and suggest that it be implemented as an optional module in future releases of the MM toolkit.

Background

In order to get an understanding of the various scenarios in which MM is being applied, we conducted a literature review to identify all reports of MM usage in the past three years. We identified 8, 13 and 9 articles within the proceedings of MEDINFO 2004, AMIA 2005 and AMIA 2006 respectively, all of which directly used MM for their respective tasks. We ignored the articles which made secondary use of MM via other NLM applications like *SemRep*. We also did not analyze journal articles since conference publications often get extended to journal articles.

Recent uses of MetaMap

Recently, MM has been used for several different applications both within and outside the NLM. Whalen et al. used MM to identify UMLS concepts for use in medical textbook summarization [5]. Mary et al. report the usage of MM for identifying genes and proteins in scientific articles related to molecular biology [6]. Chapman et al. used MM to index emergency department reports by UMLS terms related to respiratory findings [7]. Hsieh et al. tried to identify UMLS terms used by patients in email messages using MM [8]. Kim et al. indexed descriptive clinical text from a web-based repository of dermatology cases to MeSH terms using MM[9]. Bernhardt et al. identified diagnoses and causes of death and mapped them to MeSH terms using MM [10]. Lacson et al. identified nouns and replaced then with the UMLS semantic labels through MM [11]. Wedgwood used MM to map medical terms to the UMLS as a module within a medical question answering system [12]. Gay et al. indexed the title and abstracts of biomedical articles using MM [13]. Niu et al. identified UMLS semantic categories for analyzing polarity information in medical text [14]. Kahn et al. annotated radiology images with patient demographics and UMLS concepts generated using MM for a digital searchable library [15]. Mirhaji et al. used MM to identify single UMLS concepts found in chief complaint data [16]. Silfen et al. processed biomedical publications using MM to identify UMLS con-

cepts [17]. Peace et al. used MM to index concepts in nursing procedure manuals [18]. Meystre et al. customized the target dataset of MM adapted to 80 targeted medical problems and used it to identify a predefined list of problems [19].

Baud et al. used MM to analyze the variability of Medical terms [20]. Leroy et al. tried to disambiguate word senses using UMLS concepts identified by MM and their semantic types[21]. Hagerty et al. processed clinical guidelines using MM [22]. Tringali et al. used MM to process gastrointestinal endoscopy reports [23]. Meng et al. identified anatomy-related concepts from surgical reports with MM [24]. Coonan coded clinical narratives to UMLS concepts using MM [25]. Yetisgen-Yildiz et al. used MM as a phrase chunking tool [26]. Zhou et al. identified UMLS concepts in documents related to five health domains using MM [27]. Crowley et al. used MM as a coding module of an intelligent tutoring system [28]. Lieberman et al. used MM to extract concepts from chief complains entered into an ambulatory medical electronic medical record [29]. Wang et al. report the use of MM as an identification tool for medical terms which are later manually labeled for an information retrieval system [30]. Ogren et al. tried to annotate corpora using UMLS semantic types using MM [31]. Patel et al. attempted to mine cross terminology links in the UMLS and used MM to process their dataset [32]. Hagerty et al. used MM to identify UMLS concepts and associated concepts for encoding clinical text in a markup language [33]. Ruiz used MM to first identify English concepts and then translate them to French [34].

Need for a new module

Our examination indicated that 15 [5-19] of the 30 mentioned reports used MM primarily to identify whole concepts in the UMLS. This task does not require some of the sophisticated processing modules present in MM which are computationally complex and hence time consuming. By making a reasonable assumption that this set of articles accurately reflects the distribution of MM usage in the medical informatics community, we safely conclude that around 50% of MM applications only seek whole UMLS concepts and not multiple concept mappings. A method bypassing the time-consuming modules of MM would be very suitable for these applications. Our experiment was a step toward this goal.

Materials and methods

Our methodology consisted of four stages – 1) Identifying the data to be processed our method and by MM, 2) Processing the data by our method, 3) Processing the data with MMTx and, 4) A comparative evaluation of the two methods. The details of the individual stages follow:

Data Collection

Collection of the sentences to be indexed to the UMLS was the first stage in our experiment. We followed the steps outlined below to collect sentences.

Step 1: A set of one thousand radiology reports related to the neuro-radiology domain were collected from the existing hospital database at our institution.

Step 2: Section boundary detection was performed on the reports to break up a report into individual sections such as HEADER, HISTORY, FINDINGS, CONCLUSIONS etc. Next, sentence boundary detection was performed on the sections. Both of these modules have been previously tested, with recall and precision accuracies of over 99% within the domain of radiology [35].

Step 3: A lexical analyzer processed each sentence performing tokenization, part-of-speech tagging, semantic class tagging and word-sense disambiguation. Our lexicon categorizes tokens into twenty syntactic categories and over eight hundred semantic categories.

Step 4: We identified a list of the primary semantic classes which occur in radiology reports, such as anatomy, findings, procedures, equipment etc. Using a high-recall sentence-level parser, all target sentences were conservatively identified while sentences that obviously have no instances of our target semantic types were rejected. To illustrate, the presence of the semantic type `phys-obj.anatomy` (e.g. brain) in a sentence would automatically result in the sentence being accepted as containing a possible UMLS concept. This step is necessary so that we do not try to find UMLS concepts in sentences which might not have any. Thus, we chose to process the sentences that are highly likely to contain UMLS concepts.

Step 5: From the target set of sentences, we randomly sampled 1000 sentences for processing by our technique and by MM.

Identifying UMLS concepts

We previously prepared a normalized word index of the entire UMLS concepts tagged as English concepts and serialized the index in a database. This index is similar to the index used by MM, GSpell and other such tools developed at the NLM. Each sentence was preprocessed using the 'LuiNorm' stemming tool developed at the NLM to generate a sentence normalized at the word level [36]. We identified a list of and word types words types such as conjunctions which, when normalized are likely to give false positive errors. These words were intentionally not normalized.

Each normalized sentence was processed by our module to identify all UMLS concepts by referring to the index. The method used to identify the concepts is the longest substring matching algorithm developed by Aho and Corasick. Details of this algorithm can be found in the original publication [37]. The algorithm finds a list of longest substrings according to the parse order (left to right in this case) given a lengthy string. Thus, this method will only identify a list of lexically matching concepts but not attempt to find any partial matches, combinatorial matches, etc. The output of this module for an example sentence is shown below.

Sentence:
The liver is enlarged.

UMLS Substrings:
1. C0023884 (Liver)
2. C0442800 (Enlarged)

Using our method, all the 1000 sentences were processed and the UMLS concepts were stored in a list.

MMTx processing

We used the MM API to process the same corpus. The MM methods were used to parse the sentences into phrases, process these phrases to obtain the best final mappings, and then arrange the final mappings into a single output set. In cases where MM found multiple equally-scoring mappings, the procedure merged the common mappings, and then arranged the dissimilar candidates as a pipe-delimited list. For example, the phrase "right thalamus" returns the two "tied" final mappings shown below:

First Candidate:
1. C0205090 (Right)
2. C0039729 (Thalamic structure)

Second Candidate:
1. C0205090 (Right)
2. C1269896 (Entire thalamus)

The result was rearranged and output as two concepts where the second concept was one of two possible concepts:

1. C0205090 (Right)
2. C0039729 | C1269896 (Thalamic structure OR Entire thalamus)

Evaluation

The list of mappings produced by MM was compared to the mapping list generated by our method. A match on either token for the 'thalamus' was considered a correct hit. In addition to gathering mapping data, the times taken to process the sentences and obtain final mappings by both methods were logged. The average time taken by MM to process each sentence was 1.03 seconds whereas our module had an average processing time of 0.14 seconds.

Results

The total number of concepts identified by MM was 6859. The total number of concepts identified by our method was 6513. Of the 6513 concepts, 173 concepts were large phrases (which MM had broken down into two or more concepts). 457 terms were erroneously mapped to wrong concepts in the UMLS which account for false positives. Considering the large phrases as accounting for two concepts each and subtracting the false positives, we arrive at 6229 valid UMLS concepts identified by our method. Since our current objective is to provide a faster, equivalent system to MM, we assume that all the concepts identified by MM are correct mappings. Thus we arrive at an accuracy score of 91%. However it is important to know that MM itself has a lot of false positive issues which have been reported by several people [24]. Table 1 summarizes the results.

Discussion

Error analysis

False positives occur for to several reasons. First, the UMLS Metathesaurus contains several concepts which are orthographically identical but have different concept unique identifiers (CUIs). When such concepts are encountered, our system adopts an adhoc method of choosing one of the multiple mappings. An example of such a false positive would be the verb am mapped to the concept am (ante-meridiem).

Other types of false positives occur due to the natural ambiguity introduced when words are stemmed. Multiple words when normalized to the same stem increases the ambiguity. Replacing the stemmed word index by a regular word index can result in an increase in precision but a possible drop in recall. To illustrate, plural forms of words would still map to the same concept when normalized but would not when indexed directly. These types of errors could be potentially reduced by using a regular word index in addition to a normalized word index.

Table 1 – Results of Evaluation

	MMTX	**Normalized Lexical Lookup**
No. of Concepts Identified	6859	6513
Large Phrases		+173
False Positives		-457
Total	6859	6229

Multiple phrasal concepts

Since we do not use a shallow parser to chunk sentences into phrases, we are able to identify concepts spanning across multiple phrases. An example of this is the phrase `liver` and spleen which exists as a concept in the UMLS (C0545797) but is not recognized by MM as a single concept returning the individual concepts liver (C0023884) and spleen (C0037993). Our method is able to return the whole concept liver and spleen. While it can be argued that failure to separate the coordinated concepts is an undesirable feature, we would like to stress that our system is also not intended to correct errors present in the UMLS. A retrieval system is only as good as the content it is retrieving. Though our method is intended to be as close to MM output, it is independent of, and can be used with any lexicon other than the UMLS. The system only needs a list of concepts with corresponding unique identifiers and can perform efficient lookup. Thus, as concepts are removed from or added to each new release of the UMLS, we don't have to make changes to our system.

Methodology differences

Our technique differs from the MM approach in several ways:

1. MM tokenizes a sentence into terms and phrases whereas our method only tokenizes sentences into words.
2. To identify phrases, MM requires the services of a part-of-speech (POS) tagger to first tag the words or terms, and a shallow parser to use the tags to mark phrase boundaries and phrase head assignments. Our system bypasses the POS tagging and shallow parsing component.
3. To identify terms, MM performs a lexical lookup within the sentence to find the longest spanning terms from the SPECIALIST Lexicon. Our method does a longest spanning match to a normalized word index of the UMLS, bypassing the term tokenization and lexical lookup into the SPECIALIST Lexicon. One of the benefits to this approach is the potential of retrieving multi-phrase concepts as demonstrated previously.
4. MM has several modules to find a) matches to the terms within phrases, b) generate spelling and inflectional variants, synonyms, derivations, acronyms, abbreviations and expansions of the words in the phrase, and a recursive combination of these, to make the match. Our system does not have any of these modules.
5. MM categorizes the retrieved matches as partial matches, over matches, and matches with gaps to the phrase. This categorization requires an alignment mapping technique that is at the heart of MM. This is a technique that is computationally expensive. By default, MM filters out over-matches and matches made with concept gaps, to reduce the noise caused by such matches. Our system does not have this alignment mapping technique.
6. MM includes a ranking mechanism to evaluate the quality of the matches. It ranks 'exact' and 'near exact' matches higher, and 'partial matches' with matches made on the basis of some form of lexical variation, lower. Our method does not include a ranking of the matches returned but returns a fixed list of concepts identified in a sentence.

Our method does not return matches made with gaps, but could return partial matches. Because of these differences, it is likely that our method will do a better job than MM at quickly finding exact and normalized exact matches within a sentence. However, since our method lacks the variant generation and ranking capabilities, it is likely to miss matches based on non-inflectional lexical variation or because of word order differences. Our method will do a better job retrieving multi-phrase concepts. It is an open question whether phrase spanning concepts occur more often in a target corpus than concepts that could only be matched via non-inflectional variation, and whether the added precision is worth the extra computational costs. These issues may be highly dependent on specific domains.

Scope for improvement and future work

Previous internal studies at the NLM have shown that over 95% of the time taken by MM in processing a sentence is due to the alignment mapping algorithm. Since we avoid the alignment mapping module, the shallow parser and the part-of-speech tagger, theoretically our method should perform much faster than its current performance. We took no particular effort in optimizing our implementation of the Aho-Corasick algorithm. Our aim was to show that for specific applications, it is possible to achieve results similar to that of MM using simpler modules and by avoiding the computationally complex modules. Thus, an efficient implementation has a possibility to be a lot faster than our current system. Additionally, there are methods other than the Aho-Corasick algorithm which could be used for finding the largest substrings. The performance of those methods for this particular application needs to be explored. Finally, a more extensive evaluation of MM versus our method needs to be conducted to arrive at conventional precision, recall and f-score measures.

Conclusion

A module for MetaMap is proposed with focus on high processing speed suitable for real-time applications. This module uses components already being used in MetaMap with minor changes. This module is likely to have multiple useful applications in medical informatics.

Acknowledgments

This research was supported in part by an appointment of the first author to the NLM Research Participation Program.

References

[1] Berner ES, Detmer DE, Simborg D. Will the Wave Finally Break? A Brief View of the Adoption of Electronic Medical Records in the United States. J Am Med Inform Assoc. 2005;12(1).

[2] Andrade MA, Bork P. Automated extraction of information in molecular biology. FEBS Lett. 2000;476(1-2):12-7.

[3] Ananiadou S, Friedman C, Tsujii J. Introduction: named entity recognition in biomedicine. Journal of Biomedical Informatics. 2004;37(6):393-5.

[4] Aronson AR. Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program. *AMIA Annu Symp Proc*. Portland, OR 2001:17-21.

[5] Whalen G. Medical Textbook Summarization and Guided Navigation using Statistical Sentence Extraction. *AMIA Annu Symp Proc*. Washington, DC 2005:8.

[6] Mary V, Marquet G, Le Beux P. MeSH and Specialized Terminologies: Coverage in the Field of Molecular Biology. *Proc of MEDINFO 2004*. San Francisco 2004:530-4.

[7] Chapman WW, Fiszman M, Dowling JN, Chapman BE, Rindflesch TC. Identifying respiratory findings in emergency department reports for biosurveillance using MetaMap. *Proc of MEDINFO 2004*. San Francisco 2004:487-91.

[8] Hsieh Y, Hardardottir GA, Brennan PF. Linguistic analysis: terms and phrases used by patients in e-mail messages to nurses. *Proc of MEDINFO 2004*. San Francisco 2004:511–5.

[9] Kim GR, Aronson AR, Mork JG, Cohen BA, Lehmann CU. Application of a Medical Text Indexer to an Online

Dermatology Atlas. *Proc of MEDINFO 2004*. San Francisco 2004:287-91.
[10] Bernhardt PJ, Humphrey SM, Rindflesch TC. Determining prominent subdomains in medicine. *AMIA Annu Symp Proc*. Washington, DC 2005:50.
[11] Lacson R, Barzilay R. Automatic Processing of Spoken Dialogue in the Home Hemodialysis Domain. *AMIA Annu Symp Proc*. Washington, DC 2005:420-4.
[12] Wedgwood J. MQAF: A Medical Question-Answering Framework. *AMIA Annu Symp Proc*. Washington, DC 2005:1150.
[13] Gay CW, Kayaalp M, Aronson AR. Semi-Automatic Indexing of Full Text Biomedical Articles. *AMIA Annu Symp Proc*. Washington, DC 2005:5.
[14] Niu Y, Zhu X, Li J, Hirst G. Analysis of Polarity Information in Medical Text. *AMIA Annu Symp Proc*. Washington, DC 2005:4.
[15] Kahn CE. A Digital Library of Radiology Images. *AMIA Annu Symp Proc*. Washington, DC 2006:972.
[16] Mirhaji P, Bryne S, Kunapareddy N, Ward Casscells S. Semantic Approach for Text Understanding of Chief Complaints Data. *AMIA Annu Symp Proc*. Washington, DC 2006:1033.
[17] Silfen E, Patel CO, Mendoça E, Friedman C. ZebraHunter: Searching Rare Medical Diagnoses and Retrieving Relevant Citations. *AMIA Annu Symp Proc*. Washington, DC 2006:1094.
[18] Peace J, Brennan PF. MetaMapping the Nursing Procedure Manual. *AMIA Annu Symp Proc*. Washington, DC 2006:1060.
[19] Meystre S, Haug P. Improving the Sensitivity of the Problem List in an Intensive Care Unit by Using Natural Language Processing. *AMIA Annu Symp Proc*. Washington, DC 2006:554-8.
[20] Baud RH, Ruch P, Gaudinat A, Fabry P, Lovis C, Geissbuhler A. Coping with the variability of medical terms. *Proc of MEDINFO 2004*. San Francisco 2004:322-6.
[21] Leroy G, Rindflesch TC. Using Symbolic Knowledge in the UMLS to Disambiguate Words in Small Datasets with a Naive Bayes Classifier. *Proc of MEDINFO 2004*. San Francisco 2004:381-5.
[22] Hagerty CG, Chang J, Pickens DS, Kulikowski CA, Sonnenberg FA. Semi-automated Encoding of Guidelines. *Proc of MEDINFO 2004*. San Francisco 2004:1625.
[23] Tringali M, Rindflesch TC, Kilicoglu H, Fiszman M, Bodenreider O. Strategies for Mapping Concepts in Gastrointestinal Endoscopy Reports to the UMLS Metathesaurus. *Proc of MEDINFO 2004*. San Francisco 2004:1885.
[24] Meng F, D'Avolio LW, Chen AA, Taira RK, Kangarloo H. Generating Models of Surgical Procedures using UMLS Concepts and Multiple Sequence Alignment. *AMIA Annu Symp Proc*. Washington, DC 2005:4.
[25] Coonan KM. Utility of Semantically Constrained Automated Extraction and Mapping of UMLS Concepts from Clinical Narratives. *AMIA Annu Symp Proc*. Washington, DC 2005:930.
[26] Yetisgen-Yildiz M, Pratt W. The Effect of Feature Representation on MEDLINE Document Classification. *AMIA Annu Symp Proc*. Washington, DC 2005:53.
[27] Zhou L, Srinivasan P. Concept space comparisons: explorations with five health domains. *AMIA Annu Symp Proc*. Washington, DC 2005:8.
[28] Crowley RS. ReportTutor-An Intelligent Tutoring System that Uses a Natural Language Interface. *AMIA Annu Symp Proc*. Washington, DC 2005:5.
[29] Lieberman MI, Ricciardi TN. Using NLP to Extract Concepts from Chief Complaints. *AMIA Annu Symp Proc*. Washington, DC 2005:1029.
[30] Wang Y, Liu Z. Personalized health information retrieval system. *AMIA Annu Symp Proc*. Washington, DC 2005:1149.
[31] Ogren PV, Savova G, Buntrock JD, Chute CG. Building and Evaluating Annotated Corpora for Medical NLP Systems. *AMIA Annu Symp Proc*. Washington, DC 2006:1050.
[32] Patel CO, Cimino JJ. Mining Cross-Terminology Links in the UMLS. *AMIA Annu Symp Proc*. Washington, DC 2006:624-8.
[33] Hagerty CG, Pickens DS, Chang J, Kulikowski CA, Sonnenberg FA. Prediction in Annotation Based Guideline Encoding. *AMIA Annu Symp Proc*. Washington, DC 2006:314-8.
[34] Ruiz ME. Combining Image Features, Case Descriptions and UMLS Concepts to Improve Retrieval of Medical Images. *AMIA Annu Symp Proc*. Washington, DC 2006:674-8.
[35] Taira RK, Soderland SG. A statistical natural language processor for medical reports. *AMIA Annu Symp Proc*. Washington, DC 1999:970-4.
[36] McCray AT, Srinivasan S, Browne AC. Lexical methods for managing variation in biomedical terminologies. *Proc Annu Symp Comput Appl Med Care* 1994:235-9.
[37] Aho AV, Corasick MJ. Efficient string matching: an aid to bibliographic search. Communications of the ACM. 1975;18(6):333-40.

Address for Correspondence

Vijayaraghavan Bashyam
vbashyam @ ucla · edu

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Extracting Subject Demographic Information From Abstracts of Randomized Clinical Trial Reports

Rong Xu[*1], Yael Garten[*1], Kaustubh S. Supekar[1], Amar K. Das[2], Russ B. Altman[3], Alan M. Garber[4]

[1]Biomedical Informatics Training Program, Stanford University School of Medicine, USA; [2]Stanford Medical Informatics, Stanford University, USA; [3]Department of Genetics, Stanford University, USA; [4]VA Palo Alto Health Care System; and Center for Primary Care and Outcomes Research and Center for Health Policy, Stanford University, USA

Abstract

In order to make more informed healthcare decisions, consumers need information systems that deliver accurate and reliable information about their illnesses and potential treatments. Reports of randomized clinical trials (RCTs) provide reliable medical evidence about the efficacy of treatments. Current methods to access, search for, and retrieve RCTs are keyword-based, time-consuming, and suffer from poor precision. Personalized semantic search and medical evidence summarization aim to solve this problem. The performance of these approaches may improve if they have access to study subject descriptors (e.g. age, gender, and ethnicity), trial sizes, and diseases/symptoms studied.

We have developed a novel method to automatically extract such subject demographic information from RCT abstracts. We used text classification augmented with a Hidden Markov Model to identify sentences containing subject demographics, and subsequently these sentences were parsed using Natural Language Processing techniques to extract relevant information. Our results show accuracy levels of 82.5%, 92.5%, and 92.0% for extraction of subject descriptors, trial sizes, and diseases/symptoms descriptors respectively.

Keywords:

randomized clinical trial, information extraction, subject demographics, semantics. statistical NLP, text analysis

Introduction

To undertake informed medical decision-making, healthcare consumers must have access to accurate and reliable medical information, particularly about illnesses and their treatments [1]. Randomized Clinical Trial papers are a reliable source of information about treatments, their efficacies, and potential adverse consequences. Although generally intended for a professional medical audience, healthcare consumers routinely discover this information using online keyword-based search methods such as Google and PubMed [2]. These methods can overload the user with irrelevant information.

Personalized semantic search and text summarization are two approaches that aim to improve the precision of the search results. In the healthcare setting, these approaches entail extracting structured and machine-understandable information from documents [3].

The PICO framework [4] is designed to provide personalized medical evidence from RCT papers. It represents structured information from RCT papers and patient records related to problem, population, intervention, comparison, and outcome. The structured information extracted from the patient records serves as query context, and is matched with the structured information captured from the RCT paper. However, the PICO framework only provides a model of the clinical information-seeking process; no operational methods are provided in the framework to actually extract structured information.

RCT Bank [5], a well-known medical summarization approach, has focused on modeling information that is to be extracted from RCT papers. This model includes concepts that encode trial information such as the trial protocol, execution, and results summary. Currently, this information is manually extracted from RCT papers, and is archived in the RCT Bank. The number of RCT papers archived, however, is low, primarily because the curation process is manual.

Approaches such as RCT Bank and PICO require methods to automatically extract trial information from RCT papers, in order to assess the relevance of a trial report to a particular query. For example, the *number of participants* may be indicative of the quality of the trial; *subject descriptors* such as gender and age of the participants as well as *diseases/symptoms* can be used to stratify trials into subgroups of interest. These items of information should improve search precision, while providing a measure of the quality of these trials. In this paper, we present a novel method to extract subject demographic information from RCT abstracts.

Prior work on an automated approach for extracting subject demographics from abstracts used manually crafted pattern-matching rules [6]. The accuracy of this approach was 80% for extracting subject descriptors. Noisy perfor-

* Authors contributed equally to this work.

mance may be a source of error in subsequent steps of the search and retrieval process, so improvements in precision/recall would be valuable. Moreover, this approach only retrieves the sentences that contain the demographic information, and not the itemized information such as age and number of trial participants.

Materials and methods

To improve performance, we developed a method to automatically extract subject demographic information from RCT abstracts. We used text classification augmented with a Hidden Markov Model to identify sentences containing subject demographics (step 1 of our Methods), and subsequently parsed these sentences using Natural Language Processing techniques to extract relevant information (step 2 of our Methods).

Step 1: Identifying sentences containing subject demographics

Most abstracts from articles reporting the results of RCTs are conceptually organized into five sections (Background, Objective, Methods, Results, Conclusions). Subject demographic information is more likely to be contained in the Methods section[1]. Only 28% of all RCT abstracts, however, are explicitly structured into these five sections. In order to automatically label sentences in an abstract as belonging to the Methods section, we combined text classification with a Hidden Markov Model [7]. 3,896 structured RCT abstracts that contained Methods sections, published in 2005, were de-labeled and parsed into 46,370 sentences using a sentence parser that we developed. Each sentence was a labeled input to a multi-class (Background, Objective, Methods, Results, Conclusions) text classifier. These sentences were used for model training. We then used this trained classifier to label 4018 sentences from 250 unstructured abstracts of RCTs with these five labels.

The Methods section not only contains sentences that describe subject demographics, but also contains additional sentences that describe information such as trial design, settings, interventions, and outcome measures. Only 4% of all RCT abstracts, however, are explicitly structured into these subsections, thus making the task of identifying the subject demographic sentences difficult. In order to identify the specific sentences that contain subject demographics, we used a Maximum Entropy two-class (Subject-Demographics, Other) text classifier. 14,774 structured RCT abstracts were selected. These abstracts are all those that contained subsections explicitly labeled as Participants, within the Methods section; the sentences in these Participants sections contain the demographic information. Unigrams and bigrams that appeared in these sentences and that appeared in non-Participant sentences from the Methods section were used as features for model training. We then used this trained classifier to label the 982 predicted Methods sentences from the 250 unstructured abstracts described above with these two labels.

Step 1 was a necessary part of our method, because patients are described throughout the abstracts, however the relevant information that we desired to extract typically appears in 1 or 2 sentences within the Methods section of abstracts. Solely using presence of anchor words such as "patient", "participant", etc., to identify informative sentences, would have resulted in a high number of false positive sentences.

Step 2: Extracting subject descriptors, number of trial participants, and diseases/symptoms and their descriptors from sentences labeled as subject demographic related

Using the sentences that were classified in Step 1 as subject demographic related, we extracted three specific types of information: (1) *subject descriptors* (such as "men", "healthy", "elderly"), (2) *number of participants* (or all subgroups, when total number is not available), and (3) *diseases/symptoms* and their *descriptors*.

(1) Extracting subject descriptors:

For each subject demographic related sentence, we extracted *subject descriptors* such as "men", "diabetic", "healthy", and "elderly". These words mainly appear before/after anchor words such as "patients", "participants", or "subjects" in a sentence (e.g. "we studied 300 elderly diabetic patients"). For each sentence, these anchor words were extracted and tagged as [PATIENT] using a closed set of words approach. The set included 37 words, and is available at http://www.stanford.edu/~xurong/medinfo2007.

In order to extract the descriptors that modify the word tagged [PATIENT], we used the parse tree of the sentence that was generated by the Stanford parser [8]. The Stanford parser is an unlexicalized natural language parser developed by Klein et al. at Stanford University. It was trained on a non-medical document collection (Wall Street Journal), but has been successfully used in the biomedical domain [9]. From the word tagged as [PATIENT] in the parse tree, the unary chain was traversed to extract siblings of the [PATIENT] word. *Subject descriptors* are located in the preceding part (left-sibling) and the following part (right-sibling) of the [PATIENT] word. For example, in the parse tree shown in figure 1, the left sibling phrase of the word "patient" (which is tagged [PATIENT]) contains the subject descriptor "severely brain-injured", and the right sibling phrase is "with first-episode psychosis".

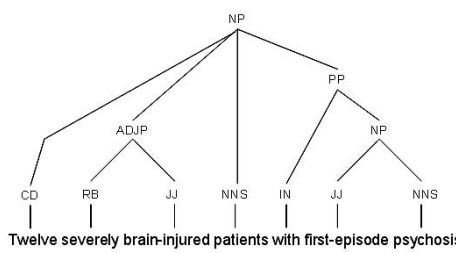


Figure 1 - Parse tree of an example subject demographics related sentence.

1 1000 randomly selected structured abstracts were manually annotated to determine sentences that contained subject demographics. 931 out of 1000 had subject demographics in the Method section.

(2) Extracting number of trial participants:

For each subject demographic related sentence, we extracted the number of trial participants. Each sentence was tagged with the Stanford parser, and words identified as "CD" (the lexical tag for numbers) were tagged as [NUMBER]. This process was augmented with a closed set of words approach, as the parser missed some words. For each sentence, these identified words were also tagged as [NUMBER]. Most sentences contain multiple numbers; however only one is to be extracted as the number of trial participants (e.g. "40 participants above age 50 who suffered from diabetes for over 2 years participated in the trial"). To identify the one ("40 participants") that contains number of trial participants, initially, [NUMBER] words such as "age 50" and "2 years" that do not pertain to the number of trial participants were filtered out using rules such as [age][NUMBER] or [NUMBER][years].

However, the filtering process, in some cases, leaves more than one word that is tagged as [NUMBER]. From this set of words, identifying the exact [NUMBER] word(s) that contain number of trial participants is non-trivial. For example: in the sentence "300 patients included 100 men and 100 children who suffered from diabetes", there are three numbers (300, 100, 100) left by the filtering processing; the correct number of trial participants is 300.

To address this complex sentence structure, (i) we manually created a grammar based on a subset of subject-demographic related sentences, (ii) this grammar was used to extract grammatical sentence structures from the rest of subject demographic related sentences, and (iii) subsequently the number of trial participants was extracted from the grammatical sentence structure identified in step ii, either directly in the case of only one [NUMBER], or by summing several numbers (in a post-processing stage) if several subgroups were identified. Each of these steps is described in detail in the following text:

(Step i) Our grammar includes 39 rules which encapsulate the different ways by which authors describe subject demographics. The lexicon of the grammar comprised of the tags [NUMBER], [PATIENTS], [INCLUDE], and several classes of words such as "are", "in". Readers can view grammar on the web at http://www.stanford.edu/~xurong/medinfo2007.

(Step ii) We extracted the grammatical sentence structures, such as "S→ [NUMBER] [PATIENT] [INCLUDE] [NUMBER] [PATIENT]", from each sentence. For example, the sentence "200 participants included 70 men" has the above-mentioned grammatical structure. To correctly extract the grammatical structure, each sentence was first transformed by replacing words with the tags in our lexicon such as [NUMBER], [PATIENT], and [INCLUDE], and omitting all others. Then, a modified Cocke-Younger-Kasami (CYK) backtracking algorithm was used to determine the longest span beginning at the sentence start, which still contains a complete sentence structure based on our grammar. For example, the sentence "The study observed 300 men who underwent over 6 operations each" is converted to "[NUMBER] [PATIENT] [NUMBER]" (corresponding to "300 men 6"), and is then traversed to extract the longest span containing a full sentence according to our grammar, which in this case is S→ [NUMBER] [PATIENT] (our grammar does not contain the rule S→ [NUMBER] [PATIENT] [NUMBER]). Thus, "300 men" is extracted, and the number of trial participants is correctly identified as 300. In this way we are able to identify which number within a sentence actually represents the number of trial participants. The significant benefit of our approach over traditional CYK based methods [10] is that not only can our approach extract grammatical structure from the whole sentences but it can also extract structure from embedded sentences within the sentence. This benefit is particularly advantageous for our problem as most of the subject-demographic related sentences have a left embedded sentence, which has number of trial participants. In the previous example, this entails extraction of "200 participants" which was the left embedded sentence that corresponds to the structure S→ [NUMBER] [PATIENT] in our grammar. The traditional CYK algorithm in this case will fail since it extracts the structure from the whole sentence, and the structure extracted for the example does not have a corresponding structure in our grammar.

(Step iii) We extracted the number of trial participants from the grammatical sentence structure identified in step ii; this was either done directly in the case of only one [NUMBER], or by summing several numbers if subgroups were identified

(3) Extracting diseases/symptoms and their descriptors:

For each subject demographic related sentence, we extracted the *diseases/symptoms* (e.g."diabetes", "rhinosinusitis") and their *descriptors* (e.g. "severe", "acute"). We extracted these using UMLS MetaMap [11], and extracted descriptors of these *diseases/symptoms* by traversing the parse tree.

UMLS MetaMap, a tool developed by the National Library of Medicine, semantically processes an input sentence by (i) parsing the sentence and then (ii) identifying the sentence phrases within it which map to a semantic type such as disease, syndrome, or symptom. MetaMap performs step (ii) by mapping the parsed sentence phrases onto concepts in the UMLS Metathesaurus, a metathesaurus of all clinically related terms (e.g. "rhinosinusitis" is mapped to the semantic type "Disease or Syndrome"). Each subject demographic related sentence was inputted into UMLS MetaMap. Sentence phrases that were mapped to the semantic type "Disease or Syndrome" or "Sign or Symptom" contain the diseases/symptoms.

In addition to identifying names of *diseases/symptoms* it is also important to extract *descriptors* of these *diseases/symptoms* not identified by UMLS Metamap. For example, in the sentence phrase "acute juvenile diabetes" MetaMap correctly identifies "juvenile diabetes" as *diseases/symptoms*, but misses the *descriptor* "acute". To extract these descriptors from the subject demographic related sentence, we generated a parse tree for that sentence. In this parse tree, the modifiers preceding the sentence phrases identified by MetaMap as *diseases/symptoms* contain the descriptors.

Results

Step 1: Identifying sentences containing subject demographics

For the task of identifying the Methods section of unstructured RCT abstracts, our five-class text classifier augmented with a Hidden Markov Model achieved an F1 measure of 91.5% on the test data. The F1 measure is a combined measure of precision and recall.

F1=(2*precision*recall)/(precision+recall)

For the task of identifying the subject demographic related sentences in the predicted Methods section of 250 unstructured abstracts, we used a two-class text classifier on the 14,774 structured RCT abstracts that contained a Participants subsection. The classifier was first trained using 60% of the 14,774 abstracts, while the remaining 40% were used to test the classifier, which achieved an F1 score of 92%. However, when testing the classifier on the predicted 982 Methods sentences from the 250 unstructured abstracts without Participants subsections, the F1 score was only 77%. We postulate that the low F1 score is due to a difference in writing style used by authors when writing a structured abstract that has an explicit Participants subsection versus one without one. To test this hypothesis, we retrained the two-class classifier on the Methods sentences from 1000 unstructured abstracts that were manually tagged either as 'Subject-Demographics' (for participant-related sentences) or 'Other'. The results of this test (F1 score = 91%) show a significant improvement over the previous results.

Step 2: Extracting subject descriptors, number of trial participants, and diseases/symptoms and their descriptors from sentences labeled as subject demographic related

Using the sentences that were classified in Step 1 as subject demographic related, we extracted three types of information.

(1) Extracting subject descriptors:

Using the method previously described in the Methods section, we extracted *subject descriptors* from each sentence. The accuracy of identification of the correct *subject descriptors* in 250 unstructured abstracts was 82.5%. In this paper, accuracy equals the percent of correctly tagged entities.

(2) Extracting number of trial participants:

Using the method previously described in the Methods section, we extracted the *number of trial participants* from 50 unstructured abstracts. The accuracy of identification of the correct *number of trial participants* in the test data was 92.5%.

(3) Extracting diseases/symptoms and their descriptors:

For each subject demographic related sentence, we used UMLS MetaMap to extract *diseases/symptoms* (accuracy = 51.0%). *Descriptors* for these *diseases/symptoms* were then extracted by traversing the parse tree (accuracy = 92.0%).

Table 1 - Accuracy measures for information extractions steps

Information Extracted	Accuracy (%)
Subject Descriptors	82.5
Number of Trial Participants	92.5
Diseases/Symptoms	51.0
Descriptors of Diseases/Symptoms	92.0

Discussion

In this paper we describe a novel method to automatically extract from RCT abstracts three types of subject demographic information: (1) *subject descriptors*, (2) *number of participants*, and (3) *diseases/symptoms* and their *descriptors*. The accuracy measures for extracting these three types of information appear in Table 1.

Our good performance relies on two novel elements used in our methods. First, we extracted the subject demographic information from the sentences in the abstract that are most likely to have that information, namely, specific sentences found within the methods section. Previous approaches examined the entire abstract and thus did not take advantage of the inherent structure within the RCT abstract. Second, we used a handcrafted grammar to extract the subject demographic information from the relevant sentences. We developed this grammar by analyzing unstructured abstracts in order to capture the diversity of ways to describe subject demographic information. This grammar-based approach is advantageous over existing approaches that use handcrafted rules, and thus are not robust to varying writing styles.

The performance of our method is not perfect, and in the following section we analyze, for each step, limitations of our methods and propose solutions that may be useful in overcoming these limitations.

Limitations in identification of subject demographics related sentences

We trained the multi-class text classifier used for classifying sentences into five sections using structured abstracts. This classifier had an accuracy of 91% when tested on unstructured abstracts. We may further improve this accuracy by training the classifier using unstructured abstracts, instead of structured, which were chosen to avoid the time-consuming process of manual tagging of unstructured abstracts.

Limitations in extraction of subject descriptors

For each subject demographic related sentence, we identified anchor [PATIENT] words and then the parse tree of that sentence was used to extract descriptors that modify the [PATIENT] words. This approach works well for sentences that have a simple structure (Accuracy =82.5%). However, this approach tends not to perform well in cases where authors tend to describe subjects using a complex sentence structure. For example, for the sentence "Forty

ASA physical status I, II, and III patients presenting for primary total hip replacement" our method extracts "III patients" as subject descriptors. This is because the comma before "and III patients" causes the parser to separate this phrase as a new noun phrase.

These parse tree errors are not surprising, because the Stanford parser used to generate the tree was trained using a non-medical text corpus. To improve the parse tree, this parser could be retrained using a relevant medical corpus, but no good corpus exists for RCTs.

Limitations in extraction of number of trial participants

For each subject demographic related sentence, we tagged words in that sentence with tags in our lexicon. All the words that were not tagged were deleted. We then used our handcrafted grammar to extract *number of trial participants* from the tagged words.

Our approach for extracting number of trial participants works very well for sentences that have the numbers in close proximity to the word tagged as PATIENT] (accuracy = 92.5%). However, our grammar tends not to perform well in cases where the numbers are not in close proximity. For example, for the sentence, "We checked 20 patients and found that 6 of the patients had …", would be tagged as "[NUMBER] [PATIENT] and [NUMBER] [PATIENT]". Since our grammar has the rule "S→ [NUMBER] [PATIENT] and [NUMBER] [PATIENT]" because this rule often allows us to identify two subgroups (e.g. "40 men and 30 women"), the grammar in this case interprets the sentence as including two subgroups of 20 patients and 6 patients, which of course is wrong, and simply results from the fact that we have the word "and" in the sentence, followed by a number and the word "patients" with words interspersed between them, that are removed due to the sentence transformation algorithm. We may avoid this problem if we can encode heuristics such as: "if there is a distance of over 10 words between the 2 numbers, we truncate the tagged sentence before the second number." However, we may then lose correct information, such as cases where two true subgroups are separated by many words.

Limitations in extraction of diseases/symptoms and their descriptors

For extraction of *diseases/symptoms*, we used UMLS MetaMap, which has some limitations. For example, MetaMap did not recognize some diseases. For those cases in which it did extract the *diseases/symptoms*, we extracted *descriptors* for these *diseases/symptoms* with high accuracy by traversing the parse tree generated using the Stanford parser. However, this parser was trained on the Wall Street Journal corpus, which does not focus on diseases and symptoms. Therefore, if the parser encounters a word that is not present in the training corpus, it will not know the correct part of speech (POS) and thus will give equal probability to all POS tags. In many cases, the parser may incorrectly tag *diseases/symptoms* as verbs. The subsequent parse will then be entirely wrong. The accuracy of this step may be further improved by training the parser using a medical corpus.

Conclusion

By focusing on specific sentences in the abstract, and subsequently using a grammar we developed, we obtained high performance in automatic extraction of subject demographics from RCT abstracts. This extracted information can be used to build systems that deliver accurate (*subject descriptors* and *diseases/symptoms* can be used to stratify trials into subgroups of interest) and reliable (*number of participants* may be indicative of the quality of the trial) information for informed medical decision-making. It should be possible to extend our method to obtain broader types of information, such as outcomes from individual RCTs, and to use the automatically extracted data for labor-intensive projects like meta-analyses.

Acknowledgments

YG and RX are supported by training grant 5T15LM007033-22 from the U.S. National Library Medicine (NLM). RBA is supported by NIH/NIGMS Pharmacogenetics Research Network and Database (U01GM61374). AMG is supported by the Department of Veterans Affairs and grants from the National Institute on Aging. The authors wish to acknowledge Christopher D. Manning for helpful discussions.

References

[1] McKeown K, Chang SF, Cimino J, et al. PERSIVAL, a system for personalized search and summarization over multimedia healthcare information. Proc of the Joint Conf. on Digital Libraries, 2001.

[2] Steinbrook R. Searching for the right search - reaching the medical literature. N Engl J Med 2006; 354: 4–7.

[3] Lin J and Demner-Fushman D. The role of knowledge in conceptual retrieval: A study in the domain of clinical medicine. Proc of SIGIR 2006.

[4] Sackett DL, Strauss SE, Richardson WS, Rosenberg W, and Haynes RB. Evidence-based medicine: how to practice and teach EBM. Churchill Livingstone, second edition, 2000.

[5] Sim I, Owens DK, Lavori PW, Rennels GD. Electronic Trial Banks: A complementary method for reporting randomized trials. Med Decis Making 2000; 20(4):440-50.

[6] Demner-Fushman D, Lin J. Knowledge extraction for clinical question answering: Preliminary results. Proc of the AAAI 2005 Workshop on Question Answering in Restricted Domains, 2005.

[7] Xu R, Supekar KS, Huang Y, Das A, and Garber AM. Combining text classification and hidden Markov modeling techniques for structuring randomized clinical trial abstracts. Proc of AMIA Symp 2006.

[8] Klein D and Manning CD. 2003. Accurate Unlexicalized Parsing. Proc of the 41st Meeting of the Association for Computational Linguistics 2003; 423–30.

[9] Huang Y, Lowe HJ, Klein D, Cucina RJ. Improved identification of noun phrases in clinical radiology reports using a high-performance statistical natural language parser augmented with the UMLS specialist lexicon. J Am Med Inform Assoc 2005; 12(3):275-85.

[10] Chappelier J, Rajman M. A generalized CYK algorithm for parsing stochastic CFG. Proc. Tabulation in Parsing and Deduction, 1998; 133-137.

[11] Aronson AR. Effective mapping of biomedical text to the UMLS Metathesaurus: The MetaMap program. Proc of AMIA Symp 2001.

Address for correspondence

russ.altman@stanford.edu

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Coupling Ontology Driven Semantic Representation with Multilingual Natural Language Generation for Tuning International Terminologies

Anne-Marie Rassinoux[a], Robert H. Baud[a], Jean-Marie Rodrigues[b], Christian Lovis[a], Antoine Geissbühler[a]

[a] *University Hospitals of Geneva, Service of Medical Informatics, Switzerland*
[b] *University of Saint Etienne, France*

Abstract

Objectives: *The importance of clinical communication between providers, consumers and others, as well as the requisite for computer interoperability, strengthens the need for sharing common accepted terminologies. Under the directives of the World Health Organization (WHO), an approach is currently being conducted in Australia to adopt a standardized terminology for medical procedures that is intended to become an international reference.*

Method: *In order to achieve such a standard, a collaborative approach is adopted, in line with the successful experiment conducted for the development of the new French coding system CCAM. Different coding centres are involved in setting up a semantic representation of each term using a formal ontological structure expressed through a logic-based representation language. From this language-independent representation, multilingual natural language generation (NLG) is performed to produce noun phrases in various languages that are further compared for consistency with the original terms.*

Results: *Outcomes are presented for the assessment of the International Classification of Health Interventions (ICHI) and its translation into Portuguese. The initial results clearly emphasize the feasibility and cost-effectiveness of the proposed method for handling both a different classification and an additional language.*

Conclusion: *NLG tools, based on ontology driven semantic representation, facilitate the discovery of ambiguous and inconsistent terms, and, as such, should be promoted for establishing coherent international terminologies.*

Keywords:

natural language generation, standardized terminology, ontology, medical procedures or interventions

Introduction

Today, one of the major challenges in clinical terminology standards is the ability to build consistent terminological systems that fit the needs of specific healthcare systems while being reusable in different countries. They should provide the basis for communication between human actors and simultaneously be computer understandable, in order to insure interoperability between systems. By definition, a clinical terminology is a set of terms, expressed in a specific language, and organized within a universe of discourse in order to reflect specific entities in clinical reality. A model that organizes these entities into a taxonomy, provides formal definitions for them and clarifies their distinctive properties is called an ontology. A terminology can then be seen as a bridge from ontology to linguistic information and as such, is strongly language-dependent. Specific purposes and links between terminologies, ontologies and natural languages must be clearly understood in order to avoid confusion [1, 2].

Sharing terminologies at regional, national and international levels requires standardization processes at two levels: information technology level and information content level. The first level concerns the choice of a language-independent modeling structure able to represent entities and relationships expressed in terms, in a way that is tractable (that is to say allowing navigation, parsing and interpretation) by a machine-based system. Formal ontological representations expressed through a logic-based language have proven to be adequate for this task [1]. Concrete implementations have come to light in recent years with the GALEN model [1] that consists of around 52,000 entities and 800 relationships, SNOMED CT [1] that describes over 366,000 entities with unique meanings that are depicted by a set of terms (descriptions), or the FMA ontology [1] that contains approximately 75,000 classes and 168 relationship types for expressing biomedical knowledge. The second level deals with what must be expressed in such a modeling structure and requires tools for disambiguating and validating the overall content. For this task, natural language generation (NLG) tools have proved their efficiency as they allow the semantic meaning of terms to be assessed both at the level of the internal representation and at the level of the natural language (NL) expressions [1, 2, 3].

The main objective of this paper is to show that the abovementioned dual approach is feasible and cost effective for the development of an international terminology for procedures. First, a description of the interest in several countries and for the World Health Organization (WHO) [1] to have an international classification for procedures

and interventions is given. Then, the flexibility of this method to add representations of new terms into the ontological structure of GALEN, and to further generate them into an additional language, is emphasized. The disambiguation and refining processes of both the intermediate representation and the initial terms are performed by comparing the generated terms with the original ones. Finally, critical points for establishing the basis of an international terminology are highlighted.

Classification of interventions: state of the art

The first International Classification of Procedures in Medicine (ICPM) was published in 1978. But, contrary to the international Classification of Diseases (ICD-9) that is used almost worldwide, ICPM was never accepted by many nations that were compelled to develop their own incompatible standards for coding procedures and interventions. The most significant work was conducted in Australia with the development of the Australian Classification of Health Interventions (ACHI) [11], in Canada with the Canadian Classification of Health Interventions (CCI) [12], in the USA with the ICD-10 Procedure Coding System (ICD-10-PCS) [13], and in France with the new French procedures classification called CCAM (a French acronym for Classification Commune des Actes Médicaux) [14].

International evaluations of these local classifications, as well as comparisons with their national adaptation of ICD (such as ICD-10-AM in Australia or ICD-10-CA in Canada), are difficult to carry out without the existence of a common accepted structure upon which agreement on standards can rely. Recent initiatives from the National Centre for Classification in Health (NCCH) – Australia, conjointly with the World Health Organization Family of International Classifications (WHO-FIC), tend to fill in this gap by proposing an international reference coding system for procedures. It is based on the existing reduced Australian International Classification of Health Interventions (ICHI), that is derived from both ACHI and ICD-10-AM classifications. Moreover, it is intended to be used by countries that do not, as yet, have their own classification of interventions.

The decision to generate this new terminology in the additional Portuguese language is not fortuitous as no widespread classification for procedures exists in Portuguese speaking countries. This paper presents the tuning and assessment of this new terminology, by first describing its language-independent representation using the GALEN methodology and then focusing on the efforts needed to adapt the NLG tools, developed in Geneva as part of the GALEN project [15], for the handling of the Portuguese language.

GALEN CCAM developments and ICHI

Following the interest of WHO and in several countries for the GALEN CCAM developments, a similar approach was adopted to assess the meaning of terms described in ICHI. Indeed, the GALEN project affords, in a language-independent manner, a high level ontology that organizes concepts and relationships in a multiple inheritance hierarchy, together with formal subsumption and multi-level sanctioning to constrain composition of sensible concepts. This common reference model for medical concepts (the so-called CORE Model) is expressed in a formal language (the so-called GRAIL formalism for GALEN Representation and Integration Language) [16]. This project thus evolved between 1996-1999 toward the GALEN-IN-USE project whose aim was to assist national coding centers in the collaborative construction and maintenance of surgical procedures classifications. In particular, the University of Saint Etienne Medical Informatics Centre (France) has applied this methodology for ensuring the quality and coherence of the new French coding system for surgical procedures called CCAM [17]. Such a methodology produces a structured, language-independent representation for each original term. This is realized by the various national coding centers that are in charge of producing semi-automatically an intermediate representation (so-called a dissection) which was subsequently nearly-automatically expended into a GRAIL expression. For the needs of the NLG tools [15], this GRAIL expression is automatically transformed into a conceptual graph (CG) [18] using Definite Clause Grammar (DCG). These semantic graphs, depicting entities or concepts linked by relationships, allow formal operations [19] to be performed in order to contract or expand parts of the graph.

The 15 chapters of the new French CCAM classification, each containing about 500 French rubrics, were successfully represented in the GALEN model, enlarged for this task with around 700 new entities. In order to disambiguate and refine the final linguistic labels of these rubrics, a generation from these semantic representations, in French, but also in English for comparisons, was systematically performed. The validating process of these generated terms against the initial ones has prompted their authors to reformulate 20% of the initial rubrics. Errors were mainly due to ambiguities in the source terms which were never detected before computer representation was performed [15, 17].

The flexibility of both the ontological representation of GALEN and the generation process to disambiguate and refine the terms, justifies its adoption by WHO for checking the ability and appropriateness of the existing reduced Australian ICHI coding system to become the international reference coding system for procedures [20]. ICHI contains approximately 1,400 procedure codes in its tabular list, as well as inclusion terms (i.e., additional procedural terms categorized to the code and giving examples of terms) that are distributed into 20 chapters. It is structured with a principal axis of anatomical site and a secondary axis of procedure type, and embraces surgical and non surgical interventions. A detailed example of the GALEN methodology is shown in Figure 1 for the handling of the procedure code 1361, belonging to chapter XV of ICHI. The first three steps afford the clarity of meaning, as information is described in a non-ambiguous way, as well as the specificity as only information relevant for interpretation and sharing is expressed. Several GRAIL expressions are produced when the input ICHI procedure code contains

more than one procedure or anatomical localization. This is the case for the ICHI code 1360 "*Application, insertion or removal procedures of head*" or the ICHI code 1364 "*Excision procedures on maxilla, mandible or temporo-mandibular joint*" whose conceptual representation results in three distinct graphs. Each conceptual graph constitutes a reliable starting point for the fourth step dealing with multilingual generation.

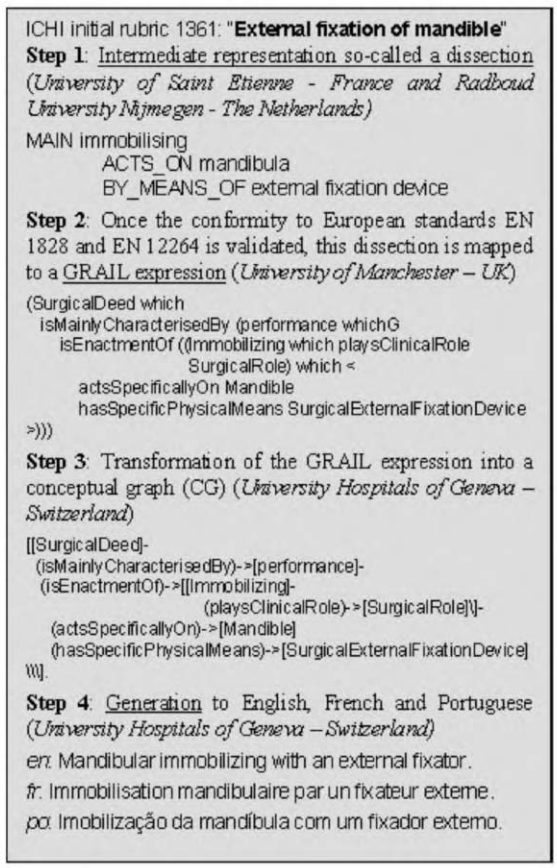


Figure 1 - GALEN methodology applied on ICHI code 1361

Portuguese generation of ICHI procedure codes

A large-scale experiment of generating natural language phrases for surgical procedures modeled in GRAIL has already been achieved within the framework of the French CCAM classification [17]. Target languages for generation were French, English, and to a minor extent Italian. The decision to generate ICHI in the additional Portuguese language was a good opportunity to scale the flexibility and adaptability of the NLG tools to cope with a new language.

The generator, developed as part of the GALEN project, mainly consists of a two-phase process. The first one is concerned with the transformation operations (i.e., expansion and/or contraction of both concept's and relationship's definitions) on the input graph. The second one is concerned with the linguistic realization and consists of selecting the syntactic structures and words that are allowable in the target language, in order to express relationships and concepts respectively. The first phase has two main goals: 1) to simplify the input graph by concealing the specific modeling styles (such as masking the concepts "*SurgicalDeed*" and "*performance*" in Figure 1); 2) to take into account the desired language style which, in this terminological context, is to produce phrases "as detailed as necessary but as concise as possible" thus preferring concise over verbose generations. Moreover, the management of these operations agrees with the linguistic realization phase as replacing an adjacent portion of graph by an explicit entity, is only performed when corresponding words exist in the target language to name this latter. Therefore, adapting the NLG tools, based on the ontological GALEN model, to an additional language, consists of adding a linguistic layer to the two main semantic components of the model, namely the typology of concepts, and the semantic rules that govern the sensible combinations of these concepts. The first linguistic task relates to the language annotation process for concepts, resulting in the creation of dictionaries. The second linguistic task concerns the annotation process for semantic rules. This consists of clarifying syntactic structures that are commonly used in a given language to support the expression of the relationships occurring between conceptual entities. In addition, grammatical rules, that define how the language entities are actually put together, must be specified such as subject/verb or adjective/noun agreements, style, capitalization, punctuation and so on.

The modular representation of data in our system, which results in a clear separation between linguistic knowledge and conceptual knowledge, greatly facilitates the handling of a new language by taking advantage of what was already made in other languages. In the present work, the linguistic features of Portuguese have been added in strong analogy with those existing for French. Indeed, French and Portuguese are two Romance languages (sometimes referred to as Romanic), as they descend from Latin, the language of the Roman Empire. As a result, they share many similar linguistic features. In particular, they have lost the inflection system of nouns to indicate cases (subject, object and so on), they have relatively rigid SVO[1] sentence structure, and they make extensive use of prepositions. Therefore, the same concepts for which an annotation was needed to generate French (likewise English) terms were also annotated for Portuguese. Moreover, all the annotations of the GALEN sensible statements were systematically modified by recopying, for Portuguese, the syntactic structures already defined for French. Finally, function words (or grammatical words) as well as general word usages were added for Portuguese by pointing out those existing for French. For instance, the

1 In linguistic typology, agent-verb-object (AVO), commonly called subject-verb-object (SVO), is a sentence structure where the agent comes first, the verb second, and the object third.

following three rules that describe the definite articles in French:

```
fct_wrd(cl_grammatical,fr,le,det(art_def,mas,sing),_).
fct_wrd(cl_grammatical,fr,la,det(art_def,fem,sing),_).
fct_wrd(cl_grammatical,fr,les,det(art_def,no,plur),_).
```

were translated for Portuguese into the following four rules:

```
fct_wrd(cl_grammatical,po,o,det(art_def,mas,sing),_).
fct_wrd(cl_grammatical,po,a,det(art_def,fem,sing),_).
fct_wrd(cl_grammatical,po,os,det(art_def,mas,plur),_).
fct_wrd(cl_grammatical,po,as,det(art_def,fem,plur),_).
```

In the same way, the convention saying that in French the adjective is generally positioned after the noun is also reutilized for Portuguese.

The obvious effect of this systematic approach is the similarity of the syntactic structures generated in each language. The example shown in Figure 1 corroborates this observation as an adjectival (adj) then a noun-complement (cpl) structures are specified in this order (corresponding to the usage weight given to the structure) in French (fr), English (en) and Portuguese (po), to support the expression of the relationship "*actsSpecificallyOn*", in the mentioned context. In particular, the following more general sensible co-occurrence pattern is applied:

```
stat_annot(22.1,cl_GeneralisedProcess,
                rel_actsOn, cl_BodyStructure,
    [en([adj,cpl,nprem]), fr([adj,cpl]), po([adj,cpl]), it([adj,cpl])],
                                    [en([adj,partPerf(affected,by)]),
fr([adj,partPerf('affecté',par)]),
    po([adj]), it([adj,partPerf(trattato,da)])]).
```

It is worth noting that the relation "*actsOn*" subsumes the relation "*actsSpecificallyOn*". Moreover, the second part of the aforementioned pattern describes the syntactic structures for the realization of the relation "*isActedOnBy*", which is the reverse relation of "*actsOn*". The fact that a noun complement, rather than an adjectival structure, is generated in Portuguese for expressing the localization in the ICHI code 1361, is due to the small coverage of the Portuguese dictionary, where only the noun "*mandíbula*" is defined as annotation for the concept "*Mandible*".

This approach by analogy must nevertheless be carefully checked. On the one hand, the fact that it is based on reusable French rules endangers the generation results that could become a French kind of Portuguese. In order to avoid this outcome, a review process of the generated terms by a Portuguese speaking person is systematically performed, thus leading to the subsequent refinement of the statement annotations, in order to accurately take into account the regular Portuguese language style. On the other hand, producing similar generated terms in different languages strengthens the congruence between generated terms and eases the quality assessment of the resulting terms. Indeed, the generated phrase can be compared, not only with the initial rubric, but also with the terms generated in similar languages.

Results

In order to rapidly evaluate the feasibility and quality of the Portuguese generation on the Australian ICHI coding system, an experiment of applying the GALEN methodology for ICHI was conducted on chapter XV that includes procedures on the musculoskeletal system. This chapter contains around 200 rubrics whose codes are extended from 1360 to 1579. Generating these procedure codes in Portuguese has required an extension of the GALEN model with 6 new concepts, as well as the creation of a Portuguese dictionary of 220 words corresponding to the annotation of 193 distinct GALEN concepts. All the sensible statements (126 in all) were extended to the Portuguese language. Finally around 70 grammatical rules, specified for French, were adjusted. One man-week was necessary to perform the dissections of the 200 initial rubrics and around the same time period was required to adapt the NLG tools for Portuguese.

The initial results for Portuguese generation are outstanding. Future progress would cover the treatment of compound words [9], as well as the growth of the Portuguese dictionary by adding more lexical variants for the annotation of concepts. The review process of the generated terms, against the original ones, pointed out inconsistencies amounting to 20% in the ICHI-Chapter XV source terms. A relevant example, discussed in [21], highlights the incompatible usage of the word "arthroplasty" in three ICHI procedure codes: 1489 (*Arthroplasty of hip*), 1518 (*Arthroplasty of knee)*, and 1519 (*Arthroplasty of knee with bone graft to femur or tibia)*. Indeed, this same word is used to denote either a limited or a complete plastic joint repair. In order to correct these inconsistencies, an improvement of the ICHI structure, using the benefits of the CCAM architecture, is recommended [22].

Discussion and conclusion

Performing comparisons of data between different countries and services, in particular for statistical reporting or cost analysis, implies sharing a common accepted terminology. It should be recognized that, to be an international standard, a terminology should rely on a sound ontology that serves as a formal basis for multiple natural languages.

The GALEN model has proved to be suitable for the computerization of clinical terminologies. Indeed, it supports accepted specifications including among others, conceptual definitions, sensible rules, and basic as well as composite entities that establish a referent language, upon which understanding and exchange of information between different parties can be performed. The total development effort for maintenance and validation is then reduced by sharing such an internal language-independent representation of entities that constitutes a common platform between various WHO member states.

Coupling such a model with NLG tools, constitutes an interesting approach that presents attractive benefits. First, the generation of nonsensical terms is prevented as the semantic representation from which the generation is trig-

gered, is defined unambiguously. The review process of the generated terms, against the original ones, serves to identify and repair inconsistencies and ambiguities both at the level of the original classification and at the level of the derived dissections. Second, this approach automates, and therefore greatly speeds up, the translation of the original classification into various languages, as the internal representation of the original rubrics is made once and serves as a basis for generating terms for all target languages. Similarities between languages are also strongly utilized and this accelerates the adaptation of the NLG tools to new Romanic languages. The fact that the generation is restricted to noun phrases also simplifies the linguistic approach while fitting precisely the terms' structure used to express health classifications. In reality, concise terms are often preferred to wordy ones. Finally, this approach avoids local cross-lingual maintenance and validation of standardized terminology, as comparable national versions of such a terminology can be generated for each country.

In the coming years, NLG tools based on ontology driven semantic structure should become the standard multilingual means to rapidly deliver, in many different countries, consistent national versions of any standardized health terminology.

Acknowledgments

We wish to thank Stefan Schultz for his Portuguese language expertise in reviewing part of the term generation performed on chapter XV of ICHI.

References

[1] Bodenreider O. Lexical, terminological and ontological resources for biological text mining. In: Ananiadou S and McNaught J, eds. Chapter 3, Text mining for biology and biomedicine. Artech House, 2006; pp. 43-66.

[2] Baud RH, Ceusters W, Ruch P, Rassinoux A-M, Lovis C, Geisbühler A. Reconciliation of Ontology and Terminology to cope with Linguistics. Accepted for MEDINFO 2007, Brisbane, Australia, 20-24 August 2007.

[3] Simon J, Dos Santos M, Fielding J, Smith B. Formal ontology for natural language processing and the integration of biomedical databases. International Journal of Medical Informatics, 2006: 75: 224-31.

[4] Rector AL, Baud R, Ceusters W, Claassen W, Rodrigues J-M, Rogers J, Rossi Mori A, vander Haring E, Solomon WD, Zanstra P. A comprehensive approach to developing and integrating multilingual classifications: GALEN's Classification Workbench. In Chute CG, ed. AMIA Annu Symp Proc. 98. Philadelphia: Hanley & Belfus, Inc., 1998: 1115.

[5] Schulz S, Hahn U, Rogers J. Semantic Clarification of the Representation of Procedures and Diseases in SNOMED® CT. In: Engelbrecht et al., eds. MIE 2005 Proc. IOS Press, 2005; pp. 773-78.

[6] The Foundational Model of Anatomy (FMA) : http://sig.biostr.washington.edu/projects/fm/index.html

[7] Traum D, Habash N. Generation from Lexical Conceptual Structures. NAACL-ANLP 2000 Workshop on Applied interlinguas: practical applications of interlingua approaches to NLP: vol. 2, pp. 52-9.

[8] Baud RH, Rodrigues J-M, Wagner JC, Rassinoux A-M, Livis C, Rush P, trombert-Paviot B, Scherre J-R. Validation of Concept Representation Using Natural Language Generation. In: Masys DR, ed.AMIA Annu symp Proc. 97. Philadelphia: Hanley & Belfus, Inc., 1997: 841.

[9] Rassinoux A-M, Ruch P, Baud RH, Lovis C. Semantic Handling of Medical Compound Words through Sound Analysis and Generation Processes. In Overhage JM, ed. AMIA Annu Symp Proc. 2000. Philadelphia: Hanley & Belfus, Inc., 2000; pp. 675-9.

[10] The World Health Organization (WHO): http://www.who.int/

[11] The Australian Classification of Health Interventions (ACHI) 5th edition: http://meteor.aihw.gov.au/content/index.phtml/itemId/335419

[12] The Canadian Classification of Health Interventions (CCI): http://secure.cihi.ca/cihiweb/dispPage.jsp?cw_page=codingclass_cci_e

[13] The ICD-10 Procedure Coding System (ICD-10-PCS): http://www.cms.hhs.gov/ICD9ProviderDiagnosticCodes/08_ICD10.asp

[14] The Classification Commune des Actes Médicaux (CCAM): http://www.caducee.net/DossierSpecialises/systeme-information-sante/ccam.asp

[15] Rassinoux A-M, Lovis C, Baud RH, Scherrer J-R. Versatility of a Multilingual and Bi-directional Approach for Medical Language Processing. In Chute CG, ed. AMIA Annu Symp Proc. 98. Philadelphia: Hanley & Belfus, Inc., 1998; pp. 668-72.

[16] Rector A. Compositional Models of Medical Concepts: Towards Re-usable Application-Independent Medical Terminologies. In: Barahona P and Christensen JP, eds. Knowledge and Decisions in Health Telematics. Amsterdam: IOS Press, 1994; pp. 109-14.

[17] Rodrigues J-M, Trombert-Paviot B, Baud R, Wagner J, Meusnier-Carriot F. Galen-In-Use: Using artificial intelligence terminology tools to improve the linguistic coherence of a national coding system for surgical procedures. In: Cesnik B, McCray AT and Scherrer J-R, eds. MEDINFO 98 Proc. Amsterdam: IOS Press. 1998; pp. 623-27.

[18] Sowa JF. Conceptual Structures: Information Processing in Mind and Machine. Reading, Massachusetts: Addison-Wesley Publishing Company, 1984.

[19] Rassinoux A-M, Baud RH, Lovis C, Wagner JC, Scherrer J-R. tuning up Conceptual Graph Representation for Multilingual Natural Language Processing in Medicine. In Mugnier M-L and Chjein M, eds. Proc. of the 6th International Conference on Conceptual Structures (ICCS'98). Berlin: Springer, 1998; pp. 390-97.

[20] Rodrigues J-M, Rector A, Zanstra PE, Baud R, Rogers J, Rassinoux A-M, Schulz S, Trombert B, ten Napel H, Clavel L, vander Haring E, Mateus C. Supporting WHO Classifications with GALEN tools. Who-family of international classifications network meeting, Tokyo, Japan, October 2005.

[21] Rodrigues J-M, Rector A, Zanstra P, Baud R, Innes K, Rogers J, Rassinoux A-M, Schulz S, Trombert Paviot B, ten Napel H, Clavel L, vander Haring E, Mateus C. An ontology driven collaborative development for biomedical terminologies: from the French CCAM to the Australian ICHI coding system. In: Hasman A et al., eds. MIE 2006 Proc. IOS Press, 2006; pp. 863-8.

[22] Zaiss A., Hanser S, Schulz S. Mapping of ICHI to CCAM Basic Coding Sysgtem. Workshop "Foundations of Clinical Terminologies and Classifications:. Timisoara, 2006.

Address for correspondence

Anne-Marie Rassinoux, PhD
E-mail: anne-marie.rassinoux@sim.hcuge.ch

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Biomedical Knowledge Discovery with Topological Constraints Modeling in Bayesian Networks: A Preliminary Report

Guoliang Li, Tze-Yun Leong

School of Computing, National University of Singapore*, Singapore*

Abstract

Serving as exploratory data analysis tools, Bayesian networks (BNs) can be automatically learned from data to compactly model direct dependency relationships among the variables in a domain. A major challenge in BN learning is to effectively represent and incorporate domain knowledge in the learning process to improve its efficiency and accuracy. In this paper, we examine two types of domain knowledge representation in BNs: matrix and rule. We develop a set of consistency checking mechanisms for the representations and describe their applications in BN learning. Empirical results from the canonical Asia network example show that topological constraints, especially those imposed on the undirected links in the corresponding completed partially directed acyclic graph (CPDAG) of the learned BN, are particularly useful. Preliminary experiments on a real-life coronary artery disease dataset show that both efficiency and accuracy can be improved with the proposed methodology. The bootstrap approach adopted in the BN learning process with topological constraints also highlights the set of the learned links with high significance, which can in turn prompt further exploration of the actual relationships involved.

Keywords:

Bayesian networks, domain knowledge, bootstrap approach, coronary artery disease

Introduction

Serving as exploratory data analysis tools, Bayesian networks (BNs) support modeling direct dependency relationships among various domain variables, which may lead to understanding or discovering the causal knowledge involved. In the last two decades, many promising methods have been proposed to learn BNs from data (1, 2). One of the main problems with the existing methods is that some links in the learned BNs are inconsistent with the domain knowledge.

Many authors have tried to combine domain knowledge in the BN learning process. Cooper and Herskovits required the complete causal ordering of the variables in the domain (1). Heckerman, Geiger and Chickering used a prior network for BN structure learning (2). These methods work well theoretically when the required systematic domain knowledge is available. In practice, however, only partial domain knowledge is available in most cases. For instance, certain variables are known as roots or leaves, or there are known links among some variables. BN learning with such partial domain knowledge is supported in packages like LibB, TETRAD and BN PowerConstructor[1].

Experience has shown that partial domain knowledge can be very helpful in improving both the efficiency and accuracy of the learned BN structures. However, systematic analyses and discussions on incorporating partial domain knowledge into BN structure learning are lacking; explicit effects and influences of different kinds of topological constraints are unknown, especially in biomedical domains. In this paper, we examine the representation of partial domain knowledge, its usage in BN structure learning and its effect on the accuracy of the learned BN structures. We base our discussions on both a canonical BN example and a case study on a real heart disease data. We hope our discussion on and examination of a set of relevant topological constraints can lead to active causal pattern discovery with experiments.

We propose two ways to represent partial domain knowledge as topological constraints in BNs: in rule format and in matrix format. The rules for partial domain knowledge are easier to derive from the domain experts. These rules can then be compiled into a matrix format, which would facilitate consistency checking in the topological constraints.

We combine the specified domain knowledge with the score-based greedy search method for BN structure learning. Experiments on the canonical Asia network example show that the topological constraints can increase the accuracy of the structure of the learned BNs, especially for the constraint sets with the undirected links in the corresponding completed partially directed acyclic graph (CPDAG) of the BN.

Experiments on a case study in coronary artery disease (CAD) show that both efficiency and accuracy of BN learning can be improved with the proposed methodology. The significance of the identified direct dependency relationship among variables is estimated with the bootstrap approach. The direct links in the learned BN with topolog-

1 LibB: http://www.cs.huji.ac.il/labs/compbio/LibB
TETRAD: http://www.phil.cmu.edu/projects/tetrad
BN PowerConstructor: http://www.cs.ualberta.ca/~jcheng/bnsoft.htm

ical constraints are statistically significant, which can in turn be used as new hypothesis for further analysis.

Materials and methods

Bayesian networks

Bayesian networks (BNs) are directed acyclic graphs (DAGs). The nodes in a BN represent the variables in the target domain. The links between nodes represent probabilistic dependencies among different variables. The node at the head of a link is called a *child*, and the node at the tail of a link is called a *parent*. Together, the graphical representation models the qualitative information in a domain – which domain variables are probabilistically dependent to each other. The quantitative information of a BN is represented as the conditional probability distribution inside each node, which represents the probability of the variable under each configuration of its parent variables. Combining the qualitative and quantitative parts, a BN can define a factorized joint probability of variables $X_1, \ldots, X_n$ in the domain:

$$p(X_1,...,X_n) = \prod_i p(X_i \mid Pa(X_i))$$

where $Pa(X_i)$ means the parents of variable X_i in the BN.

Bayesian network construction and learning

Bayesian networks can be constructed entirely from domain knowledge elicited from domain experts (3, 4). However, in many domains, only partial domain knowledge is available, which is not enough to support full BN construction. Even if there is sufficient domain knowledge, the elicitation process is usually time-consuming, especially for eliciting the numerous conditional probabilities. There are many recent research efforts to build BNs from data (see (5) and references therein). Although there are good applications in BN learning, the learning process remains computationally costly. It is also common in practice to observe that some links in the learned BNs contradict the domain knowledge. For example, in the medical domain, the variable "Age" is usually believed as one not affected by other variables. It should be a root node in a BN, but may not be so in a BN learned without taking into account of this "common sense" domain knowledge (see Figure 2 in our experiment). In order to address this problem in general, we need to properly capture the available domain knowledge in the BN learning process.

Representing domain knowledge as topological constraints in Bayesian networks

In any domain, we may have certain prior knowledge, which could come from scientific laws, common sense, expert opinions, accumulated personal experiences, etc (6). Such domain knowledge can be combined in the BN learning process to establish better structures (1, 2). Usually such domain knowledge is qualitative, but the extent of the influence between two related variables is not certain. A class of useful and important domain relationships are related to the topological structures of the BNs, and can be represented as topological constraints in the BNs. Table 1 summarizes the known topological constraints in rule format, some of which have been applied in BN learning package TETRAD, LibB, and BN PowerConstructor. These rules are easy to elicit from domain experts. However, if there are conflicts and/or cycles in the elicited domain knowledge, it is difficult to detect them in the rule representation.

To facilitate consistency checking, we propose to compile the rules of topological domain knowledge into matrix format: one matrix for the known links, one matrix for the forbidden links, and one matrix for the partial ordering, one vector for the maximum number of parents, and one vector for the maximum number of children, as summarized in Table 2. The first ten rules in Table 1 suggest the known links, forbidden links and partial orderings in Table 2. The last two suggest the limits on the number of parents and children.

Table 1 - Summary of the domain knowledge in rule format

Type of domian knowledge	Meaning
Roots	Nodes without parents. Such variables influence other variables, but are not influenced by any other variable
Leaves	Nodes without children. Such variables are influenced by other variables, but do not affect other variables.
Known links	The links are fixed before learning.
Forbidden links	Definitely there are no such links.
Partial ordering	Some variables are before some other variables in the ordering.
(Conditional) independence	Some variables are conditional independent.
Known parents	The parents of some variables are known.
Known children	The children of some variables are known.
Possible parents	The parents of some variables are restricted to a subset of variables.
Possible children	The children of some variables are restricted to a subset of variables.
Maximum number of parents	The numbers of parents of variables can be different and limited.
Maximum number of children	The numbers of children of variables can be different and limited.

Table 2 - Summary of domain knowledge in matrix format

Names of components	Meaning
Matrix_k	Matrix for knowledge links
Matrix_f	Matrix for forbidden links
Matrix_p	Matrix for partial ordering
V_maxParents	Vector for the maximum parents
V_maxChildren	Vector for the maximum children

Checking the consistency of the topological constraints

After the compilation, we need to check the consistency in the specified domain knowledge before we apply it in the BN structure learning process. The following are the five main steps in consistency checking.

1) Check whether there are cycles in the known links. There should be no cycles in the valid BN topology.

2) Check whether there are overlapping links in the known links and the forbidden links. There should be no overlapping in valid BN topology.

3) Check whether there are cycles when combining the known links and the partial orderings. If there are cycles, it means that certain paths of the known links conflict the known partial orderings.

4) Check the number of the maximum parents of each variable. The sums of the parents of the known links to each node represent the number of the known parents.

5) Check the number of the maximum children of each variable. The sums of the children of the known links from each node represent the number of the known children.

If there are inconsistencies in the topological domain knowledge, report them. Currently we do not resolve the inconsistencies computationally, and leave the work to the domain experts or further experiments.

Greedy search algorithm with topological constraints

In this work, we adopt a score-based approach with a greedy search method and the Bayesian Information Criterion (BIC) score to learn the BNs. The learning process starts from a randomly-generated BN, and continues moving to the best neighbors in the structure space until there is no improvement in the BIC score.

In the learning process, we check whether the candidate BN structures are satisfied with the topological domain knowledge. If yes, we will keep the candidates; we reject them otherwise. This helps to narrow the structure space and speed up the learning process. Figure 1 shows the pseudo code of a greedy search algorithm for BN structure learning.

Compared with the general greedy search, there are two main changes in the process. First, the initial DAG generated should be consistent with the domain knowledge. Second, an additional step filters out neighbors of the current DAG which are not consistent with the topological constraints. This step is to guarantee that the selected neighbors are consistent with the topological domain knowledge. The process of the filtering is similar to domain knowledge consistency checking.

Generate an initial DAG consistent with the topological constraints as the current DAG
Done = false
While ~Done
Generate all the possible neighbors of the current DAG
Filter out the neighbors which are not consistent with the topological constraints
Evaluate the remaining neighbor DAGs
If the best score of the remaining neighbor DAGs is better than that of the current DAG
Set the DAG with the best score as the current DAG
Else
Done = true

Figure 1 - BN learning with topological domain knowledge

Estimation of edge significance with bootstrap approach

We are also interested in the confidence level of the learned links: do these links appear in the learned BNs by chance or with statistical significance? We adopt the bootstrap approach to examine the confidence of the links in the learned model. A bootstrap approach is a statistical method to re-use the original samples. It draws the same number of samples from the original samples with replacement as a new sample set, and builds a model with the same method to analyze the original data. Such experiments repeat many times (500 in our experiments). The results from multiple experiments show the significance of the conclusions from the original samples alone. In our work, the confidence level of the learned links is the percentage of the links appeared in the repeated experiments. If a link appears with a high percentage in the bootstrap experiments, it means that the strength of the direct dependency relationship between two variables is not spurious or accidental but actually indicating strong correlations. Friedman *et al* (7) first applied the bootstrap approach to assess the significance of the edges in the learned BNs.

Results

In this section, we describe experiments on testing the effect of the topological constrains on the BN structure learning process. Since we filter out the DAGs inconsistent with the topological constraints, the learned BN will be consistent with the domain knowledge. Another effect is the change in running time when topological constraints are included. The third effect is the correctness of the learned BN structure

with different topological constraints, when compared with the original structure. Our experiments are based on a benchmark BN Asia network. We have also applied our method on a real heart disease data set without known BN structure. All the implementations are under Matlab with the support of Bayes Net Toolbox (BNT)[2]. The learned BNs are visualized with Genie[3].

Time for consistency checking

We have implemented the consistency checking in both rule format and matrix format. For comparison purpose, we randomly generated 100 topological constraint sets for BNs with different numbers of nodes from 10 to 100. The time for consistency checking in matrix format includes the time to compile the topological constraints from rule format to matrix format. On average, the consistency checking in matrix format is about 8.3 times faster than that in rule format. A possible reason is that, in the matrix format, consistency checking is done by matrix operation; but in the rule format, we need to enumerate all the possible situations for consistency checking, which would take more time.

Effects of topological constraints on Asia network

The canonical Asia network example(8) is an eight-variable BN that models the situation for determining the likelihood of a person having a disease, given his visiting history and smoking habit. In the first experiment, we applied the learning algorithm on the randomly sampled data sets from Asia Network. 1000 cases were sampled for each randomly-sampled data set, which were combined with each individual constraint set to learn a BN. There are 61 different constraint sets in this experiment, each with a single constraint (1 without constraints, 2 with one root, 2 with one leaf, and 56 with 1 links, known or forbidden). 100 different data sets were randomly-sampled. The program ran 36 hours and 6100 (=100*61) BNs were learned in total.

The results showed that the single constraint, which led to the maximum number of average correct links in the learned BNs, is a link from node "visit to Asia" to node "Tuberculosis". This link is distribution-indistinguishable. Establishing this link can lead to more correct links in the learned structure. The second most useful constraint is a link from node "lung cancer" to node "Tuberculosis or Lung Cancer". This constraint is a link in a v-structure. The third most useful constraint is a leaf node "Dyspnea". These results suggested that we need to pay more attention to certain types of topological constraints in practice.

We also experimented with multiple constraints in one constraint set. We randomly selected links in the original BN as known links or forbidden links in the constraint sets with one to seven links selected. There were 43 constraint sets in total. There were 1000 randomly-sampled cases in each data set, and 93 different data sets were sampled. Each data set was combined with every individual constraint set to learn a BN. Altogether 3999 (=93*43) BNs were learned for multiple constraints, and 453 of them were found to be the same as the original one. Combined with the results from the single constraints, where almost no learned BNs were exactly the same as the original ones, this means that more topological constraints are more likely to lead to the correct BNs. This coincides with our belief that the more the links we know, the easier the remaining links can be learned.

Heart disease data

We performed a case study on a data set collected from a Singapore hospital for coronary artery disease (CAD) study. It contains 2,949 cases of human subjects. 1,462 of the subjects were diagnosed to have coronary artery disease at the time of data collection; and the rest were healthy at the time of recruitment. The values of CAD in this work were based on the presence of at least 50% narrowing in at least one of the major coronary arteries by angiography. In the data set, aside from CAD, another ten patient profile factors were selected for our experiments. They are the subject's 1) age, 2) sex, 3) race, 4) current body mass index (CBMI), 5) smoking habit, 6) absent of hypertension or not, 7) absent diabetics or not, 8) family CAD history (FCAD), 9) family hypertension history (FHY), and 10) family diabetic history (FDM). In these ten variables, eight factors are discrete variables and two other variables, Age and CBMI, are continuous variables. These two continuous variables were discretized separately. Refer to (9) for more detailed data summary.

BNs learned with and without topological domain knowledge

The BNs learned without and with topological domain knowledge are shown in Figure 2 and 3. In Figure 2, the variables "Age", "Race" and "Sex" have parents, which is conflicting with common sense as these variables are not affected by other variables in this case. This motivated us to combine topological domain knowledge in the BN learning process.

In our work, we applied two types of topological domain knowledge. First, variables "Race", "Age" and "Sex" are specified as roots in BNs, since we know from common sense that the probability distributions of these three variables are not dependent on other patient profile information. Second, variables "FHY", "FDM" and "FCAD" precede all other variables in partial ordering, since family health history precedes the patient profile in time. However, these three family health history variables may depend on each other. The BN learned with such topological domain knowledge is shown in Figure 3. As compared to the BN in Figure 2, the BN structure in Figure 3 is more meaningful as judged by commonsense.

2 BNT: http://www.cs.ubc.ca/~murphyk/Software/BNT/bnt.html

3 Genie: http://genie.sis.pitt.edu/

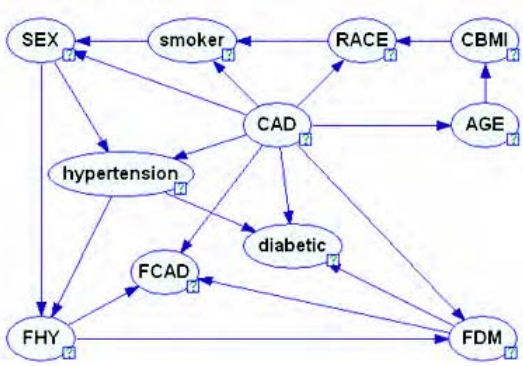


Figure 2 - BN learned without domain knowledge

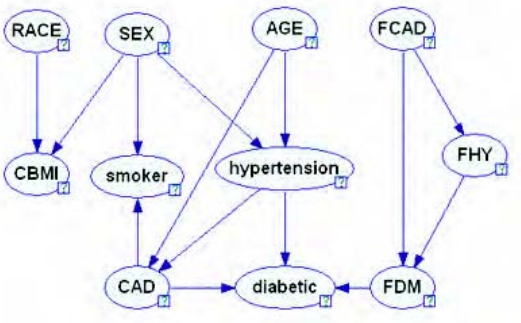


Figure 3 - BN learned with domain knowledge

Effects of domain knowledge on heart disease data

To verify the significance of the links in the learned BNs, we applied the bootstrap approach. We sampled data from the original data with replacement, and two BNs were learned from each sampled data set – one without topological domain knowledge and one with such knowledge.

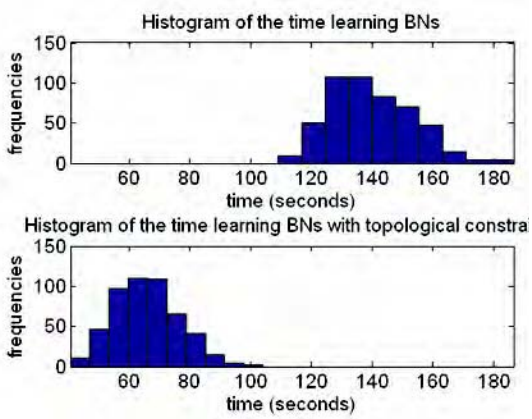


Figure 4 - Histograms of the times learning BNs with/ without domain knowledge

Figure 4 shows the histograms of the running time of the BN learning processes with and without topological constraints. As indicated, the running times with topological constraints are much shorter than those without constraints. The average running time with topological constraints is 65.9 seconds. Compared with the average running time without topological constraints (140.1 seconds), the speed-up is more than two times. However, we notice that the speed-up of the learning process is dependent on the available domain knowledge. In our example, there are 11 variables and 110 possible links (two directions for 55 pairs of variables). The available domain knowledge in our example is that three variables are roots and three variables are not affected by the other 8 variables. Based on such constraints, 54 links were forbidden. The BN structure space with the constraints was much smaller, which led to the speed-up in the BN structure learning process. If the known domain knowledge is small compared to the whole possible topological constraints, the speed-up may not be so significant.

Table 3 shows the significant pairs of variables in the learned BNs with the bootstrap approach. The results showed that almost all the links in the learned BNs were quite significant and appeared more than 60% of times in the 500 bootstrap experiments. The top pair of variables in the learned BNs with bootstrap approach and domain knowledge is Sex-Smoker, which appeared surprisingly 100% in the 500 bootstrap experiments. This pair of variables is deemed to be related based on the current domain understanding. Other evaluation methods such as chi-square and mutual information, however, did not rank it highly (data not shown here for space limit). Other highly ranked pairs of variables in Table 3 appeared in the learned BN with topological domain knowledge (Figure 3). It means that the links in the learned BN with domain knowledge are statistically significant.

Table 3 - Top occurring links in BN learning with bootstrap and topological constraints

Order	Variable 1	Variable 2	Occurrences (%)
1	SEX	smoker	500 (100%)
2	AGE	CAD	482 (96.4%)
3	CAD	smoker	469 (93.8%)
4	CAD	diabetic	453 (90.6%)
5	CAD	hypertensi on	408 (81.6%)

Note: The percentage in the Occurrences column means the percent of the links appearing in the 500 learned BNs from bootstrap approach.

Discussion and conclusion

Identifying the possible direct dependency relationships among variables is an important task in exploratory data analysis. BNs are a good tool for this purpose. There are

several advantages in BN-based data analysis: 1) easy to combine prior knowledge, 2) easy to interpret and 3) easy to deal with missing and noise data. In this paper, we have proposed two canonical formats to represent domain relationships as topological constraints in the BN learning process. In addition to improving the efficiency of the BN learning process, the experiments show the links in the learned BNs with domain knowledge are potentially more meaningful. Results from bootstrap approach show that the learned links are statistically significant. Our case study in CAD risk analysis with the learned BN allowed us to discover some interesting patterns, which have prompted further analysis.

We want to emphasize that the purpose of the analysis with BNs and domain knowledge is to identify the properties in the data with high probabilities and consistent with common sense, which can be used as new hypotheses for further analysis. The purpose of using BN and topological domain knowledge is not to replace the other traditional data analysis methods for hypothesis testing.

Our on-going work is to extend the proposed method based on more extensive experiments with more data. We have collected the genotype single nucleotide polymorphism (SNP) data for the subjects in the heart disease data set. We will incorporate the genotype data with the patient profile for further analysis. For future work, we also plan to verify patterns found by our method with new biological or clinical experiments.

Acknowledgments

We thank Dr Chew Kiat HENG for sharing with us the Heart Disease data set in this work.

References

[1] Cooper GF, Herskovits E. A Bayesian method for the induction of probabilistic networks from data. Machine Learning 1992;9:309-347.

[2] Heckerman D, Geiger D, Chickering DM. Learning Bayesian networks: The combination of knowledge and statistical data. Machine Learning 1995;20:197-243.

[3] Nadkarni S, Shenoy PP. A Causal Mapping Approach to Constructing Bayesian Networks. Decision Support Systems 2004;38(2):259-281.

[4] Druzdzel MJ, van der Gaag LC. Building Probabilistic Networks: Where Do the Numbers Come From? Guest Editors' Introduction. IEEE Transactions on Knowledge and Data Engineering 2000;12(4):481-486.

[5] Heckerman D. A Tutorial on Learning with Bayesian Networks. In: Jordan M, editor. Learning in Graphical Models. Cambridge, MA: MIT Press; 1998. p. 301-354.

[6] Cooper GF. An overview of the representation and discovery of causal relationships using Bayesian networks. In: Glymour C, Cooper GF, editors. Computation, Causation, and Discovery: AAAI Press and MIT Press; 1999. p. 3-62.

[7] Friedman N, Goldszmidt M, Wyner A. Data Analysis with Bayesian Networks: A Bootstrap Approach. In: Proceeding of the Fifteenth Conference on Uncertainty in Artificial Intelligence (UAI); 1999. p. 206-215.

[8] Lauritzen SL, Spiegelhalter D. Local computations with prababilities on graphical structures and their application to expert systems. Journal of the Royal Statistical Society, Series B 1988;50(2):157-224.

[9] Chen Q, Li G, Han B, Heng CK, Leong T-Y. Coronary Artery Disease Prediction with Bayesian Networks and Constraint Elicitation, Technical report TRC9/06, School of Computing, National University of Singapore, September 2006.

Address for correspondence

Guoliang Li
School of Computing, National University of Singapore
3 Science Drive 2, Singapore, 117543
Email: ligl@comp.nus.edu.sg

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Automatic Infection Detection System

Ove Granberg[b], Johan Gustav Bellika[ab], Eirik Årsand[a], and Gunnar Hartvigsen[ab]

[a] *Norwegian Centre for Telemedicine, University Hospital of North Norway, Tromsø, Norway*
[b] *Department of Computer Science, Faculty of Science, University of Tromsø, Norway*

Abstract

An infected person may be contagious already before the first symptoms appear. This person can, in the period of disease evolution, infect several associated citizens before consulting a general practitioner (GP). Early detection of contagion is therefore important to prevent spreading of diseases. The Automatic Infection Detection (AID) System faces this problem through investigating the hypothesis that the blood glucose (BG) level increases when a person is infected. The first objective of the prototyped version of the AID system was to identify possible BG elevations in the incubation time that could be related to the spread of infectious diseases. To do this, we monitored two groups of people, with and without diabetes mellitus. The AID system analyzed the results and we were able to detect two cases of infection during the study period. The time of detection occurred simultaneous or near the time of onset of symptoms. The detection did not occur earlier for a number of reasons. The most likely one is that the evolution process of an infectious disease is both complicated and involves the immune system and several organs in the body. The investigation with regard to isolating the key relations is therefore considered as a very complex study. Nevertheless, the AID system managed to detect the infection much earlier than what is possible with today's early warning systems for infectious diseases.

Keywords:

health intelligence, automatic infection detection, disease surveillance

Introduction

Disease outbreaks like SARS in March 2003 and the deliberate spread of *B.anthracis* in the US in 2001 have drawn attention to the need for tools to detect and monitor disease outbreaks. In august 2005, WHO provided the following assessment of the threat of an avian influenza pandemic [1]:

1. The risk of a pandemic is great.
2. The risk will persist.
3. Evolution of the threat cannot be predicted.
4. The early warning system is weak.
5. Preventive intervention is possible, but untested.
6. Reduction of morbidity and mortality during a pandemic will be impeded by inadequate medical supplies.

Early detection of contagion is important to prevent spreading of diseases. By the time the health authorities is able to identify a disease outbreak by reports, it may be too late to prevent large epidemic or pandemic outbreaks. It is therefore important that the authority receives timely and adequate indication of ongoing infectious disease activity. An important goal is therefore to reduce the time from infection, to the time when the authorities get information about infectious disease activity as much as possible.

An infection is in traditional medicine first established in accordance with a diagnosis on the day of consultation. This leaves an unexplored time gap back to the time of intrusion. We have therefore explored new approaches in order to reveal diseases at an earlier stage, e.g. syndromic surveillance. In the future management of chronic diseases, transfer of vital sensor data from patients to the public health care system is likely to become commonplace. Systems for automatic transfer of sensor data are now at the prototype stage [2], and that Electronic Health Record (EHR) systems will include improved functionality to capture sensor data is also likely.

This has aimed our curiosity in direction of how to detect infections outside clinical settings. One way of facing this goal is to use self monitoring equipment to measure indicator data. The starting point of this project has therefore been to exploit secondary use of blood glucose (BG) measurements performed daily by diabetics. These measurements are then analyzed to see if they can be used to reveal and detect infections.

We believe that it may be possible to detect an infection close to the time of contagion (pathogenic intrusion), and escalating trough the incubation time. The project was therefore designed to investigate if BG measurements are suitable as an infection indicator in public health care surveillance. This leads to the following problem definition: *Is it likely that a person's BG- level, given an infection (independent of type), would change sufficiently and early enough that monitoring of these changes can be used as a reliable indicator of an infectious disease?*

Materials and methods

This paper describes the development of the prototyped AID system for detecting infectious diseases, discusses the results from the captured clinical BG measurement data

and evaluates the experiment. The work is based on an experimental approach. An important part of the work has been input from a clinical study of nine participants, of this; two people with and seven people without diabetes mellitus. The group monitored their BG level over a period of up to 6 weeks. The results from the clinical study were used for testing of the AID system prototype. Each participant was instructed to perform measurement according to a time definition before or after a meal. One part of the group was instructed to perform measurements before the main meals and up to four times a day. The other group was instructed to measure within a time interval after each main meal, also up to four times a day.

The measured data is provided by use of a blood glucose meter; FreeStyle Mini from Abbott. The data is transferred to a Microsoft SQL Server 2005 database for EHR simulation purposes. The captured sensor data was then analyzed in a two-step detection algorithm. The first step is a moving average algorithm that main objective is to smooth out large fluctuations, especially with regard to BG measurement. Then, both the original BG measurement and the moving average data are analyzed in the detection algorithm against the clinical defined Cut-Off value. The number of measurements per participant in the study ranges from 44 to above 120 measurements. Those with diabetes provided a larger number of measurements, and from a longer period. These measurements permitted us to test run and tune the detection algorithm against BG cases with real-time quantity of fluctuations. The corresponding feedback from the participants made it possible to tune the algorithm to make detections of incidents of real infections.

The number of participant in this experiment is too low to fulfill the statistical power demands. The experiment provide useful information in order to indicate trends and reveal important signals that could be followed up in future work. However, the reader needs to be aware that the results from the experiment are preliminary.

AID system background

Current medical knowledge states that the BG values increases during an infection [3]. Based on this knowledge, we wanted to investigate whether BG measurements were suited as an indicator of infections and able to be used to predict and detect a disease outbreak at an early stage. Continuous analyses of daily blood glucose measurements may enable early detection of infections on a person level. On a population level, this may in turn enable us to prevent large outbreaks of infectious diseases, such as different strains of influenza, cholera, plague, smallpox and anthrax.

When the analyzed data indicating infections exceed a predetermined threshold of incidents, the health care authorities would be able to initiate early investigation at an early stage in order to begin treatment, prevent the disease to spread, and to prepare for vaccination etc. We have not identified other studies that use BG measurements for infectious disease surveillance purposes. There are few studies [4] linking use of point of care testing or near patient testing (NPT) [5], to outcomes [6]. Some studies have been performed on inpatients, (both diabetics and non-diabetics) with particular focus on glycemic control and its possible impact on hospital outcomes. The result shows that patients with hyperglycemia have had an increased risk of both postoperative infections and in hospital mortality [7]. The C-reactive protein (CRP) is the most commonly used marker to establish a diagnosis of infectious diseases. CRP is normally used as an infection parameter or inflammatory marker primarily to establish bacterial infections, but to a certain extent also as a marker of viral infections.

The University of Tromsø and the Norwegian Centre for Telemedicine have previous collaborated on projects investigating the use of BG measurements in different medical settings [2]. This co-operation evolved gradually into this new idea and the desire to investigate the usefulness of BG measurements, performed several times a day by diabetic patients, in order to investigate a secondary utility value. Systems that support rapid and automatic transfer of NPT outcomes are evolving. Data from NPT equipment that is used in hospital settings and integrated in practice and bedside treatment of patients is likely seldom or never transferred and reused outside the treatment of the current/local patient. Systems that automatically transmit results of NPT information to EHRs are also evolving. This development creates a need for systems able to interpret the data.

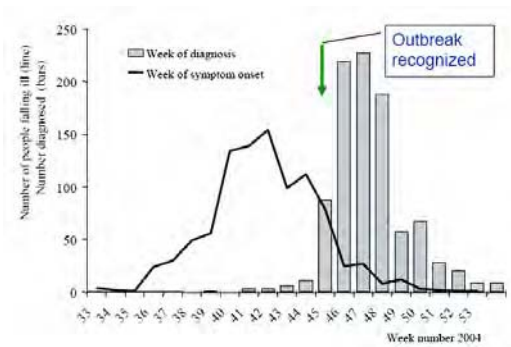


Figure 1 - Characteristic of Giardiasis disease propagation in Bergen 2004 (Graphics by Chief Physician P. Aavitsland, The Norwegian Institute for Public Health)

Figure 1 shows the Giardiasis outbreak in Bergen in 2004 where the health authorities first became aware of the scope of the disease at a time when the disease outbreak was decreasing.

Our intention is to capture self-measurements from diabetes NPT settings (that primarily is used to provide the patient with information to take necessary self-treatment actions), and to exploit and make secondary use of these measurements in disease and public health surveillance. A synergistic effect of real time analyzed NPT outcome would provide general practitioners, outpatient clinics, and public health authorities with real-time information, which may change clinical practice, and improve the ability to

quickly determine whether to take proper actions. Normal intervention in the capture of NPT-results has previously been insufficiently recorded in EHRs. Our intended use omits this previous potential loss [8] of valuable measurements which may improve health surveillance efficiency. We believe that our approach may provide solutions to some of the challenges that exist in public health. It expands the methods for and potentially the ability to detect infectious diseases and to take timely actions that could limit the scope of disease propagation.

Additionally, occurrences of bioterrorism [9] has also increased the need and focus on anti-bioterrorism and precautions towards defeating malicious use of biological warfare agents. Infectious diseases due to deliberate terror actions are no longer a question of if, but of when and where [9]. We should, when faced with such a threat, even if it at best may be of limited effect, immediately start to optimize our safety and increase preparedness.

Results and discussion

At the Norwegian Centre for Telemedicine in co-operation with the University of Tromsø, we have constructed an AID - Automatic Infection Detection system, which is a prototype system for detection of infection through monitoring blood glucose values collected from people with diabetes using a communication enabled blood glucose meter. The blood glucose meter may communicate wirelessly with a mobile phone that transfer the BG measurements to a remote system where the analysis may be performed, as described in [2].

The basis for the proposed detection system is the hypothesis that the BG level elevates when a person gets an infection [10]. The International Diabetes Federation estimates that there are currently some 194 million people around the world with diabetes [11]. Monitoring and control of blood glucose levels is critical in the management of Type 1 and Type 2 diabetes, and the patients measure their blood glucose levels up to several times a day to minimize long-term complications [12, 13]. If values from all diabetes patients in a region were instantly accessible, it would provide a foundation for extended data analysis. Extreme accumulations of infection indications in a specific geographical area could be spotted, and the public health authorities could monitor and act on such signals.

We were able to detect two cases of real infections from data provided by the nine participatns in the investigation. The detection algorithm detected all cases above the predefined Cut Off value. The time of detection occurred in parallel with the patient symptoms, i.e. the onset date. Both cases of infection detections were made from the diabetic's participant measurements. As a result of feedback from the patients we were able to check and verify that our AID system was able to detect infections even from more complex BG fluctuations. The ability of detecting an infection in a person with adequate insulin regulation is assumed to be less complicated than for patients that are not well regulated.

Figure 2 illustrates the outcome of an analysis on population level. The left table summarizes participating individuals / sensors (column 3), the corresponding number of total measurements and indications from participants in the zip code areas 9011, 9037 and 9100 on the December 27th 2005. This provides information about indications of infections. The number of indications is in this case negligible. However, we believe, given the occurrence of a serious outbreak, that the number of indications of infections would increase dramatically. When a zip code area holds an adequate number of reporting sensors representative to the population of that specific area, and when the number of indications exceeds a given threshold, this information may be a signal for the health authorities to take proper action.

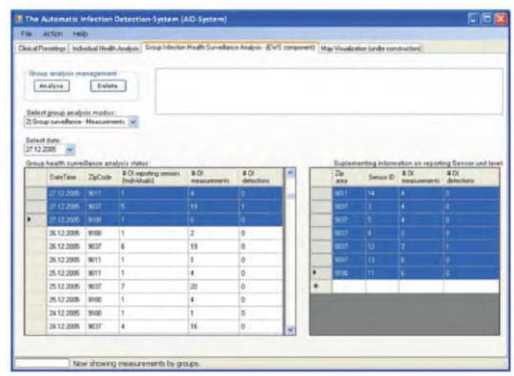


Figure 2 - Group infection health surveillance analysis

The right side shows each sensor device's recorded measurements and indications. This information provides the authorities with additional information to support decision. E.g. if a known diabetes patient that normally performs an average of 6-8 measurements each day suddenly measure a noticeable number of measurements (e.g.12 times) with a corresponding high level of detections this would support the suspicion of an infection

In our study we did not see any significant changes in the period before onset. It is advantageous to perform new experiments, and/or collaborate with hospitals in collecting continuous blood glucose values, e.g., from NPT settings. The experience and empiricism in use of BG as an infection indicator is now in the very beginning. Our preliminary results hopefully will encourage others to follow up with studies, experiments and evaluation on the topic.

Our algorithm also detected six false positives. The false detections occurred due to unregulated diabetic conditions. The false positive rate we experienced is a challenge that could potentially reduce the value of our detection system. However, because the system is based on using a mobile phone as a gateway, the mobile phone may also be used to collect additional data about the patient. We therefore recommend that the diabetic patient is sent a SMS with a question like "Are you feeling well?" or something similar, to get additional data from the patient. The diabetic patient should then provide a "yes" or "no" answer, using

SMS. This communication could form the basis for a system that can aid the diabetic patient both when the blood glucose values are too high, like when the patient has an infection, or too low, like when the patient experiences hypoglycemia. Such use of the mobile phone could strengthen the evidence for a real infection as well as aiding the patient in potentially very dangerous situations.

What may also strengthen prediction performance is including additional infectious indicators such as temperature, blood pressure, heart beat, pedometer /accelerometer or the outcome of measurements from bioinformatics equipment e.g. CRP etc. Additional registration of the patient's physical activity integrated with the patient's mobile phone could also contribute to a better prediction. E.g. if a person BG-value and temperature rise at the same time as activity drops it could support the suspicion of illness and even an infection. A confounding element is whether highly affected patients would measure their BG value or answer SMS messages. This happened to one of the non diabetic patients that became ill during the study but was unable to measure the BG value during the disease episode. Other confounding elements may be the occurrence of a simultaneous outbreak of less dangerous diseases as normal influenza.

However, we believe that data from patients with diabetes could provide early warnings of an outbreak in a population. This would give the public health authorities the opportunity to take action to limit an outbreak and its consequences for all the inhabitants in the affected area. Typical actions may be advice, vaccination and even isolation in serious cases.

Using blood glucose as an indicator of infection

There is a differece between inflammatory markers and infection markers (also called indicators). Inflammatory markers cover a more broad area including physical injuries, while indicators imply that the infection is solely caused by a pathogenic agent. Infection indicators are measurable parameters able to give certain health information of a specific condition present during an infection. One such indicator may be blood glucose. There might be a connection between different types of infectious diseases and which impact each specific disease causes in elevation of BG, and further research is needed to see if the investigation of if elevated BG condition is similar among different infectious diseases.

Type 1 diabetes patients measure their BG level several times a day in order to regulate their BG, while diabetes type 2 patients have a lower frequency in order to maintain well regulated. BG measurements is performed several times to adjust injection of insulin [14]. Different types of measurement equipment exists ranging from devices which only require a small droplet of blood (e.g. FreeStyle Mini) to continuous glucose monitoring (CGM) by subcutaneous test devices e.g. FreeStyle Navigator [15] and non-invasive devices as the GlucoWatch [16] provided by Animas Technologies, LLC.

The AID system made one detection of a confirmed real case of infection for the individual using sensor unit 13 (see. Figure 3). Vaccination of the individual using sensor unit 13 took place at the 21.Oct. 2005.

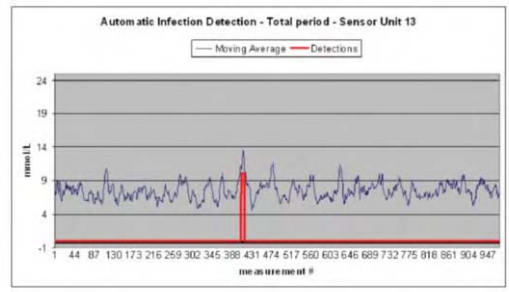


Figure 3 - Analyzed data from Sensor Unit 13 (diabetes)

Figure 3 shows The Automatic Infection Detection system for analysis of patients' sensor data. The sensor data consists of physiological parameters, in this case monitored blood glucose data. The red line detection peak correspond with the clinical defined cut-off threshold value: 10,2 mmol/l.

The system analyses and evaluates each patients BG data, and potential abnormal BG increases are detected. The evaluation seeks to eliminate random variations. In addition, the values must be higher than a predefined cut-off value before they are indicated as an infection.

Conclusion

Use of BG data holds a potential to act as an indicator of infections in health care surveillance. Use of BG as an indicator of infections seems to be suitable. BG measurement from sensor patients provide a continuous stream of data and the frequency in measurements per day opens up for further development and secondary usage of sensor data from patients and NPT measurements.

Exploiting secondary use of BG measurements to perform public heath surveillance might lower the total social- and medical economic costs of pandemics. Remote monitoring increase the health care efficiency, but maybe most important are the patient's quality of life. Our approach aims to only collect physiological data, which means that personal integrity and ethnical issue also could be safeguarded.

Our approach, seems to provide detection of infection very early.

Our main recommendation, with regard to facing timely detection and impending usage of BG measurements, is to include and correlate measured BG values against other non-invasive health indicators, such as temperature, heart rate, blood pressure and activity instruments (movement sensors, step counters, etc). We also recommend future research to perform investigations with use of non-invasive devices or subcutaneous continuous glucose monitoring (CGM) type devices. A subcutaneous CGM device like Abbott's Freestyle Navigator could provide even more realistic real-time data. Due to the possibility of developing a stronger algorithm this approach may

increase the detection reliability. In this technological area, many different solutions already exist. The challenge seems to be in part to merge all non-invasive sensors into one small wearable unit that wirelessly transfers health data via the mobile phone or internet network access node (e-house technology) into the public health services.

Independently of which way to obtain indicator data, it is important to note that there will always be some disease activity present. Before a surveillance system could make use of BG values or other suitable infection indicator data, data must be collected and be representative of the population. By expanding the number of patients providing sensor data it would be possible to cover each geographic area. Outbreaks can be discovered very early and it is possible to take proper action.

References

[1] World Health Organization. Responding to the avian influenza pandemic threat. 2005 [Accessed October 11th 2005]; Available from: http://www.who.int/csr/resources/publications/influ-enza/WHO_CDS_CSR_GIP_2005_8/en/index.html

[2] Arsand E, Walseth OA, Andersson N, Fernando R, Granberg O, Bellika JG, et al. Using blood glucose data as an indicator for epidemic disease outbreaks. Stud Health Technol Inform. 2005;116:217-22.

[3] Chase H. Understanding Diabertes - Long term Complication of Diabetes. 2002 [Accessed [2005, July 25]]; Available from: http://www.uchsc.edu/misc/diabetes/udchap22.html

[4] Hobbs FDR, Thorpe GH, Delaney BC, Earl-Slater S, Fitzmaurice DA, Howett S, et al. A review of near patient testing in primary care. Health Technology Assessment. 1997;1(5):1-244.

[5] Coiera E. Guide to Health Informatics. 2 ed: Arnold Publishers; 2003.

[6] Price CP. Point of care testing. Bmj. 2001 May 26;322(7297):1285-8.

[7] Clement S, Braithwaite SS, Magee MF, Ahmann A, Smith EP, Schafer RG, et al. Management of diabetes and hyperglycemia in hospitals. Diabetes Care. 2004 Feb;27(2):553-91.

[8] Borriello SP. Science, medicine, and the future. Near patient microbiological tests. Bmj. 1999 Jul 31;319(7205):298-301.

[9] Frist B. When every moments counts: what you need to know about bioterrorism from the Senate's only doctor / Bill Frist: Rowman & littlefield publisher, inc.; 2002.

[10] Chase HP. Long-term Complications of Diabetes. In: Harris S., editor. Understanding Diabetes 10th edn. Denver, Colorado: MGM Consumer Products; 2002.

[11] International Diabetes Federation. Diabetes Prevalence. 2003 [Accessed 2005, January 13]; Available from: http://www.idf.org/home/index.cfm?node=264

[12] Prospective Diabetes Study (UKPDS) Group. Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33). Lancet North Am Ed. 1998;352(9131):837-53.

[13] Prospective Diabetes Study (UKPDS) Group U. Effect of intensive blood-glucose control with metformin on complications in overweight patients with type 2 diabetes (UKPDS 34). 1998;352(9131):854-65.

[14] Cohen BJ, Wood DL. The Human Body in Health and Disease. 9th ed. ed: Lippincott Williams & Wilkins; 2000.

[15] HomeTestMed.Com. New Skin Patch Monitors Glucose Levels. 2005 [Accessed [2005, July 4]]; Available from: http://www.hometestmed.com/diabetes_skin_patch.asp

[16] Renard. Implantable glucose sensors for diabetes monitoring. Minim Invasive Ther Allied Technol. 2004 Apr;13(2):78-86.

Address for correspondence

Johan Gustav Bellika,
Norwegian Centre for Telemedicine,
University Hospital of North Norway,
N-9038 Tromsø, Norway
Fax: +47 77 75 40 98;
Email: johan.gustav.bellika@telemed.no)

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Risk Stratification for LDL Cholesterol using Induction Technique

Seung Hee Ho[a], Hyun Young Park[b], Yang Soo Jang[b], Sun Ha Jee[c]

[a] *Department of Health Informatics, Graduate School of Public Health, Yonsei University, Korea*
[b] *Cardiovascular Genome Center, Yonsei University Medical Center, Seoul, Korea*
[c] *Department of Epidemiology and Health Promotion, Graduate School of Public Health, Yonsei University, Korea*

Abstract

We identified the combined patterns of LDL cholesterol risk factors including biometric, environmental and genetic factors using induction technique. In this hospital based cardiovascular genome study of Korean men and women, we found that CART (classification and regression tree) was a better method to predict LDL cholesterol compared to the regression method. The CART had a better prediction ability than the multiple regression for male and female, respectively. We also identified combined patterns of LDL cholesterol risk factors and segment specific information for LDL cholesterol management using induction rules. The CART method provided more detailed results according to each segmentation and subgroup. In addition, we demonstrated how the CART algorithm could be used in risk assessment and target segmentation of LDL cholesterol management.

Keywords:

risk stratification, LDL cholesterol, induction technique, genetic polymorphism

Introduction

Because low-density lipoprotein (LDL) cholesterol is multi-factorial in origin, it is important to consider all risk factors simultaneously, including biometric, environmental, and genetic factors, when estimating an individual's LDL cholesterol risk. The relationship between gene polymorphism and LDL cholesterol has been studied by many groups [1-6]. Most past LDL cholesterol studies have focused on describing individual patient characteristics and genetic polymorphisms separately using statistical methods. Therefore, the relative importance of risk factors could not be compared together and their relative statistical significance could not be established.

With the intention of overcoming limitations of past studies, we attempted to discover the combined patterns of LDL cholesterol risk factors including biometric, environmental and genetic, using an induction technique as a data mining approach. Data mining, which is also known as knowledge discovery in databases, is a process of non-trivial extraction of implicit, previously unknown and potentially useful information from databases [7].

This study presents a classification and regression tree (CART) algorithm as an induction technique to discover the knowledge to predict LDL cholesterol risk by biometric, environmental and genetic factors. We examined the risk factors for LDL cholesterol using multiple regression and the CART algorithm, then analyzed the relationship among biometric, environmental and genetic factors for LDL cholesterol using the CART algorithm. In addition, we compared multiple regression and the CART algorithm results, and presented the practical use of rule induction for the management of LDL cholesterol.

Materials and methods

Research model

We analyzed the simultaneous relationship among biometric, environmental and genetic factors for LDL cholesterol using the CART algorithm (Figure 1).

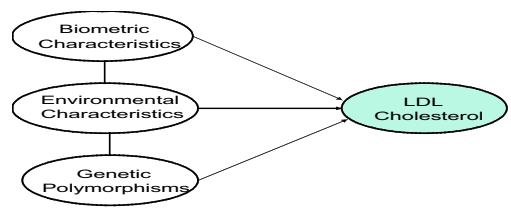


Figure 1- Framework of the analysis

Subjects

This study analyzed data from a hospital-based cardiovascular genome center in Korea. Blood samples were collected from patients from January 2001 to January 2004. In this study, 242 males and 295 females were selected. Subjects presenting with cardiac problems or receiving treatment that would affect serum lipoprotein and carbohydrate metabolism were excluded from the study.

Measurements of risk factors

This study analyzed data from a hospital-based cardiovascular genome center in

- Biometric and Environmental Characteristics
  Body mass index (BMI) was calculated as weight in kilograms divided by height in meters squared. The

concentration of LDL cholesterol was calculated using the Friedewald equation [8]. Environmental parameters including smoking, alcohol consumption and exercise, were obtained via one to one interview type questionnaire.

- Genetic Polymorphisms Testing
  Genomic DNA was isolated from whole blood using a Genomic PUREGENE® DNA Isolation Kit (Gentra Systems, Inc., MN, USA). Genotyping was performed using a polymerase chain reaction (PCR)-based method, and negative controls without DNA template were included for each reaction.

Decision tree induction: CART algorithm

Decision Trees are powerful and popular tools for classification and prediction capabilities. The attractiveness of tree-based methods in contrast to neural networks, is largely based on the concept that decision trees represent rules [9]. The CART algorithm is one of the most popular methods of building decision tree. Since its publication by Briemen [10] and associates in 1984, it has been a staple algorithm of machine-learning experiments. The CART tree is constructed by splitting data into subsets using all predictor variables to create two child nodes repeatedly, beginning with the entire data set. The best predictor is chosen using a variety of measures to reduce impurity or diversity.

Results

The effect of biometric, environmental and genetic factors on LDL cholesterol by multiple regression analysis

The effects of the independent variables to LDL cholesterol were tested (Table 1). In males, the overall model was significant (F=1.781, p = 0.032) and 12.5 percent of the variance of LDL cholesterol level is explained by independent variables (R^2=0.125). Biometric and environmental factors had significant influence on LDL cholesterol in one variable, drinking status. Drinking status was significantly related to LDL cholesterol (beta= -13.635, t-value= -2.169 , p<0.031). Three genetic polymorphisms had significant influence on LDL cholesterol, TAQ1B, I405V, and M235T.

In females, the overall model was significant (F=1.798, p = 0.028) and 10.3 percent of the variance in LDL cholesterol is explained by independent variables (R^2=0.103). Biometric and environmental factors had no significant influence on LDL cholesterol, but two genetic polymorphisms had significant influence on LDL cholesterol, APOE and APOA5(-1131T>C).

Decision tree and rules for the prediction of LDL cholesterol

We used the CART algorithm to build a decision tree for predicting the LDL cholesterol, and identified the risk factors that play important roles in explaining LDL cholesterol for the two groups, male (Figure 2) and female (Figure 3). The decision tree shows the risk factors that were significantly different between the two nodes based on t-test at 5% level.

Table 1 - Determinants of LDL cholesterol by multiple regression analysis

Covariate		Male			Female		
		beta	t-value	P-value	beta	t-value	P-value
Age		-0.299	-1.270	0.205	0.337	1.533	0.126
BMI		1.083	1.215	0.226	0.935	1.367	0.173
Smoking		-4.541	-0.738	0.461	7.674	0.984	0.326
Drinking		-13.635	-2.169	0.031	2.812	0.670	0.503
Exercise		-1.911	-0.388	0.699	-2.405	-0.548	0.584
APOE	E2/E3 E3/E3 E3/E4	 -1.549 -1.314	 -0.188 -0.133	 0.851 0.894	 11.939 22.717	 1.765 2.857	 0.079 0.005
APOA5 (1131T>C)	TT TC CC	 -0.210 -6.959	 -0.042 -0.682	 0.966 0.496	 10.946 17.070	 2.663 2.218	 0.008 0.027
CETP (C-629A)	AA CA CC	 2.940 12.862	 0.376 1.183	 0.707 0.238	 -7.498 0.716	 -1.207 0.089	 0.228 0.929
CETP (TAQ1B)	B1B1 B1B2 B2B2	 -23.954 -10.742	 -2.171 -1.176	 0.031 0.241	 7.722 7.093	 0.909 0.917	 0.364 0.360
CETP (I405V)	II IV VV	 -6.239 -19.038	 -1.022 -2.337	 0.308 0.020	 1.666 2.686	 0.310 0.370	 0.757 0.712
AGT (M235T)	MM MT TT	 32.421 6.612	 2.478 1.256	 0.014 0.211	 -2.938 -0.915	 -0.207 -0.259	 0.795 0.836
Model F statistic		1.781			1.798		
Fitness *P*-value		0.032			0.028		
R square		0.125			0.103		

In males, the mean LDL cholesterol for all subjects was 130.51 mg/dl. The first split was based on age, with younger subjects (54.5 years old) having the higher LDL cholesterol (134.99 mg/dl) than that (119.70 mg/dl) of older subjects (>54.5 years old). This indicated that age was the most important determining factor for LDL cholesterol.

Differences were observed between the two subtrees. Older subjects (>54.5 years old) are subsequently split by drinking status. Younger subjects (54.5 years old) are subsequently split by age again, with tracts having an age between 31.5 and 54.5 years old in one node and an age less than 31.5 years old in the other. Subjects between 31.5 and 54.5 years of age had higher LDL cholesterol (138.04 mg/dl) than (112.04 mg/dl) subjects less than 31.5 years of age.

I: Improvement (reduction of the within-node variance)
M.L: Mean of LDL cholesterol

Decision trees are charts that illustrate decision rules. Decision rules provide specific information about risk factors based on the rule of induction. As shown in Figure 2, the decision tree has 20 leaf nodes and 11 terminal nodes. Figure 3 demonstrates a decision tree having 18 leaf nodes and 10 terminal nodes. Each terminal node depicted in the

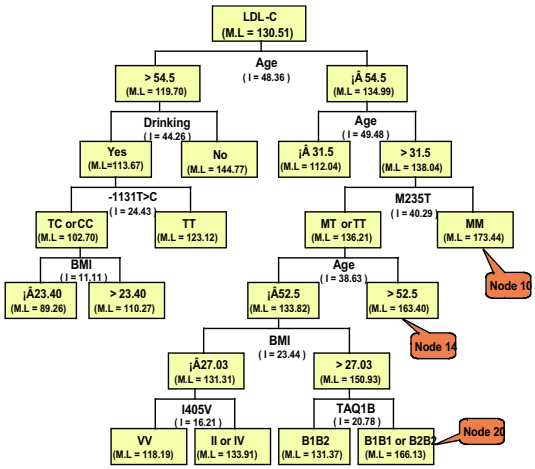


Figure 2- Male CART algorithm decision tree

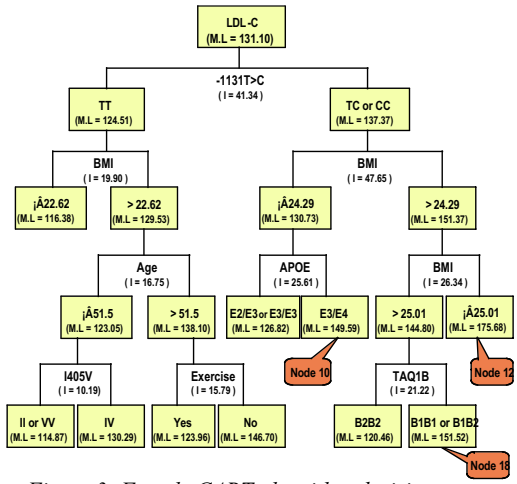


Figure 3- Female CART algorithm decision tree

decision tree can be expressed in terms of an 'if-then' rule, as follows:

For male, /* Node 10 */
If [(31.5 < Age ≤ 54.5) and (AGT-M235T genotype = MM)],
then the mean of LDL cholesterol = 173.44 mg/dl

Risk assessment and target segmentation for the management of LDL cholesterol

The gains chart produced by the decision tree can be used for as risk stratification for LDL cholesterol management. There are two parts to the gains chart, node-by-node statistics and cumulative statistics (Table 2). In regards to the target variable LDL cholesterol, which is continuous, gain charts provide node statistics relative to the mean of the target variable.

For the male group, the first node in the table, node 10, indicates the mean LDL cholesterol is 173.44 mg/dl for 7 subjects. For node 10 in males, the index score is 132.89%, indicating that the mean LDL cholesterol for this node is about 1.32 times higher than the overall sample.

Cumulative statistics demonstrate that cases with elevated LDL cholesterol can be derived by procuring the best segments of the sample. In males, if only the best node, (node 10), is reviewed, cases with LDL cholesterol means 1.32 times higher than the overall sample are reached by targeting only 3.06 % of the sample. If the next best node is included, (node 20), then cases with a 1.29 times higher LDL cholesterol mean than the overall sample are returned, from only 6.99 % of the sample.

Table 2 - Male risk stratification for LDL cholesterol by gains chart

Node	Node-by-node				Cumulative			
	Node(n)	Node(%)	Gain	Index (%)	Node (n)	Node(%)	Gain	Index (%)
10	7	3.06	173.44	132.89	7	3.06	173.44	132.89
20	9	3.93	166.13	127.29	16	6.99	169.33	129.74
14	11	4.80	163.40	125.20	27	11.79	166.91	127.89
4	13	5.68	144.77	110.92	40	17.47	159.72	122.37
18	91	39.74	133.91	102.60	131	57.21	141.79	108.64
19	7	3.06	131.37	100.66	138	60.26	141.26	108.23
8	29	12.66	123.12	94.33	167	72.93	138.11	105.82
17	18	7.86	118.19	90.56	185	80.79	136.17	104.33
5	19	8.30	112.04	85.85	204	89.08	133.92	102.61
12	16	6.99	110.27	84.49	220	96.07	132.20	101.29
11	9	3.93	89.26	68.39	229	100.00	130.51	100.00

* Node (n) : Number of cases for the node Node (%) : Percentage of the total sample cases falling into the node
Gain : Mean of LDL cholesterol for the node Index (%) : Ratio of the gain score for the node to the gain score for the total sample

Comparative assessment of multiple regression and CART

A comparison of performance for the two models is shown in Table 3. The CART algorithm performed better prediction (24.53%, 20.66%) than multiple regression(12.5%, 10.3%) for male and female groups, respectively. In the CART algorithm, error estimate is calculated as the within-node variance about the mean of the node. The total variance equals the within-node variance plus the between-node variance.

The within-node variance of this model for the male group is 975.144, while the total variance is 1292.14. The proportion of variance due to error is 975.144 / 1292.14 = 0.7547. Thus, the proportion of variance explained by the model is 100% - 75.47% = 24.53%. The within-node variance of this model for the female group is 863.255, while the total variance is 1088.03. The proportion of variance due to error is 863.255 / 1088.03 = 0.7934. Thus, the proportion of variance explained by the model is 100% - 79.34 % = 20.66%

Table 3 - Performance Comparison of Multiple Regression and the CART Algorithm

Model	Fitness	Male	Female
Multiple Regression	F statistics P-value R^2	1.781 0.032 0.125	1.798 0.028 0.103
CART	Total variance Within-node variance Proportion of variance explained by model	1292.140 975.144 24.53 %	1088.030 863.255 20.66 %

Discussion and conclusion

In this hospital-based cardiovascular genome study of Korean men and women, we found that CART is a better method to discover the knowledge to predict LDL cholesterol by biometric, environmental, and genetic factors compared with the regression method. Based on the CART method, we found that genetic polymorphisms APOA5, M235T, CETP(I405V), and CETP(TAQ1B) for male and APOA5, APOE, CETP(I405V), and CETP(TAQ1B) for female, were associated with LDL cholesterol. The genetic polymorphisms combined with other risk factors generated the rules for the prediction of LDL cholesterol.

Although genetic factors have been extensively investigated as risk factors for CVD, the results are still debatable [1-6,11]. In this present study, overall results displayed association of risk factors with LDL cholesterol which were very similar in both the CART and regression methods. However, the CART method provided more detailed results according to each segmentation and subgroup.

This study examined characteristics of the CART algorithm to demonstrate how they can be used to predict LDL cholesterol risk and provide management information from a patient database. The CART algorithm provided cumulative statistics demonstrating how efficient high LDL cholesterol cases were obtained by taking the best segments of the sample. The gains chart also provided valuable information about which segments to target and which to avoid. In addition, we presented rules that provided an occurrence relationship among the risk factors. Such information, which could not be obtained from multiple regression, can be used in examining the effects of individual risk factors on a specific segment of the target population.

There were several limitations in this study. One limitation is its reliance on data from a single hospital, and thus was biased in terms of the affluence of the subject group. In order to generalize data from a larger sample, partitioning data into training and testing set or cross validation is a pre-requisite or requirement. Sufficient validation for the generalization of our findings was not performed.

Future analyses will include an improvement to the CART algorithm. Furthermore, cost-effectiveness information will be incorporated into a data mining algorithm for each of the risk factors in order to estimate budgets for providing LDL cholesterol management services to the specific target population. This model provides a comprehensive analytic framework to construct the optimal design of clinical guidelines and healthcare policy for the prevention and management of LDL cholesterol.

References

[1] Arai Y, Hirose N, Yamamura K, Nakazawa S, Shimizu K, Deficiency of choresteryl ester transfer protein and gene polymorphisms of lipoprotein lipase and hepatic lipase are not associated with longevity. J Mol Med. 2003;81:102-9.

[2] Evans D, Buchwald A, Beil FU. The single nucleotide polymorphism .1131T>C in the apolipoprotein A5 (APOA5) gene is associated with elevated triglycerides in patients with hyperlipidemia. J Mol Med. 2003;81:645-54.

[3] Horÿ€´nek A, Vra´bl€´k M, Cÿesÿka R, Ada´mkova´ V, Poledne R, Hubacek JA. T-11310->C polymorphism within the apolipoprotein AV gene in hypertriglyceridemic individuals. Atherosclerosis. 2003;167:369-70.

[4] Liu S, Ma J, Ridker PM, Breslow JL, Stampfer MJ. A prospective study of the association between APOE genotype and the risk of myocardial infarction among apparently healthy men. Atherosclerosis. 2003;166:323-9.

[5] Chen Q, Reis SE, Kammerer CM, McNamara DM, Holubkov R, Sharaf BL, Sopko G, et al. APOE polymorphism and angiographic coronary artery disease severity in the Women's Ischemia Syndrome Evaluation (WISE) study. Atherosclerosis. 2003;169:159-67.

[6] Austin MA, Talmud PJ, Farin FM, Nickerson DA, Edwards KL et al. Association of apolipoprotein A5 variants with LDL particle size and triglyceride in Japanese Americans. Biochimica et Biophysica Acta. 2004;1688:1-9.

[7] Bose I, Mahapatra RK. Business data mining: a machine learning perspective. Information & Management. 2001;39:211-25.

[8] Friedewald WT, Levy RI, Fredrickson DS. Estimation of the concentration of low-density lipoprotein cholesterol in

plasma, without use of the preparative ultracentrifuge. Clin Chem. 1972;18:499-502.
[9] Michael JA, Gordon SL. Data mining technique: for marketing, sales and customer support. John Wiley & Sons ; 1997.
[10] Breiman L, Friedman JH, Olshen RA, Stone CJ. Classification and regression trees. Wadsworth, Belmont ;1984.
[11] Hubacek JA, Sÿkodova´ Z, Ada´ mkova´ V, La´ nska´ V, Poledne R. The influence of APOAV polymorphisms (T-1131>C and S19>W) on plasma triglyceridelevels and risk of myocardial infarction. Clin Genet. 2004;65:126-30.

Address for correspondence

Sun Ha Jee, Ph.D., MHS
Department of Epidemiology and Health Promotion,
Graduate School of Public Health,
Yonsei University, Seoul, Korea.
Tel: 82 2 2228 1523
Fax: 82 2 392 7734
E-mail: jsunha@yumc.yonsei.ac.kr

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Results from Data Mining in a Radiology Department: The Relevance of Data Quality

Martin Lang[acd], Nanda Kirpekar[b], Thomas Bürkle[c], Susanne Laumann[d], Hans-Ulrich Prokosch[c]

[a]*Workflow Research Group, Imaging Science Institute (ISI), Tübingen, Germany*
[b]*Department of Radiology, New York University Medical Center (NYUMC), New York, U.S.A.*
[c]*Chair of Medical Informatics, University Erlangen-Nuremberg, Erlangen, Germany*
[d]*Department Image and Knowledge Management, Siemens AG Medical Solutions, Erlangen, Germany*

Abstract

This work is part of an ongoing effort to examine and improve clinical workflows in radiology. Classical workflow analysis is time consuming and expensive. Here we present a purely data-driven approach using data mining techniques to detect causes for poor data quality and areas with poor workflow performance. Data has been taken from a operational RIS system. We defined a set of four key indicators for both data quality and workflow performance. Using several mining techniques such as cluster analysis and correlation tests we were able to detect interesting effects regarding data quality and an abnormality in the workflow for some organizational units of the examined radiology departments. We conclude that data-driven data mining approaches may act as a valuable tool to support workflow analysis and can narrow down the problem space for a manual on-site workflow analysis. This can save time and effort and leads to less strain for clinicians and workflow analysts during interviews.

Keywords:

data mining, data quality, workflow, radiology

Introduction

Improving workflows in clinical environments and especially the radiology workflow is gaining importance due to limited financial resources, increasing commercial pressure, changing hospital reimbursement e.g. in DRG based systems and an increasing requirement for quality management in medicine. Measuring the current workflow performance and its quality are key to any improvement efforts.

“If you cannot measure it, you cannot control it. If you cannot control it, you cannot manage it. If you cannot manage it, you cannot improve it" [1]. However, as we will show, a reliable data basis is required to determine reliable performance indicators. Poor data can delay decision making, cause bad decisions, or prevent appropriate decision making [2]. Redman [2] estimates that poor data quality costs the typical company at least ten percent of revenue. Here we present a data-driven approach using data mining techniques to detect causes for poor data quality. We will demonstrate that reliability of workflow performance measurement heavily depends upon data quality.

We define data quality according to Juran in [3], cited in [2],[4]: "Data is of high quality if it is fit for the intended uses [...] in operations, decision making and planning". Aspects of data quality comprise e.g. completeness, correctness, comprehensiveness, accessibility, timeliness, objectivity, relevancy, interpretability and understandability [4][5]. Data mining is the process of extracting knowledge from large amounts of data. Typical data mining functionalities comprise clustering, the discovery of concept/class descriptions, associations or correlations, classification, prediction, trend analysis, outliers analysis, deviation analysis, and similarity analysis [6].

Methods

Our working environment is the radiology department of New York University Medical Center (NYUMC). We focused on outpatient examinations of 2 NYUMC radiology sites during the time period 02/2006 to 08/2006. Site *A* performs ultrasound (US), computer tomography (CT), and conventional X-ray (XR) studies in three departments. Site *B* is responsible for magnetic resonance imaging (MRI). Site *A* has three, site *B* one big central unit and both have several smaller so called satellite units either in different buildings/floors or even spread across the city. All units run one radiology information system (RIS). Our source database is derived from this RIS and covered more than 60.000 examinations which are characterized by 30 attributes each. We built a data mining infrastructure in alignment with the knowledge discovery in databases process described in [6] which comprised the sequence steps data pre-processing, key indicator calculation, data mining, and pattern evaluation.

Data pre-processing

The following action sequence was performed:

- The data was loaded from an existing RIS-fed Oracle database containing the raw data into a staging database (Microsoft Access 2003).

- A pre-processing step then excluded extreme outliers and inconsistencies like dangling references, resulting in a pre-processed data base. E.g. we excluded examinations with a specific tag which did not refer to the examined body region
- The next step was data transformation (data normalization and aggregation) and data reformatting (e.g. changing date formats, etc.) We used a combination of SQL queries and VBA scripts to define the sequence of transformation and reformatting tasks. As an example we merged date and time values from two independent string type fields into one date type field.
- Finally data reduction (reduction of data volume without affecting the analytical results, like e.g. data discretization) was performed. We removed attributes such as patient identifiers or accession numbers. Furthermore we reduced the deviation from the preset duration of the examination from minutes to a value between [0;1], specifying the fraction of the preset in percent.
- Then we loaded the cleaned and transformed data into the project dedicated data warehouse (Oracle 10g) which itself is the actual data base for our mining efforts.

Calculation of key indicators (KIs)

Our primary goal was to concentrate on both data quality and workflow performance using a mixed set of key indicators. For workflow performance we defined:

- KI.1: *report-turnaround time*
  How long does it take from completion of an examination to finalization of the corresponding medical report
- KI.2: *adherence to the pre-assigned duration*
  How many minutes does the examination deviate from its scheduled duration. Alternatively the relative deviation in percent of the scheduled duration was used.

It is noteworthy, that the reliability of the key performance indicators *KI*.1 and *K*.2 of a radiology department depend on the availability and correctness of a relatively small set of timestamps. These underlying timestamps are e.g. the exam begin and completion timestamps or the report finalization timestamps. In contrast to the correctness of these timestamps (data correctness indicator *KI*.4) their availability (data completeness indicator *KI*.3) has no direct influence on the reliability of the performance indicators *KI*.1 and *KI*.2. However, a high fraction of unavailable timestamps may lead to a biased view of the performance. To measure data quality we used the following key indicators:

- KI.3: *data completeness*
  Which fraction of items has all relevant information available (no attribute values are missing or undefined)

- KI.4: *data correctness*
  Which fraction of items has "correct data" for all relevant information (no contradictory facts in the data)

Our database, derived from the operational RIS, comprised optional data from manual entry such as patient arrival time, the begin of an examination, and the patient departure time. The end time of an examination is also manually entered, but unlike the previously mentioned timestamps it is required by the system. Report creation and report finalization timestamps are auto-generated by the RIS.

Normally patient arrival (as well as patient departure) should be tracked by the front-desk personnel of the respective unit. The begin examination timestamps are maintained by radiology technologists. In detail *KI*.3 *data completeness* was calculated as the fraction of radiology examinations which do not violate one or more of the following constraints:

C3.1: Patient arrival time is available
C3.2: Begin of examination time is available
C3.3: Patient departure time is available

Data correctness was even more difficult to define. As a coarse correctness indicator we searched for studies, which contain obviously nonsense values for the timestamps. *KI*.4: *data correctness* was calculated as the fraction of items, which do violate none of the following check constraints:

C4.1: $time$(Pat. Arrival) < $^{*}time$(Exam Begin)
C4.2: $time$(Exam Complete) < $^{*}time$(Pat. Departure)
C4.3: $time$(Pat. Arrival) ≠ $^{*}time$(Pat. Departure)
C4.4: $time$(Exam Begin) ≠ $^{*}time$(Exam Complete)
C4.5: 1min < $duration$(Exam) ≤ 5$ScheduledDuration$

This reads as follows: $t_1 < {}^{*}t_2$ means t_1 is earlier than $t_2 \pm 1min$. $t_1 > {}^{*}t_2$ means t_1 is later than $t_2 \pm 1min$. Equality of two timestamps $t_1 = {}^{*}t_2$ is defined if $t_1 = t_2 \pm 1min$. Two timestamps t_1, t_2 are considered to be unequal ($t_1 \neq {}^{*}t_2$) if $t_1 < t_2 \pm 1min$. The following rule (*C*4.6) was retrospectively abandoned due to too many violations.

C4.6: $time$(Exam Completion) ≠ $^{*}time$(Pat. Departure)

Data mining

Data mining was performed on the computed key indicators *KI*.1 to *KI*.4 and the following information: patient sex, patient age, examining site, examining unit, medical category, report transcription time, report signing time, referring physician, modality type, imaging device, technologist, and responsible radiologist. We used the WEKA Software (v 3.5.3) [7] and focused on cluster analysis, outlier analysis, and classification of findings.

For cluster analysis, multiple algorithms were used, such as the Simple-EM algorithm, an expectation maximization algorithm described in [6] or the Farthest-First algorithm described in [8] and implemented in [7]. For outlier analysis we considered instances as outliers if they exceed the following range: $x_i < Q1 - 1.5\ IQR$ or $x_i > Q3 + 1.5\ IQR$. *IQR* is the Inter-Quartile Range, $Q1$ is the first quartile, $Q3$ the third. In addition we used the χ^2-test to determine if data follows a specific distribution and for testing two or more groups regarding statistical independence.

Cluster classification was done primarily by generating C4.5 decision trees as described in [9],[7]. In addition, we relied on the 1R classifier which uses the minimum-error attribute for prediction and discretizing numeric attributes see [10], and a nearest-neighbour-like algorithm using non-nested generalized exemplars as described in [11],[7].

Pattern Evaluation

The findings from the previous step must be evaluated and valuable results should be used for workflow improvement. We concentrated on the following parameters to distinguish interesting and less interesting findings [6]: Simplicity, Certainty, Utility, and Novelty. Simplicity gives information with regard to how easily a finding can be understood. It can be quantified e.g. by measuring an (association) rule length, or the size of the mined (decision) tree. Certainty provides information about the classification reliability, the accuracy, the rule strength or the discriminating weight. The utility measurement quantifies the potential usefulness of a finding and can be quantified by calculating the rule support or noise threshold. And finally, a finding is of interest if it was not previously known or surprising (Novelty).

In this project, a rough pre-selection of interesting findings was done, which then was presented to and discussed with the involved and/or affected clinicians. This sifting through the intermediate findings leads to further open questions, additional issues to be examined, and to new and deeper insights into detected problems. Finally, workflow improvements were derived from the findings in consultation with the affected participants and the corresponding supervisors.

Results

In our case the objective of data mining was to classify the data to detect outliers and to find patterns in the data, which describe the set of detected outliers. Therefore we started the mining process with the application of clustering algorithms on the whole data base, including performance and data quality indicators *KI*.1 to *KI*.4 to detect groups of data items with a high degree of similarity. In this first step we detected three different clusters. To gain more information about the characteristics of those clusters, we used the R1 classification given by [7] which came up with the following rules:

if <Site=A> then <cluster0> and
if <Site=B> then <cluster2>

Using just these rules would classify 89.29% data items correctly to the respective site. The third cluster did not prompt such a clear rule. As a consequence there must be a difference of data entered at site *A* (CT, US, XR) vs. site *B* (MRI) in terms of data quality or workflow performance. We now looked at the two sites separately and continued the same clustering using performance and data quality indicators *KI*.1 to *KI*.4.

For site *A* we could assign two clusters. Classifying these clusters derived the following rule:

if <Site is a satellite> then <cluster1>
else <cluster0>

Using this rule 97.49% of the instances would be classified correctly. The same approach applied to Site *B* resulted in similar findings and exactly the same rule, which would classify 92.68% of Site *B* instances correctly based on their satellite or non-satellite status.

Table 1: Fraction of a) correct and complete, b) incomplete (ℜKI.3), and c) incorrect (ℜKI.4) data items

	a) KI.3 and KI.4		**b) ℜKI.3**		**c) ℜKI.4**	
Dept.	**non-Satellite**	**Satellite**	**non-Satellite**	**Satellite**	**non-Satellite**	**Satellite**
A1	74,36%	29,48%	20,46%	5,20%	6,19%	68,03%
A2	79,26%	n/a	11,91%	n/a	9,04%	n/a
A3	82,56%	45,21%	5,44%	16,05%	12,67%	41,13%
B1	85,37%	79,55%	11,39%	15,53%	4,06%	5,71%

As an intermediate result at this stage we can say that there is a difference between site A and site B with regard to the data quality and/or workflow performance in the examined time period. Furthermore within the sites there is a difference between satellite units and the central unit in both cases. Now we tried to work out which of our key indicators, workflow performance (*KI*.1, *KI*.2) or data quality (*KI*.3, *KI*.4) was responsible, starting with the data quality indicators and comparing satellite to non-satellite units for sites A and B respectively. An overview of the findings is given in Table 1.

The fraction of data items entered that do not violate one of the data completeness rules *C*3.1 to *C*3.3 or the data correctness rules *C*4.1 to *C*4.5 are shown in Table 1 column a). Table 1 column b) and c) give a more detailed view. Table 1 column b) depicts the fraction of data entries having at least one incomplete timestamp (ℜ*KI*.3). Table 1 column c) (ℜ*KI*.4) gives detail about the entries violating one of the data correctness rules of indicator *KI*.4. Department A2 does not have satellite units, but was listed for comprehensiveness reasons.

We found for example that in department A1 the amount of data entries with missing values (ℜ*KI*.3) is much lower in satellite units (5,20%) compared to the central unit (20,46%). But as the indicator *KI*.4 shows, the majority of data entered at A1 satellite units is bogus (ℜ*KI*.4 = 68,03%), whereas the bogus ratio (ℜ*KI*.4) at the central unit is much lower (6,19%). Departments A3 and B1 show also a lower amount of correct entries (Table 1 column a) at satellite units compared to the central unit. But in contrast to the findings at department A1, the satellite units of A3 and B1 perform worse regarding both *KI*.3 and *KI*.4. Obviously, there are differences between satellite and non-satellite units with regard to data quality.

Based on these findings we focused on the detection of coherences between the detected data quality issues at satellite units and front desk staff, technologists, radiologists, devices, types of examinations, patient characteristics, etc.

Therefore we performed different classification analysis based on smaller and more condensed views on the original data. We computed a large set of different transient aggregations comparing satellite units with central units for site *A* and *B* respectively. These aggregated views comprised for example the averaging (*avg*) of:

- *avg*(ℜ*KI*.3) and *avg*(ℜ*KI*.4) per weekday, site, and dept.
  - *avg*(ℜ*C*3.1) per weekday, site, and dept.
  - ...
  - *avg*(ℜ*C*4.5) per weekday, site, and dept.
- *avg*(ℜ*KI*.3) and *avg*(ℜ*KI*.4) per tech., site, and dept.
  - *avg*(ℜ*C*3.1) per tech., site, and dept.
  - ...
  - *avg*(ℜ*C*4.5) per tech., site, and dept.
- *avg*(ℜ*KI*.3) and *avg*(ℜ*KI*.4) per physician, site, and dept.
- ...

The classification based on these condensed data revealed - among other findings - the following rule for site *B*, using the 1R classifier after [10].

if avg(ℜ*C*3.1 per tech, site, dept)
<4.22 or ≥ 12.17 then <is Satellite>=true
else <is Satellite>=false

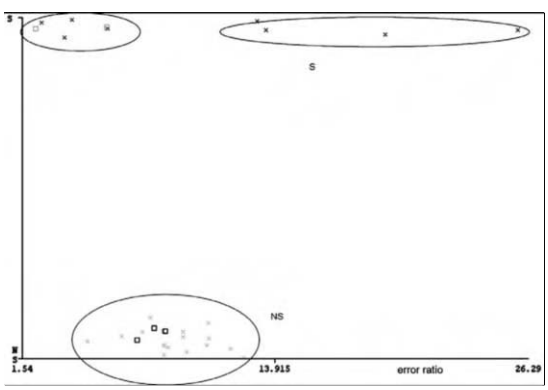


Figure 1 - Classification results based on condensed data

A visualization of the found rule is depicted in figure 1, showing correctly classified instances marked with an x and incorrectly classified instances marked with a square. This rule would classify 83.3% data instances correctly. The rule essentially states that the fraction of missing patient arrival times (ℜ*C*3.1) per technologist at satellite units is different from the fraction of missing patient arrival times at central units. Remember that we said at the beginning that the technologist is responsible for entering the begin examination timestamp but not for entering the patient arrival timestamp which should be entered by front desk personnel. Therefore there should be no correlation between technologist and the violation of indicator *C*3.1; the error ratio for *C*3.1 should be distributed uniformly. We defined the null hypothesis: "the amount of missing patient arrival timestamps should be equally distributed among technologists" and performed a confirmatory data analysis using the a χ^2-test. We detected a highly significant ($\alpha <$ 0,1%) deviation from uniform distribution for both sites *A* and *B* and rejected the null hypothesis at a level of confidence of 99.9%. Comparable results were found when comparing correctness of the patient arrival timestamp (*C*4.1 and *C*4.3) and technologists.

Discussion

Within this study we detected interesting relationships between data quality of timestamps within a hospital RIS system and staff involved in the examination procedure. This information is extremely valuable when considering workflow performance measured on the basis of those timestamps and gives some insight in the likely working procedures. For example, in the satellite radiology units it might often be the case that patient arrival and departure is not encoded by front desk staff of the respective unit. We assume that in such cases the technologist enters those timestamps himself to continue the procedure but does so incorrectly. So there seems to be a defect in the satellite workflow because technologists are forced to track patient arrival. This needs to be confirmed by local workflow analysis projects and will influence our ongoing efforts to measure and optimize workflow performance.

Data mining was performed using data taken directly out of the operative RIS database. We looked into some data quality aspects such as data comprehensiveness, data accuracy, data objectivity and timeliness of the data but did not examine aspects like interpretability, understandability and relevancy of the data within the current project. Our rules for checking data completeness and data correctness have a coarse granularity. If someone entering bogus data into the RIS were to put two minute differences between the timestamps for patient arrival, begin examination and patient departure, we would not detect this currently. This has to be taken into account when evaluating the findings.

Nevertheless, in contrast to the usual common proceeding in workflow improvement projects - which requires workflow analysts or consultants to manually analyze and asses the current workflow by interviewing clinicians or measuring performance - we propose a purely data-driven approach which requires much less effort in order to identify previously unknown areas with potentially suboptimal performance or quality for further analysis. This may well reduce time and expenditure for on-site workflow analysis by narrowing down the problem space in advance. It may also save precious clinician time for interviews during the workflow analysis. Additionally, in contrast to a plain hypothesis-based analysis, which requires the analyst to define the right hypothesis in advance, the given data-driven mining approach is able to lead to new and previously unknown findings.

We selected a set of interesting findings and discussed these with the involved and/or affected clinicians, who attached great importance to this information. For example, no one was aware of the existing differences in data quality between satellites and the main areas. However, the selection of findings for the subsequent and more in depth

mining was done subjectively within this group of informaticians and clinicians with a focus on local activities and improvement potentials. A more objective “utility” for assessing and selecting findings still requires further research and will be part of our future efforts.

Our experiences in using data mining are well aligned with [5] and [12], both stating good results in detecting areas of poor data quality. In contrast, our approach not only focuses on data quality, but also regards workflow performance on a par with data quality. To achieve reliable results during workflow performance measurement it is mandatory to start with sufficient data quality in order to avoid a garbage in - garbage out effect. Not including data quality issues in workflow performance focused improvement projects can result in bogus findings and thus lead to potential counter-productive and expensive decisions based on these bogus findings. For our performance measurement efforts we will definitely take the quality related findings into account and would highly recommend a similar approach to anyone performing such a task.

References

[1] Harrington HJ. Business Process Improvement. The Breakthrough Strategy for Total Quality, Productivity, and Competitiveness. 1st ed. New York (NY): McGraw-Hill; 1991. ISBN: 0070267685.

[2] Redman TC. Data: An Unfolding Quality Disaster. DM Review Magazine. 2004;08.

[3] Juran JM, Godfrey AB. Juran's Quality Handbook. 5th ed. McGraw-Hill Professional; 1999. ISBN: 007034003X.

[4] Luebbers D, Grimmer U, Jarke M. Systematic Development of Data Mining-Based Data Quality Tools. In: Proceedings of the 29th International Conference on Very Large Data Bases (VLDB 2003). Berlin, Germany: Morgan Kaufmann; 2003. p. 548-559.

[5] Hipp J, Guentzer U, Grimmer U. Data Quality Mining - Making a Virtue of Necessity. In: Proceedings of the 6th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery (DMKD 2001). Santa Barbara, California; 2001. p. 52-57.

[6] Han J, Kamber M. Data Mining - Concepts and Techniques. 2nd ed. Morgan Kaufmann; 2006. ISBN-10: 1558609016.

[7] Witten IH, Frank E. Data Mining: Practical machine learning tools and technique. 2nd ed. San Francisco: Morgan Kaufmann; 2005. ISBN 0-12-088407-0.

[8] Hochbaum, Shmoys. A best possible heuristic for the k-center problem. Mathematics of Operations Research. 1985;10(2):180-184.

[9] Quinlan JR. C4.5: programs for machine learning. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.; 1993. ISBN:1-55860-238-0.

[10] Holte RC. Very simple classification rules perform well on most commonly used datasets. Machine Learning. 1993;11:63-91.

[11] Roy S. Nearest Neighbor With Generalization. Christchurch, New Zealand: University of Canterbury, Christchurch, New Zealand; 2002. University of Canterbury, Christchurch, New Zealand.

[12] Dasu T, Johnson T. Hunting of the Snark - Finding Data Glitches using Data Mining Methods. In: Fourth Conference on Information Quality (IQ 1999). vol. 4. MIT; 1999. p. 89-98

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Evaluating Learning Models with Transitions of Human Interests Based on Objective Rule Evaluation Indices

Hidenao Abe[a], Hideto Yokoi[b], Shusaku Tsumoto[a], Miho Ohsaki[c], Takahira Yamaguchi[d]

[a]*Department of Medical Informatics, Shimane University, School of Medicine, Japan*
[b]*Department of Medical Informatics, Kagawa University Hospital, Japan*
[c]*Faculty of Engineering, Doshisha University, Japan*
[d]*Faculty of Science and Technology, Keio University, Japan*

Abstract

This paper presents a method to support the evaluation procedure of a data mining process using human-system interaction. The post-processing of mined results is one of the key factors for successful data mining process. However, it is difficult for human experts to completely evaluate several thousands of rules from a large dataset containing noise. We have designed a method based on objective rule evaluation indices to support the rule evaluation procedure; the indices are calculated to evaluate each if-then rule mathematically. We have evaluated five representative learning algorithms to construct rule evaluation models of the actual data mining results from a chronic hepatitis data set. Further, we discuss the relationship between the transitions of the subjective criterion of a medical expert and the performances of the rule evaluation models.

Keywords:

data mining, post-processing, human interest, rule evaluation index.

Introduction

In recent years, enormous amounts of data relating to natural science, social science, and business have been stored on information systems. People have been able to obtain valuable knowledge due to the development of information technology. Besides, data mining techniques combine different types of technologies such as database technologies, statistical methods, and machine learning methods. Moreover, data mining has been well known for utilizing data stored on database systems. In particular, if-then rules, which are obtained by rule induction algorithms, are considered to be one of the widely usable and readable outputs of data mining.

However, for large datasets with hundreds of attributes, including noise, the process often results in several thousands of rules. From such a large rule set, it is difficult for human experts to obtain valuable knowledge, which is rarely included in the rule set.

To support such a rule selection, many studies have been performed using objective rule evaluation indices such as recall, precision, and other interestingness measurements [1-3] (Hereafter, we refer to these indices as "objective indices"). Further, it is difficult to estimate the subjective criterion of a medical expert using a single objective rule evaluation index; this is because his/her subjective criterion such as interestingness and importance for his/her purpose is influenced by the amount of his/her medical knowledge and/or the passage of time.

With regard to the above mentioned issues, we have developed an adaptive rule evaluation support method for human experts with rule evaluation models. This method predicts the experts' criteria based on objective indices by re-using the results of the evaluations by human experts.

In Section 2, based on objective indices, we describe the construction method for the rule evaluation model. Then, we present a performance comparison of learning algorithms for constructing rule evaluation models in Section 3. With regard to the results, we discuss the relationship between the transitions of each subjective criterion of a medical expert and the objective performances of the rule evaluation models.

Rule evaluation support with rule evaluation models based on objective indices

In practical data mining situations, costly rule evaluation procedures are repeatedly performed by a human expert. In these situations, the useful information from each evaluation, such as focused attributes, interesting combinations of attributes/attributes and values, and valuable facts are not explicitly used by any rule selection system, but tacitly assumed by the human expert. To solve these problems, we suggest a method to construct rule evaluation models based on objective rule evaluation indices as a way to explicitly describe the criteria of the human expert by re-using the human evaluations. By combining this method with the rule visualization interface, we design a rule evaluation support tool, which can perform more certain rule evaluations with explicit rule evaluation models.

We consider the process of modeling the rule evaluations of human experts as the process that clarifies the relationships between the human evaluations and the features of the inputted if-then rules. Based on this consideration, we find that the construction process for the rule evaluation

model can be implemented as a learning task. Figure 1 shows the construction process based on the re-use of human evaluations and objective indices for each mined rule.

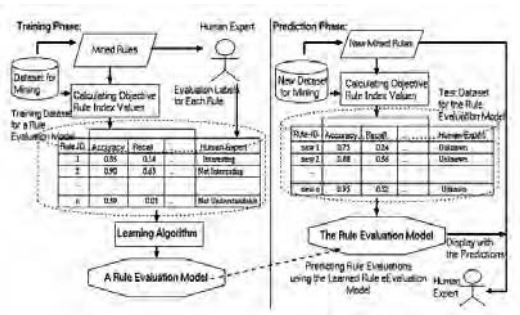


Figure 1 - Overview of the construction method for rule evaluation models

Using this framework, the human expert can evaluate a large number of if-then rules iteratively. The system supports his/her rule evaluation by presenting predictions for unevaluated rules based on a rule evaluation model learned from the evaluated rules. Thus, the human expert can avoid the evaluation of unfamiliar rules or/and uncertain rules based on his/her background knowledge from the first time evaluation process of each rule.

In the training phase, the attributes of a meta-level training dataset are obtained by objective indices such as recall, precision, and other rule evaluation values. The human evaluations for each rule are combined as classes of each instance. To obtain this data set, the human expert has to evaluate the all or a part of the input rules at least once. After obtaining the training data set, its rule evaluation model is constructed by using a learning algorithm.

In the prediction phase, the human expert receives the predictions for new rules based on their objective index values. Since the rule evaluation models are used for predictions, we need to choose a learning algorithm with high accuracy in order to solve the current classification problems.

Comparisons of learning models on chronic hepatitis data mining results

In this section, we present the results of empirical evaluations using the dataset obtained from two hepatitis data mining results [4,5].

First, we obtained a dataset for each rule set from the above mentioned data mining results. The five learning algorithms were then used to construct each rule evaluation model.

Based on the experimental results, we evaluated the availability of each learning algorithm with respect to the accuracies of the rule evaluation models, the minimum costs to construct valid rule evaluation models, and the contents of the rule evaluation models learned.

Since a rule evaluation model requires high accuracy to perfectly support a human expert in the proposed rule evaluation support method, we compared the predictive accuracies for the entire dataset and Leave-One-Out validation.

The accuracy of a validation dataset D is calculated using correctly predicted instances *Correct(D)* as $Acc(D) = (Correct(D)/|D|)*100$, where $|D|$ denotes the size of the dataset. The recalls of class i on a validation dataset are calculated by using correctly predicted instances of the class *Correct(Di)* as $Recall(Di) = (Correct(Di)/|Di|)*100$, where $|Di|$ denotes the size of the instances of class i. Furthermore, the precision of class i is calculated by using the size of the instances that are predicted as $Precision(Di) = (Correct(Di)/Predicted(Di))*100$.

We obtained the learning curves of accuracies of the learning algorithms for the entire training dataset to evaluate whether each learning algorithm can perform in the early stage of the rule evaluation process.

The accuracies of randomly sub-sampled training datasets are averaged for 10 trials on each percentage of subset.

By observing the elements of the rule evaluation models for the meningitis and hepatitis data mining results, we consider the characteristics of the objective indices that are used in these rule evaluation models.

In order to construct a dataset to learn a rule evaluation model, the values of the objective indices have been calculated for each rule by considering 39 objective indices [6], as shown in Table 1.

Thus, the dataset for each rule set has the same number of instances as the rule set. Each instance has 40 attributes, including those of the class.

We applied the five learning algorithms to these datasets to compare their performances as construction methods for the rule evaluation models.

We employed the following learning algorithms from Weka [18]: J4.8 (a C4.5 decision tree learner) [19], a back-propagation neural network (BPNN) learner [20], support vector machines (SVM) [21], classification via linear regressions (CLR) [22], and OneR [23].

*Table 1 - Objective rule evaluation indices for classification rules used in this research. **P**: Probability of the antecedent and/or consequent of a rule **S**: Statistical variable based on P **I**: Information of the antecedent and/or consequent of a rule **N**: Number of instances included in the antecedent and/or consequent of a rule **D**: Distance of a rule from the others based on rule attributes*

Theory	Index Nabe (Abbreviation) [Reference of Literature]
P	Coverage (Coverage), Prevalence (Prevalence), Precision (Precision), Recall (Recall), Support (Support), Specificity (Specificity), Accuracy (Accuracy), Lift (Lift), Levarage (Leverage), Added Value (Added Value) [2], Kloesgen's Interestingness (KI) [7], Relative Risk (RR) [8], Brin's Interest (BI) [9], Brin's Conviction (BC) [9], Certainty Factor (CF) [2], Jaccard Coefficient (Jaccard) [2], F-Measure (F-M) [10], Odds Ratio (OR) [2], Yule's Q (YuleQ) [2], Yule's Y (YuleY) [2], Kappa (Kappa) [2], Collective Strength (CST) [2], Gray and Orlowska's Interestingness weighting Dependency (GOI) [11], Gini Gain (Gini) [2], Credibility (Credibility) [12]
S	Chi-square Measure for One Quadrant (Chi-Square M1)[13], Chi-square Measure for Four Quadrant (Chi-Square M4)[13]
I	J-Measure (J-M) [14], K-Measure (K-M) [6], Mutual Information (MI) [2], Yao and Liu's Interestingness 1 based on one-way support (YLI1) [3], Yao and Liu's Interestingness 2 based on two-way support (YLI2) [3], Yao and Zhong's Interestingness (YZI) [3]
N	Cosine Similarity (CSI) [2], Laplace Correction (LC) [2], Phi Coefficient (Phi) [2], Piatetsky-Shapiro's Interestingness (PSI) [15]
D	Gago and Bento's Interestingness (GBI) [16], Peculiarity (Peculiarity) [17]

A case study on the chronic hepatitis data mining results

In this case study, we used four data mining results for chronic hepatitis, as shown in Table 2. These datasets show patterns of the values obtained from the laboratory tests of the blood and urine of chronic hepatitis patients as attributes.

First, we performed the data mining processes twice to determine the relationships between the patterns of attributes and the those of glutamate-pyruvate transaminase (GPT) as the class, which is one of the important tests required to determine the condition of each patient. A medical expert evaluated these results at different times, which were approximately one year apart. However, the expert held his interestingness concerning the significant movement of GPT values between these two evaluations. Second, we performed other data mining processes twice to determine the relationships between the patterns of attributes and the results of interferon (IFN) therapy. The expert also evaluated these results at different times, which were two weeks apart. For each rule, we assigned a label (EI: Especially Interesting, I: Interesting, NI: Not Interesting, and NU: Not Understandable) according to the evaluations provided by the medical expert.

Table 2 - Description of datasets of the chronic hepatitis data mining results

			Class Distribution			
			EI	I	NI	NU
GPT	Phase 1	30	3	8	16	3
	Phase 2	21	2	6	12	1
INF	First Time	26	4	7	11	4
	Second Time	32	15	5	11	1

Comparison of the classification performances

The results for the performances of the five learning algorithms with respect to the entire training dataset and those for Leave-One-Out are shown in Tables 3 and 4, respectively.

Most of the accuracies for the entire training dataset are higher than those estimated by simply predicting each default class. This indicates that the five learning algorithms can construct valid rule evaluation models for these datasets.

Table 3 - Accuracies (%), Recalls (%) and Precisions (%) of the five learning algorithms for the training datasets

		Acc	Precision				Recall			
			EI	I	NI	NU	EI	I	NI	NU
GPT1										
	J4.8	96.7	100.0	88.9	100.0	100.0	66.7	100.0	100.0	100.0
	BPNN	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	SVM	56.7	0.0	100.0	68.2	14.3	0.0	12.5	93.8	33.3
	CLR	63.3	0.0	66.7	62.5	0.0	0.0	50.0	93.8	0.0
	OneR	60.0	0.0	66.7	59.3	0.0	0.0	25.0	100.0	0.0
GPT2										
	J4.8	90.5	66.7	85.7	100.0	0.0	100.0	100.0	91.7	0.0
	BPNN	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	SVM	95.2	100.0	100.0	92.3	100.0	50.0	100.0	100.0	100.0
	CLR	85.7	50.0	100.0	85.7	0.0	50.0	83.3	100.0	0.0
	OneR	85.7	0.0	75.0	92.3	0.0	0.0	100.0	100.0	0.0
INF1										
	J4.8	88.5	80.0	100.0	83.3	100.0	100.0	71.4	90.9	100.0
	BPNN	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	SVM	46.2	26.7	0.0	70.0	100.0	100.0	0.0	63.6	25.0
	CLR	53.8	100.0	0.0	47.6	66.7	50.0	0.0	90.9	50.0
	OneR	50.0	0.0	50.0	50.0	0.0	0.0	85.7	63.6	0.0
INF2										
	J4.8	90.6	88.2	100.0	90.9	0.0	100.0	80.0	90.9	0.0
	BPNN	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	SVM	56.3	72.7	0.0	45.0	100.0	53.3	0.0	81.8	100.0
	CLR	65.6	63.2	100.0	60.0	0.0	80.0	60.0	54.5	0.0
	OneR	68.8	62.5	0.0	87.5	0.0	100.0	0.0	63.6	0.0

Besides, the accuracies of Leave-One-Out show the robustness of each learning algorithm. These accuracies show the predictive performances of each learning algorithm in the "Prediction Phase" in Figure 1. For GPT1 and GPT2, the medical expert is required to determine "EI" rules by himself/herself, because the learning algorithms cannot predict "EI" instances based on the training datasets. Besides, for GPT1 and IFN1, the accuracies are lower than those simply estimated by predicting default classes because the medical expert evaluates these data mining results without a specific criterion in his/her mind.

Table 4 - Accuracies (%), Recalls (%), and Precisions (%) of Leave-One-Out of the five learning algorithms

		Acc	Precision				Recall			
			EI	I	NI	NU	EI	I	NI	NU
GPT1										
	J4.8	50.0	0.0	60.0	60.0	0.0	0.0	75.0	56.3	0.0
	BPNN	30.0	0.0	12.5	50.0	0.0	0.0	12.5	50.0	0.0
	SVM	46.7	0.0	0.0	65.0	11.1	0.0	0.0	81.3	33.3
	CLR	40.0	0.0	14.3	50.0	0.0	0.0	12.5	68.8	0.0
	OneR	43.3	0.0	25.0	55.6	0.0	0.0	37.5	62.5	0.0
GPT2										
	J4.8	76.2	0.0	66.7	90.9	0.0	0.0	100.0	83.3	0.0
	BPNN	66.7	0.0	83.3	81.8	0.0	0.0	83.3	75.0	0.0
	SVM	81.0	0.0	100.0	91.7	25.0	0.0	83.3	91.7	100.0
	CLR	76.2	0.0	83.3	84.6	0.0	0.0	83.3	91.7	0.0
	OneR	81.0	0.0	66.7	91.7	0.0	0.0	100.0	91.7	0.0
INF1										
	J4.8	19.2	37.5	0.0	20.0	0.0	75.0	0.0	18.2	0.0
	BPNN	26.9	40.0	22.2	25.0	25.0	50.0	28.6	18.2	25.0
	SVM	34.6	21.4	0.0	54.5	0.0	75.0	0.0	54.5	0.0
	CLR	19.2	33.3	0.0	28.6	0.0	25.0	0.0	36.4	0.0
	OneR	19.2	0.0	11.1	23.5	0.0	0.0	14.3	36.4	0.0
INF2										
	J4.8	75.0	76.5	66.7	75.0	0.0	86.7	40.0	81.8	0.0
	BPNN	37.5	50.0	28.6	22.2	0.0	53.3	40.0	18.2	0.0
	SVM	31.3	36.4	0.0	28.6	0.0	26.7	0.0	54.5	0.0
	CLR	34.4	41.2	20.0	30.0	0.0	46.7	20.0	27.3	0.0
	OneR	68.8	60.0	0.0	100.0	0.0	100.0	0.0	63.6	0.0

Estimating the minimum training subset to construct a valid rule evaluation model

Since the construction method for the rule evaluation model requires the mined rules to be evaluated by a human expert, we investigated the minimum training samples to construct a valid rule evaluation model that works better as compared to simply predicting each default class, as shown in Table 5.

GPT1 and IFN1 require more instances than GPT2 and IFN2 in order to learn valid rule evaluation models. In particular, GPT1 requires significantly larger training instances as compared to GPT2 to construct valid rule evaluation models based on the paired t-test ($t = 0.02 < 0.05$). These results indicate that there is a difference in the criteria of human evaluation when evaluating each data mining result.

Table 5 - Minimum training instances to construct valid rule evaluation models using each learning algorithms

		J4.8	BPNN	SVM	CLR	OneR
GPT	Phase 1	30	3	8	16	3
	Phase 2	21	2	6	12	1
IFN	First Time	26	4	7	11	4
	Second Time	32	15	5	11	1

Rule evaluation models for the data mining result dataset for chronic hepatitis

Figure 2 shows the decision trees of J4.8 for each data mining result dataset for chronic hepatitis. With a decision tree, the class label (label of squared node) of an instance is predicted based on its attribute values of the instance according to each branch of the tree. Hence, these decision trees show the decision models of the medical expert from one of the objective aspects.

As shown in Figure 2, these models consist of not only indices expressing the correctness of rules but also other types of indices. This shows that the medical expert evaluated these rules with both correctness and interestingness based on his/her background knowledge.

In each problem, the variance of indices was reduced in the second data mining process. This indicates that the medical expert evaluated the second data mining results with a more certain criterion than he/she did for the first data mining process.

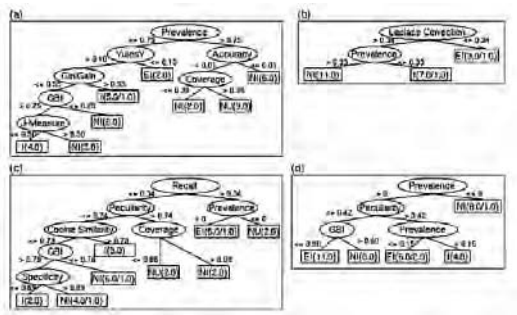


Figure 2 - Rule evaluation models constructed using J4.8 for each data mining result ((a)GPT phase 1, (b) GPT phase 2, (c) IFN first time, (d) IFN second time)

Discussion

Based on the results obtained using the data mining results of chronic hepatitis, it is observed that the performances improve with progress in the data mining processes. In particular, the prediction performances between “I” and “NI” show explicit improvement between the two data mining processes, as shown in Table 4.

The rule evaluation models constructed using the learning algorithms comprise not only indices related to the correctness of each rule but also indices related to other features of each rule, as shown in Figure 2. This result corresponds to the evaluation of the rules based on both their correctness and interestingness by human experts on the basis of their expertise. Furthermore, the variances of the objective indices used in the rule evaluation models decrease with the progress in data mining processes.

The difference between the hepatitis data mining results concerns the certainness of the criterion of the medical expert. The medical expert built up his/her criterion by using iterating data mining processes. Based on these differences, we identified the following three phases of rule evaluation: hypothesis generation, hypothesis validation, and hypothesis refinement.

Then, by considering the differences in performances and the contents of the rule evaluation models, we defined the relationship between these observable differences and the three evaluation phases, as shown in Table 6.

Table 6 - Relationship between the transitions of human criteria and the changes in rule evaluation model.

Phase	Classification Performance	Predictive Performance	Variance of Contents
Generation	low	low	high
Validation	low	improve	decrease
Refinement	high	high	low

Conclusion

In the present paper, we have described the evaluation of five learning algorithms for a rule evaluation support method; rule evaluation models based on objective indices are used to predict the evaluations for an if-then rule by re-using the evaluations of a human expert.

Based on the comparison of performances of the five learning algorithms for the dataset from the data mining results of chronic hepatitis, the rule evaluation models have achieved higher accuracies for simply predicting each default class. Considering the results of the rule evaluation models constructed for four different data mining results of chronic hepatitis, the differences in the human evaluation criteria appear as differences in the rule evaluation models with respect to both their performances and their contents. This result indicates that the proposed approach can detect differences in the human evaluation criteria as differences in the performances of rule evaluation models.

In the future, we will apply this rule evaluation support method to other datasets from various domains.

References

[1] Hilderman R.J. and Hamilton H.J. Knowledge discovery and measure of interest. Kluwer Academic Publishers, 2001.

[2] Tan P.N., Kumar V., and Srivastava J. Selecting the right interestingness measure for association patterns. In: Proc. of International Conference on Knowledge Discovery and Data Mining, KDD-2002, 2003; pp. 32–41.

[3] Yao Y.Y. and Zhong N. An analysis of quantitative measures associated with rules. In: Proc. of Pacific-Asia Conference on Knowledge Discovery and Data Mining, PAKDD-1999, 1999; pp. 479–488.

[4] Ohsaki M., Sato Y., Kume S., Yokoi H., and Yamaguchi T. Comparison between objective interestingness measures and real human interests in medical data mining. In: Proc. of the 17th International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems IEA/AIE-2004, LNAI 3029, 2004: pp. 1072–1081.

[5] Abe H., Ohsaki M., Yokoi H., and Yamaguchi T. Implementing an integrated time-series data mining environment based on temporal pattern extraction methods– A case study of an interferon therapy risk mining for chronic hepatitis, In: JSAI2005 Workshops, LNAI 4012, 2006; pp. 425–435.

[6] Ohsaki M., Kitaguchi S., Kume S., Yokoi H., and Yamaguchi T. Evaluation of rule interestingness measures with a clinical dataset on hepatitis. In: Proc. of ECML/PKDD-2004, LNAI 3202, 2004; pp. 362–373.

[7] Kloesgen W. Explora: A multipattern and multistrategy discovery assistant. In: Fayyad U.M., Piatetsky-Shapiro G., Smyth P., and Uthurusamy R. (Eds.), Advances in Knowledge Discovery and Data Mining. AAAI/MIT Press, California, 1996; pp. 249–271.

[8] Ali K., Manganaris S., and Srikant R. Partial classification using association rules. In: Proc. of the International Conference on Knowledge Discovery and Data Mining, KDD-1997, 1997; pp. 115–118.

[9] Brin S., Motwani R., Ullman J., and Tsur S. Dynamic itemset counting and implication rules for market basket data. In: Proc. of the ACM SIGMOD International Conference on Management of Data, 1997; pp. 255–264.

[10] Rijsbergen C. Information retrieval, chapter 7, http://www.dcs.gla.ac.uk/Keith/Chapter.7/Ch.7.html, 1979.

[11] Gray B. and Orlowska M.E. CCAIIA: Clustering categorical attributes into interesting association rules. In: Proc. of Pacific-Asia Conference on Knowledge Discovery and Data Mining PAKDD-1998, 1998; pp. 132–143.

[12] Hamilton H.J., Shan N., and Ziarko W. Machine learning of credible classifications. In: Proc. of Australian Conference on Artificial Intelligence AI1997, 1997; pp. 330–339.

[13] Goodman L.A. and Kruskal W.H. Measures of association for cross classification, Springer Series in Statistics, 1, Springer-Verlag, 1979.

[14] Smyth P. and Goodman R.M. Rule induction using information theory. In: Piatetsky-Shapiro G. and Frawley W.J. (Eds.), Knowledge Discovery in Databases, AAAI/MIT Press, 1991; pp. 159–176.

[15] Patetsky-Shapiro G. Discovery analysis and presentation of strong rules. In: Piatetsky-Shapiro G. and Frawley W.J. (Eds.), Knowledge Discovery in Databases, AAAI/MIT Press, 1991; pp. 229–248.

[16] Gago P. and Bento C. A metric for selection of the most promising rules. In: Proc. of European Conference on the Principles of Data Mining and Knowledge Discovery PKDD-1998, 1998; pp. 19–27.

[17] Zhong N., Yao Y.Y., and Ohshima M. Peculiarity oriented multi-database mining. IEEE Transactions on Knowledge and Data Engineering, 15(4), 2003; pp. 952–960.

[18] Witten I.H. and Frank E. Data Mining: Practical machine learning tools and techniques with Java implementations, Morgan Kaufmann, 2000.

[19] Quinlan J.R. Programs for machine learning, Morgan Kaufmann, 1993.

[20] Runmelhart D. E., McClelland J. L.: Parallel Distribute Processing, MIT Press, 1986.

[21] Platt J. Fast training of support vector machines using sequential minimal optimization. In: Schoelkopf B., Burges C., and Smola A. (eds.), Advances in Kernel Methods-Support Vector Learning, MIT Press, 1999; pp. 185–208.

[22] Frank, E., Wang, Y., Inglis, S., Holmes, G., and Witten, I.H. Using model trees for classification. Machine Learning. 32(1), 1998; pp. 63–76.

[23] Holte R.C. Very simple classification rules perform well on most commonly used datasets. Machine Learning, 11, 1993; pp. 63–91.

Address for correspondence

Hidenao Abe

89-1 Enya-cho, Izumo,

Email: abe@med.shimane-u.ac.jp

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

An Automated Data Pattern Translation Process for Medical Data Mining

Anna Shillabeer

Department of Informatics and Engineering, Flinders University of South Australia, Australia

Abstract

This paper describes the development and application of a medical pattern evaluation methodology. The ADAPT process aims to facilitate the increased use of data mining technologies in medicine by providing a method of translating the language of data mining outputs into that of the medical domain. The research behind this work was developed in response to issues raised both in the literature and in public forums regarding the difficulty of adapting data mining technologies for the unique needs of the medical domain. This paper also responds to the need for automated systems to reflect the methodologies and processes applied in the management and analysis of medical data. The solution presented here shows promise as a method of providing a bridge between traditional data mining and the needs of medical domain users.

Keywords:

automatic data processing, computing methodologies, public health informatics.

Introduction

The acceptance of automated data analysis technology in medicine has been slow to develop [1][2]. While acceptance of data mining and associated technologies is growing, they are not yet an integrated part of the medical data analysis toolkit. Many reasons have been documented for this but one of the primary issues is that the needs of the medical domain are not adequately reflected in the technologies available. This has led to the development of specialised mining tools to suit individual users or data types. Whilst this has enabled the application of the technology in specific scenarios, it has also resulted in the development of a range of similar tools which cannot be utilised outside of the purpose for which they were built. This paper presents further work towards a more flexible solution to many of the documented issues, and introduces a novel process to connect the technical data mining domain with the clinical medical domain through the translation of technical outputs into a language which can be better understood and accepted by medical professionals. The process named ADAPT (Automated DAta Pattern Translator) aims to facilitate greater access to mining technologies for all medical users, and to provide an ability to apply some of the more complex mining technologies without the risk of producing irrelevant or incomprehensible outputs. ADAPT is hereby demonstrated through its relevance to the process of diagnostic decision making, theory validation and hypothesis generation but its potential for use in other areas is high due to its intentionally generalised form.

Terminology

This paper utilises technical terminology which is defined for clarification here as it relies upon an understanding of the data mining process and traditional support metric as applied in data mining systems.

Data mining is essentially an automated data analysis process, however there are almost as many formal definitions of data mining as there are data miners. One of the more commonly quoted definitions is that it is a "non-trivial process of identifying valid, novel, potentially useful and ultimately understandable patterns in data" [3]. These patterns are often described through the standard heuristics of support and confidence.

A pattern is defined as a set of attribute values produced through a process of association rule mining, which occur equal to, or greater than, the minimum support required by the user. Patterns take a general format as shown in the following example; Condition 'A' + Treatment 'B' lead to an outcome of 'Recovery Time C' with support of 20%.

A pattern element is a part of a pattern. In the example above, Condition 'A', Treatment 'B' and Recovery Time 'C' are the three elements of the pattern. Treatment 'B' leads to a Recovery Time 'C' is also an element, as it represents only a part of the pattern.

Support essentially quantifies the frequency with which a pattern occurs within the original data source, expressed as a percentage. The minimum support threshold is generally set by the user at runtime. Further to this, elemental support is the support for an isolated element in a pattern. This concept is explained in detail in foundation work by the author [4].

Confidence quantifies as a percentage the degree to which we can rely on the presence of part of a pattern given another part of the same pattern within the data source. For example, given Condition 'A', the confidence states how often we could expect Treatment 'B' to also occur in the same pattern.

Medical diagnostic decision making

The decision making processes of medical practitioners incorporates subjective and objective aspects and diagnosis generally applies one of two methods; by exclusion or by pattern categorisation [5][6][7]. The evidence required for diagnosis by each of these methods is different. Diagnosis by exclusion involves examining the results of objective clinical tests to eliminate potential diagnoses. The alternative method of diagnosis by pattern categorisation incorporates both objective testing and subjective

experience. A diagnosis is developed by comparing the presenting patient's disease pattern and test results to known patterns of disease and treatment. This frequently requires the doctor to discern which pattern elements exert the greatest influence on the diagnostic process and use these parts to match to longer patterns representing disease progression and treatment outcomes.

For automated processes to be applied and accepted it is important that they are able to mirror or facilitate current practices. This requires the provision of knowledge which is able to be both trusted and descriptive from an objective viewpoint and informative in a way which assists in the subjective appraisal of a patient. Unfortunately many data mining outputs (patterns) are presented in a non-intuitive and frequently opaquely coded form. The above diagnostic methods suggest that there is a requirement for heuristic triggers and substantiated hypotheses which can suggest rather than dictate a course of action based upon pattern similarity. An added complication, as described in the pattern categorisation method above, is the need to be able to deconstruct the pattern into more and less influencing pattern elements upon which to base the diagnosis or treatment. If we accept this requirement, it is a logical step to move towards an automated process of identifying and describing the important elements of patterns derived from the application of data mining technologies and it is this precept that defines current work and the development of ADAPT.

The ADAPT process

Data mining outputs are currently deficient in their application to the medical domain as they are not able to identify which pattern elements directly affect the medical outcome or which have no effect and are therefore little more than confounders, although this is necessary if patterns derived from medical data are to assist in processes such as diagnosis as explained earlier. Whilst previous work by the author has provided a proof of concept discussion for achieving this, it was not formalised or applied in a structured, practical sense.

The ADAPT (Automated DAta Pattern Translator) process aims to provide guidance and structure to the application of translating technical outputs into clinically relevant patterns of diagnosis or treatment, and presents a novel approach to post-mining pattern evaluation. There are six steps identified in the process as follows:

1. Identify a set of interesting patterns.
2. Calculate the representative pattern element weightings.
3. Deconstruct the support for each pattern and determine its elemental supports.
4. Determine the positive, negative and inert elements.
5. Order patterns by their degree of representation.
6. Present patterns with their associated heuristics.

Data mining processes that have attempted to evaluate the patterns presented have focussed primarily on step 1, however approaches taken have often been unsuitable for the medical domain due to the wide range of user types and definitions of interest. Medically focussed solutions have been developed but have generally provided a user specific solution. Unfortunately this step alone presents patterns which are unsuitably formatted, and/or do not inform sufficiently well, and require analysis by domain experts to enable them to be considered for clinical application [8][9][10]. The need to supplement information content to aid understanding of the patterns produced as a result of data mining is the focus of steps 2 to 4[1]. Presentation of patterns in a logical and value driven order is the focus of steps 5 and 6. Steps 1 to 4 are described following.

Identify interesting patterns

The ability to define interest is one which has held the focus of data mining researchers since the birth of data mining and several surveys on interestingness measures have been published including [11][12][13][14]. Unfortunately these measures are usually applied in a fixed and limited manner, and there is little scope to develop a more subjective approach to defining what is interesting to a particular user. This is highly important in medicine as there is a wide range of users and user types. This need was recognised by Shillabeer and Roddick [8] who developed a system of role based pattern evaluation which allows each user to define their own subjective opinion on what is interesting and select and apply measures to apply this definition. Whilst defining and applying interest during mining runs involves a variety of statistical processes, identifying interesting patterns post-mining through ADAPT could be as simple as eliminating patterns which do not contain the particular attributes for a specific investigation as the statistical analysis may have been integrated into the mining process.

Representative element weighting

The determination of element representation has been shown to be possible through the application of a technique commonly used in the field of Information Retrieval (IR) [4]. The technique tested and shown to be applicable in this situation is known as TFIDF. TFIDF is a combination of two calculations TF which quantifies the incidence of a term in a focal document and IDF which is the incidence of a term in the document corpus. In combination these theories describe the entropy of a piece of information. The concept of information entropy was introduced by Shannon [15] who defined it as a measure of the average information content associated with a random outcome. Information entropy relates to the amount of uncertainty about an event associated with a given probability distribution. These calculations have been demonstrated to apply in the medical data mining context to reveal the level at which a pattern is representing the data set from which it was produced [4]. In effect, the application of TF can quantify the frequency with which a pattern element occurs within a particular sub-set of patterns of equal length. The concept of IDF can be applied to determine the occurrence frequency of a pattern element within the original data set. Through a comparison of these two values it is possible to show whether the patterns are representative of the source data set or if they are anomalies and therefore potentially less applicable to a general population but more applicable in a specialised population.

1 These concepts have been discussed in more detail in previous work by the author [4] and are paraphrased for clarity in following sections.

Pattern support deconstruction

Support deconstruction is a method by which we can determine the additive effect of each element in a pattern. This is achieved conceptually by calculating the support for an element before and after adding each other element to the pattern. This is done by identifying the participation of the focal element in all relationships, then subtracting the elemental support for each longer element it participates in from its traditional support. . For each focal element (F) the process can be formally denoted as follows:

$$\mathrm{FEsupport}_{\mathrm{length}n} = \mathrm{FTsupport} - \mathrm{Esupport}$$

(E denotes elemental and T traditional).

The effect of applying this formula is demonstrated in Table 1 by deconstructing a fictitious pattern with its support values.

Table 1 - Comparison of traditional and elemental support

Elements	A	AB	ABC	ABCD	TOTAL
Traditional support %	16	10	5	4	N/A
Elemental support %	6	5	1	4	16

The simplified example in Table 1 shows that the effect of adding element C to element AB is negative in that it has a lower elemental support than AB alone. In contrast adding element D to element ABC increases the elemental support of ABC, and hence denotes an increased incidence of the pattern due to the presence of D. From these examples C would be deemed a negative element and D a positive element. An inert element is one which does not significantly increase or decrease the elemental support.

This method has shown promise in defining how much each element affects the outcome described in the pattern [4]. In a real world example this could be applied to reveal the most or least effective combinations of medications or treatments for a particular condition.

Methodology

ADAPT was tested using data collected from the Royal Australasian College of Surgeons National Breast Cancer Audit[2,3].

Association rule patterns were produced from these data to demonstrate that ADAPT is able to perform well on the most problematic mining algorithm for medical data due to the potential for a large number of patterns to be presented. It is hence the area which stands to benefit most from ADAPT.

The original source data held 23,259 patient records, 53,950 separate surgical procedure records for those patients and 82 attributes with an average of 5 potential values per attribute across three tables. After combining patient records with their surgical procedures, allowing for multiple procedures performed together, then removing duplicate columns and extraneous data including dates, 42,705 records with 39 columns were mined using a simple association rule mining tool with support set at 1%.

ADAPT was applied to test the validity of four theories, and provide information relating to one area of interest described in the data dictionary provided with the data. The 5 cases analysed were:

1. "Some treatment recommendations differ by menopausal status, e.g. Ovarian ablation is not indicated after the menopause."
2. "The omission of axillary dissection should be considered only in the case of small primary tumours and is least desirable in pre-menopausal women but is standard in cases of DCIS"
3. "The presence of a significant amount of DCIS in or adjacent to an invasive tumour is a predictor of high relapse rates following wide local excision and radiotherapy.
4. "For DCIS, 'clear margins' as defined by no tumour within 1mm of any inked margins are associated with a high rate of recurrence."
5. It was also suggested that "correlations between excision technique and cosmetic result may be useful."

Interesting patterns were defined as those which contained the particular attribute values required to test the cases - as described in the results section.

Results

Association rules were developed to provide patterns relating to the five cases listed above using the attributes and values detailed in Table 2. For some cases it was necessary to apply different sets of attributes and develop more than one set of patterns for analysis to test each part of a more complex theory in isolation.

The results of the five test cases are detailed in Table 3 which shows elemental supports and TFIDF values. TFIDF was categorised on whether it fell in the upper or lower values for representativeness. Positive elements are shown in bold, negative in italic and inert in standard text and were classified as described earlier. One is denoted as both positive and negative based upon the elements which preceded it. Table 4 shows the descriptive potential of ADAPT based on these heuristics.

Overall, only one of the theories was fully supported by the data and two were fully discounted with one gaining partial support. The TFIDF value was able to indicate if this result is generally representative. Where the theory pattern was in the lower level of representativeness it suggests that the theory should not be depended upon and its foundation should be investigated further.

Discussion

Data mining has many documented benefits for the medical domain and much work is being done to provide

2 The supplied data has since been superseded and hence results detailed herein may or may not accurately reflect current data or practices within the management of breast cancer in Australia or New Zealand and hence should not be relied upon in a clinical setting.

3 Acknowledgement and thanks are hereby given to the ASERNIP-S Adelaide team for their permission to use this data.

medicine specific solutions to data analysis problems. However this work has focussed on the adaptation of technologies to manage the complexities and non-uniform nature of medical data and while the management of input is being addressed, research by the author has been unable to find evidence of a similar effort in the management and translation of outputs. This is an issue when non statisticians or computer scientists are attempting to make sense of the technology and its products. It has been reported that medical practitioners in general have little experience in statistical analysis and as a result are unable to apply or comprehend results which do not match their sphere of knowledge and understanding and as such they need assistance in translating or deciphering the meaning or content of statistical analysis which is beyond the complexity of a 0.05 p-value or confidence interval of 95% [4][16]. This has a potential two pronged effect, it puts data mining technologies and the like out of the reach of many medical practitioners and, it necessitates the application of only the most basic analytical tools thus negating the potential of the more powerful tools.

The contribution of the work presented here is in its ability to broaden the applicability and marketability for the technology by making outputs approachable to all user types and levels without minimising the complexity or range of processes available. It does not attempt to modify the data mining process as the algorithms and statistical methods available currently are generally applicable and effective for the medical domain and have been used successfully by many medical teams. However these projects are often undertaken by specialists who understand the tools and technologies or they have required manual post mining translation or interpretation of outputs by domain specialists [1][10]. The results presented here show that ADAPT is able to facilitate the ordering and presentation of complex outputs in a more intuitive language and provide the knowledge items required for decision making as described earlier. It is also able to overcome some of the more potent criticisms of the technology, for example the belief that the results of data mining are not always representative of the data set from which they were created and they are often misreported as a result [17][18][19]. Whilst ADAPT pattern descriptions are currently manually created, a future challenge will be to create an automated natural language description of patterns based on the heuristics provided by the process.

Conclusion

Results have been developed from patterns which are far simpler that the technology is capable of providing; however the theories were often binary and did not require the production of multi-faceted patterns. Data mining would not generally be applied to develop such simple patterns but it has proved to be a suitable demonstration of the process in a realistic application. Also, the focus here was on the application of the ADAPT process rather than on the application of data mining technologies and the simplistic patterns provided an unambiguous and non-complex platform from which to work. The logical next step is to process the results of more complex association rule mining through ADAPT and validate the output through empirical testing. This paper tested and demonstrated the potential for ADAPT on essentially one tailed tests, but the real power of the methodology and its parts will become evident when applied to purely exploratory data mining.

Table 2 - Pattern attributes and values for each case

Case	Focal Element A	Element B	Element C	Element D
1	Menopause Status 1	Ovarian ablation 1	NA	NA
2.1	Ax Dis. 3	DCIS 2	NA	NA
2.2	Ax Dis. 3	Menopause Status 1	NA	NA
2.3	Size > 15	Aux Dis. 0	NA	NA
3.1	Relapse 2	EIC 1	Surgery 2	Radio 1
3.2	Relapse 1	EIC 1	Surgery 2	Radio 1
4	Relapse 2	DCIS 2	Clear mar.	NA
5.1	Surgery	Symmetry3	NA	NA
5.2	Surgery	Symmetry1	NA	NA

Table 3 - Elemental supports and TFIDF heuristics

Case	A	AB	AC	AD	ABC	ACD	ABD	ABCD	TF/IDF	Theory proven/ Representative
1	59.85	0							NA	Yes/*NA*
2.1	10.01	4.61							0.1975	No/upper – generally applicable
2.2	12.88	13.04							0.1421	Yes/lower – conditionally applicable
2.3	18.04	0							0.1837	Yes/upper – generally applicable
3.1	2.27	0	0	0					NA	No/*NA*
3.2	12.62	2.27	2.44	4.18	0.53	2.04	9.91	1.46	0.0704	No/lower – conditionally applicable
4	0.22	1.03	0						NA	No/*NA*

5.1	30.98	1.02							0.1488	*NA*/lower – conditionally applicable
5.2	13.59	7.2							0.1510	*NA*/lower – conditionally applicable

Table 4 - Comparison of data mining pattern output and ADAPT description output

Case	Traditionally formatted data mining patterns for each case	ADAPT description
1	"Meno_Status":2.00, "Ovarian_Ablation":2.00 (49.48%) "Meno_Status":2.00, "Ovarian_Ablation":9.00 (9.83%)	No post-menopausal patients received an ovarian ablation.
2.1	"Insitu_Necrosed":2.00 (14.62%) "Insitu_Necrosed":2.00, "Ax_Type":3.00 (4.61%)	Axillary Dissection is equally applied and omitted in DCIS cases.
2.2	"Meno_Status":1.00 (25.92%) "Meno_Status":1.00, "Ax_Type":3.00 (13.04%)	Half of all pre-menopausal women have axillary dissection and there was no level of omission.
2.3	"Tumor_Size":20.00, "Ax_Type":3.00 (3.10%) "Tumor_Size":0.00, "Ax_Type":0.00 (10.35%)	No cases of larger tumours avoid axillary dissection.
3.1	None	Relapse is not paired with relevant attributes.
3.2	"Surgical_Event":2.00, "Status":1.00, "EIC_Status":1.00, "Radiotherapy":1.00 (1.46%)	Non relapse is paired with the theoretical pattern.
4	None	Clear margins with DCIS are not an indicator of relapse.
5.1 5.2	"Surgical_Event":5.00, "Symmetry":1.00 (1.49%) "Surgical_Event":4.00, "Symmetry":3.00 (1.02%) "Surgical_Event":4.00, "Symmetry":1.00 (15.44%) "Surgical_Event":3.00, "Symmetry":1.00 (3.53%) "Surgical_Event":2.00, "Symmetry":1.00 (24.44%) "Surgical_Event":1.00, "Symmetry":1.00 (7.20%)	Only total mastectomy was associated with a poor result. Open biopsy has ranked highest for good results >50%

References

[1] Imberman SP, and Domanski B. Using dependancy/associa tion rules to find indications for computerised tomography in a head trauma dataset. Artificial Intelligence in Medicine. September 2002: Vol. 26: 55-68.

[2] Hagland M. Data mining. Healthcare Informatics Online 2004: 21:33-6. Accessed 28/3/2005.

[3] Fayyad U.M, Piatetsky-Shapiro G, and Smyth P. From Data Mining to Knowledge Discovery in Databases. AI Magazine. 1996: Vol 7 Issue 3: 37-54.

[4] Shillabeer A, and Pfitzner D. Determining pattern element contribution in medical datasets. In: ACSC, Vol.29. No.7. ACSW Frontiers 2007; pp 209-216.

[5] Merck & Co. Medical Decision Tests. www.merck.com / mmhe/print/sec25/ch300/ch300.html Accessed 14/9/2006.

[6] Frenster JH. Matrix Cognition in Medical Decision Making. In: Proceedings of the AAMSI Congress on Medical Informatics. San Fransisco: Volume 7 1989; pp 131-134.

[7] Elstein AS, and Schwartz A. Evidence Base of Clinical Diagnosis. BMJ March 2002: Vol. 324: 729-732.

[8] Shillabeer A, and Roddick JF. Towards Role Based Hypothesis Evaluation for Health Data Mining. Electronic Journal of Health Informatics 2006:1(1):1-8.

[9] Gebski V, and Keech A. Statistical methods in clinical trials. Medical Journal of Australia. 2003: 178: 182–184.

[10] Moser SA, Jones WT, and Brossette SE. Application of Data Mining to Intensive Care Unit Microbiological Data. Emerging Infectious Diseases. 1999:5:3: 454-457.

[11] Hilderman RJ, and Hamilton HJ. Evaluation of interestingness measures for ranking discovered knowledge. In: 5th PAKDD 2001. Volume 2035 of LNCS., Hong Kong, China: Springer 2001; pp247–259.

[12] Geng L, and Hamilton HJ. Interestingness Measures for Data Mining: A Survey. ACM Computing Surveys 2006; Vol.38(3): Article 9.

[13] Freitas A. On rule interestingness measures. KBS 12 1999. pp309–315.

[14] Hilderman RJ, and Hamilton HJ. Knowledge Discovery and Interestingness Measures - A survey. www.cs.uregina .ca / Research/Techreports/9904.ps Accessed 23/9/2004.

[15] Shannon C.E. A Mathematical Theory of Communication Bell Systems Technical Journal, July, Oct. 1948: Vol.27:.379-423, 623-656.

[16] Link TM, and Marz R. Computer literacy and attitudes towards e-learning among first year medical students. BMC Medical Education. June 2006: 6:34.

[17] Milloy S. Science Without Sense. The Risky Business of

[18] Public Health. Cato Institute, Washington DC. 1995.

[19] Raju S. Data Flaws. American Council on Science & Health. www.healthfactsandfears.com/high_priorities/vs/ 2003/data072303.html Accessed 23/9/2003.

[20] Smith GD, and Ebrahim S. Data dredging – bias or confounding. BMJ Dec. 2002: Vol.325:1437-1438.

Address for correspondence

Anna Shillabeer
School of Informatics and Engineering
Flinders University of South Australia
P O Box 2100, Adelaide SA 5001
Email: anna.shillabeer@infoeng.flinders.edu.au

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

A Data Mining Approach to Analyze Non-compliance with a Guideline for the Treatment of Breast Cancer

Amir R. Razavi [a], Hans Gill [a], Hans Åhlfeldt [a], Nosrat Shahsavar [a,b]

[a] *Department of Biomedical Engineering, Division of Medical Informatics, Linköping University, Sweden*
[b] *Regional Oncology Centre, University Hospital, Linköping, Sweden*

Abstract

Postmastectomy radiotherapy (PMRT) is prescribed in order to reduce the local recurrence of breast cancer and improve overall survival. A guideline supports the trade-off between benefits and adverse effects of PMRT. However, this guideline is not always followed in practice. This study tries to find a method for revealing patterns of non-compliance between the actual treatment and the PMRT guideline.

Data from breast cancer patients admitted to Linköping University Hospital between 1990 and 2000 were analyzed in this study. Cases that were not treated in accordance with the guideline were selected and analyzed by decision tree induction (DTI). Thereafter, four resulting rules, as representations for groups of patients, were compared to the guideline.

Finding patterns of non-compliance with guidelines by means of rules can be an appropriate alternative to manual methods, i.e. a case-by-case comparison when studying very large datasets. The resulting rules can be used in a knowledge base of a guideline-based decision support system to alert when inconsistencies with the guidelines may appear.

Keywords:

breast cancer, data mining, decision tree induction (DTI), guideline, postmastectomy radiotherapy (PMRT).

Introduction

The aim of postmastectomy radiotherapy (PMRT) of the chest wall and the regional lymphatics is to reduce recurrence of cancer in the chest wall and improve overall survival [1, 2]. Currently, many patients with breast cancer are treated with breast conserving surgery rather than mastectomy. However, a considerable percentage of women still require or choose the surgical procedure of mastectomy.

Clinical benefits related to PMRT depend on the treatment technique and the risk of breast cancer recurrence. Patients with a large primary tumor (T3), and those with four or more positive axillary nodes or pectoral fascia involvement are at substantial risk for locoregional recurrence, and are often treated with PMRT [3]. However, there is a trade-off between the benefits and adverse effects of this treatment. The proportion of patients who are likely to benefit from this treatment is larger when the risk for locoregional recurrence is high. Furthermore, inclusion of the heart or lungs in the target volume may result in an increased number of non-cancer deaths and a risk for a secondary cancer [3, 4]. Therefore, guidelines based on evidence obtained from different studies are used for prescribing PMRT to patients who have undergone mastectomy as their primary surgical treatment [5, 6].

In general, clinical guidelines are developed to help physicians make decisions about appropriate treatment for specific circumstances [7]. Guidelines can result in improvements in overall healthcare, including clinical practice, and can provide decision support tools for practitioners [6].

Despite their many advantages, physicians' adherence to clinical guidelines is not 100 percent. A number of reasons for this have been studied, such as physicians' disagreement with guidelines [8], poor availability [9], and low outcome expectancy on the part of clinicians [10]. Some of these barriers are also associated with patient-related obstacles [11].

Non-compliance with the guideline in some individual cases is of less general importance. These patients may have refused to accept the treatment. In some cases, physicians prescribe the treatment, despite the fact that it is not recommended or suggested by a guideline, because they believe that the treatment is appropriate for that particular patient. However, it is essential to analyze and investigate systematic disagreement between the actual treatment and the clinical guideline recommendations [12]. If repetitive patterns are identified, then they can be used as rules to alert physicians to follow the related guideline more carefully.

Additionally, there may be useful information in these disagreements. Every institution has its own experience with such treatments and these repetitive patterns can provide the writers of guidelines with new insight that may result in improvements in the guidelines. These patterns may even show that it is necessary to check more clinical information regarding a specific treatment.

Table 1 - Study variables and some statistical information about them

Variable	Valid	Missing	Mean		SD	
			BHMVs	AHMVs	BHMVs	AHMVs
Age	858	0	65.27	-	14.09	-
No. of involved LNs*	774	84	2.96	2.91	4.97	4.96
Tumor size *	847	11	26.78	26.79	16.79	16.78
Multiple tumor†	815	43	-	-	-	-
Tumor location†	858	0	-	-	-	-
Perigland growth *,†	858	0	-	-	-	-
Estrogen receptor	836	22	3.07	3.09	5.03	5.03
Progesterone receptor	835	23	3.26	3.28	5.69	5.69
S-phase fraction	717	141	8.43	8.62	5.94	6.00
DNA index	827	31	1.45	1.45	0.50	0.50
DNA Ploidy	825	33	-	-	-	-
PMRT	760	98	-	-	-	-

Abbreviations: LN: lymph node, BHMVs: Before handling missing values, AHMVs: After handling missing values, SD, Standard deviation, PMRT: Postmastectomy radiotherapy
* from pathology report, † Nominal variable

One method for finding these disagreements is to evaluate each case in terms of the guideline and compare the recommendation to the actual treatment received by the patient. When there is disagreement between the guideline and the real treatment, the reason can be analyzed by studying each case separately. The problem is that if there are too many cases, it will take too long to investigate the reason for disagreements, and therefore an automatic or semi-automatic method would be better. A method that ignores sporadic disagreements and can find repetitive patterns is preferable.

Data mining as a method for discovering meaningful new patterns and trends can be applied to the data to find repetitive patterns in disagreements with the related guideline. In this method, instead of evaluating each case with the guidelines, groups of cases that are similar are evaluated. These groups are described by a set of decision rules after a decision tree is trained with the available dataset.

In this study, we show the level of compliance in prescribing PMRT by analyzing a regional dataset from a breast cancer register. Then a data mining method is applied to the dataset to find patterns of non-compliance between the guideline and the real collected figures for PMRT. This is preceded by filtering the dataset using the guideline rules and selecting the cases that received treatment that did not comply with the guideline recommendations.

Materials and methods

In this study, a dataset during a certain time period was initially selected from a regional breast cancer register. Outliers were omitted and missing values were handled. Then a locally modified guideline for postmastectomy radiotherapy (PMRT) was selected and data were filtered in accordance with that guideline. Only cases that were not treated according to the PMRT guideline were selected. Decision tree induction (DTI) was used as the analysis method to find noteworthy patterns for inappropriate PMRT prescription.

Data source and variables

Data were collected from 962 female patients with the diagnosis of malignant breast cancer. Patients were admitted to Linköping University Hospital, and the earliest patient was diagnosed on 1 January 1990 and the last one on 29 December 2000.

Variables from the tumor marker register were linked and matched by using specific identification numbers. The tumor marker register contains information about tumor markers for breast cancer such as S-phase fraction and

receptors for estrogen and progesterone. After matching different data sources, repetitive cases were identified and omitted. There were 858 cases after this pre-processing step.

Thereafter, missing values for continuous variables were handled using the expectation maximization (EM) method [13]. This was done in SPSS. Cases with missing values for nominal variables and PMRT (98 cases) were omitted. With this step, the number of cases decreased to 759. Table 1 shows some descriptive statistics about the dataset before and after handling missing values. One case was omitted because of an incorrect age at the time of diagnosis.

The guideline

In this study, adherence to a local modification of the guideline for postmastectomy radiotherapy was studied [5]. Some important points in the guideline are shown in Table 2. However, the original guideline was changed at Linköping University Hospital in order to conform with local experiences of physicians who have worked in this field for a long time. The main modifications comprise considering age as an important factor and recommending the treatment for patients younger than 75 years, and considering PMRT for patients with any number of lymph nodes with malignant invasion.

Table 2 - Selected parts of the guideline for postmastectomy radiotherapy

Patients with four or more positive axillary lymph nodes. PMRT is recommended for patients with four or more positive axillary lymph nodes.
Patients with one to three positive axillary lymph nodes: There is insufficient evidence to make recommendations or suggestions for the routine use of PMRT in patients with T1/2 tumors with one to three positive nodes.
Patients with T3 or stage III tumors: PMRT is suggested for patients with T3 tumors with positive axillary nodes and patients with operable stage III tumors.
…

Non-compliance with the PMRT guideline

With knowledge of the actual treatment for each patient, a contingency table was drawn showing the recommendations from the PMRT guideline and whether or not the patient received this treatment. Cases that were not treated in accordance with the guideline were identified and extracted for further analysis. In order to process a comparison between the dataset and the guideline, the text guideline was manually converted to rules. The rules were applied to the dataset as SQL queries to filter cases that were not treated according to the PMRT guideline.

Data mining

In the resulting dataset, repetitive patterns of inconsistencies with the guideline were identified by a data mining method. For this purpose, decision tree induction (DTI) was used to analyze the filtered dataset. The algorithm uses information gain as a heuristic for selecting the variable that will best separate the cases into each outcome. In a decision tree, each internal node denotes a test on variables, each branch stands for an outcome of the test, leaf nodes represent an outcome, and the uppermost node in a tree is the root node. Understandable results of acquired knowledge and fast processing make decision trees one of the most frequently used data mining techniques [14].

In this study, DTI was carried out using the J48 algorithm in WEKA [15]. The J48 algorithm is the equivalent of the C4.5 algorithm developed by Quinlan [16]. Post-pruning was also done to trim the resulting tree.

Results

Table 3 is a contingency table showing the outcomes from the PMRT guideline and whether or not the treatment was done for the patient. Decision tree induction (DTI) was used to analyze 51+74 cases whose treatments were not in accordance with the guideline, with accuracy, sensitivity and specificity equal to 97.6%, 0.97 and 0.96, respectively. Despite training the decision tree with twelve variables (Table 1), just three of them (Age, Tumor size, Number of involved LNs) became important nodes in the resulting tree (Table 4).

The tree is shown in Figure 1, with four leaves and an overall size of seven nodes. In this figure, there are two numbers in parentheses in the leaves. The first number shows the number of cases who reached this leaf and the second shows the number of cases for whom the outcome was not predicted to happen. In the boxes, the numbers in parentheses indicate the number of cases assigned to that node, followed by the number of cases that are incorrectly classified as a result. Each path through the decision tree from the root node to the end leaf is a rule, and therefore the tree can be easily transformed to rules. The resulting rules are shown in Table 4.

Discussion

Increasing guideline adherence is of importance in improving the quality of care [17]. Many ongoing research efforts are focusing on the feasibility of using guidelines to increase the quality of care [18]. Finding the reasons for why adherence to guidelines is not 100% would be useful in healthcare management. However, when it comes to detecting and grouping reasons for non-compliance, it can be rather difficult to review each case separately and then categorize the reasons later. This would be exhausting, especially when the number of cases is very large, such as in national registers, and a more efficient method is therefore preferable. Svatek et al. [12] used a rather similar approach for finding non-compliance patterns with the 1999 WHO/ISH hypertension guideline. They used a lim-

ited number of cases (48 patients) to detect and visualize association rules for non-compliance patterns.

Advani et al. [19] proposed the Quality Indicator Language (QUIL) for modeling and creating a quality assessment method to score non-compliances. In contrast to specifying patient-based quality constraints, our approach is used in order to find repetitive patterns while ignoring sporadic disagreements.

By classifying patient data with similar features, an easier and faster methodology has been presented in this study. With this method, good accuracy (97.6%) can be achieved for the resulting rules partition, and rules representing outliers or exceptions can be removed by pruning. Pruning is an intrinsic part of the C4.5, and its equivalent in WEKA (J48), and it is implemented to decrease the size of the tree and make it more robust.

Table 3 - Contingency table for PMRT and decision according to PMRT guideline

		Guideline PMRT	
		-	+
Datset PMRT	-	356	51
	+	74	278

Prior to data mining, the data are filtered by the guideline in order to limit the cases to those with inappropriate treatment compared to the guideline. This is done because the aim is not to predict the prescription of PMRT in new cases but to find patterns of non-compliance with the PMRT guideline.

Table 4 - Rules transformed from the decision tree

1- IF age is younger than 74 years **and**
no LN involvement is present **and**
tumor size is smaller than or equal to 43 mm
THEN PMRT+
2- IF age is older than 74 years
THEN PMRT+
3- IF age is younger than 74 years **and**
no LN involvement is present **and**
tumor size is larger than 43 mm
THEN PMRT-
4- IF age is younger than 74 years **and**
Some LNs are involved (>=1)
THEN PMRT-

Abbreviation: LN: lymph node, PMRT: postmastectomy radiotherapy.

Non-compliance with the PMRT guideline can have different reasons, and these should be carefully studied. Rule number 4 (Table 4) concerns patients younger than 74 with some involved LNs but no PMRT prescription, which is pronounced non-compliance with the guideline.

Occasionally, non-adherence is due to the cut-offs in a guideline. For instance, if the patient is 76 years of age and the cut-off in the guideline is 75 years, should the clinician omit the treatment? If he/she prescribes PMRT for this patient, can this be considered non-adherence to the guideline? Alternatively, when the tumor size is 45 mm, and not larger than 50 mm as stipulated by the guideline for receiving the treatment, is it acceptable to give the treatment? These examples show that the clinician's flexibility in prescribing PMRT can also result in non-compliance with the guideline, but this is not pronounced non-compliance. However, fuzzy set theory can be used to handle this type of non-compliance. Representing states of variables by fuzzy sets in order to quantify them into different intervals can be a good modification of the resulting rules [20].

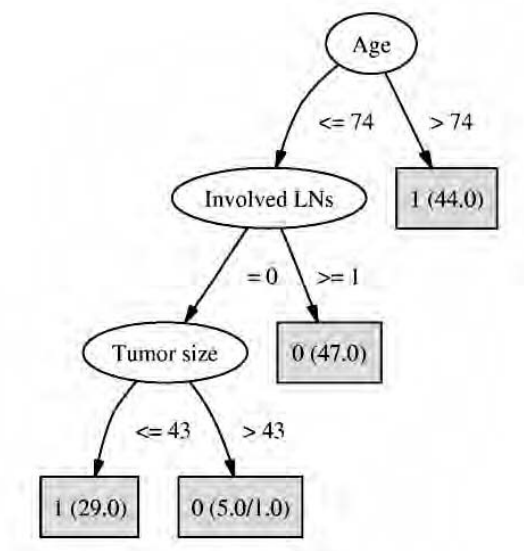


Figure 1 - The resulting decision tree.

In decision support systems, alerts play an important role in increasing the efficiency of healthcare systems [21]. Supporting clinicians in following the guidelines can be done by using the resulting rules as alerts for identifying inconsistencies between clinicians' practices and guidelines. Resulting rules from mining historical data can also be embedded in knowledge bases of guideline-based decision support systems.

In future work, researchers are going to study patient records for the same cases with inconsistencies with the guideline to determine the detailed reasons for non-compliance and any clinical consequences. The reasons will be categorized and compared with the rules from DTI. Furthermore, combining this with fuzzy set theory to analyze

non-compliance with the guideline seems to be a good modification of our current approach.Conclusion

Analyzing cases in which there is non-compliance with the guideline using decision tree induction can reveal patterns of systematic inconsistency between the actual treatment and PMRT. Grouping patients according to decision tree rules makes it easier and faster to categorize reasons for non-compliance as compared to a case-by-case analysis. By embedding the resulting rules in a knowledge base, they can also be used in constructing guideline-based decision support systems to enhance compliance with guidelines.

Acknowledgments

This study is supported by grant No. F2003-513 from FORSS, The Health Research Council in the South-East of Sweden. Special thanks to the South-East Swedish Breast Cancer Study Group for fruitful collaboration and support in this study.

References

[1] Lee JC, Truong PT, Kader HA, Speers CH, Olivotto IA. Postmastectomy radiotherapy reduces locoregional recurrence in elderly women with high-risk breast cancer. Clin Oncol (R Coll Radiol). 2005 Dec;17(8):623-9.

[2] Overgaard M, Hansen PS, Overgaard J, Rose C, Andersson M, Bach F, et al. Postoperative radiotherapy in high-risk premenopausal women with breast cancer who receive adjuvant chemotherapy. Danish Breast Cancer Cooperative Group 82b Trial. N Engl J Med. 1997 Oct 2;337(14):949-55.

[3] Fowble B. Postmastectomy radiation: then and now. Oncology (Williston Park). 1997 Feb;11(2):213-34, 39; discussion 39-40, 43.

[4] Tahir M, Hendry P, Baird L, Qureshi N, Ritchie D, Whitford P. Radiation induced angiosarcoma a sequela of radiotherapy for breast cancer following conservative surgery. Int Semin Surg Oncol. 2006;3:26.

[5] Nationella riktlinjer för behandling av bröstcancer, ver 1.1 och vårdprogram för sydöstra sjukvårdsregionen, ver 1.3. [PDF] April 2006 [cited 1996 28 November]; Available from: http://www.lio.se/us/onkologisktcentrm

[6] Recht A, Edge SB, Solin LJ, Robinson DS, Estabrook A, Fine RE, et al. Postmastectomy radiotherapy: clinical practice guidelines of the American Society of Clinical Oncology. J Clin Oncol. 2001 Mar 1;19(5):1539-69.

[7] Institute of Medicine (U.S.). Committee on Clinical Practice Guidelines., Field MJ, Lohr KN, Institute of Medicine (U.S.). Committee on Clinical Practice Guidelines. Guidelines for clinical practice: from development to use. Washington, D.C.: National Academy Press; 1992.

[8] Wall RJ, Dittus RS, Ely EW. Protocol-driven care in the intensive care unit: a tool for quality. Crit Care. 2001 Dec;5(6):283-5.

[9] Feldman EL, Jaffe A, Galambos N, Robbins A, Kelly RB, Froom J. Clinical practice guidelines on depression: awareness, attitudes, and content knowledge among family physicians in New York. Arch Fam Med. 1998 Jan-Feb;7(1):58-62.

[10] Atkinson LR, Escobar GJ, Takayama JI, Newman TB. Phototherapy use in jaundiced newborns in a large managed care organization: Do clinicians adhere to the guideline? Pediatrics. 2003 May 1;111(5):E555-E61.

[11] Cabana MD, Rand CS, Powe NR, Wu AW, Wilson MH, Abboud PA, et al. Why don't physicians follow clinical practice guidelines? A framework for improvement. JAMA. 1999 Oct 20;282(15):1458-65.

[12] Svatek V, Riha A, Peleska J, Rauch J. Analysis of guideline compliance--a data mining approach. Stud Health Technol Inform. 2004;101:157-61.

[13] McLachlan GJ, Krishnan T. The EM algorithm and extensions. New York: John Wiley & Sons; 1997.

[14] Razavi AR, Gill H, Åhlfeldt H, Shahsavar N. A Data Preprocessing Method to Increase Efficiency and Accuracy in Data Mining. In: Miksch S, Hunter J, Keravnou E, editors. 10th Conference on Artificial Intelligence in Medicine; 2005 July 23-27; Aberdeen, UK: Springer-Verlag GmbH; 2005. p. 434 - 43.

[15] Witten IH, Frank E. Data Mining: Practical machine learning tools with Java implementations. San Francisco: Morgan Kaufmann; 2000.

[16] Quinlan JR. C4.5: Programs for Machine Learning. San Mateo, CA: Morgan Kaufmann; 1993.

[17] Dickey B, Normand SL, Eisen S, Hermann R, Cleary P, Cortes D, et al. Associations between adherence to guidelines for antipsychotic dose and health status, side effects, and patient care experiences. Med Care. 2006 Sep;44(9):827-34.

[18] Milchak JL, Carter BL, James PA, Ardery G. Measuring adherence to practice guidelines for the management of hypertension: an evaluation of the literature. Hypertension. 2004 Nov;44(5):602-8.

[19] Advani A, Shahar Y, Musen MA. Medical quality assessment by scoring adherence to guideline intentions. Proc AMIA Symp. 2001:2-6.

[20] Razavi AR, Nystrom M, Stachowicz MS, Gill H, Ahlfeldt H, Shahsavar N. An approach for generating fuzzy rules from decision trees. Stud Health Technol Inform. 2006;124:581-6.

[21] Steele AW, Eisert S, Davidson A, Sandison T, Lyons P, Garrett N, et al. Using computerized clinical decision support for latent tuberculosis infection screening. Am J Prev Med. 2005 Apr;28(3):281-4.

Address for correspondence

Corresponding author: Amir R. Razavi, Linköping University, Dept. of Biomedical Engineering, Div. of Medical Informatics, University Hospital, S-58185 Linköping, Sweden.
amirreza.razavi@imt.liu.se

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Bayesian Networks for Multivariate Data Analysis and Prognostic Modelling in Cardiac Surgery

Niels Peek[a], Marion Verduijn[a,b], Peter M.J. Rosseel[c], Evert de Jonge[d], Bas A. de Mol[b,e]

[a] *Dept. of Medical Informatics, Academic Medical Center, Amsterdam, The Netherlands*
[b] *Dept. of Biomedical Engineering, University of Technology, Eindhoven*
[c] *Dept. of Anesthesia and Intensive Care, Amphia Hospital, Breda*
[d] *Dept. of Intensive Care Medicine, Academic Medical Center, Amsterdam, The Netherlands*
[e] *Dept. of Cardio-thoracic Surgery, Academic Medical Center, Amsterdam, The Netherlands*

Abstract

Prognostic models are tools to predict the outcome of disease and disease treatment. These models are traditionally built with supervised machine learning techniques, and consider prognosis as a static, one-shot activity. This paper presents a new type of prognostic model that builds on the Bayesian network methodology that implements a dynamic, process-oriented view on prognosis. In contrast to traditional prognostic models, prognostic Bayesian networks explicate the scenarios that lead to disease outcomes, and can be used to update predictions when new information becomes available. A recursive data analysis strategy for inducing prognostic Bayesian networks from medical data is presented, and applied to data from the field of cardiac surgery. The resulting model outperformed a model that was constructed with off-the-shelf Bayesian network learning software, and had similar performance as class probability trees.

Keywords:

prognosis, statistical models, Bayesian networks, thoracic surgery

Introduction

Prognostic models have become important instruments in medicine. Given information on the condition of a patient, they predict the future occurrence of medical events (e.g., complications, recurrence of disease) and outcomes (e.g., survival). These models are currently used for prediction purposes at levels that range from individual patients (where their predictions help doctors and patients to make treatment choices) to patient groups (where they support health-care managers in planning and allocating resources) and patient populations (where they provide for case-mix adjustment in evaluations of delivered care) [1,2].

Prognostic model are usually induced from historical data by applying supervised data analysis methods such as multivariate logistic regression analysis or decision tree induction. This approach has three limitations. First, supervised data analysis methods apply attribute selection before inducing a model, often removing many attributes that are deemed relevant for prognosis by users of the model (e.g., clinicians). Second, the resulting models regard prognosis to be a one-shot activity at a single, predefined point in time. In reality, however, expectations with respect to a patient's future will regularly change as new information comes available during a disease or treatment process. And third, these models impose fixed roles of predictor (independent variable, input) and predictee (dependent variable, output) to the attributes involved. This ignores the dynamic nature of care processes, where today's outcome helps to predict what will happen tomorrow.

This paper introduces a new type of prognostic model, based on the Bayesian network methodology [3,4], that overcomes these limitations. Since their introduction in the 1980s, a large number of Bayesian networks have been developed for medical diagnosis and therapy selection (e.g., [5,6]). Prognostic applications of Bayesian networks have however been rare [7]. We describe the advantages of prognostic Bayesian networks (PBNs) over conventional prognostic models, and present a dedicated algorithm for inducing this type of network from data. The algorithm is illustrated with an application in the field of cardiac surgery. We compare the predictive accuracy of the PBN that was developed with our algorithm with a Bayesian network that was developed using off-the-shelf software, and with three CART [8] models that were developed based on the same data.

Materials and methods

Prognostic Bayesian networks

Suppose that we wish to build a prognostic model for a particular surgical procedure, using historical data from patients who underwent this procedure. The dataset includes attributes regarding diagnosis, presurgical condition, type of surgery, course of the operation, postsurgical condition, and outcome (at hospital discharge) of each patient. Generally speaking, the surgery type will depend on the diagnosis, the patient's presurgical condition will influence how the operation proceeds, and together with

the type of surgery this will largely determine the patient's condition after surgery. The patient's outcome, finally, will depend on that postsurgical condition. These relations are listed in Table 1. Scenarios of complicated and uncomplicated surgeries are typically cast in terms of these variables and their relations.

Now suppose we build a prognostic model (to predict patient outcomes) from these data using a conventional supervised learning method (e.g. logistic regression analysis). This method first requires us to choose the moment of prediction, e.g. at hospital admission (prior to surgery), or just after the operation. In the first case, the attributes regarding the course of the operation and the patient's postsurgical condition are excluded from the model, as they are not available at the chosen prediction moment. Furthermore, the patient's diagnosis would be removed by attribute selection because it has no value for predicting the outcome once we know the surgery type. The resulting model would therefore predict outcome in terms of presurgical condition and type of surgery (Table 1). Similarly, the postsurgical model would just use the patient's postsurgical state to predict the outcome, as information from earlier phases would not have additional predictive value.

Table 1 – Relations between attributes in surgery example

Causal relations	diagnosis → surgery type surgery type, presurgical condition → operation course surgery type, operation course → postsurgical condition postsurgical condition → outcome
Presurgical model	presurgical condition, surgery type → outcome
Postsurgical model	postsurgical condition → outcome

Summarizing, such a conventional method forces us to reduce the prognostic task to a one-shot, static activity. It creates a simplified representation of the problem domain that supports making prediction at a single moment; updating prognostic expectations when new information comes available is not possible. Because many attributes are not included into these models, their form is couter-intuitive and often misunderstood by domain experts.

This paper puts forward a new type of prognostic model, that builds on the theory of Bayesian networks; a sample network for the surgery example is shown in Fig. 1. Because Bayesian networks allow us to explicitly model, and reason with, probabilistic independency relations, domain attributes can be included into this type of model even when their prognostic influence is indirect and depends on subsequent events. For instance, the network shown in Fig. 1 includes 'diagnosis', even though it has no direct influence on the outcome. Furthermore, Bayesian reasoning methods allow us to make predictions at each moment during the care process, thus doing justice to the dynamic nature of prognosis in clinical medicine. Using the example network, we can make outcome predictions prior, during, and after surgery, for the same patient.

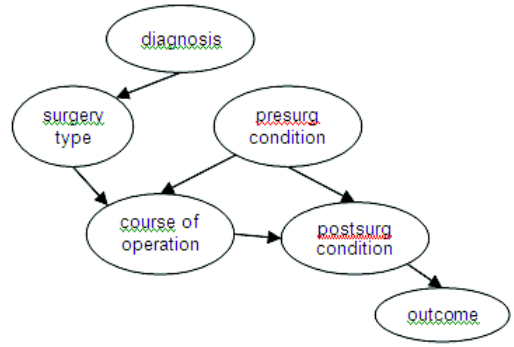


Figure 1 – Bayesian network for the surgery example

Learning Algorithm

Mathematically speaking, a Bayesian network B consists of a pair (G,), where G is a directed acyclic graph, and is a set of conditional probability functions. Each node in the graph G represents one attribute in the application domain; the arcs in the graph represent conditional independency relations between the attributes. For each attribute X_i, there exists a function $_i$ in that describes the estimated conditional probability distribution of X_i given its parents (X_i) in graph G.

In the past 15 years, various algorithms for learning Bayesian networks from data have been developed [9-11]. In these algorithms, learning the network is considered to be an unsupervised learning problem. Starting from an empty graph G, arcs are recursively added to the model to increase the likelihood

$$L(B \mid D) = \prod_{i=1}^{m} P(\gamma_i \mid D) \tag{1}$$

of the model given the data. Each time an arc $X_i \rightarrow X_j$ is added to the graph, the estimated conditional probability function $_j$ is adjusted to reflect the fact that X_j now depends on X_i. As a result, a greedy search through the space of all possible models is conduced, where preference is given to modelling strong correlations between attributes in the data. No single attribute is considered to be more important than any other attribute.

In prognostic Bayesian networks, there exist an outcome attribute, whose accurate prediction is of principal importance, and preference must be given to this prediction task during the construction of the model. For this reason, we developed a modification of the above learning scheme which starts with constructing a complete conditional probability function for the outcome attribute, and recursively builds up the conditional probability functions for the other attributes. As such, the Bayesian network learning task is approached as a set of supervised learning problems instead of a single unsupervised learning problem.

In brief, our algorithm proceeds as follows. Let $\mathbf{X} = \{X_1,\ldots, X_m\}$ denote the set of domain attributes where X_m denotes the outcome attribute. Let G be an empty graph (i.e., no arcs) with nodes corresponding to the attributes in $\mathbf{X}$. Let $I(X_i;X_j)$ denote the mutual information of attributes X_i and X_j. The algorithm now consists of the following steps:

1. Initialisation: Let Q be a priority queue with the outcome variable X_m as its only element.
2. Repeat
   i) X_i dequeue(Q)
   ii) From the non-descendants of X_i in graph G, select a predictive feature subset S_{Xi} for X_i
   iii) For each X_j in S_{Xi}, add an arc $X_j \rightarrow X_i$ to graph G,
   iv) Build a conditional probability model $_i$ for X_i using the attributes in S_{Xi} as covariates
   v) For each X_j in S_{Xi} and not in Q, enqueue(Q, X_j), with $I(X_j;X_m)$ as priority value until Q is empty.

In steps ii and iv, the dataset is used to select a feature subset and build a predictive model for attribute X_i. Any methods for feature subset selection and model building can be "plugged into" our algorithm here, provided that the resulting model provides conditional probability estimates. Examples of such methods are multivariate logistic regression analysis and class probability tree induction.

In step iii of the algorithm, the graphical part of the Bayesian network is extended with arcs leading from the attributes in S_{Xi} to node X_i. To arrive at a meaningful network model, it is possible to use knowledge of the temporal ordering of domain attributes in this step. This requires the specification of a temporally ordered set of attribute strata. Step ii of the algorithm is then modified to select attributes only from non-descendants of X_i in the graph that are not in higher temporal strata than X_i. As a results, all arcs in the graph will follow the flow of time.

Step v of the algorithm adds the selected features in S_{Xi} to the queue Q unless they were enqueued before. Their rank in the queue is determined by their prognostic value: attributes with a strong correlation to the outcome are given a high priority. Thus, the algorithm recursively builds local predictive models for the the outcome attribute and its predictors. The set of all local models eventually constitutes the quantitative part of the prognostic Bayesian network, .

Data

The learning algorithm was applied to data from 10,114 patients who underwent cardiac surgery in the Amphia Hospital, a teaching hospital in Breda, the Netherlands, between January 1998 and November 2004. This dataset contains preoperative patient characteristics (demography and cardiovasculair condition), details of the surgical procedure, measurements during the first 24h of postoperative ICU stay, occurrence of postoperative complications, and clinical outcomes (death, length of ICU stay). Table 2 lists the attributes of the dataset, grouped in temporal strata that were defined in consultation with three domain experts (PMJR, EdJ, BAdM). Strata 1-3 contain preoperative attributes, strata 4-6 details of the operation, and strata 8-10 postoperative attributes. Stratum 7 and 11 define clinical outcomes: death during surgery, death during recovery after the operation, and death during hospitalization. The latter outcome attribute summarizes the former two:

$$\texttt{hospmort} = \texttt{ORmort} \vee \texttt{postORmort}. \qquad (2)$$

Table 2 – Attributes from the cardiac surgery dataset

Stratum	Name	Description
1	age	age
2	bmi	body mass index
	diabetes	diabetes mellitus
	ejfrac	ejection fraction
	precreat	serum creatinin
	pulmhyp	pulmonary hypertension
	surtype	surgery type (intention to treat)
3	emerg	emergency surgery
4	ecctime	extra-corporeal circ. (ecc) time
5	eccacctime	ecc minus aortic cross-clamped time
6	temp	postsurgical body temperature
7	ORmort	death during surgery
8	bicmin	minimal bicarbonate (ICU)
	creatmax	maximal serum creatinine (ICU)
	fiO2	fraction inspired oxygen (ICU)
	meanbpmax	maximal mean blood pressure (ICU)
	meanbpmin	minimal mean blood pressure (ICU)
	ptt	prothrombine time (ICU)
9	ICUlos24h	ICU length of stay longer than 24h
10	cardcomp	cardiac complications
	infect	infection
	neurcomp	neurological complications
	pulmcomp	pulmonary complications
	mof	multiple organ failure
11	postORmort	death during postoperative recovery
	hospmort	hospital death

Two hundred and seventy-seven of the 10,114 patients died during hospitalization (2.7%): 66 patients died in the operation room, 196 died at the ICU, and the remaining 15 patients died at the nursing ward. From the 9,837 surviving patients, 6,232 (63.4%) were discharged from the ICU within 24 hours.

We used the method of class probability trees from the tree building methodology Classification and Regression Trees (CART) [8] as supervised learning method in our network building procedure. Compared to decision trees, class probability trees estimate the (conditional) probability distribution of outcome classes, instead of predicting the most probable class. All tree models were developed using the

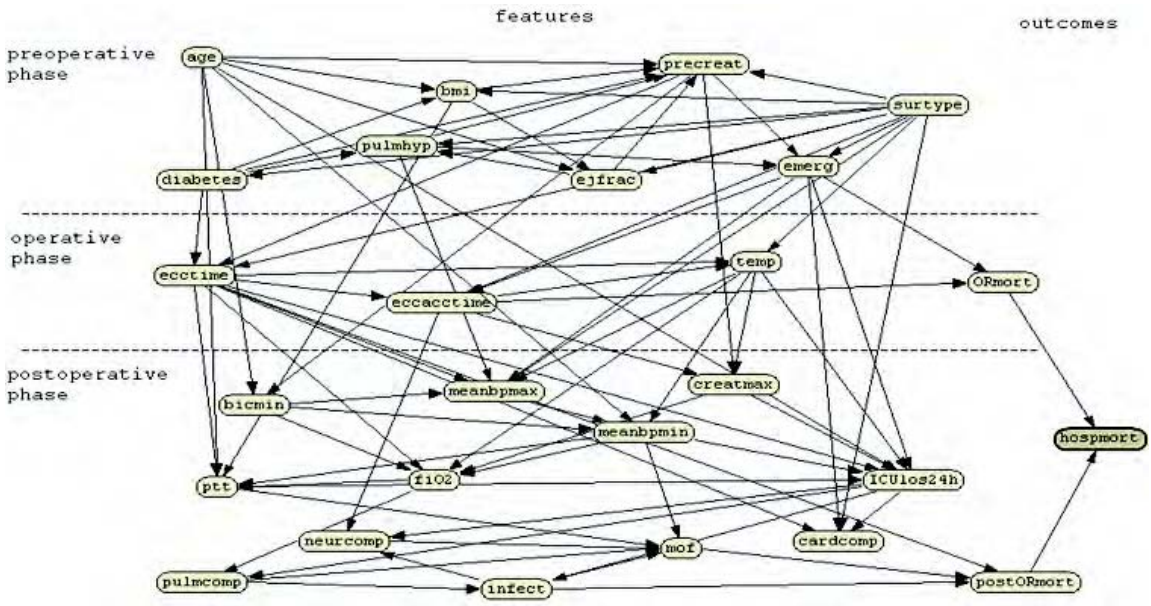


Figure 2 - the prognostic Bayesian network for the domain of cardiac surgery

S-plus library *Recursive Partitioning and Regression Trees* [12], which is an implementation of CART.

Model evaluation and comparison

To evaluate the predictive performance of the network model, the dataset was randomly split into a training set (2/3) and a test set (1/3), and three mortality predictions were made for each patient in the test set: a) using preoperative data only (strata 1-3 in Table 2), b) using preoperative data and information on the course of surgery (strata 1-6), and c) using preoperative data, surgical data, and measurements from the first 24h of ICU stay (strata 1-10).

For each of these three prediction moments, the predictive accuracy of the network model was quantified in terms of area under the ROC curve (AUC) [14]. In the evaluation of postsurgical predictions, patients who died in the operations were removed from the test dataset. In the evaluation of predictions at 24h after surgery, also patients who had died during this first postoperative day were removed.

To place the measured performance in perspective, four additional prognostic models were developed: one Bayesian network was learned using off-the-shelf software (BN Powerconstructor, [11]), and individual tree models were induced for the prediction moments a, b, and c. To enable a fair comparison, the grouping of variables into temporal strata (Table 2) was used to constrain the network topology of the BN Powerconstructor model. The tree models were developed using the class probability tree learning method described earlier, and are regarded as representative for conventional prognostic methodology.

Results

Network model

The prognostic Bayesian network for cardiac surgery is shown in Fig. 2. As appears from the model, the risks of operative death (`ORmort`) primarily depend on elective status of the operation (`emerg`) and duration of surgery (`eccacctime`). Similarly, the risks of postoperative death (`postORmort`) depend on the occurrence of multiple organ failure (`mof`) and infections (`infect`) during postoperative recovery, and again duration of surgery (`ecctime`).

The model incorporates many other relations that are easy to understand from a clinical perspective. For example, patients with low amounts of body fat (low `bmi`) may experience coagulation problems (high `ptt`), leading to an increased risk of multiple organ failure (`mof`). Cardiac surgery requires complete anaesthesia, and therefore a lengthy operation (high `eccacctime`) may yield brain damage (`neurcomp`). A low postsurgical body temperature (`temp`) increases the risks of shock (low `meanbpmin`), which may lead to multiple organ failure and an extended stay at the ICU (`ICUlos24h`).

Evaluation and comparison

When comparing the network from Fig. 2 to the network that was developed using the BN PowerConstructor (BNPC) software, some remarkable differences in the network structure appeared. The BNPC network is relatively sparsely connected with 48 arcs compared to 85 arcs in our network. Furthermore, the sub-outcome `ORmort` has no parent nodes in the BNPC network, and therefore this model cannot predict operative death better than chance.

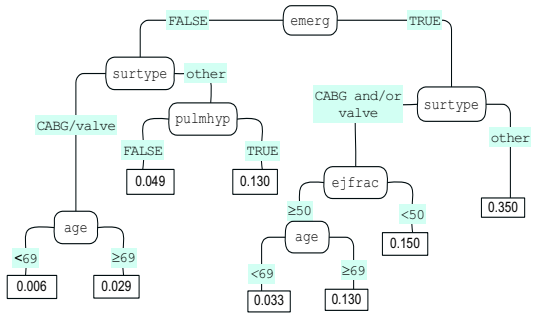


Figure 3 - Tree model for preoperative prediction

The tree model for preoperative mortality prediction is shown in Fig. 3. The model uses the attributes `emerg`, `surtype`, `pulmhyp`, `ejefrac`, and `age` for making predictions. Elective (i.e. non-emergent), uncomplicated procedures (CABG and valve surgery) on patients younger than 69 years have the lowest risk (0.6%), while emergent, complicated operations (e.g. aorta surgery) carry the highest risk (35%).

Table 3 shows the results of the evaluations of the five models on the test set. The network shown in Fig. 2, that was developed with the learning algorithm described earlier, has a moderate predictive accuracy (0.767) prior to the operation, which increases to good (0.834) at 24h after surgery. The PBNC network, however, performs inferior at all three prediction moments. The three class probability trees have similar predictive accuracy as the PBN from Fig. 2.

Table 3 - Predictive performance of models on test set.

	Bayesian networks		tree models
	prognostic learning algorithm	BN Power-Constructor	
prior to operation	0.767	0.693	0.767
at ICU admission	0.782	0.754	0.785
24h after surgery	0.834	0.787	0.834

Conclusion

PBNs overcome the limitations of traditional prognostic models by including all available attributes in the model and making their mutual relationships explicit. This allows for analyzing scenarios that include various stage of care, such as surgery, ICU stay, and subsequent recovery. Furthermore, Bayesian reasoning mechanisms implement a dynamic process-oriented view on prognosis where predictions are updated when new information comes available.

We have presented a learning algorithm PBNs in which local models are recursively learned in a top-down approach, starting at the outcome attribute. In our application, we used class probability trees as local learning method, but other supervised learning methods can be used as well. The PBN for cardiac surgery thus developed outperformed a Bayesian network that was developed using a standard method.

References

[1] Abu-Hanna A, Lucas PJF. Prognostic models in medicine. Methods Inf Med 2001;40:1-5.
[2] Wyatt JC, Altman DG. Prognostic models: clinically useful or quickly forgotten? BMJ 1995;311:1539-41.
[3] Pearl J. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. San Mateo, CA: Morgan Kaufmann, 1988.
[4] Cowell RG, Dawid AP, Lauritzen SL, Spiegelhalter DJ. Probabilistic Networks and Expert Systems. New York: Springer, 1999.
[5] Shwe MA, Middleton B, et al. Probabilistic diagnosis using reformulation of the INTERNIST-1/QMR knowledge base. the probabilistic model and inference algorithm. Methods Inf Med 1991;30:241-55.
[6] Van der Gaag LC, Renooij S, et al. Probabilities for a probabilistic network: a case study in oesophageal cancer. Artif Intell Med 2002;25:123-48.
[7] Lucas PJF, Van der Gaag LC, Abu-Hanna A. Bayesian networks in biomedicine and health care, Artif Intell Med 2004;30:201-14.
[8] Breiman L, Friedman JH, Olshen RA, Stone CJ. Classification and Regression Trees. Monterey: Wadsworth & Brooks, 1984.
[9] Cooper GF, Herskovits E. A Bayesian method for the induction of probabilistic networks from data. Mach Learn 1992;9:309-47.
[10] Heckerman DE, Geiger D, Chickering DM. Learning Bayesian networks: the combination of knowledge and statistical data. Mach Learn 1995;20:197-243.
[11] Cheng J, Greiner R, Kelly J, Bell DA, Liu W. Learning Bayesian networks from data: an information-theory based approach, Artif Intell 2002;137:43-90.
[12] Therneau TM, Atkinson EJ. An introduction to recursive partitioning using the Rpart routines, Technical Report, Mayo Foundation, 1997.
[13] Metz CE. Basic principles of ROC analysis. Sem Nucl Med 1978;8:283-98.

Address for correspondence

Dr. Niels Peek
Dept. of Medical Informatics
Academic Medical Center – Universiteit van Amsterdam
PO Box 22700
NL1100DE Amsterdam
The Netherlands

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Role of Syndromic Management using Dynamic Machine Learning in Future of e-Health in Pakistan

Aijaz Qadir Patoli

Senior Medical Officer Health department government of sindh Pakistan

Abstract

Sexually Transmitted Diseases (STDs) constitute important primary health issues in Pakistan which face inadequacy of resources required in early detection and investigative procedures for their diagnosis and treatment. Syndromic approach to management of STDs is based on the identification of a consistent group of symptoms and syndromes to classify the exact disease or infection beforehand, so that further investigations are sought for based on these initial criteria. This paper envisions the results based on two different approaches: Human and Artificial Intelligence (AI) along with some examples of on-going usage e of Artificial Intelligence in Medicine.

Pakistan is in an early stage regarding the use of informatics in health care for sustainable health system but is also under international obligation to adapt it & WHO EMRO has developed an e-Health plan for the member countries including Pakistan.

The paper presents an informatics application model for a very common but important problem & development of such e-health applications, as starting point, will certainly have positive impact the future of development of e-Health in Pakistan.

Keywords:

sexually transmitted diseases/infections, artificial intelligence, syndromic management, strategic management, machine learning and neural networks

Introduction

Pakistan like other developing countries has been ravaged by sexually transmitted diseases (STDs). In 1996, the World Health Organization (WHO) estimated that, worldwide, there are 333 million new STD infections per year. The largest number of cases occurs in South East Asia (an estimated 150 million new STD cases annually) and sub-Saharan Africa (65 million cases annually) (Adler, 1996). The serious consequences of STDs include infertility, entopic pregnancy, low birth weight, stillbirth, chronic pain in women who have had pelvic inflammatory disease, late symptomatic syphilis, and mutilating lesions for individuals suffering genital ulcers (WHO, 1991). Further consequences include perinatal infections, ophthalmia neonatorum, spontaneous abortion, cervical cancer, and congenital infections (Costello, et al. 1994).

Perhaps the worst consequence, however, is the increased risk of the transmission of HIV with STDs. Improved management of STDs by primary health care workers in the Mwanza region of Tanzania reduced incident cases of HIV-1 infection by 42% in the treatment communities compared to controls. Ghys, et al. (1997) have shown that shedding of HIV in vaginal secretions is strongly correlated with gonoccocal (odds ratio: 1.9, C.I. 1.2-3.0) and chlamydial infections (odds ratio: 2.5, C.I. 1.1-5.8) and with cervical or vaginal ulcers (odds ratio: 3.9, C.I. 2.1-7.4). Furthermore, treatment resulted in a decrease of the genital shedding of the HIV virus from 42% to 21% (Ghys, et al. 1997). Finally, Wasserheit (1997) reviewed a number of studies examining the risk of HIV seroconversion relative to different STDs

Syndromic management is based upon the premise that there are strong correlations between syndromes (signs and symptoms) and the presence of a particular disease.

Thus, the health care provider diagnoses and treats STDs based primarily on the history and physical diagnosis, rather than upon evidence obtained from laboratory tests or microscopy. The WHO algorithms, eight of which are used in the papers reviewed here, involve the diagnosis of urethral discharge, genital ulcer disease, vaginal discharge and lower abdominal pain.[1.]

In a country like Pakistan, patients in need of an immediate diagnostic approach to their problems require the doctors to assemble a solution within as small a time as is possible andconducting test that would otherwise require a huge sum of money.

dset(VG) = {NG, CT, TV, CA, GV}
where,
VG = Vaginal discharge
NG = N. gonnorthoeae
CT = C. trachomatis
TV = Trichomonas vaginalis
CA = Candida albicans
GV = Gardnerella vaginalis

Fortunately enough, the symptoms and signs caused by a number of STDs are similar to be easily identified clinically. A group of such symptoms and signs, easily recognizable under clinical light is called a *syndrome*.

A combination of signs, symptoms, or phenomena seen in association with each other alerts the physician to compile

an early responsive strategy to work with throughout the course of the treatment of patient. This paper will focus more on an automated version of the course of treatment that would not only assist the *STI* specialists, physicians, or Venereologists but, also common Health Service Providers and Family Physicians Development of such application can be the first step & impart growth stimulating influence on the future of e-Health in Pakistan.

Discussion

Syndromic management for early detection of STDs/ STIs[2].

Sexually Transmitted Infections (STIs) commonly occur together. It is acceptable to manage a number of STIs at the same time in countries where there are poor diagnostic facilities. Doing so is cheaper and more effective, as the diagnosis is about the symptoms and not the cause of the disease. Such management is called *Syndromic Management*.

The pervasive stigma associated with sexually transmitted infections (STIs) inhibits open discussion, and leads to widespread misconceptions and unfounded fears of diagnosis and treatment. Syndromic management partly addresses these concerns by allowing for far less invasive diagnostic procedures[1].

Using Syndromic management, health workers and health service providers make a diagnosis from picking up and observing a collection of signs and symptoms. Evaluations are most likely to mean that a person might be suffering from one of a group of STIs that are *ALL* treatable with *ONE* treatment (usually consisting of more than one drug). This is very much like a broad-spectrum antibiotic that is given because it is effective against more than one microorganism.

dset(VG) = {NG, CT, TV, CA, GV}
dset(ON) = {NG, CT, OPO}
dset(Obs.) = dset(VG) U dset(ON)
where,
VG = Vaginal discharge
ON = Opthalmia neonatorum
NG = N. gonnorthoeae
CT = C. trachomatis
TV = Trichomonas vaginalis
CA = Candida albicans
GV = Gardnerella vaginalis
OPO = Other pyogenic organisms

Physicians declare Syndromic Management techniques as a cost-effective way of clinically identifying STI related diseases. In this technique, all possible syndromes associated to Sexually Transmitted Infections (STIs) may be classified beforehand. The list being small is put up in atabular format for doctors to construct associations with diseases, *refer to Table 1*.

Judgments regarding diseases caused by multiple syndromes in a single patient are evaluated based on a list as mentioned in Table 1. The common procedure of identifying the exact cause in question is based on unifying or intersecting the common causes or sets of diseases *(dset)*.

So, if a patient comes with a complaint of *Vaginal Discharge*, the *dset* for that particular syndrome is obtained from such a list. The doctor then perceives that the patient might have a disease from within that certain set of diseases.

However, if the same patient arrives to the doctor with another complaint regarding yet another syndrome, the doctor might have multiple sets of diseases, $dset_2$.

Common perception would lead the physician to understand that the probable causes would be an amalgamation of both the sets, a theory that medical investigations live

Table 1 - List of common STI syndromes along with their associations

STI Syndrome	Common Causes *(dset)*
Urethral discharge	Neisseria gonnorthoeae, Chlamydia trachomatis
Genital ulcer disease	Treponema palidum, Haemophilus ducreyi, Herpes simplex virus
Vaginal discharge	N. gonnorthoeae, C. trachomatis, Trichomonas vaginalis, Candida albicans, Gardnerella vaginalis
Pelvic inflammatory disease	N. gonnorthoeae, C. trachomatis, anaerobic bacteria, other pyogenic organisms, T.B.
Opthalmia neonatorum	N. gonnorthoeae, C. trachomatis, other pyogenic organisms
Bubo	H. ducreyi, C. trachomatis pyogenic infections
Scrotal swelling	N. gonnorthoeae, C. trachomatis, other pyogenic organisms, viruses and surgical conditions

by for the classification of the syndromic data available for STI compilation.

Once a circle has been etched around the potential STIs, they may be treated accordingly. Identifying a list of potential causes of disease is, still, the first step. Properly diagnosing it is another leap forward in the whole process. A little intelligence helps here.

Artificial intelligence

Alan Turing (1912-1954) was a mathematician, a philosopher, a hacker, and a visionary. He was the father of computer science and the founder of the field of artificial intelligence

Turing was the first scientist to believe that computers could be intelligent some day. In 1950, he wrote his classic paper: *'Computing machinery and intelligence'*

In his paper, Turing presented one simple question – *'Can machines think?'* His answer was *'yes'*. But by *'think'*, Turing actually meant, *'Can a machine make people believe that it can think'*.

Humans tend to think of themselves as the only intelligent creatures in the world. It's hard to accept the fact that language is a skill like any other, because if that is so, you are not the only ones who can acquire it.

Alan Turing had a lot to say about the relationship between computers and humans. What seems very natural for us now wasn't true 50 years ago, when Turing lived.

The different research branches in the field of Artificial Intelligence (AI) are in many ways analogous to various functions of the brain.

The foundation for the philosophy of AI was the infamous Turing Test, defined by Turing in 1950. He held that the computers by time would be programmed to acquire abilities rivaling human intelligence.

As part of his argument Turing put forward the idea of an *'imitation game'*, in which a human being and a computer would be interrogated under conditions where the interrogator would not know which was which, the communication being entirely by textual messages. Turing argued that if the interrogator could not distinguish them by questioning, then it would be unreasonable not to call the computer intelligent.

To sustain a live conversation, a machine should be able to effectively understand the different questions put forth and learn from each and every one of them. In his paper, Turing introduced a simple ideal conversation, useful to grasp the concept at hand.

Q: Do you play chess?

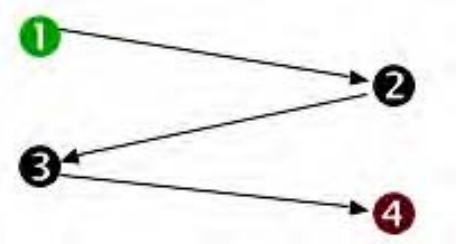
A: Yes.

Q: I have K at K1, and no other pieces. You have only K at K6 and R at R1. It is your move. What do you play?

A: (After a pause of 15 seconds) R-R8 mate.

In this simple example, a computer was asked if it knew how to play chess. The computer said, "Yes". But when a complicated text message is put in the next statement, it was obvious that the computer would think that the statement was related to the previous and it certainly was. The computer understood that it was being asked of a chess move. Such ability of a computer to understand and relate two separate entities after collecting relevant information is called *Dynamic Machine Learning ability.*

Machines that gather information follow a certain path to reach to the acquired information. Such course is referred to as simply *'path'* here, is a series of *'information blocks'* or *'nodes'* linked together.



When asked if the computer played chess, the computer responded affirmative. The computer makes up four information blocks: (1) Human asks about Chess, (2) Chess is game that can be played, (3) Human knows how to play chess, (4) I can play chess too.

Eventually only the two nodes, 1 and 4 would matter. Number 1 being the question and number 4 being the perceived evaluation. (1) Human asks about Chess, (4) I can play chess too corresponds to the reply of the computer in the Turing experiment.

Q: Do you play chess?

A: Yes.

Computers these days are intelligent enough to come to such a conclusion eventually, but the real matter at hand is not coming to the conclusion, but coming to it *'fast'*. For a situation where there are over a thousand nodes, interwoven into a web of *'information blocks'*, a computer might confuse itself into choosing wrong paths and making incorrect decisions.

Choosing the right beginning and ending node is half the challenge. Choosing how to traverse through various others to reach to the end in the shortest possible course and time is what is to be considered at first. Finding the correct shortest path requires the use of *Path-finding algorithms*.

Automated syndromic management

Bringing together all concepts talked about in this paper, we look at Syndromic Management under a new light here. We have come to terms with just how to identify a set of probable causes for the symptoms and syndromes reported for/by a patient.

A step forward in the process would be to pick the correct cause out of the set of all the other causes. Say, the final *dset(Obs.)* obtained in section 2 by unifying *dset (VG)* and *dset (ON)*, we get around six possible common causes for the STI syndromes: Neisseria gonorrhea, Chlamydia trachomatis, Trichomonas vaginalis, Candida albicans, Gardnerella vaginalis, and other pyogenic organisms.

Possible identification of these few diseases helps us to etch a domain around them and to find which the probable cause is.

Now, Syndromic management goes one step further to actually traverse through all the appropriate causes to pin-point the exact cause for the particular STI syndrome in inspection.

A proposed system here would appoint each cause as an *'information block'* in a series of webbed blocks. Each disease would be treated independently and randomly of the other in such a way that as each disease gets treated for, an indication of whether the disease was actually present in the patient gets marked as positive, and vice-versa.

Each positive node would be aligned in a path. Using path-finding algorithms on the given data, an appropriate path is identified for the disease. Such paths serve as a basis for a machine to identify a path. Several such paths are kept in an information storage unit, from which and where they can be obtained and summarized to form information useful for further such cases.

As for the machine, the machine should learn at every consecutive analysis of the patient using its Machine Learning abilities, such as if such a scenario occurs again, the machine would know how to respond to it in faster response time.

Artificial intelligence in medicine

The potential of AI in medicine has been expressed by a number of researchers. Hoong (1988) summarized the potential of AI techniques in medicine as follows:

- Provides a laboratory for the examination, organization, representation and cataloguing of medical knowledge.
- Produces new tools to support medical decision-making, training and research.
- Integrates activities in medical, computer, cognitive and other sciences.
- Offers a content-rich discipline for future scientific medical specialty.

Many intelligent systems have been developed for the purpose of enhancing health-care and provide better health care facilities, reduce cost and etc.[3]

Role in sustainable health systems[4]

Syndromic Management of STDs imparts a positive impact on the objectives of the sustainable Health System.

Strong focus on primary healthcare & prevention, accessibility to health by all irrespective of place time and other handicaps, expertise through expert systems with optimal technology use & appropriate resources usage.

It also facilitates semantic interoperability and among the systems.

Conclusion

Developing Applications that use automated Syndromic Management of STDs will not only bring dramatic improvements in achieving the objectives of the strategy but also will also serve the first stone for use of eHealth in Pakistan where yet an initiative in this regard has to be take.

Such development is eHealth in Pakistan will also aid Knowledge Management operations which encompass certain vital issues like digital divide & e-Health.

With such existing glaring opportunities Pakistan and those countries, which have not yet been able to determine change management process in Health care delivery system, can have uplift.

To start with Pakistan may develop her National Knowledge management strategy in line with WHO EMRO. Such move will open the avenues of e-Health development in Pakistan and capability to development of strategies for sustainable health system in light of recommendations of 05th Regional Conference WHO EMRO in Cairo in June in the current year.[5&6]

Acknowledgments

1. Dr. Rashid ul Bashshur Director Telemedicine Systems Michigan University USA.
2. Dr. Najeeb Al Shorbaji WHO EMRO Cairo.
3. Prof: Evelyn J. S Hovenga Faculty of Business & Informatics Central Queensland University Australia.
4. Ms. Tove Sorensen Head WHO Collaborating Center Tromso Norway

References

[1] "When is Syndromic Management of Sexually Transmitted Diseases Useful? An Analysis of the Literature" Catherine Gergen, Victoria Wilkins, Pratyma Ragunathan, Ph.D., Julia Walsh, M.D. February 2000.

[2] "Syndromic Management of Sexually Transmitted Diseases using Dynamic Machine Learning and Pathfinding algorithms" Arun Reginald ACCP, Dr. Aijaz Qadir Patoli MBBS, MBA, DISM. First International Conference on Information and Communication Technologies Karachi Pakistan 2005.

[3] "Artificial Intelligence in Medical Application: An Exploration" Wan Hussain Wan Ishak, Fadzilah Siraj School of Information Technology, University Utara Malaysia, 06010 Sintok, Kedah, MALAYSIA.

[4] Hovenga E.J.S, Coiera E, Lovis C (2007) Building a Sustainable Health System, Invited Survey paper for the IMIA Yearbook of Medical Informatics 2007.

[5] WHO's Eleventh General Program for Work 2006-2015 adapted in World Health Assembly May 2006; "A Global Agenda for Health" .

[6] Conclusion & Recommendations of the 05th Regional Conference on e-Health June 2006 Dr. Najeeb Alshorbaji.

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Combining Lexical and Semantic Methods of Inter-terminology Mapping Using the UMLS

Kin Wah Fung, Olivier Bodenreider, Alan R. Aronson, William T. Hole, Suresh Srinivasan

National Library of Medicine, Bethesda, Maryland, USA

Abstract

The need for inter-terminology mapping is constantly increasing with the growth in the volume of electronically captured biomedical data and the demand to re-use the same data for secondary purposes. Using the UMLS as a knowledge base, semantically-based and lexically-based mappings were generated from SNOMED CT to ICD9CM terms and compared to a gold standard. Semantic mapping performed better than lexical mapping in terms of coverage, recall and precision. As the two mapping methods are orthogonal, the two sets of mappings can be used to validate and enhance each other. A method of combining the mappings based on the precision level of sub-categories in each method was derived. The combined method outperformed both methods, achieving coverage of 91%, recall of 43% and precision of 27%. It is also possible to customize the method of combination to optimize performance according to the task at hand.

Keywords:

Unified Medical Language System, controlled terminology, inter-terminology mapping.

Introduction

The need for mapping between biomedical terminologies commonly arises when data encoded in one terminology is reused for a secondary purpose that requires a different system of encoding. Imagine an electronic patient record system that captures clinical information using SNOMED CT codes. It will be a big efficiency gain if ICD9CM and CPT codes can be generated automatically for billing purposes. For this to happen, mappings from SNOMED CT to ICD9CM and CPT are required.

A lot of research has been done on the creation of inter-terminology mappings by algorithmic methods. Generally speaking, the algorithms can be divided into lexically-based or semantically-based methods [1-7]. Lexical methods rely on the lexical properties of terms in a terminology. Terms are first normalized or broken down into components before they are compared and matched. On the other hand, semantic methods find matches by utilizing semantic links between terms from the terminologies being mapped.

The UMLS, with over 130 terminologies represented in one common format, is a useful resource for creating inter-terminology mappings. We have previously reported on the use of the UMLS to create semantic mappings between two clinical terminologies [7]. In this report, we describe the use of a UMLS-based tool (MetaMap) to generate lexical mappings between the same terminologies. The performance of the two methods is compared and one way to combine the two methods is described.

Semantic mapping (IntraMap)

The IntraMap algorithm (a modification from the Restrict to MeSH algorithm) makes use of semantic relationships between UMLS concepts to find mappings [7, 8]. Starting from the source concept (the UMLS concept containing the term in a source terminology from which mapping is sought), the algorithm looks for target concepts (UMLS concepts containing terms in the terminology being mapped to) which are related to the source concept either through synonymy or explicit mapping relationships provided by some source terminologies. Failing to find a target concept, the search will widen by using ancestors of the source concept as starting points to look for target concepts. If that fails again, ancestors of the children of the source concept and finally, ancestors of the siblings of the source concept will be used for target concept searching.

Lexical mapping (MetaMap)

MetaMap is a program developed by the NLM to map biomedical text to concepts in the UMLS [9, 10]. The algorithm of MetaMap is as follows: the input text is first parsed into noun phrases. For each phrase, lexical variants are generated. A candidate set of all UMLS concept names containing at least one of the variants is retrieved. Each candidate is then evaluated using a linguistically principled evaluation function. Finally, complete mappings are constructed by combining candidates involved in disjoint parts of the phrase. If one uses the terms in one terminology as the input text and limits the mapping to UMLS concepts containing terms from a specific target terminology, one can use MetaMap to find inter-terminology mappings. MetaMap operates on English terms from the UMLS Metathesaurus.

Selected for best paper award.

Methods

To evaluate the two mapping methods, we used them to find mappings from SNOMED CT terms to ICD9CM terms. As the gold standard for comparison, we used the January 2004 version of the SNOMED CT to ICD9CM mappings provided by the College of American Pathologists. All the mappings from one SNOMED CT term to a single ICD9CM term in the gold standard were used (84% of all mappings). The version of UMLS used was 2004AA which contained the same version of SNOMED CT as used in the gold standard. For semantic mapping by IntraMap, we set the target terminologies to ICD9CM or MTHICD9 (a source that provides additional entry terms to ICD9CM codes). The explicit mapping relationships from the source 'SNOMEDCT' were ignored; otherwise all mappings in the gold standard would be found because the gold standard mappings were also included in 2004AA. For lexical mapping with MetaMap, the fully-specified English names of the SNOMED CT concepts were used as input strings. The option of 'term processing' was turned on to bypass parsing of the terms into component phrases. Other settings of MetaMap were the same as the default mode on the MetaMap website. The target terminologies were limited to ICD9CM or MTHICD9. While no semantic type restriction was used in this experiment, we show in the discussion that restricting the output of MetaMap to semantic types of interest can improve precision.

The mappings found by the two mapping algorithms were compared to the gold standard in terms of their coverage (percentage of SNOMED CT terms for which mappings were found), recall (the percentage of mappings in the gold standard that were found) and precision (the percentage of found mappings that agreed with the gold standard). The overlap between the two sets of mappings was analyzed. A method of combining the two sets of mappings based on the precision ranking of the sub-categories was derived and the improvement in performance was analyzed.

Results

Semantic mapping alone

As reported in [7], among the 66,382 SNOMED CT terms that were used, IntraMap managed to find ICD9CM mappings for 57,293 terms (86.3% coverage). Overall recall was 43.3% and precision was 22.1%. On average, there were 2.3 mappings found per SNOMED CT term. The precision of the sub-categories of mappings were: mapping by synonymy 78.4%, mapping by explicit mapping relationships 50.1% and mapping by ancestor expansion 9.2%. The mappings found by children and sibling expansion were too small in number to warrant further consideration. The results are summarized in Table 1.

Table 1 - Overall and sub-category performance of semantic mapping by IntraMap

	Sub-category of mapping			Overall
	Synonymy	Explicit mapping	Ancestor expansion	
Coverage	19.5%	17.0%	47.2%	86.3%
Recall	16.6%	13.0%	13.0%	43.3%
Precision	78.3%	50.1%	9.2%	22.1%
Mapping per term	1.1	1.5	3.0	2.3

Lexical mapping alone

MetaMap was able to find mappings for 44,452 out of 66,382 SNOMED CT terms (70.0% coverage). The overall recall and precision was 28.4% and 14.7% respectively. There were on average 2.9 mappings found per SNOMED CT term. Among those mappings that were considered to be perfect matches (MetaMap score of 1000), the precision was 85.8%. For those SNOMED CT terms with no perfect matches found, if we only used the top ranking mappings (the mappings with the highest MetaMap score), the precision was 22.6%. The results are summarized in Table 2.

Table 2 - Overall and sub-category performance of lexical mapping by MetaMap

	Sub-category of mapping		Overall
	Perfect mapping	Top mapping	
Coverage	9.7%	57.2%	70.0%
Recall	8.6%	18.3%	28.4%
Precision	85.8%	22.6%	14.7%
Mapping per term	1.0	1.4	2.9

Overlap between the two sets of mappings

A total of 29,468 mappings (distinct pairs of SNOMED CT and ICD9CM codes) were common to both sets of mappings. This represented 22.6% and 22.9% of all IntraMap and MetaMap mappings, respectively. This set of common mappings covered 35.7% of the SNOMED CT terms, with recall and precision of 22.5% and 50.8% respectively. The mappings that were only found in one algorithm but not the other were higher in coverage but lower in precision (Table 3).

Table 3 - Mapping performance according to the method of mapping

	Both IntraMap and MetaMap	Only from IntraMap	Only from MetaMap
Coverage	35.7%	57.4%	51.9%
Recall	22.5%	20.8%	5.9%
Precision	50.8%	13.7%	3.9%
Mapping per term	1.2	2.6	2.9

Altogether 13,797 correct mappings were found by semantic mapping alone and missed by lexical mapping. Among these, mappings found by synonymy, explicit mapping relationship and ancestor expansion constituted 9%, 48% and 43% respectively. One example was the mapping from SNOMED CT term '3072001: Hormone-induced hypopituitarism' to ICD9CM term '253.7: Iatrogenic pituitary disorders'. The failure of MetaMap to find this mapping was expected as it was unlikely that the similarity in meaning between 'hormone-induced' and 'iatrogenic' could be detected by lexical matching alone.

On the other hand, there were 3,906 correct mappings that were found by lexical mapping but missed by semantic mapping. Among these, one was deemed to be a perfect match in MetaMap. This was unexpected as almost all perfect lexical matches were genuine synonyms and should normally belong to the same UMLS concept. If so, the mapping should be identified by IntraMap as well. Indeed, this was an anomaly caused by an editing error in the UMLS. In 2004AA, the ICD9CM term '241.9: Unspecified nontoxic nodular goiter' was assigned to a UMLS concept (C1313958) which was different from the UMLS concept (C1318500) containing the SNOMED CT term '190236006: Non-toxic nodular goiter'. This error has since been corrected and the two terms now belong to the same UMLS concept (C1318500). All the other cases in which the correct mapping was found only by MetaMap had less than perfect MetaMap scores. On inspection of a small sample, many of these were mappings from a narrower to a broader concept. One example was the mapping from the SNOMED CT term '67600007: Vascular-biliary fistula' to the ICD9CM term '576.4: Fistula of bile duct' by way of the synonym 'biliary fistula' in the same UMLS concept. This mapping was not found by IntraMap because the two UMLS concepts containing the two terms were not linked by any hierarchical or mapping relationships in the UMLS.

Combining the semantic and lexical mapping sets

Since semantic and lexical mapping are fundamentally different approaches, they are orthogonal and thus can be used to validate and complement each other. To make use of both sets of mappings simultaneously, we derived a method based on the precision level of each sub-category of mapping. First we created a 'precision ladder' according to the precision of each sub-category of mapping. (Table 4)

Table 4 - Precision ladder according to precision of each sub-category of mapping

Rank	Sub-category	Precision
1	M-PM (MetaMap perfect match)	85.8%
2	I-S (IntraMap synonymy)	78.3%
3	C-O (Combined overlapping)	50.8%
4	I-EM (IntraMap explicit mapping)	50.1%
5	M-TM (MetaMap top score)	22.6%
6	C-IO (Combined IntraMap only)	13.7%
7	I-AE (IntraMap ancestor expansion)	9.2%
8	C-MO (Combined MetaMap only)	3.9%

Next we pooled all the mappings together and arranged them in descending order of precision. If the same mapping appeared in more than one sub-category, only the one in the highest ranking sub-category was kept. If there were multiple mappings for the same SNOMED CT term, only the one with the highest ranking was kept and the alternative lower ranking mappings were discarded. The combined set contained 107,172 mappings for 60,454 SNOMED CT terms (coverage 91.1%). The overall recall and precision of the combined set was 42.9% and 26.6% respectively.

The fact that the mappings were arranged in the order of precision allowed us to further fine-tune the way in which the mappings could be used. By setting different cut-off points on the precision ladder (i.e. ignoring mappings below a certain sub-category) we could obtain different combinations of coverage, recall and precision. As expected, the further down we go on the precision ladder, the higher the coverage and recall but the lower the precision. (Table 5)

The mapping performance did not change further with inclusion of mappings ranking lower than rank 6. This was expected because the mappings in rank 7 (IntraMap mappings found by ancestor expansion) were already included in higher sub-categories (ranks 3 and 6). Mappings from rank 8 did not contribute to the overall performance because rank 1 and rank 5 already covered every SNOMED CT term for which a mapping was found by MetaMap.

The F-score (the harmonic mean of precision and recall) is frequently used as an overall indicator of performance. We calculated the F-scores for each cut-off point with equal weight to precision and recall by the following formula:

$$\text{F-score} = (0.5/\text{precision} + 0.5/\text{recall})^{-1}$$

Looking at the F-scores, rank 4 is the optimal cut-off point if one wishes to optimize equally on recall and precision.

Table 5 - Mapping performance according to the cut-off point on the precision ladder

	Cut-off point on the precision ladder							
	Rank 1	**Rank 2**	**Rank 3**	**Rank 4**	**Rank 5**	**Rank 6**	**Rank 7**	**Rank 8**
Coverage	9.7%	19.5%	38.1%	50.8%	71.0%	91.1%	91.1%	91.1%
Recall	8.6%	16.6%	24.4%	34.0%	38.2%	42.9%	42.9%	42.9%
Precision	85.8%	78.4%	52.0%	51.6%	39.3%	26.6%	26.6%	26.6%
Mapping per term	1.0	1.1	1.2	1.3	1.4	1.8	1.8	1.8
F-score	0.16	0.27	0.33	0.41	0.39	0.33	0.33	0.33

Discussion

The two mapping algorithms

Semantic and lexical mapping algorithms have their own strengths and weaknesses. In general, semantic mapping is considered to be more precise. This is confirmed by our results (precision of IntraMap 22.1%, precision of MetaMap 14.7%). However, for semantic mapping to work there needs to be a pre-existing knowledge base to provide the semantic linkages. In our study, the UMLS is used as the knowledge base, containing over a million concepts and tens of millions of semantic relationships. The performance of semantic mapping depends heavily on the density and quality of these relationships. If there is no semantic relationship linking two concepts a mapping cannot be found.

On the other hand, lexical mapping does not depend on a pre-existing knowledge base. One can perform lexical matching based solely on the lexical properties of the terms from the terminologies being mapped. However, the mapping algorithm of MetaMap does utilize resources from the UMLS which include the rich collection of synonyms and the SPECIALIST lexicon, which have undoubtedly bolstered its performance.

Further refinement

There are possible ways to fine-tune the performance of each of the mapping algorithms. For IntraMap, it is possible to selectively exclude the relationships contributed by certain source terminologies, if such relationships are less likely to result in correct mappings. Restricting the extent of ancestor expansion (e.g., limiting just to level 1 and 2 ancestors) has been found to improve precision [7].

In MetaMap, there is a built-in option to restrict the target concepts by semantic type (STY). In the original run, no restriction on STY was used. In a subsequent run, we restricted target concepts to 18 STYs related to diseases or findings (e.g. Finding, Disease or Syndrome, Acquired abnormality and Congenital abnormality). This restriction was appropriate because the SNOMED CT terms being mapped were all disorders or findings. As ICD9CM also contained terms for procedures, by excluding STYs like Therapeutic or preventive procedure and Diagnostic procedure some incorrect mappings could be avoided. One example of such error was the mapping of the SNOMED CT term '168943003: Abdominal aortogram abnormal' (a finding) to the ICD9CM term '88.42: Aortography' (a procedure). In the re-run, 2,588 mappings that were present before were dropped because of the STY restriction. Among these dropped mappings, most of them (97%) turned out to be incorrect. This resulted in a small increase of precision from 14.7% to 14.9%.

Apart from STY, one can further refine the mappings by term type (TTY) information in the UMLS. In ICD9CM, only the lowest level terms (the leaf nodes) are valid for coding. Therefore, in the gold standard, all mappings are to the lowest level terms. In the UMLS, these lowest level terms are given TTYs of PT (preferred terms) while higher level terms have TTYs of HT (hierarchical term). However, the two mapping algorithms did not distinguish between ICD9CM PT and HT terms. If we discard all mappings to HTs, the precision would increase from 22.1% to 32.1% for IntraMap and from 14.7% to 25.0% for MetaMap, without impact on recall.

Enhancing performance through combination

The prospect of combining semantic and lexical mapping is particularly exciting. We showed that by pooling all the mappings together according to their anticipated precision ranking, while removing duplicates and lower-ranking alternative mappings, the overall mapping performance was significantly better than either mapping method used alone. Another advantage of combining the mappings by the precision ladder approach is the possibility of adjusting the recall-precision profile of the combined mappings to suit the task at hand. For instance, if the task is automatic code translation, one would prefer a mapping algorithm with high precision. One way to achieve this is by taking only the first two rungs of the precision ladder which will give highly precise mappings (precision close to 80%) to about 20% of SNOMED CT terms. On the other hand, a more likely use case of the mapping algorithms is to suggest candidate mappings to human editors to assist them in their task of creating mappings. In that situation, one will need a mapping algorithm with high coverage. If one takes

every rung from the precision ladder, one will find candidate mappings for over 90% of SNOMED CT terms, with a precision of 27%. However, in this particular use case, even the incorrect candidate mappings may serve some useful purpose. If we compare only the first three digits of the ICD9CM codes in the candidate mappings and the gold standard mappings, the precision jumps to 49.5%. This means that one out of two of the candidate mappings will either be exactly correct or will bring the editors closer to the correct mapping.

Generalization

Except for validating the method, there was little need for creating a mapping between SNOMED CT and ICD9CM, since one already exists as part of the SNOMED CT distribution. However, the strategy presented here can be applied to virtually every pair of terminologies in the UMLS. In future work, we plan to test the algorithms on other terminologies that do not already have known mappings e.g. between SNOMED CT and MeSH.

Conclusion

Semantic and lexical mapping between two terminologies can be done using resources available in the UMLS. The performance of semantic mapping is generally better than lexical mapping. When the two sets of mappings are combined according to the anticipated level of precision of individual mappings, the overall performance is better than either algorithm used alone. One further advantage of this approach of combining semantic and lexical mapping is the possibility of customization of the trade-off between coverage, recall and precision according to the task at hand.

Acknowledgments

This research was supported by the Intramural Research Program of the National Institutes of Health, National Library of Medicine.

References

[1] Rocha RA, Rocha BH, Huff SM. Automated translation between medical vocabularies using a frame-based interlingua. Proc Annu Symp Comput Appl Med Care 1993:690-4.

[2] Cimino JJ, Barnett GO. Automated translation between medical terminologies using semantic definitions. MD Comput 1990;7:104-9.

[3] Barrows RC, Jr., Cimino JJ, Clayton PD. Mapping clinically useful terminology to a controlled medical vocabulary. Proceedings - the Annual Symposium on Computer Applications in Medical Care 1994:211-5.

[4] Dolin RH, Huff SM, Rocha RA, Spackman KA, Campbell KE. Evaluation of a "lexically assign, logically refine" strategy for semi-automated integration of overlapping terminologies. Journal of the American Medical Informatics Association 1998;5:203-13.

[5] Masarie FE, Jr., Miller RA, Bouhaddou O, Giuse NB, Warner HR. An interlingua for electronic interchange of medical information: using frames to map between clinical vocabularies. Comput Biomed Res 1991;24:379-400.

[6] Rocha RA, Huff SM. Using digrams to map controlled medical vocabularies. Proc Annu Symp Comput Appl Med Care 1994:172-6.

[7] Fung KW, Bodenreider O. Utilizing the UMLS for semantic mapping between terminologies. AMIA Annu Symp Proc 2005:266-70.

[8] Bodenreider O, Nelson SJ, Hole WT, Chang HF. Beyond synonymy: exploiting the UMLS semantics in mapping vocabularies. Proceedings / AMIA Annual Symposium 1998:815-9.

[9] Aronson AR. Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program. Proceedings / AMIA Annual Symposium 2001:17-21.

[10] MetaMap website http://skr.nlm.nih.gov/.

Address for correspondence

8600 Rockville Pike, MSC-3826
Bethesda, MD 20894
USA
E-mail: kwfung@nlm.nih.gov

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Biomedical Vocabularies – the Demand for Differentiation

Josef Ingenerf, Siegfried J. Pöppl

Institute of Medical Informatics, University of Lübeck, Germany

Abstract

The need of biomedical vocabularies is well known for various tasks, e.g., supporting structured data entry, decision support and electronic data exchange as well as retrieval and statistical evaluation of the data. Due to a considerable diversity of artifacts like interface terminologies, classifications and thesauri it seems to be reasonable to demand for a massive reduction and, finally, to end up with just one unique "multi-purpose world terminology". Concept-based reference terminologies like SNOMED CT® might be candidates for that idea. Unfortunately, the above mentioned kinds of vocabularies cannot be replaced because of their specific purpose-dependent nature. Their mutual distinctive characteristics are outlined and compared with inherent purpose-less concept systems that are based on formal approaches like description logics. For supporting interoperability the different kinds of purpose-dependent vocabularies can and should be improved by mappings to machine-processible reference terminologies. As a side-effect, this paper may contribute to the meta-terminology in this field of medical terminology.

Keywords:

vocabularies, controlled, terminology, semantics, medical informatics/standards

Introduction

In the nineties, several studies have been conducted for evaluating biomedical vocabularies, basically focusing on domain-completeness. Cimino´s desiderata, published in 1998, stimulated a discussion of qualitative criteria [1]. In ISO/TC 215 Working Group 3 this work gave reason to the creation of the standard ISO/TC 17117 "Controlled health vocabulary - Structure and high-level indicators" [2] that is actually revised with a new title "Criteria for the Categorization and Evaluation of Clinical Terminologies". In this paper, the term "biomedical vocabulary" is used as a general term subsuming specific variants like classifications, terminologies and thesauri.

The characteristics of a vocabulary influence its utility and appropriateness in clinical applications. Cimino [1] and ISO/TC 17117 ([2]: scope) refer to health terminological systems (terminologies) that are primarily designed to be used for clinical concept representation. The set of desirable criteria as a guideline for evaluating terminologies are based on one dominant characteristic, namely "concept orientation". Concepts are the basic units of a clinical terminology with unique and non-ambiguous identifiers. Different but synonymous terms are denoting the same concept. This leads to non-redundancy, non-ambiguity, non-vagueness and internal consistency.

Amongst others, the demand for domain-completeness gives rise to compositional approaches. Patient data should be represented as detailed as necessary suitable for clinical use. In compositional terminologies concepts are expressed in terms of at least two other concepts, ideally joined by relations. Precoordinated concepts are represented already within the terminology with an equivalent single unique identifier (non-semantic codes). The full power of compositional terminologies is given by the support for creating new postcoordinated concepts. Without postcoordination no terminology can escape from the phenomenon of exponential explosion when enumerating all combinations of concepts [3]. Consequently, compositional terminologies cannot longer be delivered as books. According to the slogan "defining concepts instead of building hierarchies" software (classifier) is needed for computing the hierarchy and relations from concept definitions "on the fly". For doing that, formal terminologies shall use suitable representation formalisms (*e.g.*, description logics) and terminological models (schemata, ontologies) for enabling complete and correct inferences of conceptual knowledge.

In Figure 1 reference terminologies like SNOMED CT® (Systematized Nomenclature of Medicine - Clinical Terms) [4] are shown as one type of biomedical vocabularies. Due to the primary functions shown in Figure 1 their value to support health information management has been widely recognized. They are required for precise data representation, accurate interpretation of data and interoperability among information systems exchanging such data. At the same time they can serve as an Interlingua by providing mappings to all other biomedical vocabularies. Sensibly coordinated vocabularies benefit from such a mapping by getting support for tasks like maintenance, error detection, re-engineering and translation.

There are a lot of other concept oriented terminologies that theoretically should conform to the criteria mentioned above as well. Most of them are domain specific and have several structural limits, *e.g.,* LOINC® [5]. As they have almost no reference characteristics they are candidates for

mappings to them. They are not discussed in this paper in more detail.

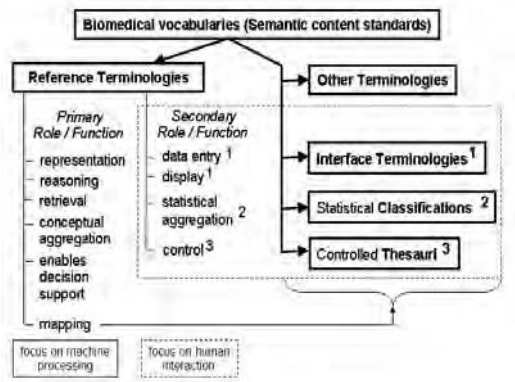


Figure 1 - Types of biomedical vocabularies (modification of Annex A2 in the actual version of ISO 17117 [2])

In principle, reference terminologies have the ability to serve also the secondary functions shown in Figure 1. However, there are intrinsic conflicts between the needs of users (*e.g.*, clinical pragmatics) and the needs of machines (*e.g.*, logical strictness), see chapter "Reference terminologies". For that reason there are other types of vocabularies that are specialized in supporting those functions.

- Interface terminologies, see corresponding chapter => structured essentially with respect to "terms", with colloquial phrases for data entry by clinicians, *e.g.*, MST (Minimal Standard Terminology) [6].
- (Statistical) classifications, see corresponding chapter => structured essentially with respect to "classes", with guaranteeing comparable aggregations, *e.g.*, ICD (International Classification of Diseases) [7].
- (Controlled) thesauri, see corresponding chapter => structured essentially with respect to "subjects", with control mechanisms for using keywords reliably, *e.g.*, MeSH (Medical Subject Headings) [8].

In the next chapter these types of vocabularies[1] are discussed with a focus on features distinguishing them from concept based terminologies. After that, the results are summarized, followed by a discussion. To make it very clear: This paper is not about questioning the importance of reference terminologies in general. Just the ability to support the secondary functions in Figure 1 is critically analyzed opposed to specialized types of vocabularies being in place for a very long time.

Different types of biomedical vocabularies

Early work on the differentiation of biomedical vocabularies has focused on levels of structural complexity from hand-crafted fixed hierarchies based on "enumeration" to computed dynamic concept systems based on "formal definitions" [10, 11]. Differentiating types of vocabularies has a long history only with respect to terminologies and classifications [12]. Due to an increasing maturity of reference terminologies this is still an ongoing issue [13]. The different nature of interface terminologies and thesauri opposed to reference terminologies has been recognized, but studied much less systematically.

Reference terminologies

Terminologies are defined as a system of concepts with terms used for naming concepts. As mentioned earlier, the demand for expressiveness gives rise to the emergence of formal compositional terminologies. Based on suitably defined concepts the subsumption hierarchy can be created automatically. Table 1 gives just an impression of a concept definition based on description logics used by SNOMED CT® [4]. The precoordinated concept "397825006" has already been classified into the concept hierarchy. All those concepts are delivered with unique identifiers, descriptors and mappings (*e.g.*, to ICD). Postcoordinated concepts are built analogously by refining value restrictions (*e.g.*, Pyloric antrum) and adding relations (*e.g.*, Causative agent) with value restrictions (*e.g.*, Helicobacter pylori). With that, they also can be classified, *e.g.,* as a subconcept of the available concept "89662003 Helicobacter-associated pyloric ulcer". However, there are still challenges with respect to identifiers, descriptors and mappings that have to be generated automatically, *i.e.*, for coded data exchange.

*Table 1 - **Example** from SNOMED CT® [4]*

CONCEPT: 397825006 Gastric ulcer
DESCRIPTORS (TERMS): Gastric ulcer (1777426014, preferred), Stomach ulcer (1785985013), GU - Gastric ulcer (1785986014), Gastric ulceration (1785987017)
DEFINITION: Fully defined by ... - *Is a* 64572001 Disease (disorder) - Group - *Associated morphology* 56208002 Ulcer - *Finding site* 69695003 Stomach structure
INFERRED SUPERORDINATE CONCEPTS (by a classifier) - 29384001 Disorder of stomach - 40845000 Gastrointestinal ulcer

Besides questions of expressiveness and tractability of the used logical representation formalism it becomes clear that the approach is heavily dependent on suitable definition and organization principles, especially with respect to the top-level hierarchy. The logic used by terminologies like SNOMED CT® guarantees only that if the premises are true then the conclusions will be true. For achieving suitable concept models, ontological principles have to be applied thoroughly like avoiding isa-overload (*e.g.*, Helicobacter pylori *isa* Infectious agent) and epistemologically

1 All mentioned vocabularies are included in UMLS (Unified Medical Language System) [9].

loaded terms (*e.g.*, Ulcer without bleeding *or* Tuberculosis, histologically specified) [14].

The compliance with ontological principles again corresponds with re-usability, *i.e.*, one of the main targets of reference terminologies with respect to the primary functions in Figure 1. However, this should not be confused with the "multi-purpose" characteristic. Quite reciprocally, the support of the secondary functions is weakened. For them pragmatic clinical conventions, epistemological terms for the observation of reality instead the reality itself and clinical knowledge instead of terminological knowledge has to be considered.

Interface terminologies

Interface terminologies (entry or point of care terminologies) are used for clinical data entry. They "interface" between clinician's own unfettered, colloquial expressions and machine-processable conceptual expressions. They range from simple pick-up lists to structured terms with attributes and relations that are optimized for an easy and meaningful data entry. The desiderata for interface terminologies [15] differ from Cimmno's desiderata for reference terminologies [1] mentioned in the introduction, *e.g.*; with much more focus on terms and language instead of conceptual knowledge.

*Table 2 - **Example** from MST [6]*

Headings	Terms	Attributes	Attr.Values	Sites
Excavated lesions	Ulcer	Number	Single (Solitary), Few, Multiple	Site(s)[other table]
		Size	Largest dia-meter in mm	
		Bleeding	Yes: Spurting/ Oozing, No	
Mucosa	Ulcerated mucosa	Continuity	Discontinuous/ Continuous	Site(s) [other table]
		Bleeding	Yes/No	
There are more tables for anatomy, extend/limits/reasons for examination (*as below*), diagnostic/therapeutic procedures. The tables are linked suitably.				
Extend and limits	Limitation	Reason	Patient unstable: Specify	Site(s)

Table 2 shows selected entries for recording findings in the colon or the stomach. A subtle distinction is made between ulcer findings difficult to anticipate in reference terminologies: "*Endoscopists felt that there may be a conceptual distinction between ulcers that are multiple and mucosa that was ulcered diffusely* [6]. Particular attention has been paid to include just those terms, attributes and values that avoid ambiguity and misuse. In contrast, reference terminologies would produce enormous noise. Hierarchies like "Diagnostic procedures: Biopsy (Purpose: Helicobacter pylori test) / Chromoscopy (Type: Staining)" reflect local needs. Furthermore, metric data (*e.g.*, Size) and epistemological terms (*e.g.*, Patient unstable) have to be regarded. Diagnoses and procedures are represented primarily as precoordinated terms. The linkages from findings to sensibly selected procedures and diagnoses cannot be inferred by terminological knowledge. They represent clinical knowledge, a consensus between experts or local factors. These and other features of interface terminologies have been studied by Rosenbloom et al. [15]. They postulate desiderata being in conflict with reference terminologies:

- Presence of relevant assertional knowledge
- Balance between pre- and postcoodination
- Adequate synonym coverage

These findings motivated the author of this paper and his colleagues not to realize a data entry system directly linked to the GALEN terminology [16]. Although the GALEN approach is quite innovative in providing "sensible statements" the data entry system works in two phases. First, the GALEN terminology enables the creation of data entry forms enhanced by multilingual language generation and code conversion. Second, the resulting items have to be rearranged, rephrased and enhanced due to local needs. More realistic, interface terminologies are already existent. Of course, they are not perfect [17]. For that reason, they do profit from a mapping to reference terminologies like GALEN [16] and SNOMED CT® [18].

Statistical classifications

Statistical classifications (aggregation or report terminologies) support the comparison of data within populations over time and between populations at the same point in time, as well as the compilation of nationally consistent

data. They have well known unique characteristics. The basic components are classes as clusters of objects in a mutual disjoint and exhaustive partition of a given domain. For guaranteeing exhaustiveness of classes, residual classes "Other ..." are introduced at every stage of subdivision. Disjointness is enforced by special instructions, amongst others inclusion and exclusion criteria.

*Table 3 - **Example** from ICD-10 [7]*

K25	Gastric ulcer
K25.0	acute with haemorrhage
K25.1	acute with perforation
K25.2	acute with both haemorrhage and perforation
K25.3	acute without haemorrhage and perforation
K25.4	chronic or unspecified with haemorrhage
...	...
K25.9	unspecified as acute or chronic, without haemorrhage or perforation

The example in Table 3 illustrates a typical combination of criteria at the same level. The special interest in haemorrhage and perforation as co-morbidities of ulcers according frequencies and practical relevance leads to the integration of these and other criteria of interest into one category. This kind of mono-hierarchical subdivision of all diseases is not without problems, *i.e.*, the loss of information (*e.g.*, with K25.4 the criterion "chronic" is lost) or the problem of context-dependent interpretations of residual classes.

This and other findings are well known so that classifications should not be used as input, but output systems [19, 20]. The crucial question is whether aggregation for all purposes can be carried out by conceptual aggregations in a reference terminology or whether there is a need for mappings to statistical classifications. Stuart-Buttle et al. [21] investigated classifications with their essential features:

- Stability over time *(enabling long-term comparison)*
- Exhaustiveness *(concentrating on usual phenomena)*
- Mutual exclusivity *(avoiding double counting)*

It is quite clear that classified data can only be as good as the original clinical record. Here, terminologies should be used for representing data as detailed as the clinician is willing to enter it. It remains to be seen if the frequent updates of terminologies and the variability of ways to enter the same data affect their potential for statistical aggregations. Inherent non-disjoint concepts in multiple hierarchies have to be restricted to single hierarchies and freezed for a fixed period of time. The process of bridging fine grained information collected for patient care to coarse grained reporting for management or planning may be obvious to humans, but it must be made explicit for machines. Crucial concerns are stated by Rector [3].

At least the context of patient care is almost not possible to be made as explicit as necessary. Due to the purpose-dependence of classifications a relevant amount of context is already reflected in the structure of classifications and even more important, in the rules for assisting users in the selection of information and categories. For the same reason, mapping concept representations to statistical classifications is inherently an interactive task, involving clinical judgements based on the whole episode that relates to a patient [19].

Controlled thesauri

Thesauri are used as controlled vocabularies in the stricter sense of the word for indexing and retrieval purposes in library and information science. The basic components of thesauri are subjects [22], also called descriptors, (main) headings or keywords. They are related by three relationships "Broader-than" (*BT*), "Related-to" (*RT*) and "Used-for" (*UF*). The relationships and other means (*e.g.*, annotations, scope notes) are used to avoid variation between those indexing and those retrieving data. Analogous to classifications, the purpose of matching documents with information needs is reflected in the structure of well chosen descriptors and (entry) terms. Misunderstandings due to ambiguity, synonymy and inconsistent use of language should be minimized.

*Table 4 - **Example** from MeSH [8]*

D005767 Gastrointestinal Diseases	
BT D013272 Stomach Diseases	*: D-codes are identifiers*
BT D010437 Peptic Ulcer *	*: for descriptors.*
BT D013276 **Stomach Ulcer**	*: Codes for placing them*
BT D010437 Peptic Ulcer *	*: into hierarchies are*
BT D013276 **Stomach Ulcer**	*: not shown here.*
* RT (see also) D000897 Anti-Ulcer Agents, ...	

Precoordinated descriptors like "Stomach Ulcer" in Table 4 should be used by MeSH coding instructions, although "Stomach" and "Ulcer" exist independently. The descriptor is placed in two hierarchies for helping the human user in discovering the keyword by top-down searching the MeSH-hierarchy. Synonymous (entry) terms like "Gastric ulcer" are added only when their use in the user community is as reliable, predictable and comparable as possible. Beyond the combination of descriptors, MeSH offers so-called qualifiers (*e.g.*, "/therapy") and major/minor tags for indicating the actual content of a document. However, compared with concept based terminologies, it is neither possible nor desirable to define compound descriptors by employing non-ISA-relationships as "Located-in" and to classify them automatically into the hierarchy.

There are ontologically valid hierarchical relations like "Stomach Diseases *BT* Stomach Ulcer" and also debatable relations like "Stomach *BT* Pylorus" or "Diagnosis *BT* Diagnostic Errors". Opposed to ISA-relations between concepts, *BT*-relations between descriptors are defined in a completely different manner:

"Should a search for documents dealing with A find all (or most) documents dealing with B? If yes, A is broader than B (and conversely, B is narrower than A)." [23]

The reader is invited to look up the MeSH term "Vocabulary, controlled". All kinds of biomedical vocabularies are included as narrower terms because the average user (*e.g.*, of MEDLINE) and even the experts in this field would not conform reliably to the subtle differences outlined in this paper.

Often, thesauri are referred to as concept-based vocabularies. However, the main idea of language control resides solely at the level of descriptors aggregating concepts. Available mappings to a concept layer are helpful for maintenance or translation [23]. However, it is not visible to MeSH users. The following characteristics differentiate thesauri from reference terminologies that are discussed in more detail in [24].

- Language control *(more than just offering concepts)*
- Carefully chosen descriptors *(minimizing ambiguity)*
- Hierarchies with respect to relevance *(≠ subsumption)*

Results/Discussion

The specific features distinguishing interface terminologies, classifications and thesauri from concept based reference terminologies have been presented in the last chapter. Separating primary and secondary functions in Figure 1 seems to be the key for distinguishing reference terminologies from the other types of biomedical vocabularies. In order to satisfy the primary functions they have to conform to logical and ontological principles that, reciprocally, weaken their potential to support also the secondary functions. At the same time it is rather nonsensical to apply ontological principles to interface terminologies[2], statistical classifications and thesauri [24].

The different types of biomedical vocabularies challenge the task of mappings between them for enabling interoperability. With the help of reference terminologies the number of mappings can be minimized, as shown in figure 2.

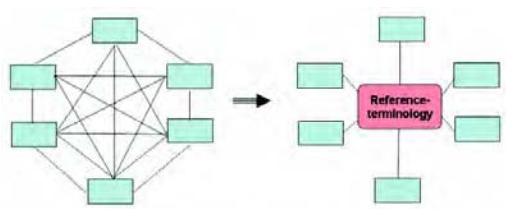


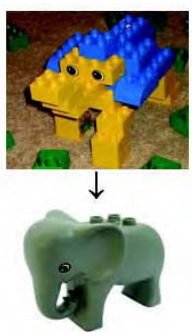
Figure 2 - Reference terminologies

Besides the reasons outlined in the last chapter the value of a reference terminology is to keep the number of mappings between biomedical vocabularies small, *see* figure 2.

2 As "terminology" is used as a short form for concept oriented (reference) terminologies we prefer the term "Interface vocabulary". However, it is uncommon to do so.

As the borderlines between the types of vocabularies are quite tricky the results of this paper will be clarified by three short analogies. Analogies might demonstrate the agreement of phenomena in some respects but they do not work for others. Looking to the first analogy reference terminologies are specialized in supporting the primary functions in Figure 1. At the same time they behave like generalists when looking to the secondary functions in Figure 1. Only the latter are sketched by the three analogies. If available, the secondary functions should be carried out by specialized vocabularies mapped to reference terminologies for reasons outlined in the last chapter.

- IT world
  In information technology and other fields the generalists/specialists tradeoff is well known. Specialized software systems are favored especially by users for supporting their particular needs. General solutions offering mostly limited functionality are favored by central divisions responsible for integrating all systems. Both approaches will co-exist so that a careful coordination of them is needed permanently.
- Lego world
  Compositional terminologies behave like Lego. Basic building blocks can be used for virtually all purposes according to given instructions. Due to pragmatic reasons users often favor extra items like animals although they can be created also with "pure Lego" in principle. However, the extra items are badly suited for reusing.



- Animal world
  Given gazelles (interface terminologies) being able to run very well, dolphins (classifications) specialized in swimming and eagles (thesauri) that are very good in flying. Then there are also animals representing multipurpose reference terminologies: ducks. They are skillful in all these tasks, but within clear limits; see also the animal school fable.

References

[1] Cimino JJ. Desiderata for controlled medical vocabularies in the twenty-first century. Methods Inf Med 1998: 37 (4-5): 394-403.
[2] SO/TS 17117. Health informatics - Controlled health vocabularies - Vocabulary structure and high-level indicators, 2002, *see* www.tc215wg3.nhs.uk.
[3] Rector AL. Clinical Terminology: Why it is so Hard? Meth Inf Med 1999: 38: 239-252.
[4] SNOMED CT: www.snomed.org.
[5] LOINC: www.regenstrief.org/medinformatics/loinc/.
[6] MST: www.omed.org/index.php/resources/re_mst/.
[7] ICD: www.who.int/classifications/icd/en/.
[8] MeSH: www.nlm.nih.gov/mesh/meshhome.html.
[9] UMLS: www.nlm.nih.gov/research/umls/.
[10] Ingenerf J. Taxonomic vocabularies in medicine: the intention of usage determines different established structures. Proc. MEDINFO 1995: 136-139.

[11] Rossi Mori A, Consorti F, Galeazzi E. Standards to support development of terminological systems for healthcare telematics. Meth Inf Med 1998: 37 (4-5): 551-563.
[12] Coté RA. Editorial: Ending the classification versus nomenclature controversy. Med Inform 1983. 8: 1-4.
[13] Roberts R, Innes K, Walker S, Scott P. Avoiding a war of words (and numbers): the uncertain future of terminologies and classifications. J Ahima 2004: 75 (9): 26-32.
[14] Spackman KA, Reynoso G. Examining SNOMED from the Perspective of Formal Ontological Principles: Some Preliminary Analysis and Observations. Proc. KR-MED, 2004: 72-80, *see* http://CEUR-WS.org/Vol-102/.
[15] Rosenbloom ST, Miller RA, Johnson KB, Elkin PL, Brown SH. Interface terminologies: facilitating direct entry of clinical data into electronic health record systems. J Am Med Inform Assoc 2006. 13 (3): 277-288.
[16] Birkmann C, Diedrich T, Ingenerf J, Rogers J, et al.. A Formal Model of Diabetological Terminology and Its Application for Data Entry. Proc. MIE, 1997: 426-430.
[17] Tringali M, Hole WT, Srinivasan S. Integration of a standard gastrointestinal endoscopy terminology in the UMLS Metathesaurus. Proc AMIA Symp 2002: 801-805.
[18] Min Z, Baofen D, Weeber M, van Ginneken AM. Mapping OpenSDE domain models to SNOMED CT. Applied to the domain of cardiovascular disease. Meth Inf Med 2006: 45 (1): 4-9.
[19] Bowman S. Coordinating SNOMED CT and ICD-10: Getting the Most out of Electronic Health Record Systems. J Ahima 2005: 76 (7): 60-1.
[20] Ingenerf J, Giere W. Concept oriented Standardization and Statistics oriented Classification: Continuing the Classification versus Nomenclature Controversy. Meth Inform Med 1998: 37 (4-5): 527-539.
[21] Stuart-Buttle CDG, Ashely JSA, O´Neil M. Clinical Vocabularies - Their Potential as Classifications. Proc. AMIA Symp, 1997: 892-896.
[22] Welty CA, Jenkins J. Formal ontology for subject. Data & Knowledge Engineering, 1999. 31: 155-181.
[23] Nelson SJ, Johnston D, Humphreys BL. Relationships in Medical Subject Headings (MeSH), chapter 11. In: CA Bean, R Green R (eds.). Relationships in the Organization of Knowledge. New York: Kluwer, 2001: 171-184. *see* www.nlm.nih.gov/mesh/meshrels.html.
[24] Ingenerf J, Linder R. Ontological Principles are not Applicable to all Types of Biomedical Vocabularies. Meth Inf Med 2007: forthcoming.

Address for correspondence

Dr. Josef Ingenerf
Institute of Medical Informatics,
University of Lübeck
Ratzeburger Allee 160,
23538 Lübeck,
Germany
Email: ingenerf@imi.uni-luebeck.de

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Development of a Taxonomy for Health Information Technology

Brian E. Dixon, M.P.A. [a,b], Atif Zafar, M.D. [a,b,c], Julie J. McGowan, Ph.D. [a,b,c]

[a] *AHRQ National Resource Center for Health Information Technology, Maryland*
[b] *Regenstrief Institute, Inc., Indiana*
[c] *School of Medicine, Indiana University, Indiana*

Abstract

Taxonomies provide schemas to help classify entities and define the relationships between them. Early computing enabled the development of ontologies and Medical Subject Headings (MeSH), the first modern classification of medical terminology as applied to medical literature. Later developments, such as MEDLINE, expanded MeSH to include a number of medical informatics terms. However, a lack of specificity in MeSH and other existing informatics taxonomies for terminology used to describe the growing field of health information technology (health IT) created the need for the development of a specialized taxonomy. Experts associated with the Agency for Healthcare Research and Qualitys (AHRQs) National Resource Center for Health Information Technology (NRC) created and evaluated a taxonomy for health IT, to enable users of a public health IT Web site to efficiently identify resources within an online, searchable repository.

Keywords:

controlled vocabulary, internet, information systems, medical informatics applications, computer communications networks

Introduction

Taxonomy is derived from two Greek words: taxis, meaning order or arrangement, and nomos, meaning law. Common definitions include a focus on the laws of classification and natural history, including the relationships among entities organized in a hierarchical structure [1][2].

While taxonomies have historically been applied to a variety of different content areas, the focus remained on the classification of those areas and the hierarchical structure that explained vertical relationships. During the twentieth century, computers enabled the development of ontologies, data models that represent concepts and facilitate understanding of relationships, particularly in the definition of ambiguous concepts and their concomitant associations.

The science of taxonomy and its related ontologies enabled the development of Medical Subject Headings (MESH) and its classification structure (e.g., tree), the first true classification of medical terminology as applied to the worlds medical literature. Its genesis was in the early work of John Shaw Billings and the publication of *Index Medicus* in 1879. However, the first major expansion of the terminology did not come until 1958, recognizing the specialization of the profession of medicine [3].

The creation of the specialized taxonomy for medicine enhanced the ability of users to retrieve information from printed materials. However, without the use of a computer to facilitate the application of the classification schema, albeit through punched cards and high-end photographic equipment to print the newly reconstituted *Index Medicus*, the foundation for the current MESH would not have been laid [4].

The computerized MEDLINE followed over a decade later, with its MESH expanded to include a number of tangential concepts. The vocabulary and its hierarchical relationships have been developed over time and have incorporated a number of medical informatics terms.

However, there is a lack of specificity of concepts and granularity in the health information technology (health IT) area, particularly as it relates to current practices in electronic health record (EHR) implementation. The MESH vocabulary also lacks terminology to encompass dynamically growing areas like health information exchange and consumer health informatics.

In trying to develop a searchable knowledge core for the Agency for Healthcare Research and Qualitys (AHRQs) National Resource Center for Health Information Technology (NRC) Web site, a major first step was the creation of a taxonomy to define the interrelationships among concepts around health IT planning, implementation, and evaluation.

Because health IT bridges medical informatics and hospital and clinical practice organizations, it became imperative in the development of a taxonomy to understand the scientific and sociological issues impacting health IT. MESH offered some concepts, but the major sources of the taxonomy development came from divergent vocabularies. This supports the premise that a good taxonomy must be developed on the basic principles of a hierarchical and interrelated classification schema, searchable within a Web environment.

Materials and methods

In March 2005, a panel of experts with a wide array of experiences in implementing health IT systems was convened to design a national, searchable Web site for persons interested in the implementation and evaluation of health information technologies. These experts, drawn from medical informaticians across the United States, agreed that a basic system of categorization would be useful to organize content into a searchable knowledge base. The experts also expressed a desire for the taxonomy to be used as a method for educating visitors about common terminology used in the field of health IT.

The panel began design of the taxonomy using a structured brainstorming exercise. The brainstorming involved thinking through the various types of content the Web site would feature, which stakeholders the Web site would target, and what topics best matched the goals of the Web site designers, stakeholders, and potential users. A preliminary confluence of ideas was slowly organized into logical buckets by the end of the brainstorming session. The buckets top level terms and the terms contained therein branches as used in most controlled vocabularies comprised the first draft of the taxonomy.

The panel next looked to a number of existing taxonomies to refine each top level term and its branches. The EHR Mind Map, created by the Healthcare Information Management Systems Society (HIMSS), provided some useful terms, but it was found to be narrowly focused on EHR systems, neglecting technologies such as telemedicine and health information exchange [5]. The panel also considered the ACMs Computing Classification System (CCS), but it consists primarily of computer science terms (e.g., programming languages, mathematical theory, and data structures) whereas the target audience for the Web site is clinical and health administration leaders engaged in health IT implementation activities [6]. Finally, the panel considered MESH but found it too lacked sufficient granularity to cover the complete technical (e.g., EHR, HIE, telemedicine) and sociological (e.g., adoption, process redesign, project management) spectrums that encompass the field of health IT [7].

To create a taxonomy for health IT, the panel created a draft using terms from the three major taxonomies it examined. The group then organized the list into clusters of related terms under broad categories, and it began to search for additional terms that would fit into its categories. The organized list became version 1.0 of the taxonomy.

Since the panels original meeting, the taxonomy has been expanded and refined. A first round of edits came after a second expert review was conducted by experts associated with the NRC who could not attend the original expert panel meeting. A second round of edits came when two of the authors (Dixon and Zafar) used the taxonomy to classify journal articles and other resource documents in the online library the taxonomy was developed to support. A third set of modifications was made when two NRC experts (Dixon and Samarth) applied the taxonomy to the more than 120 health IT grants and contracts awarded by AHRQ within the last two years (FY04 and FY05).

The taxonomy was also evaluated on two subsequent occasions to ensure comprehensiveness and usability. A formal review of the taxonomy was conducted in late 2005 by the large panel of experts on staff with the NRC. The recommendations of the panel were implemented as part of the development version of the AHRQ health IT Web site.

In early 2006, a second type of evaluation took the form of usability testing, a critical component in software and Web development. Health care and IT professionals from leading non-profit and government organizations were asked to use an online, interactive version of the taxonomy to search for health IT resources. Testing was conducted at the National Library of Medicine (NLM) usability lab in Washington, D.C. User experiences were observed, recorded, and analyzed by NLM and NRC staff.

Based on recorded user experiences, the taxonomy and its appearance on the Web site were altered. For example, a minor category EHR Systems was renamed Implementation of Health IT after some users confused it with the EHR minor category. Users also reported that they did not associate the term clinical decision support with the EHR minor category, so it was renamed Electronic Health. Other users commented on the taxonomys appearance, noting that they would prefer to see terms such as electronic health records and health information exchange as entry points (e.g., major categories) to the taxonomy. Instead of altering the major dimensions of the taxonomy to accommodate user requests, prevalent terminology was summarized together under its parent branch for display on the Knowledge Library main page.

Results

The taxonomy is organized into six major and 28 minor categories (high-level terminology) with two additional sub-levels for controlled vocabulary terminology (discrete, classifiable concepts) and a dictionary for controlled vocabulary terminology synonyms. Based on the results of the usability testing, a subset of the 400+ terms and synonyms were selected to drive terminology browsing of the online resources.

When users click on a term, a search is performed in the background, and results are displayed for the user to browse through. Resources from the core collection, those tagged by an NRC expert, are promoted to the top of the result list. Partner contributions, online resources aggregated from a variety of health IT-focused organizations, fill out the result list and are sorted based on the prevalence of the selected term within the resource.

Of the 6872 online resources available through the AHRQ health IT Web site, 702 (10.2%) reside in the core collection. Table 1 presents the major and minor categories along with the number of core collection resources belonging to them.

Table 1 - Taxonomy for health IT and volume of core collection resources

I. Organizational Strategy (171)	A. Financial (18)
	B. Planning (44)
	C. Process Change (18)
	D. Implementation of Health IT (77)
	E. Policy (14)
II. Technology (170)	
	A. Mobile (3)
	B. Infrastructure (1)
	C. Security (2)
	D. Standards (34)
	E. Electronic Health (57)
	F. Telehealth (9)
	G. Health Information Exchange (51)
III. Value (227)	
	A. Research (76)
	B. Evaluation Outcomes (151)
IV. Laws and Regulations (35)	
	A. Sample Legal Documents (2)
	B. Privacy (16)
	C. Security (7)
	D. Government (10)
V. Organizations (54)	
	A. Professional Societies (12)
	B. Payers (8)
	C. Governmental (14)
	D. Nonprofit Organizations (15)
	E. Magazines (5)
VI. Operations (35)	
	A. Governance (14)
	B. Project Management (12)
	C. Systems (5)
	D. Dissemination (4)

Organizational strategy

Technology is implemented in order to solve problems and improve processes. Strategies designed to address specific problems and processes will ensure the successful implementation of technology within organizations.

- Financial strategies create initial and sustained revenue for the deployment and maintenance of technology.
- Strategic planning helps technology deployment meet the needs of an organization and ensure that technology solutions adhere to the mission and vision of an organization.
- Strategies for process change give implementers tools for addressing the human component of technology adoption.
- Implementation of health IT strategies provide roadmaps to help organizations procure and adopt technologies. In health care, these strategies should be crafted to ensure interoperability and the protection of privacy for personal health information.
- Policy strategies address legal issues, such as federal and state regulations related to the sharing of patient data across state lines. Strategic policies can also include those designed to ensure technology meets certification requirements and accreditation criteria.

Technology

Technology in this taxonomy is broadly defined, and its definition includes individual components such as hardware (e.g., PDA, tablet PC) and complete systems (e.g. laboratory information system - LIS, computerized physician order entry - CPOE).

- Mobile computing is important to health care since the workforce is by-and-large mobile, traveling from room-to-room or clinic-to-hospital.
- Infrastructure is a critical component of all technology projects as it describes how data is communicated.
- Security is a major priority in health care as we strive to provide the right amount of data to the appropriate group of individuals providing care to a patient.
- Standards enable interoperability, the ability for two disparate information systems to seamlessly share data. Many standards are available in the market, however, so distinguishing between them becomes important when exchanging information.

- Electronic Health is a broad concept that encompasses complete systems (e.g., PACS, EMR, CPOE), functions (e.g., clinical messaging, results reporting), and components (e.g., knowledge base, dashboard, rules engine).
- Telehealth describes technologies deployed that allow care to be provided remotely, including remote monitoring of patients in their homes.
- Health Information Exchange (HIE) involves the exchange of patient data between providers and disparate clinical information systems. HIE includes various labels (e.g., LHII, NHII, NHIN), policy elements (e.g., governance, data sharing agreements), technical components (e.g., architecture, interface engine), and applications (e.g., research, public health reporting).
- Consumer Health encompasses technologies that engage patients in the delivery and management of their health.

Value

Evaluation is a core discipline in health IT as the tolerance for mistakes is extremely small. Research and outcomes are also important to demonstrate return-on-investment (ROI) and IT system impacts on patient safety and the quality of delivered care.

- Research includes study designs, data analysis techniques, evidence based medicine (EBM) practices, institutional research board (IRB) policies, and grant writing resources.
- Evaluation outcomes focuses on results of published studies and outcomes of non-academic lessons learned by professionals in the field in a variety of areas (e.g., safety, quality, ROI, patient satisfaction, etc.).

Laws and standards

Laws and standards affect the development, implementation, and adoption of health IT.

- Sample Legal Documents includes a variety of legal documents developed by health care organizations across the nation.
- Privacy includes information on protected health information laws and standards, such as HIPAA.
- Security contains resources on administrative, physical, and technical security as well as security audits.
- Government includes resources on state and federal regulations as well as government standards related to the national health information network (NHIN).

Organizations

There are many entities involved in the field of health IT, and each organization has an agenda and a role to play in development and adoption of technology in health care. It is important to understand these and the ability to tell the difference between entities.

- Professional societies are member organizations which advocate on behalf of their membership, advance professional ethics, and strive to set standards for the industry.
- Payers include those organizations that reimburse for health care expenditures.
- Governmental organizations include state and federal agencies that support health IT initiatives.
- Nonprofit organizations are independent entities who seek to educate, develop, promote, and research the advancement of technologies for use in health care.
- Magazines include news and trade publications which seek to report on the health IT industry.

Operations

Management of operations is critical to the prolonged use of technologies in health care.

- Governance includes organizational policies related to human and technological resources.
- Project management consists of resources to help manage the implementation and maintenance of health information technologies.
- Systems operation includes resources to help with day-to-day maintenance and operation of health IT systems.
- Dissemination is crucial primarily to those conducting research in the field of health IT. However, sharing best practices and successful implementations can be strategic for many organizations.

The taxonomy covers the broad spectrum of topics discussed in the field of health IT. The controlled vocabulary terms are too numerous to publish here, but many of them can be found on the AHRQ health IT Web site, http://healthit.ahrq.gov. User feedback will contribute to the taxonomys future development.

Discussion

The need for a taxonomy for health IT was driven by the creation of a Web site containing knowledge-based resources to support health IT planning, implementation, and evaluation. No single source existed which could provide a contextual vocabulary with the breadth and depth necessary for efficient and effective access of the online resources. The process to create such a taxonomy involved panels of health IT experts identifying and refining terms, hierarchies, and cross-references. The taxonomy was then evaluated for thoroughness and usability.

The formal reviews and usability testing emphasized the need for robust synonymy within any taxonomy. For example, Electronic Health is the term chosen by our experts for a minor category under Technology. Synonymy is critical to ensure that documents which use terms such as electronic medical records and computer-based patient records will be appropriately selected and ranked when searches are performed against the online library.

Usability testing revealed a strong need for the development of inter-relationships between orthogonal concepts. We found that users not only searched for documents related to computerized physician order entry, but they desired CPOE related documents in specific practice settings (e.g., ambulatory, inpatient). We further found that users desired to distinguish between documents related to

small practices and large medical centers. Based on this feedback, we intend to develop interrelated concepts to address these needs.

Conclusion

While health information technology embodies a set of concepts familiar to medical informaticians and others in the health IT field, its concepts are foreign to many in the health care field and present barriers to effective planning, implementation, and evaluation of electronic health record systems and other health information technologies. Because of the need to provide a national resource for health IT information, and to make this knowledge repository readily accessible, a health IT taxonomy was developed through a series of iterative processes.

The health IT taxonomy provides a hierarchical and interrelated classification schema for use by those engaged in the development, implementation, and evaluation of health information technologies. Its use on the AHRQ health IT Web site enables users to efficiently access the resources created and aggregated by NRC staff. Application of the taxonomy to other NRC activities and external knowledge stores is planned, and we believe it will create better understanding of common terms and their interrelations.

Just as the field of information technology is rapidly changing, the health IT taxonomy must remain dynamic. In addition to continued application of it to other components of the AHRQ health IT web site, additional evaluation is planned. This taxonomy is offered to anyone creating a searchable knowledge repository, and we hope that future development will become a living process with contributions from all who use it.

Acknowledgements

This work was supported by the Agency for Healthcare Research and Quality (AHRQ) as part of the AHRQ National Resource Center for Health Information Technologys (Contract No. 290-04-0016) project to build and maintain a free, public Web site for those interested in the implementation and evaluation of health IT. We would also like to thank the following NRC colleagues who contributed greatly to the development of the taxonomy: Davis Bu, MD, Center for Information Technology Leadership, Mark Frisse, MD, MBA, Vanderbilt Center for Better Health, Anita Samarth, ASTECH Consulting, and Jack Starmer, MD, Vanderbilt Center for Better Health.

References

[1] Bruno D, Richmond H. The truth about taxonomies. Info Mgmt J. 2003 Mar/Apr: 44-53.

[2] The Columbia Encyclopedia. 6th Ed. [homepage on the Internet] Classification; c2001-05 [cited 2006 Mar 9]. Available from: http://www.bartleby.com/65/cl/classifi.html.

[3] National Library of Medicine. Introduction and history. Bull Med Libr Assoc. 1961;49Pt2:1-7.

[4] National Library of Medicine. Operational history. Bull Med Libr Assoc. 1961;49Pt2:64-71.

[5] Healthcare Information Management Systems Society [homepage on the Internet]. EHR mind map: a comprehensive resource guide. c2003 [updated 2004 Jan 21; cited 2006 Mar 3] Available from: http://www.himss.org/content/mindmaps/EHR/index.htm.

[6] Association for Computing Machinery [homepage on the Internet]. The ACM Computing Classification System [1998 Version]. c2006 [cited 2006 Aug 11] Available from: http://www.acm.org/class/1998/.

[7] MeSH Browser [homepage on the Internet]. U.S. National Library of Medicine [2006 MeSH]. [cited 2006 Sep 5] Available from: http://www.nlm.nih.gov/mesh/MBrowser.html

Address for Correspondence

Brian E. Dixon, M.P.A.
Regenstrief Institute, Inc.
410 West 10th Street, Suite 2000
Indianapolis, IN 46202
U.S.A.
(317) 423-5582
bdixon@regenstrief.org

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

A Practical Approach to Advanced Terminology Services in Health Information Systems

Maria Laura Gambarte [a], Alejandro Lopez Osornio [a], Marcela Martinez [a], Guillermo Reynoso [b], Daniel Luna [a], Fernan Gonzlez Bernaldo de Quiros [a]

[a] *Medical Informatics Department, Hospital Italiano of Buenos Aires, Argentina*
[b] *Conceptum, Buenos Aires, Argentina*

Abstract

As the medical informatics field evolves, new functions appear as the focus of interest; a more advanced management of terminology is one of them. Using comprehensive and detailed terminology to represent clinical rules in computer systems, associated with patient information, would allow clinical software to provide patient specific recommendations or alerts. In order to uniform data collection through our HIS, and lay the foundations for future clinical decision support systems, we decided to move from our previous classification- based medical record into new terminology services built around Snomed CT, Spanish Language Version. This paper describes the characteristics of our Terminology Server.

The most important achievements of our new terminology system are the centralization of knowledge representation, using a much more detailed terminology system. Clinical data entered at any place of the institution and level of care, is represented uniformly through the whole health information system.

Keywords:

information storage and retrieval, classifications, terminology services

Introduction

Until recently, implementing vocabulary control in medical informatics implied selecting the most appropriate classification for the specific clinical scenario. This was a reliable solution for the intended objectives, like epidemiological description of reality, billing or simple chronic diseases detection on large populations.

As the medical informatics field evolves new functions appear as the focus of interest; a more advanced management of terminology is one of them. New Health Information Systems (HIS) need to store information in a way it can not only be interpreted by human users, but also by computer powered knowledge management rules (1). Encoding clinical rules in computer systems, associated with patient information, would allow clinical software to provide patient specific recommendations or to focus health professional attention to situations that may otherwise be overlooked (2).

The purpose of this paper is to describe our approach to this advanced knowledge representation system in the electronic medical records at the Hospital Italiano of Buenos Aires, with emphasis on data entry, storing and extraction processes.

Background

The most important new functions of complex knowledge representation systems are interoperability and clinical decision support.

Current HIS are designed to link all levels of care, like the primary care office, nursing stations, intensive care units, rehabilitation, etc. Vocabulary control in this new setting gets more complicated, and as each part of the system uses a different classification or list of terms we run the risk of getting into a Tower of Babel situation. Vocabulary control tools must have the ability to serve all settings, and provide an adequate level of detail to each one, as general concepts for ambulatory reasons for encounter or nursing diagnoses, as well as very specific concepts such as specialist registries.

Clinical information systems are expected to improve the quality of patient care. This can be achieved by linking clinical information stored in their databases with current scientific recommendations to bring decision support to physicians, or by selecting high-risk populations where population level interventions proved to bring clear benefits. These features require a strict vocabulary control of patient data and clinical rules data.

Old Classification-based coding systems, common in current Electronic Health Records (EHR) software, are not up to these new challenges. They fail to provide inter-setting validity, they have not the adequate granularity for different levels of care and they are also unsuitable for linking knowledge rules for clinical decision support systems.

To provide solutions to these new functions of the medical record, an advanced knowledge representation system is needed. Implementing this tool in institution-wide health information systems represents a great challenge.

There are some examples of the state of the art terminology systems with knowledge-based representations of medical concepts, like Snomed CT (3), Unified Medical Language System (UMLS) (4), the Medical Entities Dictionary (MED) (1), Nebraska Medical Center (5) or Kaiser Permanente Convergent Medical Terminology (6). Nebraska and Kaiser are based on Snomed CT.

Snomed CT is a nomenclature originally created by the College of American Pathologists (CAP), and is evolving into an international Standards Development Organization (SDO). This is currently regarded as the most advanced initiative in knowledge-based representations with clinical application. Each of the 300,000 terms included are defined using relationships with other terms, creating a powerful semantic network. Simple algorithms allow easy finding of sub types, super types and attributes of any given term. Snomed CT data model allows the continuous extension of the nomenclature, adding new terms, always following the same compositional concept representation model, called Description Logics (7, 8).

These characteristics give Snomed CT a great role as the first building block of a terminology implementation, as has been proposed in previous experiences, either private as Kaiser Permanente or Cerner (9), and governmental as the cases of health departments of the United Kingdom, Australia, and the United States.

Chute et al. outlined three functions required in a terminology system (10), as it was previously described by the U.S. Computer-based Patient Record Institute (CPRI) during the National Conference on Terminology in November 1997. These three functions or layers are:

Entry Terminology: this is the user interface, the terms used to interact with users during data entry process

Reference Terminology: this is the format used to store data, knowledge information should be stored in this layer

Aggregate Terminology: different data formats outputs for user analysis

Snomed CT provides functionalities in these three layers. As indicated before, it has every important feature of a *reference terminology*. Snomed CT also includes several descriptions that can be used as an *entry terminology*. Finally, Snomed CT has a standard cross mapping model; the official distribution includes data for mapping to ICD-9CM (ICD - International Classification of Diseases). Additional ICD-10 cross map data has also been developed in the UK, but is not as widely available, and more new cross maps are under development, like ICPC-2 or NANDA. Snomed CT is included in UMLS (11), which provides some degree of relationship with many other terminologies. These mappings provide the *aggregate terminology* features to Snomed CT.

In order to uniform data collection all through our HIS, and lay the foundations for future clinical decision support systems, we decided to move from our previous classification-based medical record into new terminology services built around Snomed CT.

HL7 (Health Level 7) organization has published a Common Terminology Services (CTS) standard specification. At this stage of the project we decided to create terminology services to provide a solution to the local terminology needs, in a second stage a CTS compatible API (Application Programming Interface) will be created as a new layer that will allow standard access to our services.

Setting

The Hospital Italiano of Buenos Aires is a 650-bed non-profit university hospital located in Buenos Aires, Argentina. More than 150,000 ambulatory visits and 3,000 hospitalizations are registered every month. It is affiliated with a Health Maintenance Organization (Plan de Salud) that takes care of a population of 140,000 patients.

Since 1998 a full scale HIS has been gradually implemented, including ambulatory Electronic Medical Record (EMR), inpatient discharge summary, administrative systems, scheduling systems, inpatient tracking systems, pharmacy systems and complementary studies report and visualization. Several health informatics standards had been implemented, including HL7, CDA Version 2, ICD-9CM, DRG, ICD10, and ICPC.

Previous vocabulary control system consisted in professionals entering free text descriptions, which were later assigned classification codes by a group of coders (12).

Design objectives

Our main objective was to design a new terminology system, whose objectives can be related to each of the functions of the terminology system previously described (entry, reference and aggregate terminology).

Entry terminology functions:

- Provide an institutional vocabulary for all user interfaces, so our professionals interact with familiar terms, including local jargon and preferences.
- Provide concept lookup functions, with loose lexical matches and options, to be used during data entry process of new items in a problems list or similar user interfaces.
- Provide short pick-lists definitions for more structured data entry in specific use templates, with a short list of valid entries, like occupation, level of consciousness, etc.
- Provide navigational tree interfaces for structured data entry in more complex templates, like cause of liver failure, etc.
- Provide different preferred terms for the same concept in different settings.
- It should include the ability to accept new terms from the user, in case a concept or description is not represented.
- It should detect inappropriate terms for being too general or not valid in a subset, as in the case of problem lists entries Emergency Consultation or Cardiology, etc.

Reference terminology functions:

- The entry terminology should be represented in the reference terminology (Snomed CT Spanish Language Version).
- New concepts should be created for institutional terms that cannot be represented with a standard Snomed CT code.
- The system should provide tools to take advantage of the knowledge stored in Snomed CT relationships, like obtaining more refined or more general terms.
- The system should provide means of updating to new versions of Snomed CT without loosing information, nor requiring large amounts of manual updates.

Aggregate terminology functions:

- Provide output to several standard classifications:
- ICD-9CM (diag & procedures)
- ICD-10 (diagnosis)
- ICPC-2 (diagnosis) (International Classification of Primary Care)
- ATC (drugs) (Anatomical Therapeutic Chemical Classification)
- Local billing nomenclatures
- Aggregate data according to Snomed CT hierarchies

All desired functions should run on a centralized software and data structure, referred to as a Terminology Server. The Terminology Server should provide these functions to all existing applications in the Health Information System in the form of Web Services.

A terminology maintenance software application should also be developed to administrate the institutional terminology, its relationship with Snomed and the mappings.

System description

Entry terminology

The institutional entry terminology is composed of concepts and descriptions. We use Snomed definition of these terms (13), where concepts represent distinct clinical meanings and descriptions are a phrase used to name a concept.

The development of the entry terminology is described with more detail in other paper in process.

Our institutional entry terminology can be divided in several subsets; examples of these are:

- Problems list terminology
- Procedures terminology
- Findings in chest radiography
- Administration routes for drugs
- State of consciousness description
- Physical examination subset
- Liver Failure Diagnosis

Some subsets are very large, including thousands of concepts (i.e. the problems list subset). Others are short lists (i.e. the liver failure subset). Each subset was designed in order to be used as the entry terminology in a specific scenario.

Concepts are defined only once, regardless of its inclusion in more than one subset; therefore, accessing liver cirrhosis from the problems list or from the Liver failure subset brings the user to the same concept.

The process of adding concepts to the entry terminology and organizing them in subsets is manual. This is done by trained coders that were previously working with the same information in secondary coding using classifications (12).

Construction of the problems list subset was one of our biggest challenges. We decided to base our work on the historic database of our EHR with more than 2 million free text inputs since 1998. All problems list entries and discharge notes were processed to extract all different textual descriptions, this strings were originally entered with a 50 character limit. We considered that these texts, entered by our own professionals in a completely unconstrained way, would be representative of the local natural language, including abbreviations and jargon. A manual depuration process, assisted by string normalization functions, led to the creation of the Problems List subset with 24,800 different concepts, with 110,000 descriptions in total.

Other subsets were created using arbitrary lists of concepts selected by the clinical terminology team with user input, like the case of Liver failure or Routes of administration subsets.

New concepts or descriptions were accepted from user interfaces and stored for manual evaluation.

The data model for the entry terminology was the standard Snomed CT data model for concepts, descriptions and subsets (14). New concepts and descriptions were added to the standard Snomed CT distribution following official Snomed rules for creating institutional extensions.

Reference terminology

Implementing the reference terminology involves establishing a very detailed terminology as the format in which clinical information will be ultimately stored. Reference terminology is invisible to the users: therefore, it is necessary to define the relationships of each concept included in the entry terminology with the concepts included in the reference terminology.

We used Snomed CT Spanish Language Version as the reference terminology, but is important to note that all different language versions of Snomed CT share the same concepts and relationships. During the translation process only new descriptions are added.

Both entry and reference terminologies were stored following the Snomed data model, and using Snomed tools to represent the concepts of the entry terminology. Snomed CT defines concepts by its relationships with others, so we created new relationships as part of our Snomed CT extension.

Snomed CT has around 300,000 concepts, but in a clinical setting, health professionals usually use very detailed expressions, adding modifiers to general concepts, like mild ankle sprain. To prevent the exponential growth of the nomenclature, Snomed CT avoids including such level of combination with modifiers, providing the general concepts (ankle sprain), the possible modifiers (mild) and the rules to correctly relate them (using the has severity relationship). Any new concept can be represented using this post-coordination technique, creating more detailed subtypes of existing Snomed CT concepts.

Around 33% of the concepts included in the Problems List subset could be directly mapped with existing Snomed CT concepts; the other 77% needed the addition of one or more modifiers (post-coordination) in order to fully represent the meaning of the entry terminology concept. This rate of post-coordination was dictated by a very permissive policy allowing the use of any term requested by the users, often very specific or personalized. The total of 24,800 concepts was represented with 45,000 new relationships.

Aggregate terminology

The official Snomed CT cross maps model was implemented, a multi-classification interface was created as part of the Terminology Maintenance Software to visualize, test and modify mappings from Snomed to different classifications.

An SQL algorithm was designed (Oracle SQL) to aggregate concepts according to knowledge stored in Snomed CT relationships, like all kinds of diabetes, including diabetes complications and excluding maternal and neonatal diseases. These queries are maintained from a module in the Terminology Maintenance Software.

Terminology server

We developed a set of software services to provide the accesses to the new terminology system to other applications:

Lexical Concept Lookup: introducing a keyword retrieves exact matches or a set of similar concepts from a given subset.

Short List Retrieval: using the subset identifier, retrieves a list of concepts to include in a pick list in user interfaces.

Navigational Tree Retrieval: using the subset identifier, retrieves a list of concepts and its relationships to create tree selector in user interfaces.

New Concept Proposal: this service accepts new concepts and adds them to the pending list in the Terminology Maintenance Software.

Map to Classification: given any entry terminology concept identifier and any available classification provides the resulting code in such classification.

Terminology maintenance software

The Terminology Maintenance Software includes the following modules:

- Entry Terminology Administration: allows the creation of new concepts, description assignment and modeling of each concept with Snomed CT.
- Subset Administration: creation of new subsets, addition and removal of concepts from the subsets, defining hierarchies for tree interfaces.
- Pending Concepts or descriptions: all proposed new concepts or descriptions are stored in a list, waiting to be evaluated and modeled, ordered by the number of proposals
- Cross Maps Administration: existing cross maps can be visualized, edited and tested using this module
- Data Extraction Rules Administration: A software interface to visualize and update Snomed based data extraction queries.

Invalid terms administration

In each subset, professionals usually try to enter terms that are not valid for later use, like trying to write Cardiology in the problems list. We would like the doctor to record the proper diagnosis or reason for encounter instead.

In order to reject these terms we tag them and add an information text so the professional understands the coding guidelines of the institution. This module provides the tools for tagging these terms and editing the information.

Status report

Four trained modelers are maintaining the interface terminology, modeling pending concepts or descriptions, running routine quality control checks, and maintaining subsets.

We created an ad-hoc automatic process to recode all historic data in our clinical repository, using string matching algorithms; more than 2,200,000 entries were processed. Around 85% of the original texts received a concept code of the new entry terminology, 10% of them were recognized as invalid entries: therefore, 75% were finally mapped to Snomed CT.

The coding services are used on-line by our ambulatory and inpatient medical record, receiving around 55,000 requests each month.

Mapping to ICD-9CM from Snomed is not fully functional. It is currently running with an approximate effectiveness of 50%, and is under revision by the modelers. Previous manual codification processes are still in use until cross maps system performance improves. The use of the ICD-9CM cross-map will be addressed in a future paper.

Data extraction from the clinical information repository, using Snomed-based aggregation functions, runs daily to feed lists of target patients to a Chronic Disease Patient Registry (15). Program managers referred a great improvement in patient detection rates as compared to the previous classification-based procedure.

Discussion

Snomed CT provided an excellent structure for the initial organization of our terminology system, even when documentation not always is fully comprehensive and some proposed models will be subjected to changes in the near future due to Snomeds technical board revisions. Snomed CT has been regarded as too complex to use in real world applications; we think that properly defining the entry terminology and cross maps are essential steps in its utilization.

The task of creating an institutional entry terminology demands a lot of work, but provides an excellent service to the users, and also isolates the terminology system from Snomed CT changes in newer versions. Local concepts will always be valid, and in the worst case a correction of modeling against Snomed CT would be required.

We found that Snomed CT cross maps data to ICD-9CM is still not adequate for clinical use in our setting, requiring additional manual work on the maps. This may be caused by a different use of the classification in Argentina and the United States.

Our clinical data extraction process, using rules based in Snomed CT knowledge data, is very effective; however, these rules should be revised for each new Snomed CT version, as changes in hierarchies and models may affect its effectiveness.

Further reduction of manual classification coding will require adjustments of mapping specifications and user interface changes, aimed to reduce the number of new concepts proposals, and enforcing the selection of existing terms. Due to acceptability issues, we have always tried to minimize user interface constrains, thus implementation of these changes will be a slow process.

The most important achievements of our new terminology system are the centralization of knowledge representation, using a much more detailed terminology system. Clinical data entered at any place of the institution, and level of care, is represented uniformly through all the health information system.

References

[1] Cimino JJ. From data to knowledge through concept-oriented terminologies: experience with the Medical Entities Dictionary. J Am Med Inform Assoc. 2000 May-Jun;7(3):288-97.

[2] Munsen M SY, Shortliffe E. Clinical Decision-Support Systems. In: Shortliffe E CJ, editor. Biomedical Informatics: Springer; 2006.

[3] Wang AY, Sable JH, Spackman KA. The SNOMED clinical terms development process: refinement and analysis of content. Proc AMIA Symp. 2002:845-9.

[4] Humphreys BL, Lindberg DA, Schoolman HM, Barnett GO. The Unified Medical Language System: an informatics research collaboration. J Am Med Inform Assoc. 1998 Jan-Feb;5(1):1-11.

[5] Warren JJ, Collins J, Sorrentino C, Campbell JR. Just-in-time coding of the problem list in a clinical environment. Proc AMIA Symp. 1998:280-4.

[6] Dolin RH, Mattison JE, Cohn S, Campbell KE, Wiesenthal AM, Hochhalter B, et al. Kaiser Permanente's Convergent Medical Terminology. Medinfo. 2004;11(Pt 1):346-50.

[7] Spackman KA. Normal forms for description logic expressions of clinical concepts in SNOMED RT. Proc AMIA Symp. 2001:627-31.

[8] Spackman KA, Campbell KE. Compositional concept representation using SNOMED: towards further convergence of clinical terminologies. Proc AMIA Symp. 1998:740-4.

[9] Cerner. Cerner and SNOMED Enhance the Sharing of Global Electronic Health Information 2004. (http://www.cerner.com/public/AboutCerner_2a.asp?id=3476)

[10] Chute CG, Elkin PL, Sherertz DD, Tuttle MS. Desiderata for a clinical terminology server. Proc AMIA Symp. 1999:42-6.

[11] Fung KW, Hole WT, Nelson SJ, Srinivasan S, Powell T, Roth L. Integrating SNOMED CT into the UMLS: an exploration of different views of synonymy and quality of editing. J Am Med Inform Assoc. 2005 Jul-Aug; 12(4):486-94.

[12] Luna D, Bernaldo de Quiros FG, Garfi L, Soriano E, OFlaherty M. Reliability of secondary central coding of medical problems in primary care by non medical coders, using the International Classification of Primary Care (ICPC). Medinfo. 2001;10(Pt 2):300.

[13] Snomed. SNOMED Clinical Terms User Guide: College of American Pathologists; 2005.

[14] Snomed. SNOMED Clinical Terms Technical Reference Guide: College of American Pathologists; 2005.

[15] Dawidowski A, Figar S, Langlois E, Marchetti M, Baum A, Montenegro S, et al. Sensibility of a chronic disease management program software to identify diabetic patients. 8th SIS; 2005 September 2005; Santa F, Argentina: Sociedad Argentina de Informtica e Investigacin Operativa (SADIO); 2005.

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Toward the Interoperability of HL7 v3 and SNOMED CT: A Case Study Modeling Mobile Clinical Treatment

Amanda Ryan[a], Peter Eklund[b], Brett Esler[c]

[a]*Faculty of Commerce, University of Wollongong, Australia*
[b]*Centre for Health Services Development, University of Wollongong, Australia*
[c]*Pen Computer Systems, Australia*

Abstract

Semantic interoperability in healthcare can be achieved by a tighter coupling of terminology and HL7 message models. In this paper, we highlight the difficulty of achieving this goal, but show how it can become attainable by basing HL7 message models on SNOMED CT concepts and relationships. We then demonstrate how this methodology has been applied to a set of clinical observations for use in the ePOC project, and discuss our findings.

Keywords:

SNOMED, HL7, interoperability, terminology

Introduction

One of the biggest obstructions to communication in healthcare occurs when there are multiple ways of describing a single concept. Using a standard terminology of clinical concepts is the first step to unambiguous information exchange. If the primary clinical data entered into a system is structured using a standard expressive vocabulary, this data can then be used for secondary purposes in later stages of patient care, such as search of and selective presentation of data, decision support, and statistical evaluation [1].

In the course of constructing the Kaiser-Permanente *Convergent Medical Terminology* (CMT), Dolin, et al have reviewed three major terminologies and described their weak points [2]. The terminologies they employ are *SNOMED CT*, laboratory *LOINC*, and *First DataBank* drug terminology. In particular, with respect to SNOMED CT they conjecture that standardizing certain subsets of concepts for use in HL7 messages would greatly enhance interoperability, but first some research would have to be conducted into where the HL7 RIM and SNOMED CT overlap, and some formal guidelines be developed.

In [3] a proposal and possible solution to ambiguous communication involving basing HL7 message models on SNOMED CT concepts and relationships was discussed. Further to this, SNOMED CT concepts corresponding to a set of clinical observations have been mapped to HL7 message models for use in the *e*POC (*electronic* Point of Care) project, and as a proof-of-concept.

The *e*POC project is a multi-phase, iterative R&D project with a research focus involving the development of a prototype Personal Digital Assistant (PDA) based *point-of-care* information delivery system for The Ambulatory Care Team (TACT), Northern Illawarra, Australia [4]. TACT has a requirement to enter data at the point-of-care about a patient condition and the clinical activities carried out during the ambulatory visit. In its implementation, *e*POC addresses the difficulties of federating disparate data, HL7 messaging, implementation of clinical terminology, and the limitations of the mobile platform within a practical community clinical health service environment.

The terminology we are concerned with in the *e*POC project is SNOMED CT, as we are focusing on clinical data. Laboratory and drug codes are outside the scope of the *e*POC project.

Our preliminary findings from mapping clinical observations data required by TACT to HL7 based on SNOMED CT concepts and their relationships are discussed in the following sections.

Methods

HL7 Version 3

Version 2 of the HL7 standard is widely implemented; however it has been increasingly difficult to support health information interoperability in today's rich computing environment using this methodology. Version 3 of the HL7 standard incorporates a new paradigm for information representation and messaging in comparison to HL7 version 2 in terms of a new information model and intrinsic extensibility [5]. The Reference Information Model (RIM) is the cornerstone of the HL7 Version 3 development process. The RIM is an object model in the form of a large UML-like representation of clinical data and is a powerful abstract model of healthcare [1, 5, 6].

The difficulty arises when trying to represent clinical concepts and constructs within the HL7 framework. This is where SNOMED CT comes into play as a standard terminology for use in conjunction with the information model.

SNOMED CT

SNOMED CT is a universal health care terminology, the aim of which is to make health care knowledge usable and accessible wherever and whenever needed [7]. The SNOMED CT core terminology provides a common language that enables a consistent way of capturing, sharing and aggregating health data across specialties and sites of care. SNOMED CT contains more than 344,000 active concepts: the most comprehensive clinical vocabulary available in any language, covering most aspects of clinical medicine [8].

Clinical observations

Eight clinical observations are made by TACT clinical staff every time they visit a patient. These clinical observations were chosen as the starting point in the *e*POC project for sending messages containing clinical information from the central database to the PDA and vice versa.

The eight clinical observations are:

- Temperature
- Pulse
- Respiration
- Blood Pressure
- Oxygen Saturation (O_2Sat)
- Weight
- Blood Sugar Level
- Urinalysis

After meetings and discussion with TACT clinical staff and domain experts, SNOMED CT codes and data structures for each of these observations were ascertained. From this point, HL7 messages were modeled based on the relationships between the chosen SNOMED CT concepts.

SNOMED-CT concepts and relationships

According to Markwell [9], statements about clinical findings can be divided into two categories: those in which there are two clearly distinct facets (example 1) and those which are often captured as a single "nominalized" expression (example 2):

- Pulse rate = 82 beats/minute (1)
- Has a fracture of his right femur(2)

Looking at (1), the two facets are separated by the equal sign (=); the first being the property which was observed (Pulse rate) and the second the result of the observation (82 beats/minute).

Clinical findings in the second category (2) raise a plethora of issues in where, or even whether, to split the different concepts represented by the statement. Some of the choices are to split them as shown in examples 3 and 4:

- Property = right femur, finding = fracture (3)
- Omit property, finding = fracture of right femur (4)

In (3), the clinical finding has been split into two-facets in an attempt to model the simpler type of finding. In (4), the concepts represented in the clinical finding have not been split at all. It can be seen where problems may arise when different people use different methods of representation, especially in situations where there are many concepts combined to create one finding, hence causing many more forms of representation, e.g., "the patient has a family history of myocardial infarction".

All the clinical observations in our set of eight are of the two-faceted form (except urinalysis which is made up of many two-faceted observations). With this in mind, it was decided not to use SNOMED CT concepts from the *Clinical Finding* hierarchy as the description codes for the property to be recorded, instead it was decided that it is more effective to represent a clinical finding in this context as a SNOMED CT concept from the *Observable Entity* hierarchy, plus a value. This is in line with the two-faceted model, the first facet being the Observable Entity and the second facet its observed value.

Concepts from the Clinical Finding hierarchy are then used in decision support to describe inferred concepts taken from the recorded observation codes and their associated value. For example, if the value recorded for a patient's blood pressure is above a certain point, the concept of *high blood pressure* from the Clinical Finding hierarchy will also be recorded against the patient. We will not go into much detail about decision support, as discussion of the inferred finding aspect of our model is not the subject of this paper.

Figure 1 shows a subset of the SNOMED CT Observable Entity hierarchy used in our messages. The top level concept in this diagram is *vital signs*. Only four of the eight clinical observations required by TACT have corresponding concepts which are subtypes of the *vital signs* concept, the other concepts (corresponding to O_2Sat, Weight, Blood Sugar Level and Urinalysis) are at the same level in the hierarchy as *vital signs*, and are all subtypes of *observable entity*. The "vital signs" subset is used as an example here for simplicity and readability. The circles in the diagram represent the concepts and the arrows depict a *subtype* relationship, the opposite direction being the *IsA* relationship.

The procedures followed to make these observations have corresponding concepts under the SNOMED CT *Procedures* hierarchy, with near-identical relationships. Figure 2 shows a subset of the Procedure hierarchy used in our messages.

HL7 messages based on SNOMED CT structures

One of the findings from Dolin, et al. [2] in the CMT project is that a tighter coupling of SNOMED CT and HL7 would greatly enhance interoperability between healthcare systems. They argue that different implementations of HL7 could use different code sets and coding schemes. This problem can start to be solved by doing away with the HL7 tables and using SNOMED CT code subsets exclusively for all code-type fields, thus setting a standard coding scheme for use throughout the model. The two exceptions to this rule are the HL7 *classCode* and *moodCode* fields, although the allowable codes for these two fields are already defined by standard, non-extensible code sets within HL7, so this will not create a problem.

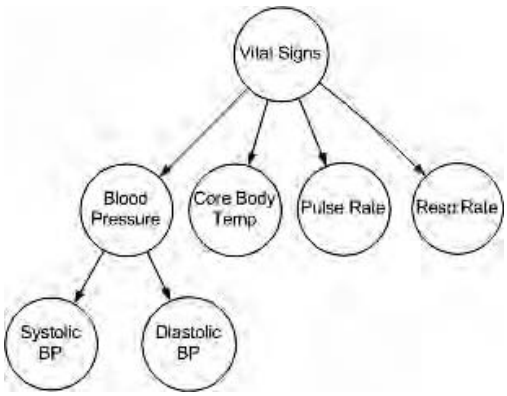


Figure 1 - Hierarchy of a subset of the SNOMED CT codes used in our Clinical Observation. The arrows depict a subtype relationship, the opposite direction being the IsA relationship.

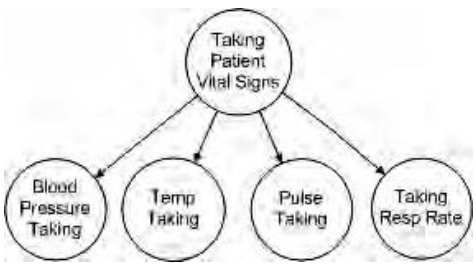


Figure 2 - Hierarchy of SNOMED CT procedure codes corresponding to the subset of observations shown in Figure 1. The arrows depict a subtype relationship, the opposite direction being the IsA relationship.

The eight clinical observations are represented in the HL7 model by instances of the Observation specialization of the Act class. Issues involving the use of terminology in the *Observation.code* and *Observation.value* fields have caused much discussion and debate [9], as discussed in the previous section. However, in the case of the two-faceted clinical finding, the representation is straight-forward. In this case, the *Observation.code* field contains the SNOMED CT code for the observation (or property), and the *Observation.value* field contains the measured value (or finding) for the observation.

The procedures shown in Figure 2 are put into the *Observation.methodCode* field to further enrich the model, by capturing both of the SNOMED CT hierarchies within the one HL7 Act "hierarchy".

The inferred findings discussed briefly in the previous section are put into the *observationInterpretation* field, but no relationships can be inferred between the observations based on this field.

Future SNOMED CT to HL7 mapping will also include the *targetSiteCode* field of the HL7 Observation class, in which the body site will be represented by a SNOMED CT *Body Structure* code.

SNOMED CT hierarchy represented in HL7

The SNOMED CT *IsA* relationship can be considered in two ways when representing it in the information model for clinical observations.

The first is that the parent (destination of the *IsA* relationship) is a *collector* (list) of entries of subtype clinical observations. This SNOMED CT concept hierarchy is preserved by the *code* field in the Observations and the HL7 *component* and *componentOf* ActRelationships. For example, in Figure 1, *blood pressure* is a subtype of *vital signs* and can be said to have the relationship *IsA* with `vital signs`. The SNOMED CT *IsA* relationship corresponds to a HL7 *componentOf* relationship, for this field.

Likewise, the relationships between the Procedure concepts for each Observation in the *methodCode* field are also preserved. For example, in Figure 2, *blood pressure taking* is a subtype of *taking patient vital signs* and can be said to have the relationship *IsA* with `taking patient vital signs`. Thus, the HL7 *componentOf* relationship also represents the *IsA* relationship for this field.

The second way of considering the SNOMED CT *IsA* relationship is to think of the parent as a *generalization* of the subtypes and thus defining a commonly supported pattern of supported specializations. The specializations constrain the allowed use of attributes, and allowed value domains/sets for attributes. This includes coded attributes in the specialization classes being restricted to valid SNOMED CT hierarchies. This has the effect of describing a HL7 v3 model hierarchy of HL7 templates with a messaging model being used as the base such as HL7 Care Provision Domain [10] messaging or Clinical Document Architecture [11].

Figure 3 shows an excerpt of the HL7 model based on the vital signs subset, along with the hierarchies shown in Figures 1 and 2, with arrows showing where the codes for these concepts fit into the HL7 model.

This message model has been implemented in the *e*POC project for use in communication between the PDA and the central database. Field trials of *e*POC are scheduled for May, in which TACT clinical staff will employ the use of the PDA with the *e*POC application in their daily home visits.

Conclusions

By basing HL7 message structures on SNOMED CT relationships a tighter coupling between data model and terminology can be created. The preliminary application of this model for simple observation data has shown that it is possible and applicable. Careful thought is being put into applying the model for more complicated data, i.e. the "nominalized" type of data shown in example (2).

The work shown here is just the beginning of mapping SNOMED CT relationships to HL7 message structures. We are moving toward a semi-automated system for creating HL7 messages on the fly, based on SNOMED-CT concepts selected by clinicians at the point-of care. We say

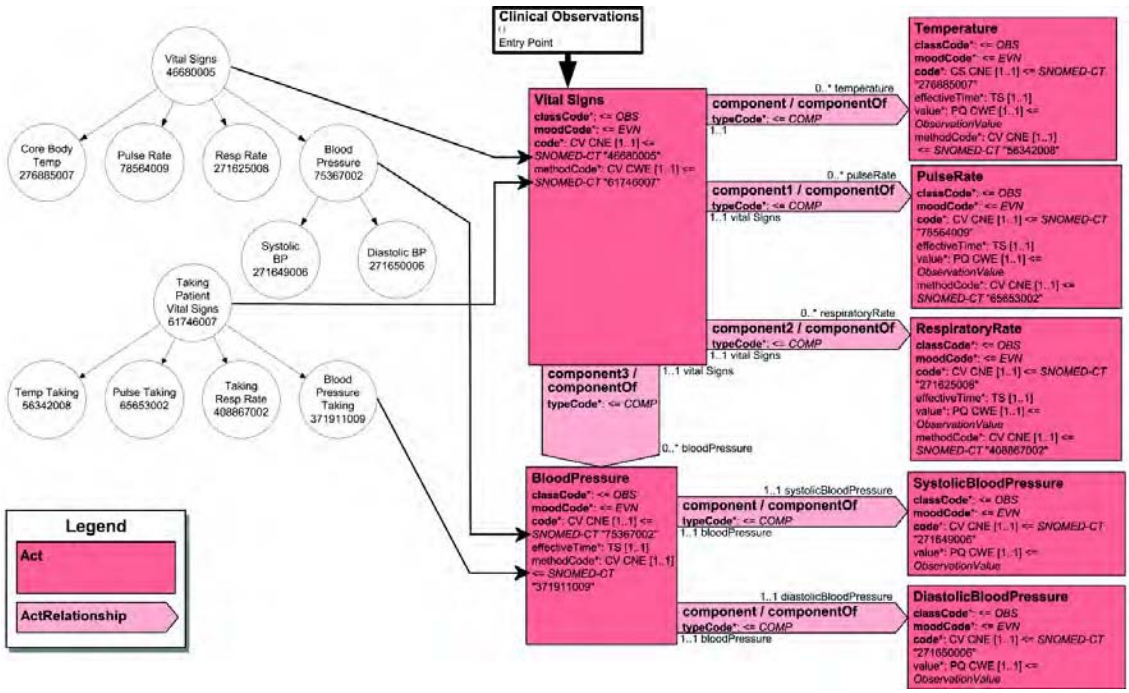


Figure 3 - Excerpt of HL7 model based on SNOMED CT Relationships shown in Figures 1 and 2. The ActRelationships correspond to the arrows in the SNOMED CT figures.

"semi-automated" because we would like the messages to be generated with as little human input as possible, but realize that achieving full automation of this process is highly unlikely. An example rule based on the clinical observation data presented here is that a SNOMED CT Observable Entity must be represented in HL7 as an Observation specialization of the Act class, with the SNOMED CT concept ID populating the Observation.code field.

Future directions of research will focus on furthering semantic interoperability between Healthcare systems using what we see as the best technology in the area.

Acknowledgments

The ePOC PDA Project is an Australian Research Council (ARC) Linkage Grant with collaborative funding, research and development by the University of Wollongong, Flinders University and University of South Australia, South Eastern Sydney Illawarra Area Health Service (SESIAHS) and a leading Australian Health Informatics company, Pen Computer Systems Pty Ltd.

References

[1] Engelbrecht R, Ingenerf J, and Reiner J. Relevance of Terminological Standards and Services in Telemedicine. In Zielinski K, Duplaga M, and Ingram D, eds. Information Technology Solutions for Healthcare, Springer, 2006.

[2] Dolin RH, Mattison JE, Cohn S, Campbell KE, Wiesenthal AM, Hochhalter B, LaBerge D, Barsoum R, Shalaby J, Abilla A, Clements RJ, Correia CM, Esteva D, Fedack JM, Goldberg BJ, Gopalarao S, Hafeza E, Hendler P, Hernandez E, Kamangar R, Khan RA, Kurtovich G, Lazzareschi G, Lee MH, Lee T, Levy D, Lukoff JY, Lundberg C, Madden MP, Ngo TL, Nguyen BT, Patel NP, Resnick J, Ross DE, Schwarz KM, Selhorst CC, Snyder A, Umarji MI, Vilner M, Zer-Chen R, and Zingo C. Kaiser-Permanente's Convergent Medical Terminology. In Fieschi M, et al., eds. Medinfo 2004, Amsterdam: IOS Press.

[3] Ryan A. Towards Semantic Interoperability in Healthcare: Ontology Mapping from SNOMED-CT to HL7 version 3. In Orgun MA and Meyer T, eds. Conferences in Research and Practice in Information Technology (CRPIT): Vol 72, Australian Computer Society, 2006.

[4] Sargent J, Eklund P, Ryan A, Burgess L, Cooper J, Alcock C, and Ryan D. Mobile Information Access and Diffusion in Ambulatory Care Service Settings. In Proceedings of HIC 2006, Sydney, Australia, 2006.

[5] http://www.hl7.org. Health Level 7.

[6] Vizenor, L. Actions in Health Care Organizations: An Ontological Analysis. In Fieschi M, et al., eds. Medinfo 2004, Amsterdam: IOS Press.
[7] http://www.snomed.org. SNOMED International.
[8] http://www.nlm.nih.gov/research/umls. Unified Medical Language System (UMLS), United States National Library of Medicine.
[9] Markwell D. Representing Clinical Findings with HL7v3 and SNOMED CT. White Paper for discussion by SNOMED International Standards Board and Concept Model Working Group, and HL7 Modeling and Methodology, TC Vocabulary TC and TermInfo Project, 2006.
[10] http://www.hl7.org/v3ballot/html/domains/pc/hmrepc.htm. HL7 v3 Ballot Care Provision Domain.
[11] http://www.hl7.org/v3ballot/html/infrastructure/cda/cda.htm. HL7 v3 Ballot Clinical Document Architecture.

Address for correspondence

Amanda Ryan
PO Box U124
University of Wollongong
Wollongong, NSW 2500
Australia
Email: ajr883@uow.edu.au

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

A Web-based SNOMED CT Browser: Distributed and Real-time Use of SNOMED CT During the Clinical Research Process

Rachel Richesson, PhD[a], Asif Syed, MD[b], Heather Guillette, MS[a], Mark S. Tuttle, FCAMI[c], Jeffrey Krischer, PhD[a]

[a]Pediatrics Epidemiology Center, University of South Florida College of Medicine, Tampa, FL, USA
[b]SNOMED International, College of American Pathologists, Chicago, IL, USA [c]Apelon, Inc., Alameda, CA, USA

Abstract

To facilitate the use of standard terminologies in clinical research data collection, members of the Rare Disease Clinical Research Network have developed tools to support study investigators and research staff to code clinical research data using SNOMED CT at the point of research. This tool is customized to help the user find appropriate SNOMED CT concepts quickly, and has implications for the successful implementation of data standards to facilitate high quality research data. This paper gives an overview of an automated tool for accessing, searching, and navigating SNOMED CT real-time, at distributed and remote clinical study locations. Also, the features of the tool that enable complete data are presented, as well as possible metrics for evaluation in terms of compliance, consistency, and reliability of coding.

Keywords:

SNOMED CT, terminology tools, clinical research informatics, data collection

Introduction

The Rare Diseases Clinical Research Network (RDCRN) consists of ten clinical research consortia and a central Data and Technology Coordinating Center (DTCC). The RDCRN conducts numerous research studies and clinical trials in many diverse and rare diseases. One goal of the RDCRN is to encourage data sharing and use of standardized data. The RDCRN is committed to the use of data standards, including SNOMED CT, the Consolidated Health Informatics (CHI) initiative's recommended standard for diagnoses, problem lists, procedures, and anatomy. The Standards Committee of the RDCRN endorses the use of structured data over free-text wherever possible, and supports the use of standardized terminologies or classifications for data coding. The DTCC promotes data standards and tools for efficient study implementation and accurate, complete, and timely data capture. The researchers of the RDCRN are distributed at over 46 sites in the U.S. and abroad, and require tools to facilitate remote, real-time, free access to SNOMED CT for searching and browsing concepts relevant to data collection in diverse studies in rare diseases. This paper describes one such tool and offers preliminary evaluation data.

One of the biggest obstacles in the implementation and widespread adoption of standard terminologies is lack of appropriate tools to access, display, search, and navigate terminologies of varying structures. [1-10] The need for these tools is even greater when the terminologies are large or complex. [9, 11-13] The literature on evaluating terminology coding tools is sparse [14] although many have suggested that interfaces for coding support tools can be an important confounding factor in inter-rater reliability studies. [11, 15-18] Many features can be incorporated into tools to assist users in finding desired concepts thereby increasing the efficiency and compliance of these tools. [19-21] The SNOMED CT Browser described here is the first type of many tools that are required for widespread adoption of SNOMED CT and other terminological standards in distributed clinical research settings.

Informatics principles and tools are entering into the clinical research domain [22-24] and one target area is the use of electronic data capture to increase the speed and quality of data collection in distributed research studies. In general, the items (or questions) on clinical research case-report forms (CRFs) are of 2 types: 1.) Structured, predefined elements [the majority] can be coded once for the entire life of the study. These questions and coded answers do not change for the duration of the study *(e.g., "Claudication? present/absent", ... "WBC count = ____", "Currently on prednisone? yes/no".)* 2.) Open-ended questions [the minority], however, require dynamic coding. The content of open-ended questions is not static for the study, and cannot be predicted before the start of the study. *(e.g., include: "List medications: ____" , "List Allergies: ____" , "Cause of Death: ____" , "Specify Other Findings:____" .)* In the case of open-ended questions, however, the space of potential answers is huge, perhaps infinite. In theory, the coding should be more accurate if it is generated by the individual identifying the concepts or making the observations.

A tool was needed to support real-time SNOMED CT coding at distributed clinical research sites to allow researchers to code concepts themselves at the point of research. This tool facilitates access to standard terminol-

ogy, such as SNOMED CT, "on demand" and can enforce the idiosyncratic rules of each terminology. Additionally, these tools should reduce the time and complexities of the coding tasks, thereby reducing the burden of this strategy on the research team. This paper describes this tool, its implementation status, and preliminary evaluation metrics.

Methods

The DTCC and Apelon, Inc. developed a web-based SNOMED CT Browser that enables distributed clinical research staff to code clinical research data from distributed clinical research settings. This tool was developed to address the following requirements: real-time access to SNOMED CT (both for searching and browsing) for multiple users at distributed sites, little training available, low cognitive burden on user, intuitive for clinical research staff with no SNOMED CT expertise, embedded within existing on-line data collection tools, and efficient return time of selected SNOMED CT concepts.

SNOMED CT browser description

This SNOMED CT browser operates within the context of automated protocol management tools developed by the DTCC to support all RDCRN studies. The browser is embedded in the online Case Report Forms (CRFs) to facilitate coding on demand by clinical research staff. The browsers connect to remote terminology host (Apelon, Inc.) via Web Service. Search terms, filters and other supporting information are tagged in XML, and then transferred to Application Service Provider via SOAP. The remote terminology host uses U.S. National Library of Medicine's public use version of SNOMED CT. The custom interface allows distributed researchers to search for specific terms or to browse SNOMED CT's native multi-hierarchical structure.

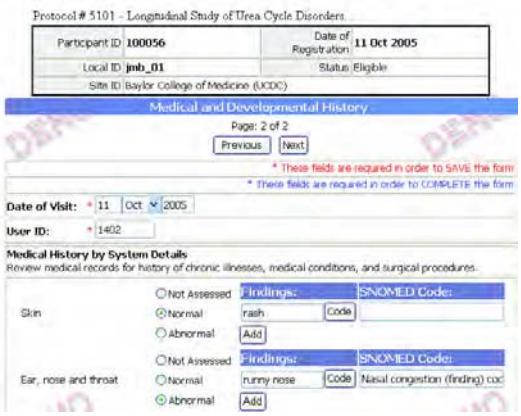


Figure 1 - Context of SNOMED CT browser.

The SNOMED CT browsers are located next to the questions on electronic Case Report Forms.

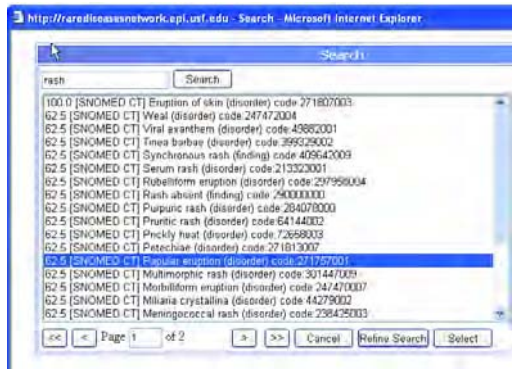


Figure 2 - SNOMED browser – original search results window

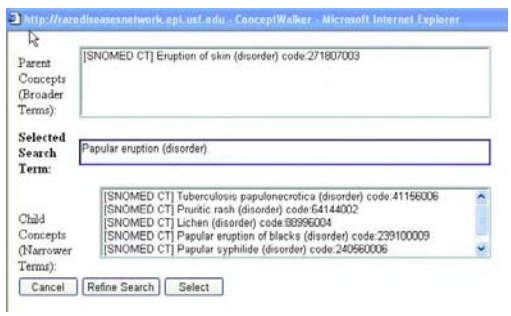


Figure 3 - Browsing broader and narrower concepts window

From the online CRF (Figure 1), the user enters a search string and selects the code button. A new window displays all potential SNOMED CT concept matches. The browser searches SNOMED CT synonyms (including abbreviations) to help search concepts. For example, if the term congestive heart failure is desired, one could type the entire search term or the abbreviation "CHF". [Note: this is part of the SNOMED CT terminological content.] The browser uses an algorithm that ranks returned codes based upon the number of search terms (or synonyms) found compared to the number of words in each returned concept. The returned concepts can span over multiple pages which the user can view. (Figure 2)

If the user does not see the desired concept, any one of the returned concepts can be selected for a "refined search", allowing users to browse related SNOMED CT concepts. A new window (Figure 3) displays all the SNOMED CT codes that are more specific to or direct parents of the refined search concept, and is divided into three parts:

- "Parent Concepts (Broader Terms)" - are displayed in the top most part of the page
- "Selected Search Term"- displays the concept selected for "refine search" in the previous page
- "Child Concepts (Narrower Terms)"- are displayed in the lower part of page below selected search term

The user can further refine the search by selecting a concept in the "Child Concepts" textbox and then hitting the "Refine Search" button. Every time a particular concept and "Refine Search" button is selected, the user will see a refined set of possible parent and child SNOMED CT concept codes.

Features

Certain search features of the SNOMED browser enable users to limit the number of returned SNOMED CT concepts. Users can enter multiple words (with AND, OR, or NOT) to limit a search, or search for exact phrases using quotes (e.g., "heart attack"). A wildcard character is also supported. Codes selected from the browsing tool are populated into the online case report form for later viewing. The stored data includes the SNOMED CT concept name and ID number, and the version of SNOMED CT accessed. The original search strings are stored for later analyses and quality control activities. To limit the number of returned SNOMED CT concepts and to increase likelihood of accurate coding by SNOMED CT novices, only context-relevant subsets of SNOMED CT are returned to users. These subsets are defined by the informaticist at the DTCC and the filters are transparent to users.

To increase compliance, the DTCC considered making the actual SNOMED code a required field for saving and submitting the CRF data. However, because of the chance that desired concepts in rare disease research might not yet be part of SNOMED CT, this feature would frustrate researchers and hinder data collection. Under development are system features to remind the user that they have not selected a SNOMED CT code and a feature to allow them send new code or help requests directly to the DTCC informaticist at the time of data collection.

Results

Implementation status

As of January 2007, the RDCRN has produced 285 electronic case report forms (CRFs) containing more than 26,000 questions for 27 clinical research protocols. Most CRF questions are in the form of structured static questions, which are coded at the DTCC once for the life of the study. A fraction of the CRF questions are open-ended (e.g., "list abnormal skin findings") and appropriate for this real-time remote SNOMED CT coding tool. Approximately 300 SNOMED browsers are included in these RDCRN CRFs. Because a single informaticist at DTCC reviews all case report forms prior to study implementation, there is confidence that these 300 SNOMED CT browsers represent all open-ended questions on case report forms for which SNOMED CT is the appropriate data standard.

Each browser has customized criteria for sub-setting SNOMED CT – i.e., limiting returned search terms by one or more SNOMED CT Parent Concepts based on the context of the CRF question (Table 1).

Table 1. Examples of context-dependent SNOMED CT sub-sets.

CRF Question:	SNOMED CT returns restricted to:
Reason for visit/ admission	Clinical finding (finding)
Kidney and urinary tract history	Urinary system finding (finding)
Muscles, bones and joints history	Musculoskeletal finding (finding)
Nervous system and development history	Neurological finding (finding) Developmental disorder (disorder) Mental disorder (disorder)

Of the 300 SNOMED CT browsers that are embedded within the RDCRN data collection, approximately 75% represent current findings (e.g., context is physical exam, clinical assessment forms) and 25% represent historical findings (e.g., medical history context).

Over 150 investigators and research staff representing over 20 studies in the Rare Disease Network have been trained on this tool. The training on the SNOMED CT Browser Tool is part of a day-long training on the clinical research protocol, and consists of a 15 minute overview of SNOMED CT and a 10 minute instruction on the tool and search techniques. Anecdotally, the tool is intuitive to users and well-accepted.

Future evaluation data sources

Evaluation of this SNOMED CT Browser tool includes assessing whether it is actually used (compliance) and whether the user-coded data is accurate (quality). SNOMED Coding Compliance Reports are generated for the DTCC to monitor quality and compliance. These reports identify and enumerate where free-text data was entered into the browsing tool but no SNOMED CT codes saved. This helps DTCC staff identify and educate users who are submitted free-text data but not applying SNOMED CT codes, and to investigate whether failures to add SNOMED CT codes to free text clinical finding data result from non-compliance to difficulties navigating SNOMED CT to missing clinical research content in SNOMED CT. Using the proportion of SNOMED CT-eligible open-ended data that is actually coded as a measure, comparisons of compliance can be calculated between users, sites, and disease/research domains can be facilitated using these data.

The data on original search strings and user-selected SNOMED CT concepts are being saved for future quality control analysis. These data can be analyzed to research both the "appropriateness" (as determined by a SNOMED

CT expert physician) of the selected SNOMED CT code, and the level of specificity of selected codes compared to the original search string (e.g., "inspiratory wheeze" as a selected SNOMED CT concept is more specific - by only one parent relationship - than the search string "wheeze"). These same quality data could be used to answer questions about variation in level of specificity across users in controlled experiments using test cases. Such questions are critical to address in distributed research settings where multiple researchers are responsible for study data, yet we see few comparisons of coding specificity in the literature. [11, 16] These questions of coding quality and appropriateness are best addressed by those with dual training in both SNOMED CT and in the clinical research domain of interest. The DTCC is in the process of identifying individuals with clinical domain expertise for extended SNOMED CT training and extended quality control roles.

Discussion

In theory, the coding of clinical research data should be more accurate if it is generated by the individual identifying the concepts or making the observations, and tools for coding at the point of research can be important aides for standards implementation. Quality coding requires some understanding of the terminological structure of SNOMED CT, and the browser application described here make the conceptualization of the SNOMED CT structure and term navigation intuitive for the user. Additionally, these browsers can be designed to subset only "reasonable" concepts from the terminology, reducing the cognitive burden and time for the user, and possibly increasing the quality of coded data.

Researchers need "terminology on demand" – reliable, fast access to updated data standards at the point of clinical research. Tools that allow customized views and subset the terminologies can reduce search time and coding burden, and perhaps increase compliance.

Because SNOMED CT is a concept-based terminology, multiple synonyms can be associated with a given concept. This allows local variation in terminology use, but stores the final study data in a uniform way. The use of SNOMED CT and supporting tools can improve data consistency because all clinical centers collect same data in the same format. Management of terminology support (including tools development and maintenance and licensing) is centralized, and the management of terminology server is outsourced. Tools that facilitate intuitive access to complex terminologies reduce the need for training on the terminology structure. Quality of coding and consistent use of this tool should be examined in future. Although the infrastructure and data sources for these types of exploration exist, dual experts (rare disease domain and SNOMED CT) will be required.

A great challenge in permeating data standards such as SNOMED CT in distributed clinical research settings is that the proper use of the standards requires domain knowledge *and* knowledge of the standard. It is more realistic to give the proper standards/terminology training to researchers with domain expertise than the reverse. However, compliance depends on DTCC ensuring a negligent task/work/time burden on the researchers, and maximizing their perception of the value of standard data coding.

Conclusion

The usability and intuitive design of this tool eliminates the need for researchers to study the SNOMED CT terminology structure, and could increase compliance and quality of coding. These tools should reduce the time and complexities of the coding tasks, thereby reducing the burden of standards use on the research team. The browser provides researchers the means to obtain complete, accurate, and timely data which is consistent across multiple clinical centers. In rare diseases, this can increase the efficiency of the research process and eliminate wasted resources, including the time and effort of the study participant resulting from incomplete, inaccurate, or missing data. Future research should focus on quality and reliability of selected SNOMED CT concepts using data sources described above.

Acknowledgments

The project described was supported by Grant Number RR019259 from the National Center for Research Resources (NCRR), a component of the National Institutes of Health (NIH). Its contents are solely the responsibility of the authors and do not necessarily represent the official views of NCRR or NIH. The authors wish to thank the Office of Rare Diseases, and Ken Young and Michael Abbondondolo, USF, for their assistance developing this tool. Terminology hosting services provided by Apelon, Inc. The authors wish to thank Rich Mathes, Apelon, Inc., for his participation in this project.

References

[1] CDC, *Information Needs and Uses of the Public Health Workforce - Washington, 1997-1998.* Morbidity and Mortality Weekly Report, 2000. **49**(6): p. 118-120.

[2] Elkin, P.L., et al., *Guideline and Quality Indicators for Development, Purchase and Use of Controlled Health Vocabularies.* International Journal of Medical Informatics, 2002. **68**(1-3): p. 175-186.

[3] Fitzmaurice, J.M., K. Adams, and J.M. Eisenberg, *Three Decades of Research on Computer Applications in Health Care.* JAMIA, 2002. **9**(2): p. 144-160.

[4] Koop, A. and R. Mosges, *The use of handheld computers in clinical trials*, in *Controlled Clinical Trials*. 2002. p. 469-480.

[5] Neces, R., et al., *Enabling Technology for Knowledge Sharing.* AI Magazine, 1991. **12**(3).

[6] Cimino, J.J., *Terminology Tools:State of the Art and Practical Lessons.* 2001.

[7] Campbell, K.E., et al., *Scalable methodologies for distributed development of logic-based convergent medical terminology.* Methods Inf Med, 1998. **37**(4-5): p. 426-39.

[8] Douglas, K. and M. Nubie, *System design challenges: the integration of controlled vocabulary use into daily practice.* Stud Health Technol Inform, 1997. **46**: p. 167-71.

[9] Rosenbloom, S.T., et al., *Review Paper. Interface Terminologies: Facilitating Direct Entry of Clinical Data into Electronic Health Record Systems.* J Am Med Inform Assoc, 2006. **13**: p. 277-288.

[10] McDonald, C.J., et al., *The Regenstrief Medical Record System: a quarter century experience.* Int J Med Inform, 1999. **54**(3): p. 225-53.
[11] Chiang, M.F., et al. *Reliability of SNOMED-CT Coding by Three Physicians using Two Terminology Browsers*. in *American Medical Informatics Association Annual Symposium*. 2006. Washington, D.C.
[12] de Lusignan, S., *Codes, classifications, terminologies and nomenclatures: definition, development and application in practice.* Inform Prim Care, 2005. **13**(1): p. 65-70.
[13] Bowie, J., *Four options for implementing SNOMED.* J AHIMA, 2004. **75**(7): p. 30-3; quiz 35-6.
[14] Chute, G.C., S.P. Cohn, and J.R. Campbell, *A Framework for Comprehensive Health Terminology Systems in the United States.* Journal of the American Medical Informatics Association, 1998. **5**(6): p. 503-510.
[15] Fung, K.W., et al., *Integrating SNOMED CT into the UMLS: an exploration of different views of synonymy and quality of editing.* J Am Med Inform Assoc, 2005. **12**(4): p. 486-94.
[16] Rothschild, A.S., H.P. Lehmann, and G. Hripcsak, *Inter-rater Agreement in Physician-coded Problem List*, in *American Medical Informatics Association*. 2005: Washington, D.C.
[17] Bakken, S., et al., *Evaluation of the clinical LOINC (Logical Observation Identifiers, Names, and Codes) semantic structure as a terminology model for standardized assessment measures.* J Am Med Inform Assoc, 2000. **7**(6): p. 529-38.
[18] Chiang, M.F., et al., *Representation of ophthalmology concepts by electronic systems: adequacy of controlled medical terminologies.* Ophthalmology, 2005. **112**(2): p. 175-83.
[19] Tuttle, M.S., et al., *Navigating to Knowledge.* Meth Inform Med, 1995. **34**: p. 214-231.
[20] Tuttle, M.S., et al., *Metaphrase:An Aid to the Clinical Conceptualization and Formalization of Patient Problems in Healthcare Enterprises.* Methods of Information in Medicine 1998, 1998.
[21] Chute, C.G., et al., *Desiderata for a Clinical Terminology Server.* 2004.
[22] Altman, R.B., *The Interactions Between Clinical Informatics and Bioinformatics: A Case Study.* J. Am. Med. Inform. Assoc., 2000. **7**(5): p. 439-33.
[23] Brandt, C.A., et al., *Approaches and Informatics Tools to Assist in the Integration of Similar Clinical Research Questionnaires.* Methods of Information in Medicine, 2004. **43**: p. 156-162.
[24] Rindfleisch, T.C. and B.L. Brutlag, *Directions for Clinical Research and Genomic Research into the Next Decade. Implications for Informatics.* Journal of the American Medical Association, 1998. **5**: p. 404-411.

Address For Correspondence

Rachel Richesson, PhD, MPH
USF College of Medicine, Department of Pediatrics
3650 Spectrum Blvd., Suite 100
Tampa, FL 33612
richesrl@epi.usf.edu

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Using SNOMED CT® as a Reference Terminology to Cross Map Two Highly Pre-coordinated Classification Systems

Steven H. Brown[ab], MD, Casey S. Husser, MD[c], Dietlind Wahner-Roedler, MD[c], Sandra Bailey[a], Linda Nugent, RHIA[a], Karla Porter, RHIA[a], Brent A. Bauer, MD[c], Peter L. Elkin, MD[c]

[a]Department of Veterans Affairs, United States of America
[b]Department of Biomedical Informatics Vanderbilt University, United States of America
[c]Mayo Clinic College of Medicine, United States of America

Abstract

We hypothesized that SNOMED CT, a granular formal reference terminology, could be used to assist in the creation of a valid crosswalk between two administrative classifications: the International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) and the U.S. Veterans Benefits Administration (VBA) disability code set. To establish a baseline, we created an ICD-9-CM terminology server and directly mapped textual descriptions of the VBA disability codes to ICD-9-CM. We next mapped ICD-9-CM and the VBA Disability codes to SNOMED CT. The SNOMED CT mappings were matched across classification systems and terms from related concepts were displayed for an expert coder's review. We report the rate of direct ICD-9-CM to VBA Disability Code mapping (26%), the eventual success of the SNOMED CT based crosswalk (95%) and the rate at which the reviewer had to add codes to complete the mapping (99%). The method using the SNOMED CT crosswalk provided significantly better coverage than the ICD-9-CM direct mapping alone (Pearson Chi Square test; $p<0.001$). We conclude that SNOMED CT can be a useful adjunct to direct mapping between administrative classifications.

Keywords:

terminology, semantics, International Classification of Diseases, Systematized Nomenclature of Medicine, United States Department of Veterans Affairs

Introduction

Networked information systems and communications standards make the mechanics of data transfer (i.e. syntactic interoperability) routine. The transfer of fully meaningful information (semantic interoperability) is a much more difficult challenge. Semantic interoperability requires comparable data, the basis for the practice of evidence-based medicine. Many clinical questions will never be addressed in a randomized, blinded, controlled clinical trial. The best hope for providing clinicians answers to these questions lies with a deeper understanding of the clinical record. Recording information at the granularity with which we practice medicine holds the promise to provide the data needed to gain an improved understanding of what constitutes the "best practice" of medicine

Compositional vocabularies are one potential answer to the problem of providing enough content completeness to be clinically useful. The degree to which a compositional mechanism can provide coverage for the concepts used by clinicians at the point of care is not completely understood. The data from the LSVT trial provides a benchmark of 58% content coverage by a large set of atomic and pre-coordinated concepts.[1] We previously reviewed SNOMED CT's ability to represent 115 complex oncology drug indications containing 1527 unique concepts.[2] We found that SNOMED CT was capable of representing 86.3% of the concepts completely and 60.1% of the linking semantics. Of the errors, 50.5% were drug names that were unknown to the terminology and 40.5% were specific treatments. SNOMED CT covered 64% of concepts used in a general medical evaluation form.[3] The majority (74.4%) of terms were compositions, and appropriate linking semantics were present in SNOMED CT for 80% of a randomly selected subset.

Content coverage alone is not enough to achieve comparable data and semantic interoperability. A major problem is that information is often coded in different terminologic systems and translation is required. Idiosyncratic pre-coordinated terminologies have been used for focused administrative purposes since long before computers became commonplace. As new business needs emerge, such idiosyncratic pre-coordinated terminologies may need to be mapped to national and international administrative code sets such as ICD-9 or ICD-10. Accurate mapping may be difficult because complex pre-coordinated terms may not have exact morph-syntactic equivalents in the corresponding terminology. We suggest that one way to create a mapping between pre-coordinated classifications is to use a granular reference terminology as an intermediary.

In this study our goal was to provide a mapping from the Veterans Benefit Administration's (VBA) Disability codes to all possibly appropriate ICD-9-CM codes. The VBA disability codes are used to classify veterans' disabilities

for the purposes of compensation and pension and to track healthcare delivery to disabled veterans. We evaluated the use of SNOMED CT v1.0 as an intermediary reference terminology to aid in the mapping process.

Materials and methods

We created an initial automated mapping between ICD-9-CM and the VBA Disability codes without using SNOMED CT as an intermediary by using the Multi-threaded Clinical Vocabulary Server (MCVS) concept-based indexing toolset.[4] To achieve this mapping, we loaded the MCVS with ICD-9-CM as the base terminology and serially mapped the VBA Disability Codes (target terminology). When necessary, the MCVS tools attempt to produce compositional expressions using terms from the base terminology to represent the original pre-coordinated term from the classification being mapped. [5, 6]

We used the same MCVS process to create an automated mapping of the terms from SNOMED CT (base terminology) and the ICD-9-CM and VBA Disability classifications (target terminologies). Figure 1 demonstrates the MCVS tool set mapping of the VBA Disability term "Vagus nerve Neuritis" using SNOMED CT as the base terminology. This mapping has two definitional elements – a disorder (neuritis) and a body structure (vagus nerve).

We categorized all mapped SNOMED CT definitional elements as either Kernel concepts, modifiers, qualifiers, severity, or laterality.[7] These distinctions were hand assigned by hierarchy to the SNOMED CT reference terminology (ple). Kernel concepts are defined as a main point of the concept, modifiers change the meaning of a term in a clinical sense (e.g., stage, grade), qualifiers change the meaning of a term in a temporal or administrative sense (e.g. history of, recurrent), severities and laterality follow the SNOMED CT hierarchies for these concepts.

We created an automated "crosswalk" between the target terminologies by joining ICD-9-CM and VBA Disability terms based on common SNOMED CT definitional elements. The first step in the automated matching process was to identify equivalent terms based on shared kernel concepts. At least one kernel concept match was required to join ICD and VBA terms. Matches for terms sharing kernel concepts were subsequently refined by identifying common modifiers, qualifiers, severity, and laterality. The output of the pair-wise crosswalk between these three classifications (figure 2) was then reviewed by two expert reviewers (clinicians), and when they disagreed a third reviewer was employed to establish consensus. When no exact match was available, the reviewers used a browser capable of displaying each of the target terminologies and determined whether 1) There was no appropriate match between the two terminologies (hence SNOMED CT legitimately could not and should not have connected the two terminologies or 2) b) The was an appropriate match between the two terminologies and SNOMED CT did not provide the appropriate mapping. In the case when SNOMED CT failed to map two synonymous concepts, a failure analysis was employed that noted what was missing in SNOMED CT that would have facilitated the mapping and as well as which classification contained the unrecognized content and/or semantics. All failures were double checked by the reviewers using a SNOMED CT terminology browser.

A web-based system (four-tier architecture with a relational database back end) was designed and implemented to facilitate the review and grading of the acceptability of the crosswalk. The system showed the reviewer the original terms, their matches and the compositional expressions to which they mapped in SNOMED CT. The interface was available to the reviewers to view and grade the output of the systems matches to the VA disability terminology and to ICD-9-CM. We also provided a set of Web based services and interfaces to allow coders to review the crosswalk between the VA disability terminology and to ICD-9-CM using SNOMED CT. This interface allowed the coders to pick the correct mapping or approve the automated mapping as warranted by the review.

The results are reported from a SNOMED CT perspective. True positives are concepts from the classifications that were found to have at least one valid match between the two classifications using SNOMED CT. The mappings were reviewed by an expert ICD-9-CM coder who had access to VBA staff in the case that questions arose regarding a VBA code's meaning. A correct mapping rate of 93% was arbitrarily judged to be acceptable accuracy for clinical use. This rate is similar to or better than many of the sensitivities for common confirmatory diagnostic tests.[8]

VBA Code	VBA Term	SNOMED Mapping
8310	Vagus nerve Neuritis	Neuritis (disorder) [84299009] [has Finding Site] Entire vagus nerve (body structure) [362466001]

Figure 1 - SNOMED-CT Mapping of a VBA Disability term

Crosswalk Mappings			Direct Mapping
VBA Term	SNOMED Mapping	ICD9 Concept	ICD-9 – VBA Term
Abscess, brain (8020)	Abscess brain [60404007]	INTRACRANIAL ABSCESS (disease) [324.0]	Abscess brain
Accommodation, paralysis of (eye) (6030)	Accommodation paralysis of (eye) [68158006]	PARESIS OF ACCOMMODATION (disease) [367.51]	eye

Figure 2: Crosswalk examples. VBA terms and ICD-9 concept are joined based on common SNOMED CT mappings. The corresponding direct mapping of VBA term to ICD-9 concept was successful for brain abscess and not successful for accommodation paralysis.

Results

SNOMED CT version 1.0 as loaded into the MCVS toolset contained 333,325 concepts and 874,391 terms. ICD-9-CM as loaded included 20,069 concepts and 29,958 terms. The investigators added 712, 294 synonyms to SNOMED CT and 36,731 synonyms to ICD-9-CM within the MCVS tool set to improve automated mapping performance. VBA Disability code terms numbered 752.

Based on automated mappings from the MCVS, 25.5% (192/752) of the VBA Disability codes mapped to a single SNOMED CT concept and 560/752 (74.5%) required a compositional SNOMED CT expression. Of the 560 VBA Disability code terms that were mapped to multiple SNOMED CT concepts, 250 required 2 concepts, 142 required 3 concepts, 64 required 4 concepts, 49 required 5 concepts, 22 required 6 concepts, 15 required 7 concepts, 8 required eight concepts, 5 required 9 concepts, and 5 required 10 concepts.

The MCVS provided an automated mapping to ICD9-CM for 359 of the 752 VBA Disability codes. Of these, 259 mapped to a single ICD-9-CM concept and 100 required multiple ICD-9-CM codes linked with Boolean operators. Of the 100 VBA Disability code terms that were mapped to multiple ICD-9-CM expressions, 84 required 2 concepts, 9 required 3 concepts, 4 required 4 concepts, 2 required 5 concepts and 1 required 7 concepts.

The ICD-9-CM direct map correctly mapped 26% of the VBA codes after expert human review. The addition of the automated SNOMED CT crosswalk to the direct map provided at least one suggested code in 98% of the cases. At least one of these suggested codes was considered an appropriate match by an expert reviewer in 95% of the cases.

The expert coder added additional codes to almost all of the suggested mappings (99%) and removed many inappropriate suggestions. When we compared the rate of direct ICD-9-CM mapping to that of the ICD-9-CM direct map plus the crosswalk the later provided significantly more coverage (Pearson Chi Square test; $p<0.001$).

Examples of mappings between base and target terminologies

Case 1: One to One Match
VBA Disabilty Term: Thrombo-angiitis obliterans (Buerger's disease)
SNOMED Term: Thromboangiitis obliterans (disorder) [52403007]
ICD-9-CM Term: THROMBOANGIITIS OBLITERANS (BUERGERS DISEASE) (disease) [443.1]

Case 2: One to Many Match
VBA Disability Term: Disseminated intravascular coagulation with renal cortical necrosis
SNOMED Terms: Disseminated intravascular coagulation (disorder) [67406007] [AND]
Renal infarction (disorder) [45456005]
ICD-9-CM Terms: ACUTE RENAL FAILURE WITH LESION OF RENAL CORTICAL NECROSIS (disease) [584.6]
[AND] DEFIBRINATION SYNDROME (disease) [286.6]

Case 3: Match Requiring Clinical Judgment
VBA Disability Term: Dementia of unknown etiology
SNOMED Terms: Dementia (disorder) [52448006] [AND]
[UNKNOWN] . Etiology (attribute) [134198009]
ICD-9-CM Terms: PRESENILE DEMENTIA (disease) [290.1]
DEMENTIA IN CONDITIONS CLASSIFIED ELSEWHERE (disease) [294.1]
PRESENILE DEMENTIA, UNCOMPLICATED (disease) [290.10]
SENILE DEMENTIA, UNCOMPLICATED (disease) [290.0]
All four ICD-9-CM terms were assessed as appropriate by an experienced coder.

Case 4: Partial Match

VBA Disability Term: Paralysis of musculocutaneous nerve
SNOMED Terms: Paralysis (finding) [44695005] - [has Finding Site]
Entire musculocutaneous nerve (body structure) [181019000]
where [has finding site] is a SNOMED relationship.
ICD-9-CM Terms: INJURY TO MUSCULOCUTANEOUS NERVE (disease) [955.4]
which only approximates the meaning of the original input string.

Discussion

SNOMED CT was useful in providing suggested mappings for a crosswalk between the VBA administrative terminology and ICD-9-CM. The SNOMED CT based crosswalk provided significantly more correct mappings than a direct ICD-9 mapping alone. Despite this improvement, additions and deletions based on expert human review were required to complete the mapping.

Mapping between two highly pre-coordinated administrative classifications is an inherently difficult task for several reasons. First, administrative classifications are likely to have been developed to meet different business needs and the existence of reliable mappings should not be assumed. Second, implicit pre-coordination is more difficult to manage algorithmically than explicit definitions and coordinations grounded in formal logic. Third, classifications are designed to group granular entities under high-level concepts based on organizing principles. Different classification systems are likely to have subtly or even significantly different underlying organizing principles that impact meaning and make cross-mapping difficult.

The MCVS crosswalk was designed to sensitively identify potential mappings between VBA Disability codes and ICD-9-CM. Due to this design choice and the fact that ICD-9-CM and the VBA Disability code systems have considerable meaning that is only defined in the user's manuals, our expectation of rejecting numerous inappropriate partial matches was realized. Furthermore, these classification systems are designed for different purposes. The authors believe the automated mapping system would yield improved results if the additional knowledge in the user's manual were included in the mapping system's rule base. We did not do a comparative study regarding how long it would have taken to perform the same task without the crosswalk but we believe that this method linked with a good user interface significantly shortened the review time. Further, we believe that the automated mapping suggestions helped the reviewer to be more complete as it found matches across hierarchies that may have been missed if not cued to the reviewer by the crosswalk system.

SNOMED CT should be considered as a useful reference terminology for providing a set of suggested matches using the crosswalk method outlined in this manuscript to map between and among administrative classifications.

Acknowledgement

This work was funded in part by a grant from the National Library of Medicine (LM06918) and grants from the Centers for Disease Control PH000022 and HK00014. The authors would like to thank Raju Kalidindi and Luka Bajzer for their technical assistance with this project.

References

[1] Humphreys BL, McCray AT, Cheh ML. Evaluating the coverage of controlled health data terminologies: report on the results of the NLM/AHCPR large scale vocabulary test. J Am Med Inform Assoc 1997;4[6]:484-500.

[2] Brown SH, Bauer BA, Wahner-Roedler DL, Elkin PL. Coverage of oncology drug indication concepts and compositional semantics by SNOMED CT. AMIA Annu Symp Proc 2003:115-9.

[3] Brown S, Elkin P, Bauer B, Wahner-Roedler D, Husser C, Temesgen Z, Hardenbrook S, Rosen bloom S. SNOMED CT®: Utility for a General Medical Evaluation Template. American Medical Informatics Association Fall Symposium 2006.

[4] Elkin PL, Brown SH, Husser CS, Bauer BA, Wahner-Roedler D, Rosenbloom ST, Speroff T. Evaluation of the content coverage of SNOMED CT: ability of SNOMED clinical terms to represent clinical problem lists. Mayo Clin Proc 2006;81[6]:741-8.

[5] Elkin PL, Bailey KR, Ogren PV, Bauer BA, Chute CG. A randomized double-blind controlled trial of automated term dissection. Proc AMIA Symp 1999:62-6.

[6] Elkin PL, Brown SH. Automated enhancement of description logic-defined terminologies to facilitate mapping to ICD9-CM. J Biomed Inform 2002;35(5-6):281-8.

[7] Elkin PL, Brown SH, Carter J, Bauer BA, Wahner-Roedler D, Bergstrom L, Pittelkow M, Rosse C. Guideline and quality indicators for development, purchase and use of controlled health vocabularies. Int J Med Inform 2002;68(1-3):175-86.

[8] Ioannidis JP, Salem D, Chew PW, Lau J. Accuracy of imaging technologies in the diagnosis of acute cardiac ischemia in the emergency department: a meta-analysis. Ann Emerg Med 2001;37[5]:471-7.

Address for Correspondence

Steven Brown MD, 2100 West End Avenue, Nashville TN.
steven.brown@va.gov

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Practical Issues in Using SNOMED CT as a Reference Terminology

Senthil K. Nachimuthu, MD, Lee Min Lau, MD, PhD

Department of Biomedical Informatics, University of Utah, Salt Lake City, Utah, USA
3M Health Information Systems, Salt Lake City, Utah, USA

Abstract

SNOMED CT® was created by the merger of SNOMED RT (Reference Terminology) and Read Codes Version 3 (also known as Clinical Terms Version 3). SNOMED CT is considered to be among the most extensive and comprehensive biomedical vocabularies available today. It is considered for use as the Reference Terminology of various institutions. We review the adequacy of SNOMED CT as a Reference Terminology and discuss the issues in its use as such. We discuss issues with content coverage of various clinical domains, data integrity and validity, and the update frequency of SNOMED CT, and why SNOMED CT alone is not adequate to serve as the Reference Terminology of a healthcare organization.

Keywords:

SNOMED CT, reference terminology, implementation issues.

Introduction

SNOMED CT® (Systemized Nomenclature of Medicine – Clinical Terms) is one of the most extensive and comprehensive biomedical vocabularies currently available, with 370,702 concepts (US Edition, core distribution, July 2006). It is recommended for use as the provider of data dictionary for several electronic health information systems. The SNOMED CT Technical Reference Guide January 2006 release mentioned that "SNOMED CT is the most comprehensive clinical reference terminology available in the world."[1] The July 2006 release of this document redefines it as "SNOMED CT is a comprehensive clinical terminology that provides clinical content and expressivity for clinical documentation and reporting."[2]

We examine the issues concerning use of SNOMED CT as the exclusive reference terminology for an electronic clinical information system. These issues include the coverage of concepts in various clinical domains included in the terminology, data consistency and the frequency of update. We examine these issues, and provide comparisons of several clinical domains of content present in SNOMED CT and another reference terminology. We use a set of objective criteria to evaluate SNOMED CT, and these criteria and methods may be used to evaluate any clinical vocabulary for its use in a clinical information system.

These three analyses have different objectives and assess different aspects of SNOMED CT. Hence, the materials and methods, results and discussion are presented individually for these three analyses for the sake of coherence.

1. Coverage of content

Materials and methods

All the active concepts in SNOMED CT (US Edition, July 2006 release) were included in the first experiment. A reference terminology created by 3M Health Information Systems, namely the 3M Healthcare Data Dictionary (3M HDD) was used to aid these comparisons.

The 3M HDD is a concept-based reference terminology that serves as the interlingua for multiple biomedical vocabularies which are mapped to it. The content included in the HDD is derived by importing various standardized terminologies such as SNOMED CT, LOINC, ICD-9-CM, CPT, HCPCS, First Databank NDDF, etc. The HDD also includes various legacy (homegrown) vocabularies such as the US Department of Defense lab, pharmacy and other clinical codesets and those from various other community and university hospitals.

Two subsets of the HDD were created for these comparisons – one including content provided by SNOMED CT alone, and another including content from the rest of the HDD excluding SNOMED CT. These two HDD subsets will here onwards be referred to as "*SNOMED CT ONLY dataset*" and "*NON SNOMED CT dataset*". The reason that the rest of the HDD was used for comparison is to provide a real world equivalent of concepts present in reference terminologies of very large enterprises. Similar results may be expected by comparing the content coverage of SNOMED CT with that of the enterprise reference terminology of any large multi-facility multi-specialty hospital organization, or to a standard reference terminology such as the UMLS.

SNOMED CT includes various domains of clinical concepts. The content of SNOMED CT is organized into 19 top-level hierarchies shown in Table 1. Various subhierarchies of concepts are organized under these nineteen top level hierarchies. The fully specified name of each SNOMED CT concept includes a suffix which roughly corresponds to the clinical domain that concept belongs to.

The suffix helps to disambiguate between two concepts that might belong in different top level hierarchies, but have the same textual name. For example, the suffix can help to differentiate between a disease and a finding at a glance without having to query its location in the SNOMED CT hierarchy. Two such examples are Endometriosis (disorder) vs. Endometriosis (body structure) and Aspirin (substance) vs. Aspirin (product).

Table 1 - SNOMED CT top-level hierarchies

Body structure	Procedure
Clinical finding	Qualifier value
Environment or geographical location	Situation with explicit context
Event	Record artifact
Linkage concept	Social context
Observable entity	Special concept
Organism	Specimen
Substance	Staging and scales
Physical force	Pharmaceutical / biologic product
Physical object	

Table 2 provides a partial listing of suffixes. Each top-level hierarchy may contain concepts with one or more suffixes.

Table 2 - SNOMED CT concept suffixes (partial list)

body structure	organism
disorder	specimen
procedure	substance
event	occupation
finding	product

We chose some of the clinical domains denoted by the top-level hierarchies or concept suffixes listed in tables 1 and 2 to compare the extent of coverage. We compared the concepts present in *SNOMED CT ONLY dataset* with the *NON SNOMED CT dataset*. The following results reflect comparisons of several domains of content included in SNOMED CT with NON SNOMED CT content included in the real world clinical information systems of various hospitals, both home-grown and vendor-developed. Examples of concepts present in these datasets are given for some of the domains to better explain the results.

Results

a. Pharmaceutical ingredients

The SNOMED CT Substance top-level domain contains pharmaceutical ingredients under the 'Drug or medicament (substance)' subhierarachy. This is a clearly definable clinical domain without any ambiguity, and hence this domain was used for comparison. It is more constrained and universal, and it is included in the core SNOMED CT data. This is in contrast with pharmaceutical *products* which vary from one country to another. In case of the US, pharmaceutical products are included in the SNOMED CT US Drugs Extension, which is not available free of cost, rather than the core US Edition.

Comparison was done by querying the specific domain of content in the NON SNOMED CT dataset and the SNOMED CT ONLY dataset. The results of comparison of various 'levels' of pharmaceutical ingredients are presented in Table 3. First Databank's National Drug Data File (NDDF) was one of the primary pharmacy content providers for the 3M HDD.

Comparisons were made at the level of Therapeutic class (Antibiotics, Antipyretics, Sedatives, etc.), Drug Class (Penicillins, Aminoglycosides, Calcium Channel Blockers, etc.), Ingredients (Diclofenac, Ampicillin, etc) and Ingredient-salt (Diclofenac Sodium, Diclofenac Potassium, Ampicillin Sodium, etc) levels. At all these levels, the NON SNOMED CT dataset had significantly better coverage for pharmaceutical ingredients than the SNOMED CT ONLY dataset. For example, 'Ketanserin' and 'Moxaverine' are some concepts present in NON SNOMED CT dataset but not in SNOMED CT ONLY dataset.

Table 3 - Pharmaceutical ingredients

DOMAIN	NON SNOMED CT	SNOMED CT ONLY
Therapeutic Class	134	22
Drug Class	873	118
Ingredient	7,265	2,647
Ingredient Salt	3,167	1,135

b. Microorganisms

Microorganisms constitute another domain of content that can be clearly defined, and hence the coverage of this domain was compared between SNOMED CT ONLY dataset and NON SNOMED CT dataset. We compared coverage of the entire microorganisms domain as well as some subdomains.

The following results in Table 4 show that the SNOMED CT ONLY dataset had about 80% to 90% content coverage compared as that of the microorganisms domain of the NON SNOMED CT dataset. For example, 'Heterosporium

species' is a concept present in NON SNOMED CT dataset but not in SNOMED CT ONLY dataset.

Table 4 - Microorganisms

DOMAIN	NON SNOMED CT	SNOMED CT ONLY
Microorganism	13,766	11,516
Bacterium	8,080	7,393
Fungus	1,545	1,321
Parasite	3,128	2,899
Virus	1,959	1,835

c. Problems

The clinical problems domain is among those that are not standardized and are highly variable. This domain will undergo much refinement over time. However, this domain is often used for picklists or dictionaries to manage electronic problem lists. This clinical domain is becoming increasingly important as clinical information systems try to implement problem-oriented documentation and maintain a longitudinal or historical record of patients' problems. Several authors have found that SNOMED CT has around 80% to 90% coverage of problems included in medical problem list domains of individual enterprise reference terminologies.[3][4][5]

We compared the problems domain of SNOMED CT with that of our reference terminology, which is the union of multiple enterprise reference terminologies, and our results indicate approximately 75% coverage. This domain is highly subjective and variable; however, the results still show very significant differences. The high variability of the domain is reflected in the concepts. For example, the NON SNOMED CT dataset contains the concept 'Intercostal bulging' which is not present in the SNOMED CT ONLY dataset.

Table 5 - Problems

DOMAIN	NON SNOMED CT	SNOMED CT ONLY
Problem	48,971	36,633

d. Specimens

Clinical specimens are an important part of medical records, and are an essential component of most biochemistry, pathology, microbiology and immunology lab orders. The following table shows the numbers of clinical specimens present in the NON SNOMED CT dataset and the number of specimens in SNOMED CT ONLY dataset. SNOMED CT ONLY dataset contained a significantly larger number of specimens compared to the NON SNOMED CT dataset, which is not surprising considering that SNOMED was created by the College of American Pathologists. For example, the concepts 'Liver cyst fluid' is present in NON SNOMED CT dataset but not in the SNOMED CT ONLY dataset. On the contrary, the concept 'BCG site swab' is present in SNOMED CT ONLY dataset but not in the NON SNOMED CT dataset.

Table 6 - Specimens

DOMAIN	NON SNOMED CT	SNOMED CT ONLY
Clinical specimen	823	1,048

e. Body Structure

The body structure domain contains concepts denoting various normal and abnormal anatomical structures at the organ, organ system or tissue level. The results in Table 7 show that the NON SNOMED CT dataset has significantly higher coverage compared to the SNOMED CT ONLY dataset.

Table 7 - Body structure

DOMAIN	NON SNOMED CT	SNOMED CT ONLY
Body Structure	34,257	31,999

f. Procedures

The procedures domain includes medical and surgical procedures performed for diagnostic or therapeutic procedures, administrative procedures (admission, discharge, transfer, billing, etc.), and so on. Procedures form an important part of several orders, clinical or administrative. The following results in Table 8 show that the NON SNOMED CT dataset contained significantly higher number of procedures compared to SNOMED CT ONLY dataset.

Table 8 - Procedures

DOMAIN	NON SNOMED CT	SNOMED CT ONLY
Procedure	59,311	53,052

Discussion

The results above show that SNOMED CT contains an extensive coverage of concepts but still lags behind a reference terminology which serves as the superset of several standardized and legacy vocabularies. Part of this difference is due to the differences in granularity. But a reference terminology will need to provide concepts with

varying granularities, and may not expect the users to perform the composition-decomposition themselves. Lack of concepts with differing levels of granularity causes problems with internal data storage as well as external data exchange. Some of the differences are due to the absence of several concepts in SNOMED CT which are present in other standard or local vocabularies. These results show that continuous authoring and updating of concepts will be required to keep SNOMED CT comprehensive and up to date for its use as a reference terminology.

Studies by other authors have also shown that SNOMED CT has anywhere between 30% to 90% coverage for various clinical domains.[6][7][8] These studies have each compared a single domain of content between SNOMED CT and a single enterprise reference terminology. Comparing multiple domains of content between SNOMED CT and the superset of multiple reference terminologies also yields similar results as shown above by this experiment.

Furthermore, we have not discussed several domains of content which are not covered by SNOMED CT, such as billing and reimbursement vocabularies, lab test names (provided by LOINC, etc. This denotes that several vocabularies need to be used together to complement each other and to build the enterprise reference terminology of a healthcare organization.

2. Data consistency

Materials and methods

The SNOMED CT 'core' consists of three tables – concepts, descriptions and relationships. We used the core tables provided by SNOMED CT and the US English dialect subset (both obtained from the US Edition, July 2006 release) to assess data consistency in SNOMED CT. We examined the compliance of the descriptions table to the constraints defined in SNOMED CT Technical Reference guide and User guide.

Table 9 - SNOMED CT representations

Representation	Type
Fully Specified Name	Scar (morphologic abnormality)
Preferred Term	Scar
Synonym	Scar tissue
Synonym	Fibrous scar
Synonym	Cicatrix

These constraints guide the importing of SNOMED CT into the reference terminology of an institution, use of SNOMED CT as the exclusive reference terminology of an institution, creating subsets, applying extensions to the core content, versioning and updating, and so on. The subset mechanism is especially important as SNOMED CT is too large to be used directly for a clinical application. SNOMED CT Subsets are often created for specific clinical domains or applications.

According to the SNOMED CT User Guide, each concept has one fully specified name, one preferred term and zero or more synonyms.[9] The fully specified name gives the definition of a concept, and is more explanatory than its preferred term. Preferred term is a more concise name which is used in clinical records. Synonyms are concise forms as well, and may be used as alternatives for the preferred term. An example is given in Table 9 – all these representations denote a single concept. We encountered some problems and inconsistencies while importing the SNOMED CT representations (textual names or surface forms of concepts) into the 3M HDD. We present the results below.

Results

The US Edition of SNOMED CT contains both American and British English representations for several concepts. The representations table in the core distribution denotes both the US (en-US) and UK (en-GB) English representations with an "unspecified type" for several concepts and a preferred term is not defined in these cases. The SNOMED CT US Dialect Subset redefines one of these representations as the Preferred Term and the rest as synonyms (often including the en-GB representation) for US English.[10] We first imported the core representations table into a relational database, and then applied the US Dialect Subset to it.

After applying the US dialect subset to the core descriptions table for the July 2006 US Edition, we found that 10,852 concepts had multiple preferred terms, often with different statuses (active, retired, limited, etc). The correct preferred term needs to be selected based on the concept status in these cases. Not all of these 10,852 concepts' preferred terms could be automatically selected based on the concept status and/or description status alone. 335 concepts had multiple preferred terms with the same status. E.g. ConceptId *103497003*, "*Streptococcus penumoniae 3 (organism)*" [sic] has two preferred terms in a "current" status.

Several of these occurrences are in direct violations of data integrity constraints defined in the SNOMED CT User Guide. These caused issues in automated importing SNOMED CT data into our reference terminology. This was overcome by defining one of the preferred terms or synonyms as the preferred term for some concepts by one of the authors. This task was labor intensive, and required several hours of the domain expert's time.

Discussion

The above results show that SNOMED CT needs stricter quality analysis before it is released to the end users. Failure of data consistency and violation of such constraints will lead to problems in importing and using the terminology by the target organizations. More studies of data consistency and integrity need to be done which involve other components of SNOMED CT. However, the authors

wish to mention that SNOMED CT is one of the well designed terminologies, and has less data consistency issues than many other standard vocabularies.

3. Frequency of update

Updates to SNOMED CT content are done through the regularly scheduled versioning mechanism. A new version is published once every six months, in January and July of every year. SNOMED has a web-based request mechanism for users and contributors to request changes. These changes include new concepts, representations, subsets, or extension namespaces through which third parties can extend the content in SNOMED CT. The users can also request modifications to existing content. However, these changes are incorporated and published once every six months through the regular release mechanism. Each new version of SNOMED CT has introduced anywhere between 2,500 and 8,000 new concepts. Statistics of changes in various core SNOMED CT tables over the years have been studied by other authors.[11]

Most organizations update their data dictionaries on a continuous basis, and some enterprise reference terminologies are updated as often as every day. This makes it harder to use SNOMED CT as the exclusive reference terminology for an organization. This can be overcome by creating local extensions to SNOMED CT via the SNOMED CT extension mechanism. The extension may be reconciled with the subsequent SNOMED CT version by inactivating the concepts in the extension and superseding them with equivalent concepts that are added to SNOMED CT. This process needs specialized expertise in vocabulary and ontology creation and maintenance, and requires contribution from subject matter experts in several content domains. An alternative is to use an enterprise reference terminology which includes SNOMED CT as one of the vocabulary sources along with other vocabularies.

Conclusion

The above analyses show that SNOMED CT does not have the necessary coverage to be the exclusive content provider for the enterprise reference terminology of an organization. SNOMED CT has one of the most extensive coverage among biomedical vocabularies, and has a well defined ontology and semantic relationships. These features make SNOMED CT as an essential component of an enterprise reference terminology. However, it needs to be complemented with other standard vocabularies, and in most cases, 'homegrown' vocabularies to build a reference terminology. SNOMED CT also has some data inconsistency issues which need manual intervention for implementation. We expect that these issues will be addressed in future releases. SNOMED CT also has a slower update schedule than would be required for the maintenance of a production enterprise reference terminology.

Due to these reasons, we conclude that most healthcare organizations will need further work to implement or adapt SNOMED CT for their requirements and use it in combination with other biomedical vocabularies, and may not be able to use it as a plug-and-play vocabulary.

Acknowledgements

The authors thank Shaun C. Shakib, MPH for his valuable help with this study.

References

[1] SNOMED CT® Technical Reference Guide. January 2006 Release. Jan 2006: 22.
[2] SNOMED CT® Technical Reference Guide. July 2006 Release. Jul 2006: 22.
[3] Elkin PL, Brown SH, Husser CS, Bauer BA, Wahner-Roedler D, Rosenbloom ST, Speroff T. Evaluation of the content coverage of SNOMED CT: ability of SNOMED clinical terms to represent clinical problem lists. Mayo Clin Proc. 2006 Jun;81(6):741-8.
[4] Penz JF, Brown SH, Carter JS, Elkin PL, Nguyen VN, Sims SA, Lincoln MJ. Evaluation of SNOMED coverage of Veterans Health Administration terms. Medinfo. 2004;11(Pt 1):540-4.
[5] Wasserman H, Wang J. An applied evaluation of SNOMED CT as a clinical vocabulary for the computer-ized diagnosis and problem list. AMIA Annu Symp Proc. 2003;699-703.
[6] Brown SH, Elkin PL, Bauer BA, Wahner-Roedler D, Husser CS, Temesgen Z, Hardenbrook SP, Fielstein EM, Rosenbloom ST. SNOMED CT®: Utility for a General Medical Evaluation Template. Proc AMIA Symp 2006: 101-105.
[7] van der Kooij J, Goossen WT, Goossen-Baremans AT, de Jong-Fintelman M, van Beek L. Using SNOMED CT codes for coding information in electronic health records for stroke patients. Stud Health Technol Inform. 2006;124:815-23.
[8] Kim H, Harris MR, Savova G, Chute CG. Content cover-age of SNOMED-CT toward the ICU nursing flowsheets and the acuity indicators. Stud Health Technol Inform. 2006;122:722-6.
[9] SNOMED CT® User Guide. January 2006 Release. Section 3.2; Jan 06: 11.
[10] SNOMED CT® Technical Reference Guide. July 2006 Release. Jul 06: 83.
[11] Spackman KA. Rates of change in a large clinical terminology: three years experience with SNOMED Clinical Terms. AMIA Annu Symp Proc. 2005:714-8.

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Statistical Selector of the Best Multiple ICD-coding Method

Eiji Aramaki[a], Takeshi Imai [a], Masayuki Kajino [b], Kengo Miyo [a], Kazuhiko Ohe [a]

[a] *The University of Tokyo Hospital, Japan*
[b] *Japan Research Group for Medical Ontology, Japan*

Abstract

The International Classification of Diseases 10th version (ICD-10) is one of the standard and most important disease classifications. Since computerized ICD-10 coding systems have drawn a great deal of attention in the medical field, a great number of different coding systems have been proposed. Thist paper proposes a hybrid architecture of different coding systems. First, given an input disease name, three coding systems output codes with their confidence scores. A C4.5-based system selector then selects the best output by using both input statistics and the confidence score from each system. The experimental results demonstrated that the selector significantly boosts the overall performance (+3.4 points).

Keywords:

ICD-10, International Classification of Disease codes, coding, decision trees, natural language processing

Introduction

International Classification of Diseases 10th version (**ICD-10**) [1] was endorsed by the 43rd World Health Assembly in 1990 and came to be used by World Health Organization (WHO) member countries from 1994. ICD-10 has become the international standard diagnostic classification for all general epidemiological purposes and many health management purposes, such as the analysis of the general health circumstances of population groups and monitoring of the incidence and prevalence of diseases [2].

Therefore, many medical doctors or coders (who are the special staff for the ICD coding) have selected a suitable disease code for each patient. As pointed out in a previous study [3], the coding task is an important semantic issue because it requires a very high level of understanding of the meaning of the patient data.

In such a situation, in order to reduce the heavy burden of coding, a number of automatic coding systems/methods have been proposed. For example, systems that output a code from a discharge summary [4-6], a coding system for autopsy reports [7], a coding system for input disease names [8], coding frameworks [9] [10], and a machine learning coding for a diagnostic text [11], have been proposed. Although there are numerous system variations, the approaches are basically classified into two types.

One approach is to directly capture the relation between an input and an ICD category [4-7][9][10]. In the present study, we refer to such systems as **direct systems**.

However, direct systems often suffer from complex language phenomena. For example, given the term "*leg eczema*", which should be classified into an ICD-10 code **[L309]** "*inflammation of the skin*", in order to realize successful coding, a direct system requires knowledge such as:

(1) "*Leg*" has "*skin*",
(2) "*Eczema*" is-a-kind-of "*inflammation*".

In order to obtain such knowledge, most systems utilize an ontology/knowledge-base. However, it is difficult to fully cover such knowledge.

Another approach [8] [11] is to retrieve the most similar coding example from a database, and output its code as is. We herein refer to such systems as **example-based systems.** The example-based system easily realizes the above coding, if the system can find a coding example, as follows:

(Input) "*Leg eczema*",
(Coding-example) "*Foot eczema*"-> **[L309].**

However, such a good example is not always available because of two difficult problems: (1) capturing the similarity of a precise example is difficult, and (2) the number of examples is not sufficient for covering all ICD categories.

In summary, both approaches have different strengths and weaknesses. Therefore, the present paper proposes a hybrid architecture, which selects a suitable system for each input. First, given an input, three coding systems output codes with their confidence scores. Then, a selector based on a C4.5 decision-tree [12] decides the best output by using information from both the confidence score of each system and input statistics.

The experimental results demonstrated that the selector significantly boosts the overall performance.

It should be noted that previous studies use various input types/forms, such as a discharge summary [4-6], an autopsy report [7], and a disease name [8]. In this paper we use a disease name as a system input, since it is the most informative clue in the coding task.

Selected for best paper award.

Although the proposed system does not depend on language, we conducted experiments using disease names in Japanese.

Material

This section describes ICD10 [1]. The ICD10 has a tree structure, in which a leaf code consists of one alphabetical character <A-Z>, three digit numbers <1-9>[1] and its description. We call the description an **ICD-term**.

Figure 1 shows an example of an ICD10 code with its ICD term "*Cholera, unspecified*". The first alphabetical character <A> indicates a top level category, and the second <0> indicates the second level category. In the same manner, the third and fourth numbers indicate the third and fourth level categories.

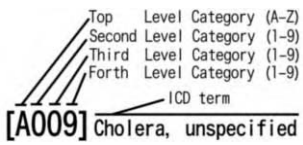


Figure 1 - Example of ICD-10 code and ICD-term

Methods

The task to be solved is to output a suitable ICD-code (**code**) for an input disease name (**input**). The idea of the proposed method is to dynamically select the best output from three systems. We first explain the three coding systems of the present study and then present the system selector.

Direct system

As mentioned above, most direct systems utilize knowledge-base/ontology. However, such resources are not always available in Japanese and other non-English languages. Therefore, the direct system of the present study utilizes ICD-terms, because ICD-terms are available as a knowledge-base in native languages in most countries.

The present direct system captures the similarity between an input and the ICD-terms. First, this system searches all of the ICD-terms, and extracts the ICD-term that is most similar to an input. For calculating the similarity, we used two measures: (1) a BM25 [13] and (2) an edit-distance [14].

BM25: The BM25 (*simBM25*) is a word-based similarity, which is employed in a state-of-the-art information retrieval system, the Okapi-system [15], which is defined as follows:

$$sim_{BM25} = \sum_{t \in T} (W_d \times W_q),$$

where

$$W_d = \frac{(k_1 + 1)tf}{k_1((1-b) + b \times dl/avdl)},$$

$$W_q = \log \frac{N - n + 0.5}{n + 0.5}.$$

In this formula, *T* is the set of words appearing in both the input and the ICD-term, *tf* is the number of occurrences of a word *t*, *dl* is the length of an ICD-term, *avdl* is the average length of the ICD term, *N* is the total number of ICD-terms, *n* is the number of extracted ICD-terms, and k_1 and *b* are the constants determined from the preliminary experiments[2]. For details, see [13].

Edit-distance: The edit-distance similarity (*simED*) is a character-based similarity, which is based on the minimum number of point mutations required to change an input into an ICD-term, where a point mutation involves: (1) changing a character, (2) inserting a character, or (3) deleting a character. For details, see [14].

The final similarity is a weighted sum of *simBM25* and *simED*, as follows[3]. We refer to this similarity as a **direct-sim**.

$$\textit{direct-sim(input, ICD-term)} = W_1\, simBM25 + W_2\, simED\,.$$

The final output code is decided by the code of the most similar ICD-term.

Example-based system

The example-based system retrieves the most similar coding example from a database (in experiments, we used 15,551 coding examples) and outputs the code of the retrieved example. In addition, in this similarity calculation, a weighted sum of BM25 and the edit-distance was employed. We refer to the similarity between an input and a coding-example as an **example-sim**.

The final output code is decided "as is" by the code of the most similar example.

Combination system

In preliminary experiments, the example-based system demonstrated high performance in rough classification, judging only the matching of the first and second digits. Considering this example-based advantage, we designed another system, referred to as the **combination system**.

The combination system first decides the first and second ICD-code using the example-based system. For example, if the example-based-system outputs **[L309]**, then **[L3**]** is the scope of the following searches. A detailed code (third and fourth digits) in the scope is then decided by a direct system.

These system frameworks are illustrated in Figure 2.

1 Some diseases have one more digit (one alphabetical character and four numbers). Since such codes are rare (7.6%), the present paper does not handle them.

2 We set $k_1 = 1.5$ and $b = 0.75$ in the experiments.

3 We set $W_1 = 0.9$ and $W_2 = 0.1$ in the experiments.

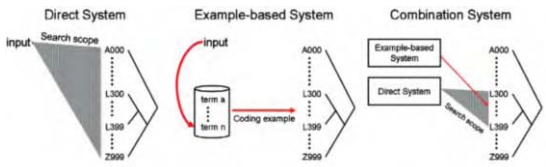


Figure 2 - Workflows of three systems

Statistical selector

The selector decides the suitable system for each input. Here, the selector uses two types of information (features): (f1) information from each system and (f2) information from an input. The detailed features are described in Table 1.

By using a training-set, the selector learns the relations between the features and which system is correct. In this learning, we utilized C4.5 [12], because it clearly shows the effectiveness of each feature.

Note that if the selector estimates that two or more systems are correct, we give the following priority: EXAMPLE-BASED > COMBINATION > DIRECT. For example, if the selector estimates that both COMBINATION and DIRECT are correct, then a COMBINATION output is employed.

In the same way, if the selector estimates that all systems are incorrect, an EXAMPLE-BASED output is employed. This priority is based on preliminary experimental results.

Table 1 - Features for system selector learning

(f1)	**Max Direct-sim:** The maximum direct-sim in the *k*-best ICD-terms*.
	Min Direct-sim: The minimum direct-sim in the *k*–best ICD terms.
	Example Code Variations: the number of second level category variations in extracted *k* coding-examples. For example, given [L309] [L300] and [B100], this value is 2 ([L3**] and [B1**]).
	Max Example-sim: The maximum direct-sim in the *k*–best coding-examples.
	Min Example-sim: The minimum direct-sim in the *k*–best coding-examples.
(f2)	**Input Length:** The number of the input words.
	Input Headword: The headword (last word) of an input. (We assume a head-final principle).
	Dictionary: Whether the input is found in the dictionary entry.

* In experiments, we set k=10.

Experiments

Experimental settings

To investigate the performance, we used two sets of materials: (1) a set of coding examples [16] consisting of 15,551 pairs (Japanese disease name and its corresponding ICD code), and (2) 995 disease names, which are input in several Japan

hospitals. We divided (2) into two sets as follows: a training-set (for selector training) and a test-set (for evaluation).

Both sets were annotated with their correct codes (gold standard). If the gold standard code is ambiguous for only the disease name, we removed it. The number of codes in each set is shown in Table 2.

To acquire the Japanese word units[4] in a term, we used a morphological analyzer developed in a previous study [17].

Table 2 - Number of coding examples

	# of codes
Coding-example	15,551
Training-set	495
Test-set	500

In the evaluations, we used three measures (1-3) and compared the following four systems (a-d), as follows:

1. **EXACT**: The exact (three digits) match.
2. **ROUGH**: Only the first and second digits match, e.g., **[L309]** and **[L303]**.
3. **10-BEST**: The system outputs 10 candidates, which are judged as to whether they contain the gold standard.

a) **DIRECT**: The direct system.

b) **EXAMPLE**: The example-based system.

c) **COMBINATION**: The combination system.

d) **PROPOSED**: The proposed system, selecting the best from the above three systems.

Results

The results are shown in Table 3. For two measures, PROPOSED showed the highest accuracy. In particular, in EXACT, the proposed system demonstrated significantly higher accuracy than the other systems in the McNemar test [18] ($p = 0.05$).

On the other hand, the ROUGH accuracy of the PROPOSED system is slightly (although not significantly) lower than EXAMPLE-BASED system. One reason for this is that the selector performed training based on the EXACT metrics.

Table 3 - Results

	(1) EXACT	(2) ROUGH	(3) 10BEST
a) DIRECT	33.6% (168)	52.8% (264)	61.2% (306)
b) EXAMPLE-BASED	63.6% (318)	**81.6% (408)**	65.0% (352)
c) COMBINATION	52.4% (262)	74.2% (371)	81.6% (408)
d) PROPOSED	**67.0% (335)**	80.4% (402)	**83.2% (416)**

* Numbers in bracket indicate the number of the correct outputs. The underlined value in (d) indicates significantly higher values than (a), (b) and (c).

Table 4 - Frequencies of appearance of each system in the proposed hybrid system (PROPOSED)

	Selected frequency
DIRECT	67
EXAMPLE-BASED	223
COMIBINATION	210

4 A word boundary in Japanese language is not explicit.

Table 5 - Output examples

Input disease name [SELECTED SYSTEM] coding example	Gold standard (a correct code & its ICD term)
HCV antibody test positive [EXAMPE-BASED] HBc antibody test positive [R768]	[R768] immunological injury of serum
E incompatibility [EXAMPLE-BASED] D incompatibility [O360]	[O360] maternal care for RH allogeneic immunity
hemolytic streptococcus infection [COMBINATION] coliform bacillus infection [A498]	[A491] chain coccus infection
type-LL Hansen's disease [COMBINATION] Type-I Hansen's disease [A300]	[A305] leper mass type lepra
palpebra tumor mass [DIRECT] *	[H029] palpebra trouble
lips of mouth sore [DIRECT] *	[K130] lips of mouth disease

* In the case that DIRECT is selected, a coding example is not used.

Table 4 shows the frequency of each selected system, and Table 5 shows some output examples.

Discussion

We then investigated an important clue for the selector. In a decision tree produced by the selector, the first branch is "*if the max example-sim > X, use EXAMPLE-BASED*", and the second branch is "*if example code variation < X, use EXAMPLE-BASED*". Based on this observation, we can see that the selector provided several opportunities for the EXAMPLE-BASED system.

For more detailed discussions, we investigated each system performance using the following measures:

Precision = # of correct outputs / # of outputs,
Recall = # of correct outputs / # of test-set.

In this investigation, a system will reject an output if the similarity is less than various thresholds. We examined two systems: (1) the DIRECT system (which relies only on direct-sim) and (2) the EXAMPLE-BASED system (which relies only on example-sim) (Figure 3).

As shown in the figure, EXAMPLE-BASED generally showed higher performance, but in some parts (top-left in the figure) DIRECT is better than EXAMPLE-BASED. Considering the potential of EXAMPLE-BASED, the strategy of the selector, which gives the first chance to EXAMPLE-BASED, is reasonable.

Finally, this figure also shows that we can freely control the system precision and recall. This can be a strong advantage from a practical point of view.

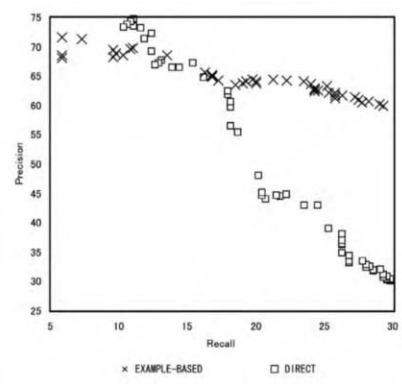


Figure 3 - Precision and recall of two systems (DIRECT and EXAMPLE-BASED)

Related studies

Thus far, a number of studies have investigated automatic coding [4-11, 19, 20, 22-25]. Most of these studies have focused on methods for knowledge representation such as a semantic frame-based ICD representation [9], a conceptual graph-based representation [5], a semantic markup [10], and an analytical index [19]. It is difficult to compare these systems to that of the present study for three reasons: (1) their inputs were different from ours, (2) we dealt with the Japanese language, and (3) more importantly, the present study focuses on the hybridization of systems. We can expect that the present system will have higher performance by incorporating the advantages of their methods.

Another important aspect of coding studies is the coding algorithm such as frequency statistics [20], a Naïve-bayes classifier [11], and an open registry algorithms [22]. Although various approaches have been proposed, each system has strengths and limitations, as pointed out in a study [23], motivating the proposed hybrid approach.

Note that some studies focused on a limited domain. For example, Tagliabue[22] dealt with a cancer domain, and Schnitzer [23] concentrated on a child injury domain. Such a limited domain enables a more finely-tuned system to be developed, leading to higher performance. In the future, we intend to incorporate such domain-limited systems into the proposed hybrid framework.

At present, automatic coding accuracy systems have not yet been perfected. However, Hohnloser [4] reported that their coding system improves human coding quality. In addition, several systems have been investigated in hospitals for years ([24]: five years, [25]: one year). In the future, we intend to examine the performance of the proposed system in a real-world environment.

Conclusion

The present paper proposed a hybrid architecture of different coding systems. First, given an input disease name, three coding systems output codes with their scores. A C4.5-based selector then decides the best output by using information from both the input and each system. The experimental results demonstrated that the selector significantly boosts the performance. In the future, we will attempt to incorporate more approaches/systems into our hybrid framework and will perform a system test in a real environment.

Acknowledgments

This research was supported in part by a Grant-in-Aid for Scientific Research through the Japan Society for the Promotion of Science (Project Number: 16200039, F.Y.2004-2007 and 18700133, F.Y.2006-2007) and by the Research Collaboration Project with Japan Anatomy Laboratory Co. Ltd.

References

[1] World Health Organization (WHO). International Statistical Classification of Diseases and Related Health Problems. 1992. http://www.who.int/classifications/icd/en/

[2] Alexander S, Conner T, Slaughter T. Overview of inpatient coding. Am J Health Syst Pharm. 2003; 60(21 Suppl 6):S11-4.

[3] Ceusters W, Lovis C, Rector A, Baud R. Natural language processing tools for the computerised patient record: present and future. In P. Waegemann (ed.) Toward an Electronic Health Record Europe, Proceedings, 1996;:294-300.

[4] Hohnloser JH, Puerner F, Soltanian H. Improving coded data entry by an electronic patient record system. Methods Inf Med. 1996; 35(2):108-11.

[5] Delamarre D, Burgun A, Seka LP, Le Beux P. Automated coding of patient discharge summaries using conceptual graphs. Methods Inf Med. 1995; 4 (34): 345-351.

[6] Rossi CR, Alberti V, Mancino G, Flor L, Martello T, Poeta L, Lise M, Diana L, Puppini G. Comparison between Manual and Automatic Coding of Medical Record Statistical Cards at a University hospital. Med Inform (Lond). 1993; 18(1):53-9.

[7] Riddick L, Long WB, Copes WS, Dove DM, Sacco WJ. Automated Coding of Injuries from Autopsy Reports. Am J Forensic Med Pathol. 1998; 19(3):269-74.

[8] Eiji Aramaki, Imai Takeshi, Kajino Masayuki, Miyo Kengo, Ohe Kazuhiko. Automatic ICD-coding for Long Disease Names based on Information Retrieval Methods (in Japanese). Joint Conference on Medical Info. 2006: 212.

[9] Fabry P, Baud R, Ruch P, Le Beux P, Lovis C. A frame-based representation of ICD-10. Stud Health Technol Inform. 2003;95: 433-438.

[10] Simon Hoelzer, Ralf K Schweiger, Joachim Dudeck. Transparent ICD and DRG Coding Using Information Technology: Linking and Associating. J Am Med Inform Assoc.2003; 10: 463-469.

[11] Serguei V S Pakhomov, James D Buntrock, Christopher G. Automating the Assignment of Diagnosis Codes to Patient Encounters Using Example-based. J Am Med Inform Assoc.2006; 13: 516-525.

[12] Quinlan JR. Simplifying decision trees. International Journal of Man-Machine Studies. 1987; 27 (1): 221-234.

[13] Robertson SE, Walker S, Jones S, Hancock-Beaulieu MM, Gatford M. Okapi at TREC-3. Proceedings of the 3rd Text Retrieval Conference. 1995: 109-126.

[14] Levenshtein V I. Binary Codes Capable of Correcting Deletions, Insertions and Reversals. Doklady Akademii Nauk SSSR. 1965; 163 (4): 845-848.

[15] Robertson SE, Sparck-Jones K. Relevance Weighting of Search Terms. Journal of the American Society for Information Science. 1976; 27 (1): 129-146.

[16] Hatano K, Ohe K. Information Retrieval System for Japanese Standard Disease-Code Master Using XML Web Service. American Medical Informatics Association (AMIA) Symposium 2003: 597-602.

[17] Sadao Kurohashi, Toshihisa Nakamura, Yuji Matsumoto, Makoto Nagao. Improvements of Japanese Morphological Analyzer JUMAN. Proceedings of the International Workshop on Sharable Natural Language Resources. 1994: 22-28.

[18] Gillick L, Cox SJ. Some Statistical Issues in the Comparison of Speech Recognition Algorithms. Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing. 1989: 532-535.

[19] Christine Bouchet and Olivier Bodenreider and Francois Kohler. Integration of the analytical and alphabetical ICD10 in a coding help system. Proposal of a theoretical model for the ICD representation. Medinfo. 1998; 9 (1): 176-179.

[20] Deimel D, Hesselschwerdt HJ, Lusznat A. Computer-assisted classification of ICD-9/10 and IKPM in compliance with the new federal health care regulation-practice-oriented solution for trauma surgery and orthopedics. Unfallchirurg. 1995; 98(10):545-50.

[21] McCallum A, Nigam K. A Comparison of Event Models for Naive Bayes Text Classification. AAAI/ICML Workshop on Learning for Text Categorization. 1998:41-48.

[22] Tagliabue G, Maghini A, Fabiano S, Tittarelli A, Frassoldi E, Costa E, Nobile S, Codazzi T, Crosignani P, Tessandori R, Contiero P. Consistency and accuracy of diagnostic cancer codes generated by automated registration: comparison with manual registration. Popul Health Metr. 2006; 4: 10.

[23] Schnitzer PG, Ewigman BG. Child Injury Deaths: Comparing Prevention Information from Two Coding Systems. J Pediatr Psychol. 2005; 30(5):413-23. Epub 2005 Feb 23.

[24] Hess T, Deimel D, Fischer R, Duchow J. Computer-based organization and documentation in orthopedics. A 5-year experience. Orthopade. 1999;28(3):212-7.

[25] Deimel D, Hesselschwerdt HJ, Heisel J. Simple and valid ICD9-/10- and IKPM coding using an electronic data-assisted coding system "do it"-experiences after one years use. Z Orthop Ihre Grenzgeb. 1997; 135(6):528-34.

Address for correspondence

Eiji Aramaki, The University of Tokyo Hospital;
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan
+81-3-5802-7115;aramaki@hcc.h.u-tokyo.ac.jp

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Creation and Evaluation of a Terminology Server for the Interactive Coding of Discharge Summaries

Hernán Navas, Alejandro Lopez Osornio, Analía Baum, Adrian Gomez, Daniel Luna, Fernan Gonzalez Bernaldo de Quiros

Department of Medical Informatics, Hospital Italiano de Buenos Aires, Argentina

Abstract

Free text entry versus structured data has been proposed as models in data entry in health information systems. A new user interface was developed with the objective of improving data capture. It also implemented a modification of the discharge summary data entry user interface that allowed the selection of already coded terms from a local terminology in the context of an inpatient electronic medical record. This software interacts online with a terminology server to provide feedback on data entry to clinical users in order to automatically code data. To evaluate the impact of this new software, we measured user satisfaction and the impact on autocodification rate. The new system had good acceptance from the users who ranked it high using QUIS (Questionnaire for User Interaction Satisfaction) and the auto codification rate improved from 61.5% to 88.39%.

Keywords:

automatic data processing, user-computer interface, computerized medical record

Introduction

For a long time there has been a discussion regarding the use of free text versus structured text for data entry that later must be codified [1]. Free text has the advantage of allowing physicians to express themselves freely, but the disadvantage is the need for a laborious codification process to allow data analysis. Structured text allows a quick codification process but has the disadvantage of being time-consuming for the physician and constrains expression to the level of detail of the selected entry terminology.

Since 2000, physicians at the Hospital Italiano of Buenos Aires have used an inpatient electronic medical record for creating the discharge summary using free text. The discharge summary is a structured abstract of the hospitalization episode where data are registered for caring and management purposes [2].

This paper describes our experience when introducing a more restrictive user interface that will require users to select terms from an existing list. The new system should have functions that can facilitate migration from the previous unconstrained text entry model and author kit is designed to assist in preparing the submission. It is an exact representation of the format expected by the editor.

Objectives

The objectives are to:

- Develop and implement a modification of the discharge summary data entry user interface that allows the selection of already coded terms from a local terminology
- Measure the usability of the new software prior to its implementation will be measured
- Assess the impact automatic codification rate after implementation.

Methods

Information contained in discharge summary is structured in several domains. This structure has the purpose of collecting all the necessary information to group episodes using diagnostic related groups (DRG). The different domains are shown in Table 1.

In each of these fields, the physician entered free text descriptions. The previous version of the discharge summary software tried to automatically code the entered text using the terminology server. If the term did not match an existing entry in the local terminology, it was addressed to the terminology team for secondary manual codification. The terminology team reviewed all the discharge summaries, assigned ICD-9CM codes and manually grouped them into a DRG [3].

The Terminology Server (TS) [4] is software composed of a local interface vocabulary (thesaurus) mapped to a reference vocabulary, SNOMED CT. The thesaurus is a list of terms created from almost 2 million free text inputs extracted from the clinical data repository. The terms included in the thesaurus are divided into concepts (real clinical entities) and descriptions (different ways of naming these clinical entities). The TS also has capabilities to reject invalid terms already flagged as not appropriate for the intended use.

The TS was used for several months in the background without interacting with the users, coding exact matches and ignoring invalid and new terms [5].

Table 1 - Description of discharge summary fields

Domain	Description
Principal diagnosis	Syndrome or disease that best explains the patient's condition on admission to hospital
Other diagnosis	Diagnosis made during hospitalization not related to the principal diagnosis or not caused by it
Clinical history	Diseases suffered by the patient prior to, and not related to, the current hospitalization
Surgical history	Procedures experienced by the patient prior to, and not related to, the current hospitalization
Co-morbidities	Preexisting diseases, currently active, that generate actions by the physician
Complications	Conditions that appear during the episode that are related to the principal diagnosis or procedures
Principal procedure	The main invasive procedure undergone by the patient, related to the principal diagnosis
Other procedures	Procedures that do not meet criteria to be considered a principal procedure

New requirements

It was established that changes in the discharge summary interface would enforce the selection of terms already included in the TS by interacting with the user during data input. The TS should also provide interactive information for refining concepts. This feature of the TS is achieved using semantic information included on SNOMED CT, navigating the sub-types/super-types hierarchy.

The new user interface should:

- Provide automatic coding for known terms
- Offer more detailed instances of a concept for refinement
- Suggest similar terms as alternatives for unknown terms.

Interactive model

Every time the physician enters a term, two situations can take place:

- Automatic codification: when there is an exact match, the entry term is automatically codified using the code of the institutional vocabulary term (e.g. measles)
- Non-autocodifiable text: refers to the situation where the physician enters text that is unknown, invalid or can be refined. It also takes into account spelling mistakes and synonyms, all contained in our server. The three different variants are outlined below.

Unknown text

If the physician enters a term that is not included in the institutional thesaurus, the system offers the user a list of possible valid terms so can they can select one of them. If the physician considers none of these terms exactly reflects the intended meaning, it is possible to choose to save the original text for manual revision.

Invalid text

If the physician enters a term that is manually marked as 'non-valid for this domain', the system alerts the user and offers, in some cases, an alternative valid term but does not allow the invalid term to be saved. If the system doesn't offer valid alternatives, the physician must enter a new term.

Refinable text

The text is related to a valid concept but there are more specific concepts available. The system displays more detailed options so the physician can choose one and make a more specific entry (e.g. arterial hypertension). The text is valid but the system will offer more specific options such as primary arterial hypertension or secondary arterial hypertension.

Impact of the new system

The impact of the automatic codification process was easily measured through the ST records using rates of automatic coded terms of May 2006 and June 2006, prior and after the software implementation respectively.

Evaluation of usability

We decided to evaluate the usability of the software prior to its implementation. We choose 5 physicians from different specialties to represent the users of the discharge summary software. A resident physician of the Medical Informatics service observed while these physicians used the software in their usual workplace. After observation, the physicians were asked to complete the QUIS (Questionnaire for User Interaction Satisfaction) Spanish version 7.0. QUIS was developed and validated by the Human-Computer Interaction Laboratory of Maryland University [7]. We used the brief version of the questionnaire due to the limited time available with the physicians. QUIS uses the Likert scale from 1 to 9. It is divided into 12 parts, 5 of which were not applicable in our setting. Part 1 evaluates the experience with the system; part 2 the experience in general when using computers; part 3 examines the general impression of the user; part 4, the screen design; part 5, the terminology used; part 6 makes reference to the learning capacity; and part 7 evaluates system qualities such as velocity, consistence and facility to use it.

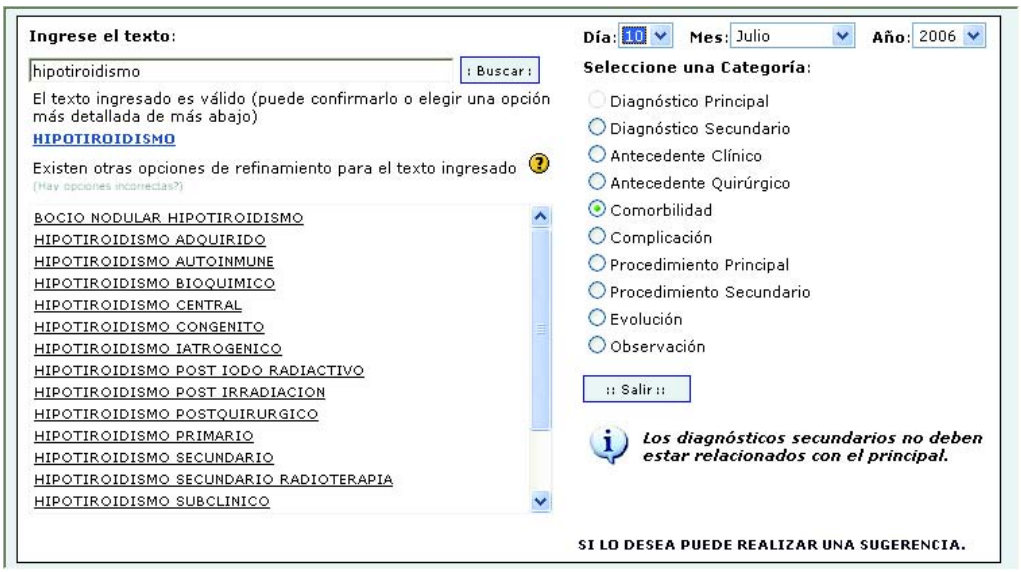


Figure 1 - New interactive coding application

Implementation

The software was successfully implemented in June 2006. The software was programmed in Java, as are all our electronic medical records. The access to the terminology services was made through store procedures programmed in PL-SQL in an Oracle 10g base. The user interface screen is showed in Figure 1.

Results

Using the previous discharge summary software, only 12,830 (61.5%) terms were automatically coded from a total of 20,862 terms in one month. During the month after the implementation of the new user interface, 16,826 (88.39%) were automatically coded from a total of 19,036 terms, as shown in Figure 2.

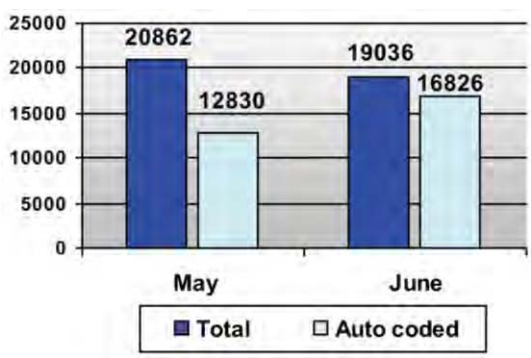


Figure 2 - Automatic code rate pre and post implementation

In spite of the modifications made to the interface, resident physicians were completely satisfied and none needed help while using it. QUIS results are shown in Figure 3.

Part 1: Experience with the system. Users had at least 1 or 2 years using the system, between 1 and 4 hours weekly.

Part 2: Prior experience. All users had prior experience using Windows OS ® and were users of previous clinical applications including discharge summary, order entry and results manager.

Part 3: Users general impression. Users rated the discharge summary screen as good, stimulating, easy to use, adequate and flexible.

Part 4: Screen design. Users found screens to be easily readable, with a clear sequence. Highlighted elements in screen were of medium utility, as well as screen format.

Part 5: Terminology and information about the system. Users found the terminology to be consistent and appropriate. The on-screen messages were clear and consistent. The system usually informed about what it was doing. The error messages were helpful.

Part 6: Learning. The users think the system is easy to learn and tasks were done mostly without complications. Names and instructions were easy to remember.

Part 7: System capacity. Users found the system to have a good working speed. It was easy to correct mistakes. The usability was directly related with the user's level of experience.

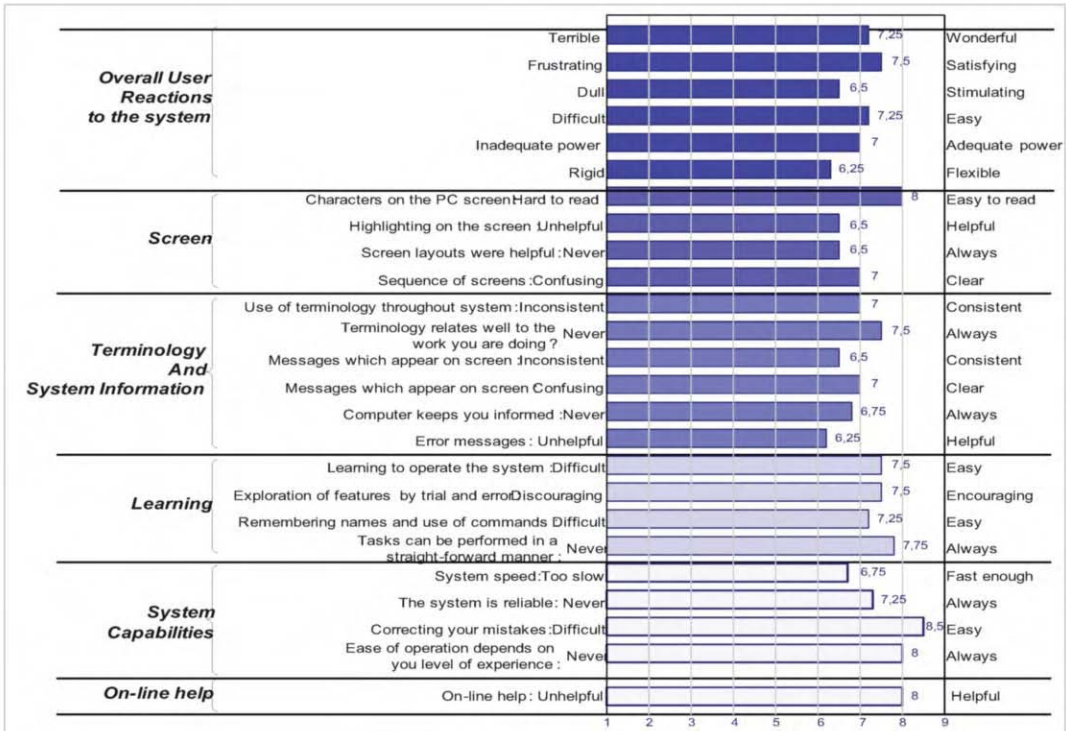


Figure 3 - QUIS results

Discussion

Health care organizations currently pay attention to systems usability [8, 9] and this can be carried out with few resources. Systems usability can be evaluated by a trained person with a notepad and five users who are representative of the users' population [10]. We decided to use QUIS as a validation instrument because we wanted to ensure before implementing the new system that medical residents would not have any problems using the new interface, and that it would not interfere with their complex daily tasks.

We think this improvement relies greatly on the important amount of text (synonyms, abbreviations and local jargon) available in the terminology server, and the relationships established between the different terms through the use of SNOMED CT.

The availability of online consultation about the terminology and input terms created acceptance among users, and led us to maximize the benefits of free and structured texts.

After going through the process of manual revision of new terms, a new version of the user interface will further restrict the data entry process, and disallow the clinical use of unknown terms. A mechanism to propose new terms will be maintained, but new terms will not be assigned to a patient until they are approved by the terminology team.

References

1. Middleton B, Renner K, Leavitt M. Ambulatory practice clinical information management: problems and prospects. Healthc Inf Manage 1997; 11(4):97-112.
2. Gonzalez Bernaldo de Quiros F, Soriano E, Luna D, Gomez A, Martinez M, Schpilberg M, Lopez Osornio A. Desarrollo e implementación de una Historia Clínica Electrónica de Internación en un Hospital de alta complejidad. In: 6to Simposio de Informática en Salud - 32 JAIIO; 2003; Buenos Aires, Argentina: Sociedad Argentina de Informática e Investigación Operativa (SADIO).
3. Lopez Osornio A, Luna D, Bernaldo de Quiros FG. Creación de un sistema para la codificación automática de una lista de problemas. In: 5to Simposio de Informática en Salud - 31 JAIIO; 2002; Santa Fe, Argentina: Sociedad Argentina de Informática e Investigación Operativa (SADIO).
4. Lopez Osornio A, Gambarte ML, Otero C, Gomez A, Martinez M, Soriano E, Luna D, Otero P, Clusella MM, Mitre MG, Reynoso G, Gonzalez Bernaldo de Quiros F. Desarrollo de un servidor de terminología clínico. In: 8mo Simposio de Informática en Salud - 34 JAIIO; 2005; Santa Fé, Argentina: Sociedad Argentina de Informática e Investigación Operativa (SADIO).
5. Luna D, Lopez Osornio A, Plazotta F, De los Ríos E, Montenegro S, Martinez M, Paris M, Bernaldo de Quiros FG. Creation of an Web Based DRG Grouper. Technol Health Care 2004; 12(5):375-6.
6. Wang SJ, Bates DW, Chueh HC, Karson AS, Maviglia SM, Greim JA, Frost JP, Kuperman GJ. Automated coded

ambulatory problem lists: evaluation of a vocabulary and a data entry tool. Int J Med Inform 2003; 72(1-3):17-28.
7. University of Maryland (UMD). QUIS: The Questionnaire for User Interaction Satisfaction. Available at: http://www.cs.umd.edu/hcil/quis Last accessed: Jul 2006.
8. Sittig DF, Kuperman GJ, Fiskio J. Evaluating physician satisfaction regarding user interactions with an electronic medical record system. Proc AMIA Symp 1999:400-4.
9. Hortman PA, Thompson CB. Evaluation of user interface satisfaction of a clinical outcomes database. Comput Inform Nurs 2005; 23(6):301-7.
10. Nielsen J. Why you only need to test with 5 users. Available at: http://www.useit.com/alertbox/20000319.html Last accessed: Jul 2006.

Address for correspondence

Hernan Navas, MD
Department of Medical Informatics
Hospital Italiano de Buenos Aires
Gascón 450
(C1181ACH) Buenos Aires – Argentina
e-mail: hernan.navas@hospitalitaliano.org.ar

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Thesaurus Anomaly Detection by User Action Monitoring

Jeferson L. Bitencourt[a], Píndaro S. Cancian[a], Edson J. Pacheco[a], Percy Nohama[a], Stefan Schulz[b]

[a]*Pontifical Catholic University of Paraná (PUCPR), Curitiba, Brazil*
[b]*University Hospital, Freiburg, Germany*

Abstract

The construction and maintenance of a medical thesaurus is a non-trivial task, due to the inherent complexity of a proper medical terminology. We present a methodology for transaction-based anomaly detection in the process of thesaurus maintenance. Our experiences are based on lexicographic work with the MorphoSaurus lexicons, which are the basis for a mono- and cross-lingual biomedical information retrieval system. Any "edit" or "delete" actions within these lexicons that undo an action defined earlier were defined as anomalous. We identify four types of such anomalies. We also analyzed to which extent the anomalous lexicon entries had been detected by an alternative, corpus-based approach.

Keywords:

thesaurus, quality control

Introduction

A thesaurus is a collection of semantically related terms, used to represent document contents with the purpose of improving document classification and content retrieval [1,2,3]. Especially in Medicine and Health, there is an ever-growing volume of texts, scientific publications, electronic textbooks, web-based health information for laypersons, and last but not least, a huge amount of text in electronic patient records.

Medical terminology is complex, rapidly changing, and highly productive. New acronyms, abbreviations, single and multi word compounds are continuously generated, and English terminology increasingly permeates non-English medical documents. Multilingualism is an important issue here, due to the global tendency of using English as the primary language of research, whilst the local idioms are used for patient-related documentation and communication [4]. All these factors hamper the use of simple text retrieval techniques such as popularized by Web search engines. So we claim that text retrieval in the biomedical domain should be supported by an underlying thesaurus that covers intra- and interlingual synonymy.

Thesaurus engineering is an iterative process, generally involving numerous domain experts. Their activities are generally directed by guidelines, but also encompass individual arbitrariness. Controversies especially come up with relation to boundary decisions such as:

- Whether to consider or not a term pertinent to the domain and therefore relevant for the thesaurus;
- Whether to include composed or derived terms if their components or base forms are already in the thesaurus;
- Whether to consider two terms as synonymous;
- Which other semantic relations to use between terms;
- Which senses to be needed to be distinguished when dealing with ambiguous terms;
- Whether to consider additional senses which are of marginal importance for the domain.

The most common method to track modifications in a database is the procedure register or log [5]. The log system has the property of recording all state changes. In this work maintenance problems in a medical thesaurus are analyzed, and a method is proposed which helps to detect anomalous change patterns in order to increase the efficiency of the thesaurus building process.

Material and methods

The subword approach

The application context of this work is given by *MorphoSaurus*, a large multilingual thesaurus for clinical medicine [6,7]. The main difference between MorphoSaurus and any other medical thesaurus such as MeSH [8] is that the inclusion of lexical entries into the repository is strictly guided by criteria of semantic atomicity. This resulted in the introduction of a new kind of lexical item, the so-called *subword* entry, assuming that neither fully inflected nor automatically stemmed words constitute the appropriate granularity level for lexicalized content descriptions [9,10].

The building of a subword thesaurus requires a scrutiny of the following aspects of language:

- Orthographic variations such as in *oesephagus* vs. *esophagus*;
- Derivations that modify a word's part of speech and/or its meaning, e.g.: *diabetes, diabetic, antidiabetic*;
- Proper string delimitation of what constitutes a valid subword entry, e.g. *nephr, nephro*, or both;

Original Document	Orthographic Normalization	Morphological Segmentation	Semantic Normalization
High TSH values suggest the diagnosis of primary hypothyroidism while a suppressed TSH level suggests hyperthyroidism.	high tsh values suggest the diagnosis of primary hypothyroidism while a suppressed tsh level suggests hyperthyroidism.	high tsh value s suggest the diagnos is of primar y hypo thyroid ism while a suppress ed tsh level suggest s hyper thyroid ism.	**#up# tsh #value#** #suggest# **#diagnost# #primar# #small# #thyre# #suppress# tsh #nivell#** #suggest# **#up# #thyre#** .
Erhöhte TSH-Werte erlauben die Diagnose einer primären Hypothyreose, ein supprimierter TSH-Spiegel spricht dagegen für eine Schilddrüsenüberfunktion.	erhoehte tsh-werte erlauben die diagnose einer primaeren hypothyreose, ein supprimierter tsh-spiegel spricht dagegen fuer eine schilddruesenueberfunktion.	er hoeh te tsh - wert e erlaub en die diagnos e einer primaer en hypo thyre ose, ein supprim iert er tsh - spiegel spricht dagegen fuer eine schilddrues en ueber funktion.	**#up# tsh - #value#** #permit# **#diagnost# #primar# #small# #thyre#** , **#suppress# tsh - {**#mirror# **#nivell#}** #speak# **#thyre# #up#** #function# .
A presença de valores elevados de TSH sugere o diagnóstico de hipotireoidismo primário, enquanto níveis suprimidos de TSH sugerem hipertireoidismo.	a presenca de valores elevados de tsh sugere o diagnostico de hipotireoidismo primario, enquanto niveis suprimidos de tsh sugerem hipertireoidismo.	a presenc a de valor es elevad os de tsh suger e o diagnost ico de hipo tireoid ismo primari o, enquanto niveis suprimid os de tsh suger em hiper tireoid ismo.	#actual# **#value# #up# tsh** #suggest# **#diagnost# #small# #thyre# #primar#** , **#nivell# #suppress# tsh** #suggest# **#up# #thyre#** .

Figure 1 – Morphosemantic indexing by MorphoSaurus

- Composition that builds complex word forms out of simple ones, such as *high blood pressure* (multi word composition) or *hyperprebetalipoproteinemia* (single word composition);
- Synonyms and translations: *nephro*, *renal*, *rim*, *kidney*, *niere*...;
- Acronyms: *AVC*, *ECG*, *DPOC*, *AIDS*,... and their possible expansions;
- Proper Names, such as substance names, trade names, or eponyms: *Diclofenac*, *Viagra*, *Parkinson,...*
- Ambiguous terms, i.e. terms which have different meanings, such as *head*, *ventricular,...*

The MorphoSaurus system

Particularly in the medical sublanguage we observe complex word forms such as in *`pseudo|hypo|para|thyroid|ism', 'append|ectomy'*, or in *'tooth|ache'*. The MorphoSaurus system does not register these complex terms, as long as their meaning can be derived from their components. Instead it focuses on so-called *subwords* as lexicon entries, which are registered together with attributes such as language and subword type (word stem, prefix, suffix, infix, invariant word). Instead of medical *terms*, MorphoSaurus includes the *building blocks* of medical terms.

In the MorphoSaurus database, each entry is assigned to one synonymy class. In the beginning of the thesaurus building process, there is one class for each entry, but in the following classes are increasingly fused, containing not only intralingual synonyms but also interlingual translations. Each synonym class is identified by one language-independent descriptor, called MorphoSaurus identifier (MID). Every MID represents a unique meaning. MIDs are further divided between those which are relevant for indexing (here marked by #) and those which have grammatical functions only (e.g. auxiliary verbs, inflection suffixes, conjunctions, etc., marked by %).

For instance, the MID *#liver={liver, leber, hepar, hepat, figad, ...}* represents the sense of "liver", whereas the MID *%ness={-ness,-idade ,-keit}* groups semantically irrelevant suffixes, just as *%be={be, is, are, sao, e, ist, ...}* groups forms of the verb "to be", which should be ignored for document indexing.

MorphoSaurus provides two semantic relations between equivalence classes:

- *has_sense*: This relation links ambiguous MIDs to their senses, e.g., the MID *#head* is linked to the MIDs *#caput* and *#boss* by *has_sense*;
- *has_word_part*: This relation links an MID having a composed meaning to each of its parts, e.g., the MID *#myalg={"myalg-", "mialg-",...}* to both *#muscle={myo-, muscle, mio, muscul, ...}* and *#pain={pain,-algy, -algi, -algia, dor, schmerz,... }*. The reason for this is to deal with composed meanings even in cases where a compound cannot be properly dissected.

The MorphoSaurus indexing engine converts text into into an interlingual representation by a three-step procedure, as

depicted in Fig. 1 for a parallel text. In the first step, The first step deals with general and language specific *orthographic normalization*. In a second step, the text is split into sequences of semantically plausible subwords, which are then checked for morphological plausibility using a finite-state automaton (*Morphological Segmentation*). Thus invalid segmentations such as segmentations without stems or ones beginning with a suffix are rejected. Finally, each meaning-bearing subword is replaced by its MID. This constitutes the interlingual output representation of the system. The bold MIDs in Fig. 1 are evidence of the high interlingual correlation on the level of semantic atomicity.

Pragmatics of thesaurus building in MorphoSaurus

Modeling decisions for a subword thesaurus are more complex than decisions for word-based thesauri. Thus, the proper delimitation of what constitutes a lexical entry is often subject to argument. These boundary decisions are even influenced by the performance of the word segmentation algorithm. The following example may provide a flavor of the daily problems in the maintenance of MorphoSaurus: The word *nephrotomy* used to be erroneously divided into *nephr|oto|my* instead of the correct *nephr|o|tomy.* This was solved by the inclusion of *nephro* as a synonym of *nephr* into the thesaurus, which may be considered linguistically unsound, but which produced the required result. Another controversial issue is how to deal with utterly short stems. For instance, the MID *muscle* can be erroneously extracted from many words due to the frequency of the character combination "my". The decision made in this case was not to include *my,* but to add the common compounds (*myosis, myalgy, ...*) instead and to use, in a second step, the *has_word_part* relations as introduced above in order to represent their meanings.

Anomaly detection in logged data

With the purpose to collect data about the changes in the MorphoSaurus database during a given time interval we collected 86 backups of the MorphoSaurus database for the period between July 14th, 2005 and March 30th, 2006, representing regular intervals of approximately three days. A script extracted all alterations for each single data object. This data volume then served as a basis for the detection of anomalies.

We here introduce the notion of thesaurus management anomalies as sequences of such actions done by the thesaurus curators that consume effort without any positive impact on the quality of the thesaurus.

We distinguish the following four anomaly types:

- A relationship anomaly is defined as a sequence of editing steps in which a semantic relation (*has_sense or has_word_part*) between two equivalence classes is first eliminated and later restored,
- A type anomaly is defined as a sequence of editing steps in which a lexicon entry is first moved from one equivalence class to another one and later happens to be moved back into the original class,
- A delimitation anomaly is defined as a sequence of editing steps in which the string delimitation of a lexicon entry is modified in a first step but is later restored to its original form,
- A permanence anomaly is defined as a sequence of editing steps in which an existing lexicon entry is first deleted, but recreated in a later phase.

According to these anomaly types, the whole body of logged user data from the backup data was analyzed.

Corpus-driven error detection

For further validation of anomalies, we used the following additional resources:

- Related corpora analysis: in a parallel study, the MorphoSaurus database was checked against real data in order do detect imbalances of MID distribution. In particular, this study used closely related corpora in different languages. All these texts were indexed by the MorphoSaurus system (yielding an output as shown in Fig.1), and the distribution of MIDs was pairwisely compared. The hypothesis was that anomalous MIDs exhibited the most unequal distribution patterns.
- Discussion Forum: a list of ill-distributed MIDs (ranked by degree of disproportion and overall frequency) was used as a basis for manual cleansing of the thesaurus by our team of lexicographers. Each of these MIDs was submitted to a thorough analysis, and the result was discussed in an online forum. Whenever the lexicographers reached a consensus, the MID entry was modified accordingly.

Joining the anomaly data from the log analysis with the data from the forum then provides a measure of relevance of the anomaly data.

Results

Analysis of anomalies

The log analysis yielded a total of 146 anomalies. Interestingly, there were many multiple occurrences in the case of relationship anomalies. So, we identified 23 MIDs

Analysis of problems addressed in the discussion

which exhibited a relationship anomaly more than once during the observation period. Counting the multiple occurrences only once, we got 99 anomalies, as shown in Table 1.

The discussion forum focused on English, Portuguese, and German and addressed 325 problems as shown in Table 2.

Comparison between anomalies and discussed problems

So far, the acquisition of anomaly data was completely independent from the analysis of problem discussions. Bringing together both sources we can now analyze the anomalous MIDs which were picked up in the discussion.

The data show that 36 out of 99 anomalous MIDs were not addressed in the forum discussions, and that forum discussions covered far more cases than could be identified by the log analysis, cf. values in parentheses in Table 1.

An inverse view is given by Table 2. Here the values in parentheses show the frequency of those MIDs which had also been spotted by the log analysis.

Finally, Table 3 gives a closer view of multiple occurring anomalies.

Table 1 – Number of occurrences of anomalies found by log analysis. Numbers in parentheses indicate anomalies also picked up in the discussion forum

Anomaly Type	Count
AR - Relationship anomaly	76 (28)
AT – Type anomaly	18 (18)
AD - Delimitation anomaly	0 (0)
AP - Permanence anomaly	5 (4)

Table 2 – Number of occurrences of problems in the discussion forum. Numbers in parentheses indicate the cases where the same problem was posted to the forum and had been identified by the log analysis independently

Problem Type	Count
PR - An expected relation between an ambiguous MID and MIDs (*has_sense* type) or an expected expansion relation (*has_word_part* type) was missing.	86 (24)
PC - Entries assigned to one MID did not cover all languages.	80 (6)
PU - The same sense is represented by two unrelated MIDs.	70 (8)
PM - Lexicon entries are assigned to one MID diverge in meaning.	11 (1)
PL - Language specific entry do not translate to other languages.	11
PO - Orthographic errors.	16 (1)
PI - Similar senses are represented by two unrelated MIDs, one of them of the type "excluded from indexing".	31
PD - Errors caused by incorrect subword delimitation	16
PS - Errors caused by incorrect functioning of the segmentation engine.	4

Table 3 – Anomalies AR: multiple changes related to one MID found by log analysis (left col.). Number of MIDs which exhibit multiple changes (right col.) Numbers in parentheses indicate the anomalies picked up in the discussion forum

# Changes	Count
2	4 (1)
4	16 (6)
5	2 (2)
6	2 (0)
7	23 (10)

Discussion

The correct handling of lexical ambiguities is the most error-prone step in the thesaurus management process [11]. This was evidenced not only by the frequency of the relationship anomaly (AR, cf. Table 1) but also by its occurrence in the discussions. Moreover, this anomaly was the only one where up to seven repetitions of editing occurred. This sheds light at a considerable waste of resources and lack of communication.

The assignment of subwords to an equivalence class (MID) was also subject to changes (AT anomaly), but here all cases were addressed in the discussion forum. This shows the effectiveness of the corpus-based detection of disproportions in the MID distribution and its good take-up by the lexicographers.

It was quite surprising that no anomaly of string delimitation (AD) could be observed. This is probably due to the fact that observed segmentation problems were solved by adding new string variations (e.g. "*-otomy*" in addition to "*-tomy*") and not by modifying the existing ones, in accordance with the guidelines used in the thesaurus building and maintenance process.

Finally, the awareness of the permanence anomaly was shown by its high coverage in the discussion forum. This kind of anomaly elicited the major disproportions in the corpus analysis, due to the fact that lexemes with a borderline semantic importance occur with high frequency. For instance, the phenomenon that the preposition "*from*" was assigned to the MID type "marked for indexing", and the Portuguese translation "*de*" was assigned to another MID marked as "excluded from indexing", put the first MID on top of the ranked list of lexicon disproportions, and was

therefore preferentially addressed in the discussions. These types of "*stopwords"* are frequent and equally distributed across the whole document space. Hence, they are irrelevant for distinguishing documents in text retrieval scenarios [3,12]. On the other hand they can acquire importance as context modifiers, such as in complex terms like *Removal of foreign body from stomach*, which could be matched to *Removal of stomach* in case that the preposition *from* is neglected.

The anomaly detection process would have an additional value if the anomalous MIDs would be detected in runtime so that the lexicographers would get an immediate feedback whenever an action executed earlier is undone. A new version of our lexicon editing tool MorphoEditWeb, currently under development, implements this functionality.

Conclusion

Thesaurus management is a highly dynamic process, and the kind of decisions which have to be done continuously imposes challenges on the lexicographers upon which the creation and continuing maintenance is incumbent. The waste of resources due to the phenomenon that one person undoes an action which another person has done before is considerable.

Unfortunately no guideline for thesaurus management can ever foresee all borderline cases which can only be solved by consensus. The proposed technique of user action based anomaly detection is useful to discover them and is complemented by other techniques such as the corpus-based check for disproportion of semantic identifiers. A seamless integration of such quality assessment routines in the thesaurus management tools is of utmost importance for achieving higher process effectiveness.

Acknowledgements

This work was supported by the CNPq, Brazil, (project number 550830/05-7), and by German-Brazilian cooperation project DLR/PUCPR (BRA 03/013).

References

[1] Hersh WR. Information Retrieval – A Health Care Perspective. New York: Springer; 1996.

[2] Ribeira-Neto B, Baeza-Yates R. Information Retrieval - Data Structures & Algorithms. 1st. ed. New Jersey, EUA: Prentice Hall.; 1992.

[3] Frakes WB, Baeza-Yates RA, Fox C. Lexical analysis and stoplists. Information Retrieval: Data Structures & Algorithms. Upper Saddle River, New Jersey: Prentice Hall PTR, 1992; pp. 102-130.

[4] Schulz S, Hahn U. Morpheme-based, cross-lingual indexing for medical document retrieval. in: International Journal of Medical Informatics. 2000; 58-59: 87-99.

[5] Bernstein P, Hadzilacos V, Goodman N. Concurrency Control and Recovery in Database Systems. Reading, Massachusetts: Addison-Wesley, 1987.

[6] Markó K, Schulz S, Hahn U: MorphoSaurus - Design and Evaluation of an Interlingua-based, Cross-language Document Retrieval Engine for the Medical Domain. Methods of Information in Medicine, 4/2005(44): 537-45.

[7] Mophosaurus. Freiburg, Germany: University Hospital, Medical Informatics, 2006. [cited 2006 Dec 4]. Available from http://www.morphosaurus.net.

[8] Medical Subject Headings. Bethesda, MD: National Library of Medicine, 2006. [cited 2006 Dec 4]. Available from http://www.nlm.nih.gov/mesh/MBrowser.html

[9] Markó K, Schulz S, Hahn U. Unsupervised Multi-lingual Word Sense Disambiguation via an Interlingua. Proceedings of the 20th National Conference on Artificial Intelligence (AAAI '05), Pittsburgh, Pennsylvania. 2005; pp. 1075-1080.

[10] Daumke P, Schulz S, Markó S. Subword approach for acquiring and cross-linking multilingual specialized lexicons. LREC 2006 Workshop: Acquiring and Representing Multilingual, Specialized Lexicons: the Case of Biomedicine. 2006; Genova, Italy.

[11] Soergel, Dagobert. Functions of a thesaurus / classification / ontological knowledge base. College of Library and Information Services. University of Maryland; 1997.

[12] Korfhage, RR. Information Retrieval and Storage. New York: John Wiley & Sons, 1997; p. 349

Correspondence Address

Jeferson Luiz Bitencourt - LER,
Pontifical Catholic University of Paraná.
Rua Imaculada Concei_ão, 1155 – CEP 80215-901,
Curitiba,
Brazil,
jefersonbitencourt@yahoo.com.br.

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Harmonizing Clinical Terminologies: Driving Interoperability in Healthcare

Russell A. Hamm[a], Sarah E. Knoop[b], Peter Schwarz[b], Aaron D. Block[c], Warren L. Davis IV[d]

[a]Mayo Clinic, Division of Biomedical Informatics
[b]IBM Almaden Research Center, Healthcare Informatics Group
[c]University of North Carolina - Chapel Hill, Computer Science Department
[d]Florida State University, Computer Science Department

Abstract

Internationally, there are countless initiatives to build National Healthcare Information Networks (NHIN) that electronically interconnect healthcare organizations by enhancing and integrating current information technology (IT) capabilities. The realization of such NHINs will enable the simple and immediate exchange of appropriate and vital clinical data among participating organizations. In order for institutions to accurately and automatically exchange information, the electronic clinical documents must make use of established clinical codes, such as those of SNOMED-CT, LOINC and ICD-9 CM. However, there does not exist one universally accepted coding scheme that encapsulates all pertinent clinical information for the purposes of patient care, clinical research and population heatlh reporting. In this paper, we propose a combination of methods and standards that target the harmonization of clinical terminologies and encourage sustainable, interoperable infrastructure for healthcare.

Keywords:

interoperability, terminology, standards, UMLS, HL7, unstructured text

Introduction

The most important characteristics of clinical and healthcare related documentation are precision and accuracy. As more and more of this documentation becomes electronic, computers will begin to assume part of the responsibility for ensuring document integrity. As such, clinical documentation will need to migrate from free-form physician notes to something more structured and processable by machines. The use of codes from established clinical terminologies within electronic clinical documents and the electronic infrastructure designed to support them will enhance machine processability and ensure comparable, if not improved, document precision and accuracy.

The need for clinical codes within electronic clinical documents poses a significant challenge to architects of NHIN efforts who desire to share information across healthcare institutions. Most healthcare organizations use codes today in a variety of use cases. Each organization, however, makes use of multiple clinical terminologies (including some that are organizationally proprietary) that do not necessarily align with terminologies used by other organizations. NHIN efforts, as well as integration efforts on a smaller scale (i.e., intra-institution and among regional institutions), must solve a translation problem to ensure that exchanged clinical information is as accurate and precise when presented to another institution as it is to the institution that created the information.

Translation among clinical terminologies is not a new problem. Concrete systems like UMLS [1] and LexGrid [2] have been under development for years and begin to attack this problem. In this paper, we outline the fundamental characteristics of this problem and requirements for a complete, automated solution. We first discuss the two key characteristics of any terminology, and how they can be leveraged to search a terminology and ultimately produce a mapping between codes. Next, we highlight the limitations of nave application of these search methods by discussing how additional factors impact the mapping of codes. Finally, we discuss how healthcare IT standards provide an infrastructure base that supports code mapping efforts.

Methods

Two approaches to search

In this section, we discuss various techniques that attempt to solve the problem of mapping from a *source* code (or lexical phrase) to a *target* code, which may belong to a different terminology. These techniques exploit the fact that every terminology has a notion of *structure* as well as *textual descriptions* of the concepts it represents [3]. The structure of a terminology can be represented as a graph that models the underlying relationships among different codes. The textual description is a human-readable word or phrase associated with each code. These fundamental characteristics aid in automated navigation within and among terminologies and can be exploited to find the target code which is the best match to a given source code or lexical phrase.

Each of the fundamental characteristics suggests a corresponding method for navigating a terminology: an unstructured lexical search or a structured graph search. Lexical searches are performed by matching the textual descriptions of codes in the target terminology. Structural searches are performed by navigating the relationships between codes in a given terminology. In order to direct

both approaches towards finding the best match, we introduce the notion of a *score*, which represents how similar two objects are to each other, and ranges between the values 0.0 (dissimilar) and 1.0 (equivalent).

For the lexical analysis, the score denotes the similarity between a phrase and the description of the code, or the similarity between the description of one code and another. For example, if we are searching for snake bite, then the code with the description snake bite will have a perfect score (*i.e.*, 1.0), whereas the code with the description squirrel bite will have a lower score. In an IBM prototype system for exploring mapping technologies, we used the Apache text search engine Lucene. Lucene constructs a text index for a data set and uses it to support searches for records or documents containing particular words or phrases. Lucene has the ability to ignore words that are used too frequently, such as a, an, and the, and can be configured to ignore additional domain-specific words. Furthermore, Lucene includes the ability to do *loose* matching, and retrieve documents or records whose text only approximately matches the desired word or phrase.

Between any two codes in a given terminology, there exists a path of relationships. For the structured search, the score of a path is calculated based on both the length of the path and the types of relationships traversed. For example, the path between the codes representing snake bite and bite will typically have a higher score than the path between the codes representing snake bite and heart attack, because (in most terminology systems) the codes for bite and snake bite are closely related, whereas the codes for heart attack and snake bite are very distant. Moreover, since the relationship type is a factor when scoring a path, it is possible that the path from bite to snake bite will have a higher score than the path from snake bite to bite. Intuitively, the reason for this behavior is that all snake bites are bites but not all bites are snake bites.

Scoring a path can be done in a number of different ways. In the IBM prototype implementation, we make use of an area of mathematics called *fuzzy set theory*. Fuzzy set theory is based on the notion that elements can have degrees of membership in particular sets. For our purposes, codes (our elements) are not simply classified as equivalent or dissimilar, but rather they range over a spectrum of similarity. If we assign a score to each individual link in a path based on the type of the relationship the link represents, we can use fuzzy set theoretic techniques to compute the score of the entire path. For example, if we were to use one of the fuzzy techniques of intersection on a path with three links having scores 0.8, 0.5, and 0.5, respectively, then the score for the *total path* would be 0.8*0.5*0.5 = 0.2.

In order to exploit structural graph searches to solve the code mapping problem, we must create *bridges* between the source and target terminologies. A bridge is some concept that is represented in both the target and source terminology. Once a bridge between two terminologies has been established, then it is possible to navigate between the source and target terminology via the closest bridge. Determining which concepts can be used as bridges is a non-trivial issue. In the IBM prototype implementation we used UMLS which is designed to relate different terminologies to each other. Thus, leveraging relationships defined by UMLS allowed us to construct bridges between the source and target code. UMLS recognizes many different types of relationships between concepts; we scored each of them differently. For example, we gave a child-to-parent relationship a higher score than a parent-to-child relationship. By scoring UMLS in this manner, we have constructed a means to perform a guided search over all the UMLS entries. We have found this technique of immense value in UMLS, since codes that are quite similar are often linked by some relationship, rather than being classified as the same concept.

The two notions of similarity (lexical and structural) can be used in isolation or in conjunction as methods for solving the code mapping problem. Currently, we are investigating blending these two approaches. While the unstructured or lexical approach is useful for finding a best match code, given a descriptive phrase, it is possible that we miss related concepts because they use different words for similar concepts. Structured search, on the other hand, requires a code to start navigating the underlying graph structure of the terminology, and may become stuck navigating an area of the graph that is only weakly related to the desired concept. Thus, structural search can in isolation return results that are of little value. By resorting to a lexical search when this happens, we can discover additional starting points for structural search that may more closely resemble the target code. This technique, which can be used iteratively, is similar to the technique of simulated annealing which is used for many global optimization problems. By combining these two approaches in this and other ways, we hope to construct an algorithm that minimizes their individual disadvantages for attacking the code mapping problem.

Additional factors impacting the code mapping problem

Several additional factors further complicate the code mapping problem. First, coding systems often allow several codes to be combined in order to describe a concept for which no specific code has been defined. For example, system A may have a specific code for lower-back pain, whereas system B may utilize one code for pain and another to designate its location as the lower back. In system A, the coding for lower-back pain is said to *be pre-coordinated*, whereas in system B the same concept is described with a *post-coordinated* pair of codes. Thus, in general, the mapping between codes in different systems is not necessarily one-to-one; it may be one-to-many or many-to-many.

Second, completely different mappings may apply depending on the clinical context in which a code appears. For example, the code for a stroke may appear in reference to the patients current condition, it may appear as an item in the patients medical history, or it may appear as an item in the patients familys medical history. Depending on how the source and target coding systems are structured, the desired target code for a mapping may be different in each of these three situations. Such examples illustrate the necessity of supporting multiple (partial) mappings between coding systems, together with machine-processable rules that can be used to select the proper mapping for a particular context.

Complicating factors such as these have significant impact on automated search for a matching code. Automated search, as previously described, would need to be configured in many different ways to be effective across disparate terminologies. The task of developing configurations for all of them would be daunting.

Discussion

A truly universal, sustainable, automated code matching method would need the codes presented to it in a consistent way so that it would be effective atop any terminology. Here is where we turn to healthcare-related IT standards in order to further the solution.

Impact of healthcare IT standards on implementation

The automated mapping techniques described in the previous sections assume that it is possible to represent knowledge about relevant coding systems in a machine-processable form, so that both the textual descriptions of codes and the structural relationships among codes can be explored by mapping algorithms. It is in this area that standards must play a critical role.

The ability to identify the correct concept mapping between code systems is further complicated in that vocabulary sources differ not just in content, but in structure. Vocabulary sources are often created for specific purposes, and as such the representational structure and description logic used by any individual source will be developed to meet the specific needs for that vocabulary source. Additionally, coded concepts are complex datatypes consisting of many properties and relationships to other attributes. Vocabulary sources will select a sub-set of the available properties to meet the specific needs for the source. Attributes such as *language* may be left out by a local code system, but a more mature and widely used vocabulary source may require a *language* attribute to be specified.

In an effort to make the mapping of coded concepts between terminology sources more transparent and predictable, Standards Development Organizations (SDOs) such as Health Level Seven Inc. (HL7) and informatics organizations like Mayo Clinics Division of Biomedical Informatics have developed standard interfaces and models for the specification of coded concepts and their associated properties**,** and for the representation of vocabularies in a consistent format.

The concept descriptor information model

One specification for the representation of coded concepts can be found in the HL7 Data Types Abstract Specification. The purpose of HL7s Data Types Abstract Specification is to define the meaning (semantics) of data values that can be assigned to a data element. Meaningful exchange of data requires that we know the definition of values so exchanged [4]. For the purpose of defining vocabulary elements, the HL7 Data Types Abstract Specification defines seven data types (or more accurately, data structures) that can be used to represent coded concepts and their associated properties (see. Figure 1).

These vocabulary data types include:

- Concept Descriptor (CD)
- Concept Role (CR)
- Coded Value (CV)
- Coded Simple Value (CS)
- Coded Ordinal (CO)
- Coded With Equivalents (CE)
- Character String with Code (SC)

The CD datatype is the most comprehensive of the vocabulary datatypes and includes identification attributes for coded concepts, including the unique *Code* for the coded concept*,* the *Display Name, Code System* and unique *Code System Identifier* as well as an optional *Code System Version.* To support operations such as concept mapping operations additional attributes have been specified.

The *Original Text* attribute can be used to represent the text or phrase used as the basis for the coding often reflecting content of physician notes [4].

The *Qualifier* attribute is used to specify additional codes that increase the specificity of the primary code.[4] For example, combining the concept identifier for *Family History* to the concept identifier for *Stroke* to indicate that there is a history of stroke in a patients family. Additionally, *Qualifier* can be used in concept mapping. For example, one can equate the concept *low back pain* in code system A to a combination of the concepts *pain* and *lower back* in code system B.

Translation is used to capture alternative concept descriptors in order to translate a coded concept from one code system into concepts from others. This is useful for cases where, say, both ICD9 and CPT codes are needed.

An attribute that greatly assists in concept mapping is *Equality.* This attribute allows us to directly equate one concept to another.

The HL7 Data Types can be expressed in several formats, including the Unified Modeling Language, to express data type attributes and the interrelationships between the vocabulary data types, as well as XML Schema. These formats provide a basis for syntactically interoperable transmission of data types between systems.

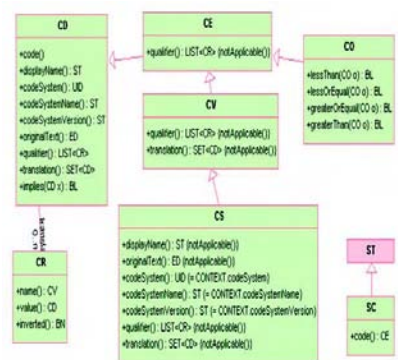


Figure 1 - The concept descriptor information model [4]

As is apparent from Figure 1, HL7s Concept Descriptor Information Model is a hierarchical model that defines specializations of vocabulary representations. These spe-

cializations act as constraints on the allowed properties of the CD data type, and are useful when communicating concept information when the semantics of a communication are agreed upon in advance, or in use case specific communication such as concept mapping.

The HL7 vocabulary data types provide a standardized syntax for communicating coded concept information between systems. Coded concepts from separate code systems can be represented in a common format, simplifying automated mapping algorithms by providing a common denominator for concept comparison.

The lexical grid

Just as the HL7 Datatypes can be used to coordinate the representation of individual coded concepts, robust vocabulary models can be constructed to define how disparate vocabularies can be represented. As terminologies are developed with specific use cases in mind, the representational structure and description logic used by any individual source will be developed to meet the specific needs for each terminology. This further complicates terminology access and mapping as programmatic access of the terminology source must be customized to the logic and structure of each source being accessed. While the semantics of individual vocabulary sources will understandably remain distinct, the ability to represent various vocabularies in a common format with a common set of tools would significantly simplify and improve concept mapping operations. This is the goal of the Lexical Grid (LexGrid).

At its core, LexGrid is a model of terminology that can represent the content of disparate vocabulary sources. Vocabularies represented in a LexGrid compliant format are available to be accessed programmatically using common tools such as HL7s Common Terminology Services Specification (CTS). CTS is an HL7 ANSI standard defining the minimum requirements for vocabulary interoperability across disparate healthcare applications. For example, when performing concept level mapping across code systems, a semantic mapping model is required to automate the mapping process. LexGrid provides a foundation and lowest common denominator for which these mapping semantics can be represented.

LexGrid, like the HL7 Vocabulary datatypes can be used to provide a common syntax for communicating across vocabulary sources. By controlling the way the semantics of a vocabulary are represented, operations such as concept mapping can be more accurately controlled, helping to reduce errors in mappings, and improve the consistency in the mapping algorithms.

Future work in combining the LexGrid model with enhanced vocabulary services APIs is underway as part of HL7s CTS 2 project. HL7s CTS 2 builds on the common set of methods and APIs specified in the CTS, with the goal of providing additional functionality of a general-purpose nature that falls outside the scope of the HL7. CTS 2 will include additional APIs for vocabulary query, authoring, and mapping. These API are only practically implementable across vocabularies when a common representation model is employed.

Conclusion

Standards such as the HL7 Datatypes and HL7s Common Terminology Services embodied by Mayo Clinics Lexical Grid, provide a consistent syntax to represent the semantic structure of a variety of existing vocabulary sources. This plays a key role in enhancing the accuracy of automated mapping techniques. The combination of search methods that exploit the fundamental properties of terminologies with standards-based data structures and APIs provide significant components of a completely automated, sustainable solution to the code mapping problem. With technology present to not only create system level but semantic level interoperability, we make a significant step towards realizing national health information networks, potentially revolutionizing healthcare for all.

References

[1] http://www.nlm.nih.gov/research/umls/umlsmain.html (Last updated: 23 February 2007).
[2] http://informatics.mayo.edu/LexGrid/index.php (Last Updated: 26 January 2007).
[3] O.Bodenreider and A. Burgun, Aligning Knowledge Sources in the UMLS: Methods, Quantitative Results, and Applications, MedInfo 2004, M.Fieschi et. al. (eds), Amsterdam, IOS Press, 2004.
[4] http://www.hl7.org (Last Updated 1 January 2007).

Address for correspondence

Russell A. Hamm
Mayo Clinic, Division of Biomedical Informatics
Email: hamm.russell@mayo.edu

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

A New Machine Learning Classifier for High Dimensional Healthcare Data

Rema Padman[a] , Xue Bai[a,b] and Edoardo M. Airoldi[b]

[a]*The H. John Heinz III School of Public Policy and Management, Carnegie Mellon University, U.S.A.*
[b]*Center for Automated Learning and Discovery, Carnegie Mellon University, U.S.A.*

Abstract

Data sets with many discrete variables and relatively few cases arise in health care, ecommerce, information security, and many other domains. Learning effective and efficient prediction models from such data sets is a challenging task. In this paper, we propose a new approach that combines Metaheuristic search and Bayesian Networks to learn a graphical Markov Blanket-based classifier from data. The Tabu Search enhanced Markov Blanket (TS/MB) procedure is based on the use of restricted neighborhoods in a general Bayesian Network constrained by the Markov condition, called Markov Blanket Neighborhoods. Computational results from two real world healthcare data sets indicate that the TS/MB procedure converges fast and is able to find a parsimonious model with substantially fewer predictor variables than in the full data set. Furthermore, it has comparable or better prediction performance when compared against several machine learning methods, and provides insight into possible causal relations among the variables.

Keywords:

Markov Blanket, Bayesian Networks, machine learning, Tabu Search, health care decision support.

Introduction

The deployment of comprehensive information systems and online databases has made extremely large collections of real-time data readily available. In many domains such as genetics, clinical diagnoses, direct marketing, finance, and on-line business, data sets arise with thousands of variables and a small ratio of cases to variables. Such data present dimensional difficulties for classification of a target variable, and identification of critical predictor variables [1]. Furthermore, they pose even greater challenges in the determination of actual influence, i.e., causal relationships between the target variable and predictor variables. The problem of identifying essential variables is critical to the success of decision support systems and knowledge discovery tools due to the impact of the number of variables on the speed of computation, the quality of decisions, operational costs, and understandability and user acceptance of the decision model. For example, in medical diagnosis and healthcare decision support, the elimination of redundant tests may reduce the risks to patients and lower healthcare costs [2]. In this study, we address this problem of efficiently identifying a small subset of predictor variables from among a large number, and estimating the causal relationship between the selected variables and the target variable, using *Markov Blanket* (MB) and *Tabu Search* (TS) approaches.

We propose a two-stage Tabu Search enhanced Markov Blanket procedure that finds a parsimonious MB Directed Acyclic Graph (DAG). This two-stage algorithm generates an MB DAG in the first stage as a starting solution; in the second stage, the Tabu Search metaheuristic strategy is applied to improve the effectiveness of the MB DAG as a classifier, with conventional Bayesian updating. Classification using the Markov Blanket of a target variable in a Bayesian Network has important properties: it specifies a statistically efficient prediction of the probability distribution of a variable from the smallest subset of variables; it provides accuracy while avoiding over-fitting due to redundant variables; and it provides both a classifier and some insight into causal relations between a reduced set of predictors and the target variable. The TS/MB procedure proposed in this paper allows us to move rapidly through the search space of Markov Blanket structures and escape from local optima, thus learning a more robust structure.

Background knowledge

A *Bayesian Network* is a graphical representation of the joint probability distribution of a set of random variables. A Bayesian Network for a set of variables $X = \{X_1, ..., X_n\}$ consists of: (i) a directed acyclic graph (DAG) S that encodes a set of conditional independence assertions among variables in X; (ii) a set $P = \{p_1, ..., p_n\}$ of local conditional probability distributions associated with each node and its parents.

P satisfies the *Markov condition* [3] for S if every node X_i in S is independent of its non-descendants and non-parents in S, conditional on its parents. The Markov Condition implies that the joint distribution p can be factorized as a product of conditional probabilities, by specifying the distribution of each node conditional on its parents. In particular, for a given structure S, the joint probability distribution for X can be written as

$$p(X) = \prod_{i=1}^{n} p_i(x_i \mid pa_i) \quad (1)$$

where pa_i denotes the set of parents of X_i; this is called a Markov factorization of P according to S.

Given the set of variables X and target variable Y, a *Markov Blanket* (MB) for Y is the smallest subset Q of variables in X such that Y is independent of $X\backslash Q$, conditional on the variables in Q. P is *faithful* to the graph S with the vertex set X if and only if there are no conditional independence relations in P other than those entailed by satisfying the Markov condition for S. If P is faithful to the graph S, then given a Bayesian Network (S, P), **there is** a unique Markov Blanket for Y consisting of the set of parents of Y, the set of children of Y, and the set of parents of children of Y.

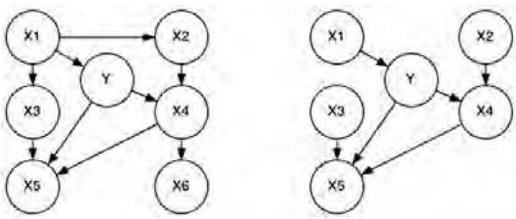


Figure 1 - (left) A Bayesian Network (S, P), and (right) the Markov Blanket for the variable Y

For example, given the two DAGs in Figure 1, the factorization of p entailed by the Bayesian Network (S, P) is

$$p(Y,X_1,...,X_6) = p(Y \mid X_1)\cdot p(X_4 \mid X_2,Y)\cdot p(X_2 \mid X_1)\cdot p(X_5 \mid X_3,X_4,Y)\cdot p(X_3 \mid X_1)\cdot p(X_6 \mid X_4)\cdot p(X_1) \quad (2)$$

The factorization of the conditional probability for the Markov Blanket of Y: $p(Y \mid X_1, \ldots, X_6)$ is the product of those (local) factors in equation above that contain the term Y:

$$p(Y \mid X_1,...,X_6) = C\cdot p(Y \mid X_1)\cdot p(X_4 \mid X_2,Y)\cdot p(X_5 \mid X_3,X_4,Y), \quad (3)$$

where C is a normalizing constant independent of Y.

Recent research in classification problems has tried to identify the Markov Blanket variables of the target class variable by filtering predictor variables using statistical tests of conditional independence [4]. Few of the proposed methods, however, have been used to generate the Markov Blanket from real-world data, and tests on real-world problems have usually involved only a small number of variables. Furthermore, while all the previous studies use the notion of Markov Blanket as a minimal set of dependent variables, none of them actually generate and retain the graphical structure of the Markov Blanket corresponding to a specific data set, nor have they used the structure for Bayesian inference in order to perform the classification. Our algorithm addresses all these limitations. Details are available in [5].

Tabu Search is a powerful meta-heuristic strategy that helps local search heuristics to explore the solution space by guiding them out of local optima [6]. Its strategic use of memory and responsive exploration is based on selected concepts that cut across the fields of artificial intelligence and operations research. It has been applied successfully to a wide variety of continuous and combinatorial optimization problems, capable of reducing the complexity of the search process and accelerating the rate of convergence. In its simplest form, Tabu Search starts with a feasible solution and chooses the *best move* according to an evaluation function while taking steps to ensure that the method does not re-visit a solution previously generated. This is accomplished by introducing *tabu restrictions* on possible moves to discourage the reversal, and, in some cases, repetition of selected moves. The *tabu list* that contains these forbidden move attributes is known as the short term memory function. It operates by modifying the search trajectory to exclude moves leading to new solutions that contain attributes (or attribute mixes) belonging to solutions previously visited within a time horizon governed by the short term memory. Intermediate and long-term memory functions may also be incorporated to intensify and diversify the search.

TS/MB model

A sketch of the algorithm is presented in Figure 2. The detailed algorithm is presented in [5]. Our algorithm first generates an *initial Markov Blanket* for the target variable. However, the initial MB may be highly suboptimal due to the application of repeated conditional independence tests and propagation of errors in causal orientation [5, 7]. Therefore, Tabu Search is applied to improve the initial MB. Four kinds of moves are considered in the procedure: edge addition, edge deletion, edge reversal and edge reversal with node pruning. At each stage, for each allowed move, the resulting Markov Blanket is computed, factored, its predictions scored, and the current MB modified with the best move. The algorithm stops after a fixed number of iterations or a fixed number of non-improving iterations.

Computational results

We tested our algorithm on two biomedical data sets [8, 9]. Table 1 provides a brief characterization of the data sets. Prostate cancer (PCA) data set concerns diagnosis of prostate cancer from mass spectroscopy of human sera [8]. Arrhythmia data set concerns classification of subjects into 8 disease categories from clinical and EKG data [9].

The parameters in our experiments are: data-splits, scoring criteria, starting solution structure, the depth of conditional independence search (***d***), and significance level (α). We use a nested, stratified cross-validation scheme [10]. In the inner layer, the procedure trains and optimizes the Markov Blanket on training data for each parameter configuration. The configuration that yields the best MB according to the scoring criterion is chosen as the best configuration. The outer layer of cross-validation estimates the performance of the optimized Markov Blanket classifier on the testing data. We report both the AUC and prediction accuracy on the testing set to evaluate the classification performance of the generated models.

InitialMBsearch (Data D, Target T, Depth d, Significance α):

/*F: the list of all the variables in D; M: a set of nodes to which edges should not be drawn; $sepSet(v_i, v_j)$: a mapping of a set of nodes s.t. $(v_i \perp v_j \mid sepSet(v_i, v_j))$; $adj(v_i)$: the set of adjacent nodes to node v_i in G; A: an edge list; $vertex(G)$: the set of vertexes in the graph G; $edges(G)$: the set of edges in the graph G:*/

/* Finding adjacency.*/
For all i, j
$vertex(G) := \emptyset$; $edges(G) := \emptyset$; $M := \emptyset$;
$F :=$ all the vertexes in the D; $sepSet(*,*) := \emptyset$;
checkedges $(T, F, M, G, d, sepSet(*, T))$;
For each $v_i \in adj(T)$
checkedges $(v_i, F, M, G, d, sepSet(*, v_i))$;
For each $w_j \in adj(adj(T))$
checkedges $(w_j, F, M, G, d, sepSet(*, w_j))$;

/* Pruning G.*/
$G =$ **Ornt** $(Y \to L_Y \leftarrow L_X)$

/* Transform into a $MBDAG$: */
$\{MB_{DAG}(Y), L\} =$ **Trsfm** (G)

/* Tabu search enhancement: */
TabuSrch $(MB_{DAG}(Y), L, Max_{Iter})$
init: $best_{MB} = curr_{MB} = MB_{DAG}$, $best_{Score} = 0$
repeat until ($best_{Score}$ does not improve for k consecutive iterations)
form $candidate_{Moves}$ for $curr_{MB}$;
find $best_{Move}$ among $candidate_{Moves}$ by **score**$(move_i)$;
if $(best_{Score} <$ **score** $(best_{Move}))$;
update $best_{MB}$ by applying $best_{Move}$ and $best_{Score}$;
add $best_{Move}$ to $TabuList$ // not re-considered in the next m iterations;
update $current_{MB}$ by applying $best_{Move}$;
return $best_{MB}$ // an MB_{DAG};

Figure 2 - A sketch of the TS/MB algorithm

Table 1 - Characteristics of Data Sets

Task	Vars.	Samp.	Var. Type	Target Var.
PCA diagnosis	779	326	Discretized	Binary
Arrhythmia diagnosis	279	417	Ordinal	8 Categories

Table 2 presents the average best-fitting classification results for PCA data and the comparison against several state-of-the-art classifiers in three different ways. Details of the classifiers and the motivation for their choice in this study are discussed in [5]. We report both the AUC[1] and prediction accuracy on the testing set as well as the size of reduction in the set of variables. The size reduction was evaluated based on the fraction of variables in the resulting models. All metrics (variable size reduction, AUC, and accuracy) were averaged over cross-validation splits.

Comparison I (columns 1 and 2) presents the results when using the *full set* of variables as input. Comparison II (columns 3 and 4) uses the *same number* of variables as identified by TS/MB, as input for all the other classifiers. These variables for the classifiers in the comparison set are selected using information gain (IG) criterion [11]. Comparison III (columns 5 and 6) uses the *exact same* variables identified by TS/MB as input variables for all classifiers. Comparisons II and III are used to test the source of the differential in the observed accuracy and AUC values in the full data set.

1 Since AUC is only applicable to binary classification problems, we report only the accuracy for Arrhythmia data.

Table 2 - Five-fold cross validation results of various classifiers on 779 peaks; on 19 peaks selected by information gain; and on the exact same 19 peaks selected by the TS/MB classifier for the PCA data. Best performance figures are in ***bold***

Input	All Peaks		Peaks selected by IG		Peaks selected by TSMB		
Method	**AUC %**	**Accuracy %**	**AUC %**	**Accuracy%**	**AUC %**	**Accuracy%**	**# Peaks selected**
MB	95.3	87.1	95.3	87.1	95.3	87.1	19
TS/MB	98.3	90.3	98.3	90.3	98.3	90.3	19
Naïve Bayes	97.5	89.3	67.5	63.2	77.4	69.4	
SVM	97.1	98.5	63.3	62.0	72.6	69.9	
Voted Perceptron	73.9	58.0	65.2	59.2	75.4	67.8	
Max. Entropy	87.4	98.8	64.7	64.1	74.9	70.9	
K-NN	96.3	88.6	65.6	58.6	72.1	65.3	
Logistic Reg.	Failed	Failed	73.6	56.4	98.1	90.0	

The results clearly indicate that TS/MB dominates on the AUC metric and performs well on the accuracy measure, except in one instance. In addition, there is an almost 98% reduction in the number of variables used for this prediction. It is also interesting to note that all competing classifiers perform better with the variables provided by TS/MB than using the IG criterion (Comparison II vs. Comparison III) on both accuracy and AUC. Figure 3 shows the MB DAG learned from PCA data that achieves the best accuracy on the testing data. All the directed edges are robust over almost all cross validation runs, with very small variation. Further study and interaction with clinicians is necessary to identify the clinical significance of these variables in actual settings.

Table 3 presents the average best-fitting classification results for Arrhythmia data and compares them against the results obtained from several state-of-the-art classifiers. Figure 4[2] shows the MB DAG that achieves the best accuracy on the testing set of Arrhythmia data. The classifier in [9] achieved an accuracy of 62%, which obviously is not sufficiently good for clinical use. The best result from TS/MB achieved average accuracy of 96.8% with a 95% reduction in the number of variables required for the prediction, a significant improvement over the earlier study.

Discussion and conclusion

On average, the TS/MB classifier reduces the set of predictor variables by at least 95% from the full set of variables. In some cases it is reduced to a sufficiently small set for entry into hand calculators, or paper and pencil decision procedures that are easy to use in clinical and other decision settings. At the same time, when compared to the state-of-the-art classification methods, the TS/MB classifier procedures excellent classification results, especially in real world applications where the cost of misclassification has significant implications. Moreover, the algorithm generates a graphical structure that represents the relationships between the variables and provides additional insight into causal discovery. These experiments, as well as more results we have obtained on data sets from other domains, such as Internet marketing and sentiment extraction, suggest that for problems where the ratio of samples to the number of the variables is small or the independence assumption is not appropriate, the two-stage MB classifier is superior in terms of the prediction performance, effectiveness in identifying critical predictors, and robustness [5].

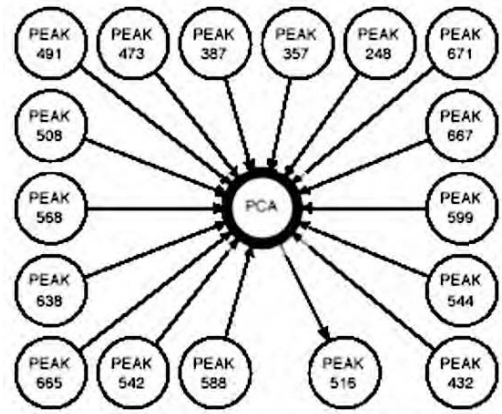


Figure 3 - The best fitting MB DAG for the PCA

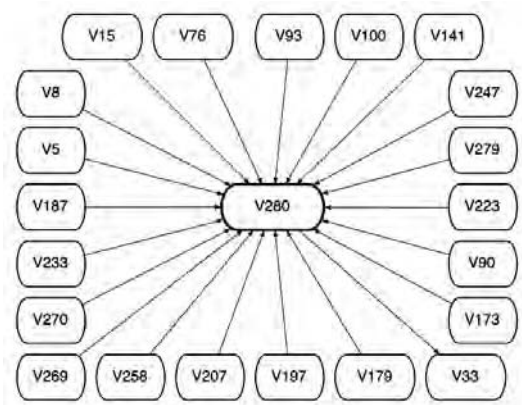


Figure 4 - The best fitting MB DAG for arrhythmia

It is possible that different Markov Blanket graphical structures that are consistent with the TS/MB classifier output would give slightly different classification results. Because any undirected and bi-directed edges are deleted after the edge orientation step, these deletions might result in suboptimal decisions. Tabu Search iteratively investigates alternative orientations and further edge additions to minimize the extent of sub-optimality. On the other hand, theoretically, two DAGs that have the same Markov factorization are Markov equivalent. In our case, two Markov Blankets can be Markov equivalent even if some edges are oriented in a different, but statistically non-differentiable way.

The results of this study need to be validated against clinicians' interpretation of the variables and their value in actual diagnosis settings. This research can also be extended to address the interesting problem of simultaneously building classifiers for all variables in a large variable data set, discovering a causal model for all variables in such data, or automatically classifying medical documents into categories. Future research will explore these issues.

2 The class variable "V280" has 8 categories, encoding 8 different diseases.

*Table 3 - Five-fold cross validation results of various classifiers using 279 variables; using 21 variables selected by information gain; and using the exact same 21 variables selected by the TS/MB classifier for the Arrythmia data. Best performance figures are in **bold***

Input	**All vars.**	**Vars. Selected by IG**	**Vars. Selected by TS/MB**	
Method	**Accuracy %**	**Accuracy %**	**Accuracy %**	**# peaks selected**
MB	77.6	77.6	77.6	**20**
TS/MB	**96.8**	**96.8**	**96.8**	**21**
Naïve Bays	57.0	57.0	57.0	
SVM	93.3	69.4	72.6	
Voted Perception	72.3	71.5	71.3	
Max. Entropy	96.4	74.7	76.4	
K-NN	71.3	76.6	75.1	
Logistic Reg.	80.9	72.0	74.0	

Acknowledgements

We thank Professor Fred Glover of the University of Colorado at Boulder, Professor Clark Glymour and Professor Peter Spirtes of Carnegie Mellon University, and Dr. Frank Wimberly at Institute for Human and Machine Cognition for valuable suggestions throughout the project; and Professor Constantin Aliferis of Vanderbilt University for providing us with the PCA data set.

References

[1] Berry M and Linoff G. Data Mining Techniques: For Marketing, Sales, and Customer Support. Wiley, 1997.

[2] Cooper GF, Aliferis C, Aronis J, Buchanan B, Caruana R, Fine M, Glymour C, Gordon G, Hanusa B, Janosky J, Meek C, Mitchell T, Richardson T, Spirtes P. An evaluation of machine-learning methods for predicting pneumonia mortality. Artificial Intelligence in Medicine 1992; 9: 107–139.

[3] Pearl J. Causality: Models, Reasoning, and Inference. Cambridge University Press, 2000.

[4] Aliferis C, Tsamardinos I, Statnikov A. Hiton, a novel markov blanket algorithm for optimal variable selection. Proceedings of the American Medical Informatics Association Annual Symposium, 2003b, 21–25.

[5] Bai X, Padman R. Tabu search enhanced markov blanket classifier for high dimensional data sets. In Proceedings of INFORMS Computing Society, pp. 337-354. Kluwer Academic Publisher, 2005.

[6] Glover F. Tabu Search. Kluwer Academic Publisher, 1997.

[7] Spirtes P, Glymour C and Scheines R. Causation, Prediction, and Search. MIT Press, 2000.

[8] Adam B, Qu Y, Davis J, Ward M, Clements M, Cazares L, Semmes O, Schellhammer P, Yasui Y, Feng Z, Wright G. Serum protein fingerprinting coupled with a pattern-matching algorithm distinguishes prostate cancer from benign prostate hyperplasia and healthy men. Cancer Research 2002; 62:3609–14.

[9] Güvenir HA, Acar A, Demiröz G, Cekin A. A supervised machine learning algorithm for arrhythmia analysis. Computers in Cardiology 1997; 24: 433-36.

[10] Weiss S. and Kulikowski C. Computer Systems That Learn. Morgan Kaufmann, 1991.

[11] Mitchell T. Machine Learning. McGraw-Hill. 1997.

Address for correspondence

Rema Padman,
The Heinz School,
Carnegie Mellon University,
Pittsburgh, PA 15213, USA.

Email: rpadman@cmu.edu

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Structuring of Free-Text Diagnostic Report

Hirofumi Fujii[a], Hiromasa Yamagishi[b], Yutaka Ando[c], Nobuhiro Tsukamoto[d], Osamu Kawaguchi[d], Tomotaka Kasamatsu[e], Kaoru Kurosaki[b], Masakazu Osada[b], Hiroshi Kaneko[b], Atsushi Kubo[e]

[a] Functional Imaging Division, National Cancer Center Hospital East, Japan
[b] Toshiba Medical Systems Corporation, Japan
[c] Medical Informatics Section, National Institute of Radiological Sciences, Japan
[d] Department of Radiology, Saitama Medical School, Japan
[e] Department of Radiology, Keio University, Japan

Abstract

Purpose: It is useful to convert free-text diagnostic reports into structured diagnostic reports by semantic analysis for the secondary investigation of their contents. In this study, we propose a system in which description units are automatically extracted to create structured text reports and we evaluated its usefulness.

Methods: We defined the rules to create description units and developed the system that can automatically extract these description units from free-text diagnostic reports. We applied this system to reports of cerebral perfusion scintigrams and obtained 5 dictionaries of description units, increasing the number of scintigrams from 100 to 500 in increments of 100. Each dictionary was used to analyze another 100 scintigrams. The results obtained using each dictionary were compared with the results of physicians' interpretation.

Results: The recall rate of this system to the physicians' interpretation increased when correlated with the number of scintigrams but with 300 cases was almost saturated at 85%.

Conclusion: We propose a semantic analysis system and show its usefulness in the semantic evaluation of the reports of cerebral perfusion scintigrams.

Keywords:

text mining, description unit, DICOM-SR, Alzheimer disease, cerebral perfusion scintigraphy

Introduction

In recent years, intrahospital information such as medical records and diagnostic images tends to be managed as electronic data [1]. And, statistical analysis of medical information and computer-aided diagnosis of imaging tests are being actively investigated by utilizing digitalization.

Reports of diagnostic imaging tests also tend to be digitally created using electronic devices such as personal computers. These electronically created reports are often used as references to evaluate intra-individual interval changes. The digital information of these reports is, however, less often used to electronically analyze the contents of the reports.

The main reason is that the diagnostic reports are generally expressed using free-text format, and it is difficult to systematically analyze the descriptive contents of these reports [2-6]. If the descriptive contents of the reports are semantically analyzed and then converted into structured text reports, they can be effectively reused [7-12]. For example, if we can use such a system, similar clinical cases can be quickly picked up from the database of previous reports. If these analyzed, structured data are saved with the format of DICOM-SR, secondary investigations using large-scale accumulated data to evaluate the tendency of findings in some entities of diseases can be easily performed [13-14].

In the present study, we have developed a system to automatically create structured text reports from the reports with free-text format, which uses a text mining tool that extracts semantic information from the contents of free-text format referring to dictionaries. We assessed the usefulness of this system by applying it to actual clinical diagnostic imaging reports of cerebral perfusion scintigraphy, which is attracting significant attention from many researchers as an important test to detect the early stages of Alzheimer disease.

Materials and methods

In this study, to convert the sentences with free-text format into structured formats, we used a text mining tool that consisted of two algorithms: the first was the morphological analysis algorithm that divided text data into units of words and the second was the structuring algorithm that extracted units of words from sentences with the free-text format. The extracted units of words, called description units, were classified into two types: finding units and

diagnostic units. The finding units included "finding", "modifier", "region", "regional modifier", and "confidence" and diagnostic units included "diagnosis",

Selected for best paper award.

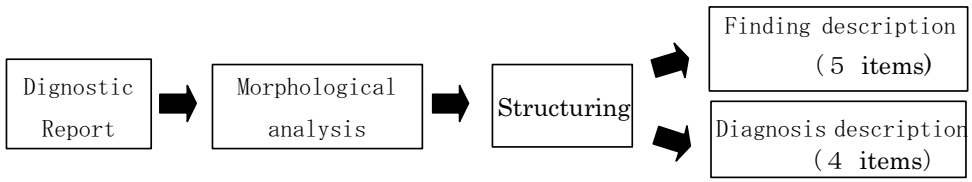


Figure 1 - Text mining processing

"region", "regional modifier", and "confidence". As requisites for description units, we defined that finding units must contain words related to "finding", "region", and "confidence" and diagnosis units must contain words related to "diagnosis" and "confidence". The concept of the structuring process is shown in Figure 1.

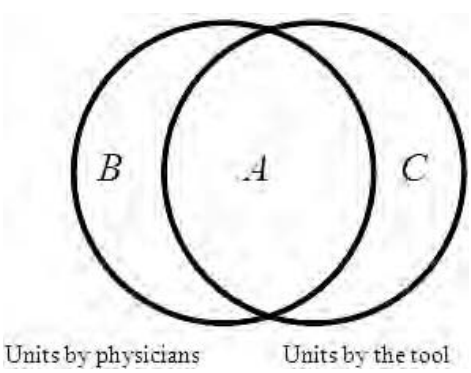
At first, to create dictionaries of description units, we manually picked up the words that can be considered description units from free-text reports of the diagnostic imaging test that we wanted to analyze.

After that, we applied this system with these created dictionaries to cerebral perfusion scintigraphy. We used the reports of this test created in the actual clinical situation; no information about structuring was given to the clinical staff who wrote diagnostic reports.

We made 5 dictionaries of description units by increasing the number of free-text reports used for picking up description units (learning reports) from 100 cases to 500 cases in increments of 100 cases.

We prepared another 100 cases of reports of cerebral perfusion scintigraphy. Each dictionary was applied to these new 100 cases and description units were automatically extracted using these dictionaries. Description units were also manually extracted using the same 100 reports by physicians. More than two doctors discussed the extraction of each unit, reaching a consensus when there were discrepancies among their opinions, and the results of physicians' analyses were considered the gold standard.

We evaluated the usefulness of this system using the following rates: the recall rate, the mismatch rate and the unexpected description unit rate (Figure 2).



A the number of identical description units automatically generated by the text mining tool.

B the number of manually generated description units that do not match description units automatically generated by our system.

C the number of automatically generated description units that do not match manually generated description units.

The recall rate is *A/(A+B)*.

The mismatch rate, that is, the false negative rate, is *B/(A+B)*.

The unexpected description unit rate, that is, the false positive rate, is *C/(A+B)*.

Figure 2 – The evaluation of the system

Results

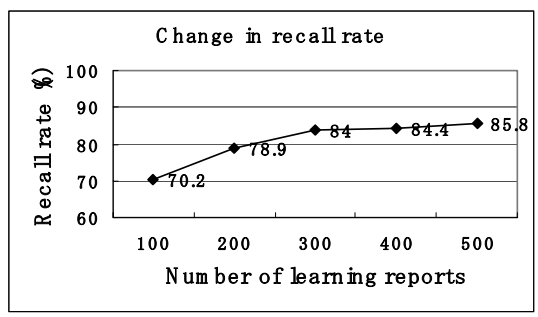


Figure 3 - Change in recall rate

Figure 3 shows the correlation between the recall rate and the number of learning reports. Although the recall rate was 70.2% when the number of learning reports was 100, it reached 85.5% when the number of learning reports was increased to 500. When the number of learning reports was 300 or more, the recall rate tended to saturate and became nearly constant. These results also suggested that there remained a certain mismatch rate even when the number of learning cases was increased.

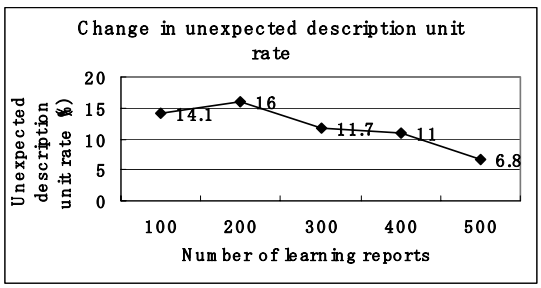


Figure 4 - Change in unexpected description unit rate

Figure 4 shows the correlation between the unexpected description unit rate and the number of learning reports. Although the unexpected description unit rate was about 15% when the number of learning reports was small, it dropped to 6.8% when the number of learning reports was increased to 500. These results indicated that increasing the number of learning reports can reduce the number of unexpected description units.

Discussion

The methods by which description units are extracted from sentences and are structured have been investigated by some researchers. Imai, et al. [15] attempted to analyze the radiological findings that appeared in the electrical medical textbook using this type of method and reported that they could successfully build a medical ontology. Anei, et al. [16] tried to apply this method to free-text diagnostic reports. They extracted "medical knowledge", which corresponds to "description units" in our study, using a "support vender machine," but, the results obtained were not excellent. The rate of accurate extraction was only 42%. The main reason for this poor result was that they did not register the extracted knowledge in the dictionary. They analyzed sentences using only the combination of the attributes of the words.

We built a system for the structuring of free-text diagnostic reports by modifying the shortcomings of previous studies stated above. And, in this study, we evaluated the usefulness of our system using the recall rate, which is the accuracy of the structuring, and the unexpected description unit rate.

The recall rate using our text mining tool improved as the number of registered words increased. This result indicates that this structuring system works well.

The recall rate almost plateaued around 85%. If it is demonstrated that this rate, 85%, is clinically acceptable in the diagnosis of cerebral perfusion scintigraphy, we can create the dictionary by analyzing the reports of as few as 300 cases and use it for structuring the free-text reports.

If a higher recall rate is required in the actual clinical situation, we must consider methods for further improvement of the recall rate.

To develop methods to improve the recall rate, we investigated why our text mining tool incorrectly extracted description units. The reasons are as follows:

1. There were some mismatched description units and unexpected description units. This could be caused by an insufficient number of registered words because this phenomenon often occurred when reports with the words that were not included in the learning reports were analyzed. To resolve this issue, we must increase the number of the learning reports. It would also be useful to provide a function permitting the operator to easily register an unregistered word while creating a report.

2. Mismatched description units were often found in the reports with misspellings or typographical errors. This phenomenon was caused by human error during the input of the text. To overcome this issue, a function for proofreading of input text data before the process of structuring could be added.

3. Mismatched description units and unexpected description units were also found when reports with grammatically imperfect sentences, such as sentences without the subject, were analyzed. The use of expressions specific to individual physicians or clinical institutions might also cause this kind of error. To overcome this issue, it would be convenient to add a function that lets the operator know the grammatical error and suggests a correction of the sentence.

To evaluate the recall rate, we also have to verify the effectiveness of description units extracted by physicians because these description units were used as the gold standard. In this study, description units were always extracted by more than two doctors with significant experience. And the discrepancies among them were resolved by discussion. Therefore, we believe that manually extracted units could be used as the gold standard although we did not calculate inter-rater agreement among the physicians.

We should also investigate methods to secondarily effectively utilize structured text reports.

The information of the reports with free-text format can be semantically managed if they are converted to structured text reports by the text mining tool. As a result, they can be reused for other purposes.

If the information of the reports is expressed by description units, it can be easily translated into another language. The Japanese that we commonly use is a language whose grammar is quite different from English; however, English is commonly used in the field of science. Thus, it is difficult to directly analyze the information of the reports written by Japanese. But if the reports were converted into a structured format with clearly expressed description units, the information could be easily translated into English. Following is an example.

•Cerebral perfusion scintigraphy report

両側頭頂葉、後部帯状回、楔前部において著明な血流低下を認めます。 典型的なアルツハイマーの画像所見です。

•Extracted description units

No	記述単位	部位(必須)	部位修飾(任意)	所見(必須)	確定度(必須)	修飾(任意)
1	所見	頭頂葉	両側	血流低下	断定	著明な
2	所見	帯状回	後部	血流低下	断定	著明な
3	所見	楔前部		血流低下	断定	著明な

No	記述単位	部位(任意)	部位修飾(任意)	診断名(必須)	確信度(必須)	-
4	診断	-	-	アルツハイマー	断定	-

Figure 5 - Extracting process example in Japanese

•Cerebral perfusion scintigraphy report

Remarkable hypoperfusion is noted in bilateral parietal lobes, posterior cingulate gyri and precuni, typical of Alzheimer disease.

•Extracted description units

No	Units	Region	Regional Mod.	Findings	Confidence	Modifier
1	Findings	Paretal	Bilateral	Hypoperfusion	Strongly positive	Remarkable
2	Findings	Cingulate gyri	Posterior	Hypoperfusion	Strongly positive	Remarkable
3	Findings	Precuni		Hypoperfusion	Strongly positive	Remarkable

No	Units	Region	Regional Mod.	Diagnosis	Confidence	Modifier
4	Diagnosis	-	-	Alzheimer	Strongly positive	-

Figure 6 - Extracting process example in English

Figure 5 shows description units extracted from a brain perfusion scintigraphy report written by Japanese using the text mining tool. Figure 6 shows the description units translated into English. As shown in these figures, the contents of Japanese diagnostic reports can be easily converted into English.

This example suggests that structuring the diagnostic reports that currently use a free-text format might be useful to discuss the information about the report with people whose native languages are different from each other. We have not, however, demonstrated yet whether sentences from foreign languages composed of converted description units can accurately indicate the meaning of the sentences of the original languages. We can now say that the meaning of description units can be understood in a foreign language without complicated translation algorithms when reports written in the original language are structured into description units using our text-mining system and these units are then converted into a foreign language.

If our structuring system can accurately translate the meanings of sentences from the original language into the second one, this system could also be applied to teleradiology between foreign countries. The more complicated the information of the report is, the more accurately speakers of other languages would discuss it with each other.

In this study, we demonstrated the usefulness of structuring reports that use a free-text format. There are, however, some limitations in our current investigation.

We studied only the reports of cerebral perfusion scintigraphy. Although this study is an important test to detect the early stage of Alzheimer disease, the findings of this imaging test are rather limited. Therefore, we might successfully analyze the reports with free-text format by a text mining tool. We must confirm the usefulness of the semantic analysis by applying our structuring system to the modalities whose diagnostic reports are more complicated, such as computed tomography.

Conclusion

Structuring of diagnostic reports in free-text format was performed using the automatic system with the text mining tool, which extracts description units. When this system was applied to the reports of cerebral perfusion scintigraphy created in the actual clinical situation, it successfully converted the reports with free-text format into the structured ones expressed by description units.

Although the usefulness of this structuring system must be validated for the modalities whose diagnostic reports are more complicated, this system could possibly compile the information obtained from the reports of many cases and to effectively reuse it.

Acknowledgment

The authors thank Barbara L. Clough, M.A., Magnus Communications, for her linguistic review.

References

[1] Akiyama M. Medical imaging information system (In Japanese). Med Clin Jpn 1999: 25: 398-402.
[2] Cohen AM and Hersh WR. A survey of current work in biomedical text mining. Brief Bioform 2005: 6: 57-71.
[3] Crasto CJ, Marenco LN, Migliore M, Mao B, Nadkarni PM, Miller P, Shepherd GM. Text mining neuroscience journal articles to populate neuroscience databases. Neuroinformatics 2003: 3: 215-237.
[4] Yeh AS, Hirschman L, and Morgan AA. Evaluation of text data mining for database for curation: lessons learned from the KDD Challenge Cup. Bioinformatics 2003: 19 (suppl 1): i331-339.
[5] Makino K. Text mining for the medical field (in Japanese). Toshiba Rev 2005: 60: 46-47.
[6] Sasaki Y. Standardization of the reports of diagnostic imaging (in Japanese). Shin-iryo 2001: 28: 106-108.
[7] Tsukamoto N. Supporting system for CT diagnosis referring to previous cases (in Japanese). Nippon Acta Radiol 1993: 53: 1415-1425.
[8] Yamagishi H, Yoshida S, Kaneko H, Osada M, Kurosaki K, Kasamatsu T, Kawaguchi O, Tsukamoto N, Ando Y, Kubo A. Structuring interpretation of radiographic reports to the unit of description by text-mining technology (in Japanese). Med Imag Tech 2005: 23: 328-332.
[9] Yamagishi H, Osada M, Kurosaki K, Kasamatsu T, Fujii H, Ando Y, Tsukamoto N. Structuring of free text diagnostic imaging (abstract, in Japanese). Jpn J Radiol Tech 2005: 61(abstract book): 169.
[10] Kasamatsu T, Tsukamoto N, Fujii H, Ando Y, Kubo A, Yamagishi H, Kaneko H, Osada M, Kurosaki K. Conversion of free-text diagnostic reports to DICOM-SR (abstract, in Japanese). Nippon Acta Radiol 2005: 65: S283-S284.

[11] Morioka CA, Sinha U, Taira R, el-Saden S, Duckwiler G, Kangarloo H. Structured reporting in neuroradiology. Ann N Y Acad Sci 2002: 980: 259-66.
[12] Anei S, Kawakami Y, and Sasai K. Application of structuring technology to diagnostic imaging database (in Japanese). Konica Minolta Tech Rep 2005: 2: 113-118.
[13] Hussein R, Engelmann U, Schroeter A, Mainzer HP. DICOM structured reporting: Part 1. Overview and characteristics. Radiographics 2004: 24: 891-896.
[14] Hussein R, Engelmann U, Schroeter A, Mainzer HP. DICOM structured reporting: Part 2. Problems and challenges in implementation for PACS workstations. Radiographics 2004: 24: 897-909.
[15] Imai K, Aramaki E, Kajino M, Miyo K, Ohe K. Building a ontology for radiological diagnosis using syntactic analysis and medical attributes (in Japanese with English abstract). Jpn J Med Inform 2005: 25: 395-403.
[16] Anei S, Kawakami Y, Sasai K. Extraction of medical knowledge that used support vector machine (SVM) from diagnostic report of natural sentence (in Japanese with English abstract). Jpn J Med Inform 2005: 25: 405-412.

Address for correspondence

Hirofumi Fujii, M.D., Ph.D.
Functional Imaging Division
Research Center for Innovative Oncology
National Cancer Center Hospital East
6-5-1 Kashiwanoha, Kashiwa 277-8577 Japan
hifujii@east.ncc.go.jp

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Semantic Issues in Integrating Data from Different Models to Achieve Data Interoperability

Rahil Qamar[a], Alan Rector[a]

[a] *Medical Informatics Group, University of Manchester, Manchester, U.K.*

Abstract

Matching clinical data to codes in controlled terminologies is the first step towards achieving standardisation of data for safe and accurate data interoperability. The MoST automated system was used to generate a list of candidate SNOMED CT code mappings. The paper discusses the semantic issues which arose when generating lexical and semantic matches of terms from the archetype model to relevant SNOMED codes. It also discusses some of the solutions that were developed to address the issues. The aim of the paper is to highlight the need to be flexible when integrating data from two separate models. However, the paper also stresses that the context and semantics of the data in either model should be taken into consideration at all times to increase the chances of true positives and reduce the occurrence of false negatives.

Keywords:

medical informatics computing, semantic mapping, term binding, data interoperability

Introduction

The field of medical informatics is growing rapidly and with it informatics solutions to improve the health care process. One of the more important issues gaining the attention of clinical and informatics experts is the need to control the vocabulary used to record patient data. Terminologies covering various aspects of medicine are being developed to bring about some standardisation of data. However, these terminologies are seldom integrated into information systems which are used to capture data at various stages of the health care process.

The paper briefly discusses a middleware application developed to map back-end data model terms to clinical terminologies. However, the main focus of the paper is to highlight the issues encountered when integrating the terminology and data models to achieve the common objective of interoperable patient data through the use of controlled vocabularies.

Archetype models, commonly referred to as archetypes, and the SNOMED CT terminology system were used to test the mapping process. Therefore, the issues discussed are focussed on these two modeling techniques.

Background

Archetype models

In this paper, archetypes refer to the *open*EHR archetypes. Archetypes are being put forward by the *open*EHR organisation as a method for modeling clinical concepts. The models conform to the *open*EHR Reference Model (RM). Archetypes are computable expressions of a domain content model of medical records. The expression is in the form of structured constraint statements, inherited from the RM [1]. The intended purpose of archetypes is to empower clinicians to define the content, semantics and data-entry interfaces of systems independently from the information systems [1]. Archetypes were selected because of their feature to separate the internal model data from formal terminologies. The internal data is assigned local names which can later be bound or mapped to external terminology codes. This feature eliminates the need to make changes to the model whenever the terminology changes.

SNOMED CT terminology

SNOMED-CT, also referred to as SNOMED in the paper, aims to be a comprehensive terminology that provides clinical content and expressivity for clinical documentation and reporting [3]. SNOMED has been developed using the description logic (DL) Ontylog [4] to allow formal representation of the meanings of concepts and their inter-relationship [5]. SNOMED concepts are placed in a subsumption i.e., 'is_a' hierarchy. Two concepts may also be linked to each other in terms of role value maps, and defining or primitive concepts[7]. The SNOMED hierarchy is easy to compute, which was the primary reason for selecting the terminology for the research. The July 2006 release of SNOMED was used for testing the mapping approach. It has approximately 370,000 concepts and 1.5 million triples i.e. relationships of one concept with another in the terminology.

Data integration issues

The aim of the research is to enable health care professionals to capture information in a precise, standardised, and reproducible manner to achieve data interoperability. Data interoperability can be defined as the ability to transfer data to and use data in any conforming system such that the original semantics of the data are retained irrespective of its point of access. Standardised data is critical to

exchanging information accurately among widely distributed and differing users.

Matching clinical data to codes in controlled terminologies is the first step towards achieving standardisation of data. The research focuses on accomplishing the task of performing automated lexical and semantic matches of clinical model data to standard terminology codes using the Model Standardisation using Terminology (MoST) system. A list of candidate SNOMED codes generated by MoST are presented to the modeler who chooses from the list the most relevant codes to bind the archetype term to. However, the matching task is made difficult because of two main reasons. First, the size and complexity of the terminologies makes it difficult to search for semantic matches. Second, the ambiguity of the intended meaning of data in both the data models and terminology systems results in inconsistent matches with terminology codes. The paper will focus on the second issue and suggest some of the solutions that were developed to resolve ambiguity of purpose and use of archetype terms with respect to SNOMED.

In this paper, the term 'modeler' is used to denote a person with clinical knowledge who is engaged in the task of modeling clinical data for use in information systems. Four modelers were used to conduct the study and provide their feedback on the results of the mapping. The second author was also involved in the evaluation process.

Issues with archetype models

Archetypes can be regarded as models of use. Archetypes based on the *open*EHR specification have four main ENTRY types i.e. Evaluation, Instruction, Action, and Observation [12]. For the research, only Observation archetypes were considered as they are the most commonly used model type with the maximum number of examples. However, there is no strict guideline used to categorise archetypes. Also all archetype terms contained in a particular archetype model do not necessarily belong to the same archetype model category. For example, the 'autopsy' archetype[1] belongs to an Observation type. However, the 'cardiovascular system' term contained in the model could belong to either the SNOMED category 'body structure' or 'finding' based on the intended meaning of the modeler. The SNOMED category 'clinical finding' is referred to as 'finding' and 'observable entity' is referred to as 'observable' in the paper.

1) *Determining the semantics of use:* The main issue encountered when looking up matches for archetype terms in SNOMED was the semantics of use. As stated earlier, there are no strict guidelines to categorise either the archetype models or the terms contained in them. The main reason being that archetype models are intended to be used as archetypical representations of a particular clinical scenario. In addition, archetype modelers do not always have SNOMED in mind when modeling clinical scenarios. For example, the recording of the apgar score of a neonate at 1, 5 and 10 minutes from birth will not only require assessment of the breathing, color, reflexes, heart rate, and muscle tone but also the total score calculated at the end of the assessment. Using SNOMED, the assessment terms can be categorised as observables or findings, among other intended semantics, while the 'total' could be a qualifier value or an observable with a value. Therefore, it is important when working with two different models that loose semantics are maintained initially unless stated otherwise. Too much reliance on the categorisation of terms in a particular model will result in fewer matches with terms categorised differently in another model. Strict adherence to categories can only be maintained when working within the same modeling environment as it can be assumed that the hierarchies are logically sound and classifiable.

2) *Determining the source of semantics:* Another issue that was commonly observed was determining the main source of the semantics of an archetype term. For example, in the 'blood film' archetype, the term 'haemoglobin' had a local meaning of 'the mass concentration of haemoglobin', shown in Figure 1. The modeler was questioned whether he would prefer a match for 'haemoglobin' or 'haemoglobin concentration'. The suggestion was that a match for the later term would be closer to its intended meaning. However, based on this suggestion, the assumption of assigning more weightage to the term definition proved incorrect in the 'autopsy' archetype. Conversely based on the above weightage, the term 'cardiovascular system' defined in the model as 'findings of the pericardium, heart and large vessels' should have resulted in 'findings' of the cardiovascular system instead of the 'body structure' itself. However, in this case the modeler stated that the use of the term was intended to serve as a label i.e. 'body structure' rather than the actual values i.e. 'findings' to be entered during data-entry. Therefore, it is advisable not to depend on any single semantic source when looking up matches for an archetype term.

3) *Spelling errors leading to incorrect or no matches:* Finally, there was the issue of resolving spelling errors in the model. For instance, 'haemoglobin' using U.K. English was spelt as 'haemaglobin' in the 'blood film' archetype. Such spelling errors can give rise to incorrect or no results.

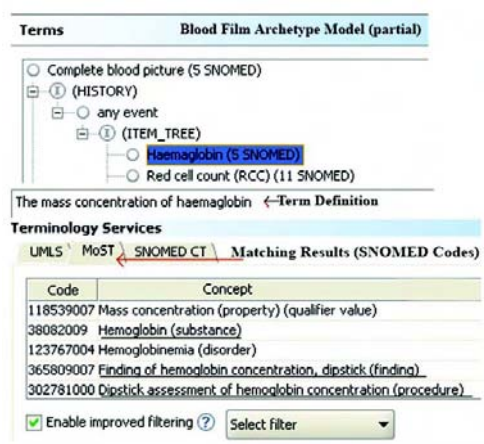


Figure 1 - Haemoglobin-related issues in blood film archetype

1 openEHR archetypes obtained from http://svn.openehr.org/knowledge/archetypes/dev/html/index_en.html

Issues with SNOMED CT

SNOMED can be regarded as a model of meaning. It is a reference terminology for clinical data that provides a common reference point for comparison and aggregation of data about the entire health care process [5]. The multi-axial hierarchy and concept definitions are easily computable to determine the category in which the concept belongs as well as its definition in the model.

1) *Discrepancies in categorisation:* Although SNOMED has a well-defined list of categories; it is not always clear what the basis is for differentiating a concept from being an observable, procedure, or finding. Concepts in the 'observable' hierarchy represent a question or procedure which can produce an answer or a result [6]. On the other hand, concepts in the 'finding' hierarchy represent the result of a clinical observation, assessment or judgement, and include both normal and abnormal clinical states. For example, 'colour of nail' is an observable whereas 'gray nails' is a finding [6]. However, the problem arises when discrepancies occur in the categorisation of certain concepts. For instance, the term 'pregnancy' occurs as a sub type of 'urogenital function' in the observable hierarchy. Clearly, a value cannot be assigned to pregnancy. However, it can have a present/absent or positive/negative value, which would then categorise it as a 'finding'. Such discrepancies in categorisation often lead to problems in correctly interpreting the semantics of a concept.

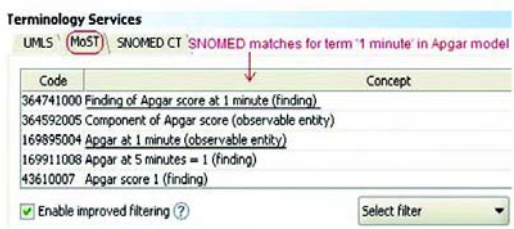


Figure 2 - SNOMED matches obtained for term '1 minute' in the Apgar archetype model

2) *Multiple representations and categorisation of similar concepts:* Another feature of SNOMED is that it allows composition or post coordination i.e. the ability to combine two or more existing concepts in the terminology to represent new meanings [5]. However, this very feature results in multiple representations of the same concept, some of which may contend with the category definitions provided in the SNOMED documentation. For instance, an 'apgar score at 1 minute' can be represented as an 'apgar at 1 minute (observable)' with a value (say 0), or with the help of a pre-existing concept 'apgar at 1 minute = 0 (finding)'. Interestingly, SNOMED has a concept 'apgar at 1 minute', which is a 'finding' as well. Its fully specified name (FSN) is 'finding of apgar score at 1 minute', as shown in Figure 2. Therefore, two concepts with similar names i.e. 'apgar at 1 minute' belong to two different categories i.e. observable and finding. Also, Figure 3 shows that the concept 'finding of apgar score' has a synonym 'observation of apgar score', which leads to concerns about the differentiation between an observation and a finding as stated by the SNOMED community. It may be easy for the human eyes to differentiate between the two and perform either post-coordination or simple semantic mapping. However, it becomes difficult to train a computer application to follow any strict rules for drawing inferences based on the documentation available. A similar example of ambiguity arising from close similarity of observables with findings is the presence of two very similar concepts e.g. 'Finding of reflex hearing response (finding) *is_a* Audiological test finding (finding)' and 'Reflex hearing response (observable entity) *is_a* Audiological test feature (observable entity)' in the July 2006 edition of SNOMED. In terms of aiding computation, it might be helpful to insert disjoint axioms in the hierarchy to enable applications mining through the large corpus of SNOMED data to differentiate between disjoint or non-similar concepts.

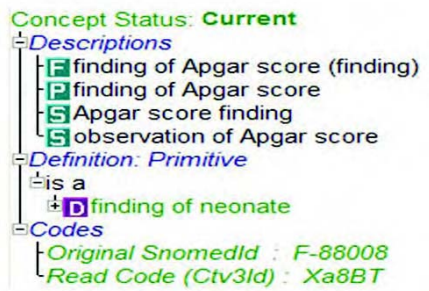


Figure 3 - List of SNOMED terms belonging to the concept 'finding of apgar score' in the July 2006 SNOMED release.

Issues arising from mapping

In the previous two sections we discussed some of the issues present in Archetypes and SNOMED. The issues highlighted in this paper mainly concern disambiguating the semantics of concepts with respect to their categorisation in both models. This issue has been found to be one of the most difficult to resolve when automating the process of finding matches of archetype terms to semantically similar SNOMED concepts[2][10][13].

1) *Grouping SNOMED categories to improve matches:* In relation to finding matches for terms of Observation type archetypes, three main SNOMED categories were identified to which several archetype terms could generally belong to. These categories were 'observable', 'procedure', and 'finding'. The assessment was made by manually inspecting the most common category (ies) to which relevant archetype term matches in SNOMED belonged. The relevance of a match was determined by the modeler. However, this assumption is not intended to mean that matches from other SNOMED categories were irrelevant or not possible. The initial hypothesis had to be modified such that archetype terms from an Observation type archetype model may not necessarily belong to a SNOMED 'observable'. 'Finding' was the next closest category a SNOMED match could belong to followed by 'Procedure'.

2) *Providing detailed SNOMED concept information to make informed decision:* When discussing the issues with archetype models, we exemplified how the archetype term and its definition helped in determining the intended use of

the term. In the 'blood film' archetype, the archetype term 'haemoglobin' found more than one SNOMED match. Among the SNOMED matches obtained, two of the results worth considering were 'hemoglobin' categorised as a 'substance' and 'hemoglobin concentration' categorised as a 'finding' in SNOMED. At first, the modelers on general lexical lookup of the results were of the opinion that the second match was more relevant. However on examining their SNOMED definitions and FSNs, it became obvious that the 'haemoglobin concentration' matches were synonyms for two different FSNs, namely, 'Dipstick assessment of hemoglobin concentration (procedure)', and 'Finding of hemoglobin concentration, dipstick (finding)', as shown in Figure 1. Both the matches were associated with the device 'dipstick', which was not necessarily the intended method to determine the haemoglobin concentration. Hence both the 'heamoglobin concentration' results were later rejected by the modelers as suitable mappings for the archetype term 'haemoglobin'. Therefore, providing the modeler with more information about the SNOMED results rather than a simple, non-informative lexical list helped in reducing incorrect mappings.

3) *Including archetype context:* Archetype models follow an object-oriented style of modeling, which means that each element in the model conforms to some entity or the entity's attribute in the Reference Model. The elements and their values in archetypes are represented in a post coordinated or composite manner. For instance, in the 'apgar score' archetype, the element 'breathing' is constrained by the values 'no effort', 'moderate effort', and 'crying', as shown in Figure 4. In a pre coordinated form the same concepts could be represented as 'no effort breathing', 'moderate effort to breathe', and 'crying or breathing normally'. Therefore, when looking up matches for archetype terms in SNOMED it is necessary to include the context in which the particular term has been used in the model. Well-stated terms can often lead to pre coordinated matches in SNOMED. However, certain terms which are verbose or ambiguous can be unhelpful in finding sensible SNOMED matches despite considering its context of use and stated meaning. Some examples are the terms 'Grimace and cough/sneeze during airways suction' as a constraint value of the element term 'reflex response' in the apgar archetype shown in Figure 4, and 'Needs help but can do about half unaided' as a constraint value of the element term 'Dressing/undressing' in the barthel index archetype. The presence of such terms in archetypes often leads to irrelevant or no matches. However, these searches expend valuable processing time to simplify the term.

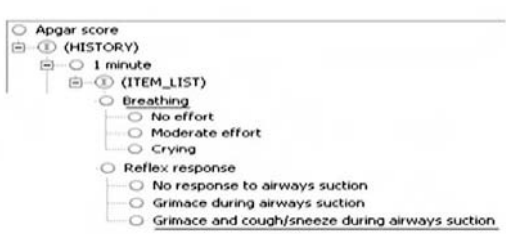


Figure 4 - Apgar score archetype with a post-coordinated 'Breathing' term and verbose 'Reflex response' values

Methodology to address issues

The MoST system was developed to perform automated matches of archetype terms to SNOMED. In the process of obtaining lexical and semantic matches, several of the issues discussed above were addressed and computable solutions were developed.

The entire SNOMED content was imported into a MySQL database and various SQL queries were run at different stages to extract the required information. The Archetype Models were parsed from their local Archetype Definition Language (ADL) format to a simpler XML format which retained only the class containment (element-value pair) hierarchy along with its data types and some other useful information. Language specific information was discarded as it was not required for the matching process. Each term (element or value) was sent to the SNOMED database to look for matches and extract their definitions. Details of the MoST matching process has been discussed in [11].

The spellings of archetype terms were checked against the GSpell spell checker provided by the National Library of Medicine (NLM). However, GSpell performed erratically when encountered with numerics such as 'apgar 1 minute', prepositions, and certain words such as 'width' or more specific archetype data labels such as 'DateTime'. A list of SNOMED and local stop words i.e. words to eliminate from a query terms to prevent too many results or superfluous spell checks, were used to improve results.

MoST utilised already established Natural Language Processing (NLP) systems to perform lexical processing. Initially, the archetype terms were sent to the Emergency Medical Text Processing (EMT-P) service, which processes raw text entries before looking up matches in UMLS [7]. Other NLP techniques used to help in constructing the search queries were word sense disambiguation using GATE [8], and English term synonyms using WordNet [9]. Local techniques applied were removal of stop words enhanced with SNOMED stop words, replacement of numeric and conjunctions with words upon unsuccessful searches, and removal or replacement of special characters and arithmetic notations. The advantage of using a resource like UMLS is that it has a large library of over a million concepts and more than 100 controlled vocabularies and classifications [10]. The semantic groupings of these concepts were used for additional semantic information on the query terms. A training data set was also used to increase the search base by including a list of clinical synonyms generated both locally as well as from the SNOMED July 2006 release.

Resolution of issues

1) *Resolving issues of semantics of use:* In order to address the issue of range of permissible categories to which matches could belong to, rules regarding inclusion of SNOMED categories such as observable, finding, and procedure were introduced. Results that were lexically or semantically similar to the archetype term were returned as results whether or not they had been clearly stated as intended categories by the modeler at the outset. It was also noted that certain other SNOMED categories such as situation (earlier known as context dependent category), qualifier value, attribute, and disorder could also be some

other categories results from which may be worth including in the final result set, if appropriate.

2) *Resolving issues of source of semantics:* The issue of ambiguity of whether the term itself or its definition and context in the model are to be weighted higher was resolved by including all result variants. This may lead to a higher rate of false positives in some cases but will reduce the number of false negatives, which is more important. Therefore, both 'haemoglobin', as well as 'haemoglobin concentration' matches was included as results if it met other filtering criteria, as shown in Figure 1. Similarly, the archetype term 'total' used to record the total apgar score resulted in both SNOMED concepts 'total (qualifier value)', and 'total apgar score (observable)'.

However, certain issues such as training the application about the guidelines to follow in eliminating results form certain SNOMED categories if the terms belonged to Observation archetypes could not be implemented in a strict form. The reason was primarily the ambiguity of term categorisation when working with two separate models that differ in their fundamental objectives. While archetype models include all terms required to record a particular clinical scenario, SNOMED models the logical position of a concept in the given domain and how it relates to other concepts in the hierarchy. All these related concepts may or may not be used by an archetype model to represent a particular clinical record.

Discussion

All the results obtained at the end of the matching process were presented to the modelers who then chose the most relevant i.e. semantically similar SNOMED result code(s) to map the archetype term to. However, not all archetype terms found relevant SNOMED codes. The MoST system thus helped the modeler to codify the local terms to a standardised terminology code. 87.4% of the SNOMED codes were found to be relevant at the end of the first phase of the evaluation involving 300 archetype terms. Details of the MoST system and the evaluation can be obtained from [11], as it is beyond the scope of this paper.

With more and more archetypes being bound to SNOMED codes, the data entry process will become more standardised. Data stored in Electronic Health Records (EHRs) will begin to conform to a single reference terminology. This will increase the ease, speed, and accuracy with which data from the EHRs will be interpreted and used by health organisations irrespective of its place of capture and storage. Interoperable data will help the interoperability of clinical information systems, ultimately leading to the safe use of data and reduction in medical errors originating from incorrect data.

Conclusion

The paper focused on the semantic issues of archetype models and SNOMED, as well as the issues encountered when automating the task of matching terms from the archetype model to SNOMED terminology codes. Although specific models were chosen to test the methodology i.e. archetypes and SNOMED CT, the approach can be tested using other similar models. However, due to the disparities in the objectives and use of the different models, care needs to be taken to address the issues arising from the disparities. Suitable solutions need to be developed to address these issues at two levels of granularity. The first level is the local level where a knowledge layer needs to be added to resolve local issues, which has been discussed in this paper. However, a second, higher level of resolution needs to be adopted when the issues concern the fundamental principles on which the models and their content are based. This will require raising a change request to each of the modeling communities.

It is important to raise issues if any real work on integrating data to achieve quick and accurate data interoperability is to be achieved. This paper hopes to start a discussion on more issues encountered by other such similar integration efforts. The work will be helpful to organisations such as the NHS in the UK who have proposed to use SNOMED CT as the standard medical terminology, and openEHR Archetypes as one of the clinical data modeling techniques.

Acknowledgement

This work is supported in part by the EU Funded Semantic Mining project and the UK MRC CLEF project (G0100852).

References

[1] Beale T, Heard S. Archetype definitions and principles. (Revision 0.6), March 2005.
[2] Bodenreider O, Burgun A, Linking the Gene Ontology to other biological ontologies, Proc. ISMB2005 SIG meeting on Bio-ontologies, 17-18, 2005
[3] Price C, Spackman K. SNOMED clinical terms. BJHC & IM-British Journal of Healthcare Computing & Information Management, 2000, 17(3):27–31.
[4] Spackman K A, Dionne R, Mays E, Weis J. Role grouping as an extension to the description logic of Ontylog, motivated by concept modeling in SNOMED, AMIA Symp 2002, 712-6.
[5] Spackman K, Campbell K, Cote R. SNOMED RT: A Reference Terminology for Health Care
[6] SNOMED Clinical Terms User Guide Jan 2006 Release.
[7] UNC Department of Emergency Medicine. EMT-P User Manual ver 2.1. http://www.med.unc.edu/emergmed/EMTP/ usermanual.pdf. Accessed July 2006.
[8] Cunningham H, Maynard D, Bontcheva K, Tablan V, Ursu C. Developing language processing components with GATE version 3 (a User Guide). University of Sheffield, 2005.
[9] Fellbaum C. WordNet: An Electronic Lexical Database. MIT Press, 1998. M.A.
[10] Sun J, Sun Y. A system for automated lexical mapping. JAMIA, 2005, 334–43.
[11] Qamar R, Rector A. Semantic Mapping of Clinical Model Data to Biomedical Terminologies to Facilitate Data Interoperability, HealthCare Computing Conference 2007.
[12] Beale T, Heard S, Kalra D, Lloyd D (2006), The openEHR EHR Information Model (Revision 5.0).
[13] Friedman C, Shagina L, Lussier Y, Hripcsak G. Automated Encoding of Clinical Documents Based on Natural Language Processing, JAMIA, 392-402, 2004.

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Medication Reconciliation Using Natural Language Processing and Controlled Terminologies

James J. Cimino, Tiffani J. Bright, Jianhua Li

Department of Biomedical Informatics, Columbia University, USA

Abstract

Medication reconciliation (MR) is a process that seeks to assure that the medications a patient is supposed to take are the same as what they are actually taking. We have developed a method in which medication information (consisting of both coded data and narrative text) is extracted from twelve sources from two clinical information systems and assembled into a chronological sequence of medication history, plans, and orders that correspond to periods before, during and after a hospital admission. We use natural language processing, a controlled terminology, and a medication classification system to create matrices that can be used to determine the initiation, changes and discontinuation of medications over time. We applied the process to a set of 17 patient records and successfully abstracted and summarized the medication data. This approach has implications for efforts to improve medication history-taking, order entry, and automated auditing of patient records for quality assurance.

Keywords:

drug therapy, computer-assisted; medication errors;
medication systems, hospital; natural language processing;
terminology; medical order entry systems;
quality assurance, health care

Introduction

The medications a patient takes are not always the medications the patient is *supposed* to take. Errors and drug interactions can occur when patients are treated without full knowledge of the medications that have been previously ordered, especially at transition points such as hospital admissions, transfers and discharge.[1]. For example, Beers and colleagues found that 83% of hospital admission histories failed to record the use of at least one medication the patient claimed to use or recorded at least one medication that the patient denied using; 46% had three or more such errors [2].

The process of creating an accurate list of a patient's medications, for the purposes of resolving discrepancies and supporting accurate medication ordering, is referred to as *medication reconciliation* (MR) [3]. Many processes have been developed for MR, such as manual audits and surveys. With the increasing availability of electronic patient records, such processes can be better integrated into the clinicians' workflow, but they remain largely manual tasks [4,5].

Poon and colleagues provide an excellent review of the problem of MR and describe a pilot test of a system they developed to support creation of a preadmission medication list (PAML) [6]. Their system extracts patient medication information from four different sources and presents it to the clinician, who then creates the PAML by consolidating and reconciling the information from the various sources into a single coherent list. The system assists the user by converting medication terms into, and then grouping them by, their generic names. The PAML was successfully used at the time of hospitalization to guide admission orders and at the time of discharge to assure that previous outpatient medications, where appropriate, were continued.

The PAML used data from multiple sources, each of which encoded medication data using a different controlled terminology. One of the challenges encountered by Poon and colleagues was the reconciliation of these different terminologies. The PAML did not use narrative text sources (such as discharge summaries or clinic notes) but had it done so, they would have also had to reconcile the free-text terms found in those reports.

In this paper, we describe an approach to automatically analyze medication information from a mixture of coded and narrative text sources, and use controlled terminology to support reasoning about MR. We describe the application of our method to a set of 17 patient records and discuss ways in which the results might be exploited for automated support for medication reconciliation in a variety of clinical situations.

Materials and methods

Our approach to MR involves the collection of patient medication data from multiple coded and narrative text sources, conversion of all data into coded form (using natural language processing where necessary), and then obtaining classification information for each medication. Patient medications can then be viewed over time, grouped by class, and organized into a matrix that can be used to identify points at which medications were introduced, changed or removed from the patients' ordered medications.

Sources of patient medication information

New York Presbyterian Hospital (NYPH) makes use of two major clinical information systems: the commercial product Eclipsys XA order entry system (Eclipsys, Boca Raton, FL) and the internally developed WebCIS documentation and repository system [7]. Eclipsys allows clinicians to order medications that are initiated and discontinued at various points during a hospitalization. Clinicians also enter medication information in narrative text, as part of the patient's discharge instructions. WebCIS provides narrative medication lists as parts of clinic notes, admission notes and discharge summaries, as well as coded medications from the outpatient medication order entry system and inpatient pharmacy system.[1] Together, the sources provide medication lists that correspond to twelve points in time related to hospital admissions, as shown in Table 1.

Table 1 – Data sources for patient medication information

Data Source	System	Data Type
Prior Clinic Note	WebCIS	Narrative
Prior Outpatient Medications	WebCIS	Coded
Admission Note[2]	WebCIS	Narrative
Admission Note Plan[2]	WebCIS	Narrative
Admission Orders	Eclipsys	Coded
Admission Pharmacy Orders	WebCIS	Coded
Active Orders at Discharge	Eclipsys	Coded
Discharge Pharmacy Orders	WebCIS	Coded
Discharge Instructions	Eclipsys	Narrative
Discharge Plan	WebCIS	Narrative
Clinic Note After Discharge	WebCIS	Narrative
Outpatient Medications after Discharge	WebCIS	Coded

Natural language processing

As noted in Table 1, the medication lists from clinic notes, admission notes, discharge summaries and discharge instructions are in narrative form. In order to convert them to coded form, we use a natural language processing system called Medical Language Extraction and Encoding system (MedLEE) [8] to parse the reports, and identify medication terms. Where possible, MedLEE provides Concept Unique Identifiers (CUIs) from the US National Library of Medicine's Unified Medical Language System (UMLS) [9].

Medication classification

All coded information obtained from Eclipsys (medication orders) and WebCIS (outpatient medication orders and inpatient pharmacy orders) are coded using our internally developed Medical Entities Dictionary (MED) [10]. The MED includes UMLS CUIs for many of its terms. Thus, we were able to obtain MED Codes for all coded data and, through UMLS mappings, for narrative data for which MedLEE provided UMLS CUIs.

Medication terms in the MED are organized into a hierarchy based on the American Hospital Formulary Service (AHFS) Codes, which classify drugs according to function and purpose [11]. For each medication term (including MedLEE-coded terms), we obtained the AHFS class code or codes, based on its location or locations in the MED hierarchy.

Evaluation

We chose a convenience sample of records for patients followed by one of us (JJC) in the NYPH medical clinic. For each patient, we identified the most recent hospital admission for which a discharge summary and at least one clinic note were available. We also obtained, where available, all WebCIS and Eclipsys medication data (as described above) that were recorded prior to, during, and after the hospitalization. Data were coded with AHFS codes, as described above. The data for each admission were organized into a matrix in which each row corresponded to a data source and each column corresponded to an AHFS code. The success of capturing, coding and organizing the data were measured at each step in the process.

Results

Patient data

A total of 70 patient records were reviewed and 30 hospitalizations were identified. Thirteen hospitalizations were excluded because they lacked a discharge summary and/or at least one clinic note prior to or following the hospitalization. Data from the remaining 17 hospitalizations were extracted from WebCIS and Eclipsys; narrative text was processed with MedLEE.

In general, MedLEE was successful at identifying medication terms in the narrative text. Manual review identified approximately 30 instances where MedLEE missed medication terms in the text. It also occasionally extracted nonmedication terms, such as "medication" and "po". MedLEE did not distinguish between medications that were being ordered and those that were being discontinued.

The twelve sources contributed data to a case an average of 12.7 times (range: 7-17), with a total of 1563 medications found (91.9 per case; range: 25-159). Table 2 shows the

1 Medication orders from Eclipsys are transmitted to the pharmacy system, converted into dispensing orders and transmitted to WebCIS.

2 In cases were admission notes were not available, admitting history was obtained (if present) from the discharge summary

numbers of medications obtained from each source for all 17 patients.

Table 2 – Data obtained from 17 patient records

Data Source	Meds	Records w/Data	Meds per Patient
Prior Clinic Note	157	17	9.2
Prior Outpatient Medications	211	13	16.2
Admission Note	102	14	7.3
Admission Note Plan	41	12	3.4
Admission Orders	88	8	11.0
Admission Pharmacy Orders	152	14	10.9
Active Orders at Discharge	93	8	11.6
Discharge Pharmacy Orders	171	14	12.2
Discharge Instructions	60	7	8.6
Discharge Plan	123	16	7.7
Clinic Note After Discharge	140	16	8.8
Outpatient Medications after Discharge	225	13	17.3

Coding medication data

Of the 623 terms it identified in the narrative text sections, MedLEE provided UMLS CUIs for 545. Of the remaining 78 medications, 30 were instances of five non-medication terms ("antiinflammatory", "cream", "lotion", "lozenge", and "po"). The other 48 were instances of eight medication terms ("INH", "MVI", "Os-Cal", "asa", "darvocet", "hctz", "niacin", and "toprol") for which UMLS CUIs were readily identified using the UMLS file MRCONSO.

Thus, UMLS CUIs were obtained for 593 of the MedLEE-identified terms, representing 169 unique terms and 165 unique CUIs. Eighty-five of these CUIs (representing 359 term instances) were found in the MED. The remaining 80 CUIs (representing 234 term instances) were mapped manually to the MED.

The coded data and the MedLEE-extracted data together comprised 1563 terms, of which 1533 were coded in the MED (444 unique terms) and the remaining 30 consisted of the instances of the 5 non-medication terms identified by MedLEE. Of the 1533 MED-coded terms, AHFS codes were available for 1517 (442 unique terms). The remaining 16 terms were instances of "oxygen" and "medication" that, while technically medication terms, do not have AHFS codes. Due to the multiple hierarchy of the MED, and multiple ingredients and/or uses of the medications, 270 terms (85 unique terms) mapped to two or more (maximum five) AHFS codes.

Creating medication matrices

The patient data for each case were grouped into an average of 22.8 AHFS codes (range: 10-37), and thus, given the twelve data sources, the matrices had an average of 273 cells (range: 120-444). Not every data source had data for every AHFS code; cases had data in 88 cells, on average (range: 28-164), with an average of 1.26 terms per cell (range: 1.09-1.55). An abbreviated example of a matrix is shown in Figure 1. The full matrix contains the original data from each source (row), along with the MedLEE abstraction, if appropriate, along with the 10 to 37 AHFS class columns.

Patient #9	201204: Anticoag-ulants	240400: Cardiac Drugs	240800: Hypoten-sive Agents	280000: Central Nervous System Agents	281604: Antidep-ressants
Prior Clinic Note	coumadin	verapamil	cozaar		cymbalta
Prior Outpatient Medications	Coumadin 5 mg Tab	Verapamil 180 mg Extended Release Tablet	Losartan Potassium 100 mg Tablet	Pregabalin 50mg Capsule	
Admission Note	coumadin	verapamil	cozaar		cymbalta
Admission Note Plan	coumadin				
Admission Orders	Warfarin Sodium Oral 10 MG	Verapamil SR Oral 240 MG	Losartan Oral 50 MG		
Admission Pharmacy Orders	WARFARIN TAB 5 MG 10 MILLIGRAM	VERAPAMIL SR TAB 240 MG	LOSARTAN POTAS-SIUM TAB 50 MG		
Active Orders at Discharge		Verapamil SR Oral 240 MG	Losartan Oral 50 MG		
Discharge Pharmacy Orders		VERAPAMIL SR TAB 240 MG	LOSARTAN POTAS-SIUM TAB 50 MG		DULOX-ETINE CAP 20 MG
Discharge Instructions					cymbalta
Discharge Plan					cymbalta
Clinic Note After Discharge	coumadin	verapamil			cymbalta
Outpatient Medications after Discharge	Coumadin 5 mg Tab	Verapamil 180 mg Extended Release Tablet	Losartan Potassium 100 mg Tablet	Pregabalin 50mg Capsule	

Figure 1 - Sample medication reconciliation matrix

The abbreviated view of the matrix in Figure 1 illustrates some of the information that can be obtained from these medication summaries. In particular, it shows instances where medications, such as Pregabalin (a Central Nervous System Agent) and Cymbalta (an Antidepressant), are listed in the outpatient note or ordering system prior to admission, but then do not appear in the admission note or orders. Cymbalta, at least, eventually appears in the discharge instructions and plan, but the Pregabalin does not, although it continues to be present in the outpatient medication system, raising the issue of whether it should be discontinued.

In addition to the appearance and disappearance of medications, the dosage of medications changes over the

course of care. In this example, we see admitting orders for Coumadin (an Anticoagulant), Verapamil (a Cardiac Drug), and Cozaar (a Hypotensive Agent) variously increasing or decreasing, relative to the preadmission orders, only to return to their original doses after discharge. The discharge plans and discharge medication orders do not address these medications at all.

Discussion

We have successfully extracted patient medication information from a variety of sources and organized it using a standard drug classification system to support a chronological summarization by medication class. This approach is similar to that carried out by Poon and colleagues as part of their PAML application [6] but differs in two important ways: medication class and chronology.

When the medications from a particular source were grouped into an AHFS class, we tended to find approximately one drug per class. This makes clinical sense, since few patients take more than one anticoagulant or antidepressant at one time. These small groupings may help support rapid assessment of the data to detect possible problems. Furthermore, the grouping by class (as opposed to Poon and colleagues' grouping by generic ingredient) accommodates changes in medication within a class and treats them as potentially acceptable. So, for example, the fact that a patient was taking one diuretic prior to admission, a different diuretic during the admission, and returned to the previous one after discharge may reflect an intentional change (perhaps due to restrictions in the hospital formulary).

The chronological arrangement of the information in the matrices acknowledges the fact that medication lists are created and maintained at points in time and that a list that was created a year ago may not be as valid as one created a week ago. Unlike Poon and colleagues, we have not yet created a medication reconciliation application, but we believe that when patient data are displayed in such an application, the relative age of the data will be relevant to those trying to resolve differences in medication lists.

Before the present methods can be employed in an MR application, some improvements to MedLEE and the MED will be needed. Some improvements in MedLEE parsing will be needed, particularly to allow it to differentiate between mentions of medications that are being continued and those that are being discontinued. MedLEE did extremely well at providing UMLS CUIs for the terms it did find - it missed only 8 unique terms out of 172 legitimate unique medication terms, for a 95% success rate. Our experience suggests that only a small effort will be needed to improve this rate.

The automated translation of UMLS CUIs to MED Codes was less successful, at slightly over 50%. This result is not due to the methodology but rather to the heretofore-low incentive for having UMLS CUIs in the MED; the mappings have not been updated for some time. However, we are confident (based on the success of our manual mapping) that the MED can be easily brought up to date and will perform this mapping function more successfully in the future.

Finally, while we believe that the matrices provide a useful organization of the data, they are probably not adequate for use directly by clinicians. Some applications, such as the PAML, will be needed to present the information to users in a way that reduces the cognitive overload that might otherwise be produced by a 164-cell matrix.

Alternatively, a program that attempts to identify discontinuities in patient medications, in order to alert clinicians at appropriate times, might use the information in the matrices. For example, when a patient is admitted and a previous outpatient medication is not ordered, the clinician might be asked, "The patient was previously taking X; do you wish to continue it?" The clinician might have a perfectly good reason for not continuing the drug, but inadvertently overlooking such information is unfortunately frequent [1]. Later, when the patient is discharged, the commonly used shorthand "continue all previous medications" is ambiguous - was X deliberately stopped or not? And, if deliberately stopped, is reinstatement really desired? A system that can track these discrepancies has great potential to reduce medication errors.

The approach described in this paper cannot, by itself, determine if a medication discrepancy represents a true error in medication ordering. Only discussion with the clinicians caring for the patient and/or chart review can identify the reasons for medication changes. However, if an ordering clinician overlooks a prior medication, or adds a previously undocumented one, it will be identified in our matrix because, at least for inpatient medications, the Eclipsys data are the gold-standard for the patient's intended therapy. Such identification is a necessary first step in preventing or correcting errors in medication reconciliation.

Conclusions

We have successfully extracted patient medication information from multiple systems and applications, and classified the information based on drug class. The resulting matrices of medications organized chronologically by class provide a new way to summarize such information and have the potential to support automated medication reconciliation.

Acknowledgments

The authors thank Dr. Carol Friedman for the use of MedLEE. This work was supported in part by National Library of Medicine Training Grant LM07079-11 and by National Library of Medicine Grant 5R01LM007593-05.

References

[1] Gleason, KM, Groszek JM, Sullivan C. Reconciliation of Discrepancies in Medication Histories and Admission Orders of Newly Hospitalized Patients. *Am J Health-Syst Pharm*. 2004;61:1689-95.

[2] Beers MH, Munekata M, and Storrie M. The accuracy of medication histories in the hospital medical records of elderly persons. *J Am Geriatr Soc.* 1990 Nov;38(11):1183-7.

[3] Institute of Healthcare Improvement. Medication Reconciliation Review. http://www.ihi.org/IHI/Topics/PatientSafety/MedicationSystems/Tools/Medication+Reconciliation+Review.htm

[4] Pronovost P, Weast B, Schwarz M, Wyskiel RM, Prow D, Milanovich SN, Berenholtz S, Dorman T, and Lipsett P. Medication reconciliation: a practical tool to reduce the risk of medication errors. *J Crit Care*. 2003 Dec;18(4):201-5.

[5] Poole DL, Chainakul JN, Pearson M et al. JHQ 177 Medication reconciliation: A necessity in promoting a safe hospital discharge. *NAHQ*. 2006 May-June 2006 www.nahq.org/journal/ce/article.html?article_id=256

[6] Poon EG, Blumenfeld B, Hamann C, Turchin A, Graydon-Baker E, McCarthy PC, Poikonen J, Mar P, Schnipper JL, Hallisey RK, Smith S, McCormack C, Paterno M, Coley CM, Karson A, Chueh HC, Van Putten C, Millar SG, Clapp M, Bhan I, Meyer GS, Gandhi TK, and Broverman CA. Design and implementation of an application and associated services to support interdisciplinary medication reconciliation efforts at an integrated healthcare delivery network. *J Am Med Inform Assoc*. 2006 Nov-Dec; 13(6):581-92.

[7] Hripcsak G, Cimino JJ, and Sengupta S. WebCIS: large scale deployment of a Web-based clinical information system. *Proc AMIA Symp*. 1999;:804-8.

[8] Friedman C, Shagina L, Lussier Y, Hripcsak G. Automated encoding of clinical documents based on natural language processing. *J Am Med Inform Assoc*. 2004 Sep-Oct;11(5):392-402.

[9] Lindberg DA, Humphreys BL, McCray AT. The Unified Medical Language System. *Methods Inf Med*. 1993 Aug;32(4):281-91.

[10] Cimino JJ. From data to knowledge through concept-oriented terminologies: experience with the Medical Entities Dictionary. *J Am Med Inform Assoc*. 2000 May-Jun;7(3):288-97.

[11] American Society of Health-System Pharmacists. *AHFS Drug Information*. Bethesda, MD: American Society of Health-System Pharmacists, 2006.

Address for correspondence:

James J. Cimino, M.D.
Department of Biomedical Informatics
Columbia University College of Physicians and Surgeons
622 West 168th Street, Vanderbilt Clinic 5th Floor
New York, New York 10032 USA
jjc7@columbia.edu

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Comparing Medical Code Usage With the Compression-Based Dissimilarity Measure

Thomas Brox Røst[a], Ole Edsberg[a], Anders Grimsmo[b], Øystein Nytrø[a]

[a] *Department of Computer and Information Science, Norwegian University of Science and Technology, Norway*
[b] *Department of Community Medicine and General Practice, Norwegian University of Science and Technology, Norway*

Abstract

It is well known that medical coding practice is inconsistent and that differences in usage may exist even at the institutional level. In this paper we introduce a novel method for investigating code usage patterns in clinical documentation corpora. By applying the Compression-based Dissimilarity Measure to calculate similarities between encounter notes, we find that certain notes can be associated with a number of different classifications and that a given classification code can be documented in fundamentally different ways. The effect is that some notes need to be understood in the context of the classification code, a finding which has implications for data mining or information extraction tasks. In addition, the method opens for a number of interesting application areas that include highlighting code use anomalies, measuring how coding practice changes over time, comparing code usage across institutions, and, perhaps most importantly, provide valuable feedback to developers of classification coding systems.

Keywords:

classification, ICPC, Kolmogorov complexity.

Introduction

The electronic patient record (EPR) is the main tool for recording, communicating and remembering information about healthcare activities. Current electronic patient records exist in a multitude of more or less structured formats, and a major share of the useful information is in the form of unstructured free-text.

For the general practice EPR, which is our main concern in this paper, the unstructured free-text comes in the shape of a sequence of encounter notes that, when taken together, constitute the patient history. Alongside the encounter notes are different types of associated information such as prescriptions, referrals, laboratory results, and, in many cases, one or more diagnosis or classification codes. Classification codes may have several purposes: The underlying reason for the encounter; the patient's health problems; the administered treatment; or simply a diagnosis of an identified disease. On the administrative and mercantile level, coding is used for tasks such as reimbursement, statistics, epidemiology, insurance claims, and resource management.

An example of a classification code is the International Classification of Primary Care (ICPC), which was designed for the particular needs of primary care [1]. An early study of ICPC usage found that GPs in general do "accurate and careful work in coding" and that "data appear to be of acceptable quality for further analysis" [2]. Later studies do, however, differ in their conclusions. Nilsson et al. [3] find a high level of completeness and correctness of diagnostic codes. Other studies for both primary [4, 5] and specialist [6-8] care calls the reliability and validity of coded data into question, though some of these studies cover coding schemes other than ICPC. Accuracy appears to be domain independent [9].

ICPC is unique in the sense that it is based on epidemiological practice from several different countries, in contrast to other expert consensus-based medical classification systems. Still, there are many known unknowns when it comes to how both ICPC and other classification systems are used in clinical practice, as highlighted by the aforementioned studies. Even though coding systems are coupled with guidelines that outline their intended usage, these may in themselves be ambiguous and open to interpretation [10]. Clinical practice is local by nature; coding practice may be equally so. Without ways of comparing and evaluating code usage, the validity of any application or analysis where coding is a mainstay may be called into question.

As a means towards investigating code usage, we have chosen to consider how the free-text content of the encounter note relates to the associated classification code. Our approach is to group encounter notes according to their textual similarity. We have two main reasons for doing so:

- To learn if we can identify cases where similar types of encounter notes can have different classification codes. In such cases, the encounter note can not be clearly differentiated from other encounter notes and can therefore not be understood without the extra information that is provided by the code.
- Conversely, to learn if we can identify cases where a single classification code is documented in different ways.

The outcome will be a measure of the relationship between text and coding which, while simple, may be extended for comparative purposes as outlined above. A secondary motivation for this approach is to follow up on earlier attempts at automated ICPC coding of general practice encounter notes [11], which indicated that a more thorough examination of the relationship between encounter note and classification code was needed.

Materials and methods

We have restricted our selection of data to patient records from general practice encoded with ICPC. The primary reason is data availability; the secondary reason is the aforementioned high quality of ICPC coding in primary care. To evaluate our method, we rely on a dataset from a medium-sized Norwegian general practice care center. The full dataset consists of the complete EPR database for the period from 1992 to 2006. In order to ensure a somewhat homogenous sample we have only considered data from the last five years. This is in part to avoid changes in coding practice that might occur over time, but also because the last major system upgrade was in 2001—this upgrade led to slightly different documentation practices that might influence the result.

Of the data available to us, we have focused exclusively on the free-text clinical encounter notes and their associated ICPC codes. Notes with more than one code were not considered, this in order to ensure a one-to-one correspondence between the encounter note and its code.

Since the data originates from a general practice care center, notes will be written by a number of authors that may differ in the way they express themselves. In particular, the primary language of the documenting physician may be in one of three slightly different languages: Standard Norwegian ("bokmål"), New Norwegian ("nynorsk") and, in a minority of cases, Danish. Syntactically, all these languages are very similar and should not confuse our choice of similarity metric too much. All in all, there are more than 20 different authors of the source material. We chose not to let authorship influence our data selection, this in order to get a broad representation of different documentation styles.

The compression-based dissimilarity measure

In order to measure the similarity between notes we have based ourselves on the Universal Similarity Metric (USM), as defined by Li et al. [12, 13]. The USM is founded on the concept of Kolmogorov complexity, or algorithmic entropy. The Kolmogorov complexity $K(x)$ of a string x is defined as the length of the shortest program for a universal Turing machine U needed to compute x [14]:

$$K(x) = \min\left\{|P|, P \text{ a program and } U(P) = x\right\} \quad (1)$$

The USM is defined as:

$$d(x,y) = \frac{\max\{K(x \mid y^*), K(y \mid x^*)\}}{\max\{K(x), K(y)\}}, \quad (2)$$

where x^* and y^* denote the shortest programs required to produce x and y. $K(x \mid y^*)$ is thus the length of the shortest program to produce x with y^* as an input. Kolmogorov complexity can be shown to be non-computable but only upper semi-computable; $K(\cdot)$ will therefore have to be estimated. To make this estimate we have employed a methodology [15], similar to [13] and [16], where $K(x)$ is approximated by $C(x)$, the size of the compressed version of x:

$$C(x) \approx \left|\text{gzip}(x)\right| \quad (3)$$

In other words: $C(x)$ is the result of applying the basic gzip compression algorithm to the encounter note that is x. The distance measure between two notes x and y can then be defined as follows [15]:

$$d_c(x,y) = \frac{C(x \mid y) + C(y \mid x)}{C(xy)} \quad (4)$$

Here, xy is the concatenation of notes x and y. $C(x \mid y)$ is the compression when first training on y and then compressing x. Based on the results from [15], we chose to use the further simplified Compression-based Dissimilarity Measure (DSM) for this paper:

$$CDM(x,y) = \frac{C(xy)}{C(x) + C(y)} \quad (5)$$

In this case, the dissimilarity is close to 1 when x and y are unrelated and smaller than 1 if x and y are related. Smaller values indicate closer relatedness. We refer to the cited papers for further proofs of these relationships.

Hierarchical clustering

The CDM can be fed directly into many standard clustering algorithms. For this experiment, we use hierarchical clustering. Clustering is a way of partitioning data into subsets, or clusters, so that the clusters have something in common—in this case the syntactical similarity of the encounter notes. With hierarchical agglomerative (bottom-up) clustering [17], clusters are gradually merged into successively larger clusters. The result can be visualized as a tree data structure known as a dendrogram.

Results

Experiments were run with datasets of respectively 50, 100, 200 and 1.000 randomly sampled notes for each code—that is, for a code to be included in the experiment it had to have at least the given number of occurrences. The experiment with 200 notes seemed to be the best compromise between code coverage and data set size — for the 1.000 note experiment, only a relatively small number of codes had more than 1.000 notes available in the data set. Accordingly, only the result from the 200 note experiment is discussed in this paper.

Figure 1 shows the results of clustering the available encounter notes in the form of a dendrogram where codes with similar encounter notes are grouped together. Codes on the left are the most similar. The major share of the

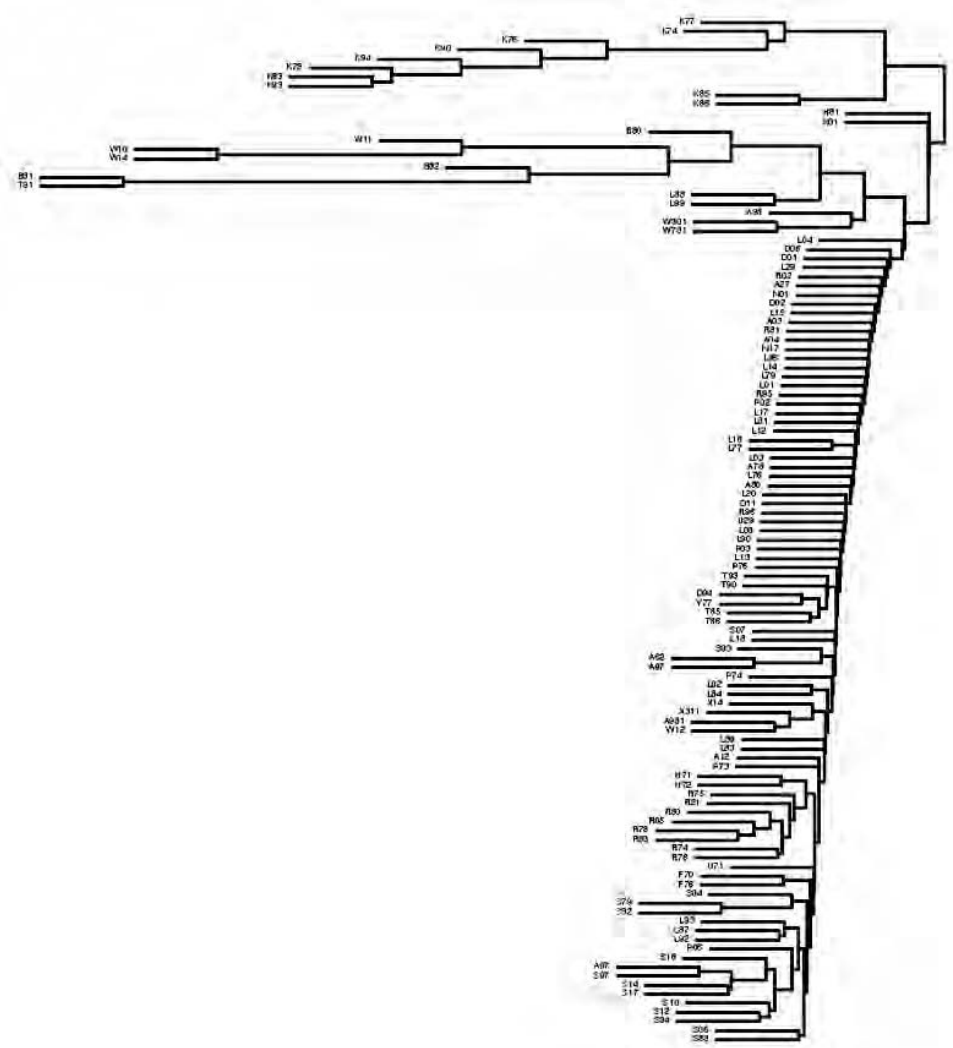


Figure 1 - Code similarity dendrogram

codes will, as seen in the long tail towards the bottom, not have much in common with other codes, signifying that there typically is something unique about the documentation for that code. The two longer branches towards the top of the graph indicate clusters of codes that are documented in a somewhat similar manner. We also notice smaller clusters towards the bottom.

Table 1 is a selection of the ICPC codes found in the most prominent clusters, based on manual inspection. The rightmost column describes the most typical text occurrence in notes with the given code.

Discussion

Referring to the manual review in Table 1, a common factor of notes being found similar would appear to be that they describe some sort of treatment that is used along with several different ICPC codes. The most prominent example is for the codes K74, K76, K77, K78, K83, K90, K93 and K94, all of which deal with different cardiovascular-related states. The combining element is, apparently, the use of the drug Marevan as an essential part of the treatment plan. A typical note will therefore describe the prescription of Marevan and not much else. There are, however, exceptions to this pattern. Take the code K83 as an example. A typical K83 note might look like this:

> Marevan 2 tbl. 7/7 Ctr. 2 weeks.[1]

Then again, a minority of the notes represent a complete break with the pattern, such as this example:

> Stressed before Christmas, occassional dizziness. 185/85 p70 reg. Cor-sbgr 4, pm 2hicr, pulm-. Pat does not want med.tmnt., promises to ”calm down”.[1]

Through manual review and counting, 20.5 % of the K83 notes used in the experiment fall into the latter atypical

1 Approximate translation from the Norwegian.

category. Thus we observe that there are internal differences in the way codes are documented. In this case, a possible explanation is that the atypical notes represent treatment milestones, such as first consultations for an emerging problem, occasions when the treatment does not go according to plan, or, simply, wherever and whenever the physician feels the need for more thorough documentation.

A similar pattern is seen for the other code clusters. Most of the clusters have the ICPC chapter code in common, but there are a few exceptions to this, such as for B81 (Anaemia, Vitamin B12/folate deficiency) and T91 (Vitamin/nutritional deficiency), where in both cases Betolvex or Vitamin B-12 is used to treat the condition. Again, the common denominator is the choice of treatment.

Table 1 - Selection of most similar codes

Code	Code description	Summary of manual note review
A97	No disease	Driving license attestation
A87	Complication of medical treatment	Wound dressing
B80	Iron deficiency anaemia	Cosmofer/Jectofer/ Vitamin B-12
B81	Anaemia, Vitamin B12/ folate def.	Betolvex/Vitamin B-12
B82	Anaemia other/unspecified	Marevan/Betolvex/ Vitamin B12
K74	Ischaemic heart disease w. angina	Marevan
K76	Ischaemic heart disease w/o angina	Marevan
K77	Heart failure	Marevan
K78	Atrial fibrillation/flutter	Marevan
K83	Heart valve disease	Marevan
K90	Stroke/cerebrovascular accident	Marevan
K93	Pulmonary embolism	Marevan
K94	Phlebitis/thrombophlebitis	Marevan
L88	Rheumatoid/seropositive arthritis	Methotrexate
L99	Musculoskeletal disease, other	Methotrexate
R78	Acute bronchitis/ bronchiolotis	Cold, fever, cough
R83	Respiratory infection other	Cold, fever, cough
S14	Burn/scald	Burns/scalding
S17	Abrasion/scratch/blister	Wound dressing/ suture removal
S79	Neoplasm skin benign/ unspecified	Mole/suture removal
S82	Naevus/mole	Mole/suture removal
S97	Chronic ulcer skin	Wound dressing
T91	Vitamin/nutritional deficiency	Betolvex/Vitamin B-12
W10	Contraception postcoital	Depo-Provera
W11	Contraception oral	Depo-Provera
W14	Contraception other	Depo-Provera

Conclusion

Through the use of the Compression-based Dissimilarity Measure in conjunction with hierarchical clustering we have identified cases where 1) the content of an encounter note can be the same for a number of different classification codes, and where 2) a single classification code can be documented in a lot of different ways. In other words, there is not always a one-to-one relationship between the classification code and its encounter note, but rather one-to-many relationships between certain codes and encounter notes, and vice versa.

One of the main implications is that encounter notes can not always be understood without their classification code, and that classification codes are sometimes only fully understood in terms of their associated encounter notes. Moreover, a classification code does not imply that the associated encounter note carries any useful information beyond the trivial. Accordingly, for a given classification code one might have to consider earlier encounter notes with the same (or a similar) code in order to find all relevant information for this code. In effect, certain encounter note/classification code pairs need to be considered in the context of their encompassing episode of care to be of use. This has implications for any kind of information extraction or data mining that is applied on primary care patient records.

Limitations

We have not attempted to run the experiment with a different clustering algorithm or a different similarity measure. While this may influence the results, our main purpose has

been to verify the existence of ambiguous codes and encounter notes.

Our view of the patient record is somewhat simplified. We have not considered the additional information that often comes with each encounter, such as lab results and prescriptions. To compare full encounters rather than notes alone, a similarity measure that considers all available information might produce different results.

Future work

In light of finding intra-code documentational variance, a natural follow-up would be to find and utilize internal clusters for single codes—that is, to identify documentational archetypes or patterns for each code and to compare these with patterns for all other codes. This might grant us a clearer picture of the different types of documentation that occurs in the primary care patient record.

The most interesting aspect of the proposed method lies in a number of potential application areas for evaluation of various healthcare documentation activities. In particular, we believe this method can prove very useful as a foundation for doing comparative studies of medical coding practice. By comparing the outcome of similarity measurements when applying the same method on different data sets, any differences may highlight variations in coding practice. In turn, this may prove to be a very useful input to those who create and maintain coding standards. Moreover, coding standards are often translated to other languages. By comparing code usage between data sets in different languages one might identify possible interpretational problems that have occurred during translation as well as differences in health care practice between countries. Other interesting approaches would be to evaluate how coding practices change over time and to compare coding practices between people and institutions. While we have only performed experiments on a data set that uses ICPC, the method should in theory be equally applicable for data sets that follow other coding standards.

Acknowledgements

Laura Slaughter, Arild Faxvaag, Inger Dybdahl Sørby and Amund Tveit all provided helpful comments, suggestions and advice.

References

[1] WONCA International. ICPC-2: International Classification of Primary Care. 2 ed: Oxford Medical Publications 1998.

[2] Brage S, Bentsen BG, Bjerkedal T, Nygård JF, Tellnes G. ICPC as a standard classification in Norway. Fam Pract 1996;13(4):391-6.

[3] Nilsson G, Ahlfeldt H, Strender LE. Textual content, health problems and diagnostic codes in electronic patient records in general practice. Scand J Prim Health Care 2003 Mar;21(1):33-6.

[4] Britt H, Angelis M, Harris E. The reliability and validity of doctor-recorded morbidity data in active data collection systems. Scand J Prim Health Care 1998;16(1):50-5.

[5] Nilsson G, Petersson H, Åhlfeldt H, Strender L-E. Evaluation of Three Swedish ICD-10 Primary Care Versions: Reliability and Ease of Use in Diagnostic Coding. Methods of Information in Medicine 2000;39(4/5):325-31.

[6] Campbell SE, Campbell MK, Grimshaw JM, Walker AE. A systematic review of discharge coding accuracy. Journal of Public Health Medicine 2001;23(3):205-11.

[7] Hoddevik GH. Diagnosis versus code. Tidskr Nor Lægeforen. 2005 Nov 3 2005;125(21):2973-4.

[8] Dixon J, Sanderson C, Elliott P, Walls P, Jones J, Petticrew M. Assessment of the reproducibility of clinical coding in routinely collected hospital activitydata: a study in two hospitals. Journal of Public Health 1998;20(1):63-9.

[9] Surján G. Questions on validity of International Classification of Diseases-coded diagnoses. Int J Med Inform. 1999;54(2):77-95.

[10] Ceusters W, Smith B, De Moor G. Ontology-Based Integration of Medical Coding Systems and Electronic Patient Records. 9th international congress of the European Federation for Medical Informatics (MIE); 2005; 2005.

[11] Røst TB, Nytrø Ø, Grimsmo A. Investigating the Value of Diagnosis Codes in the Primary Care Patient Record. In: Peek N, Combi C, editors. Proceedings of IDAMAP 06 Intelligent Data Analysis in Biomedicine and Pharmacology; 2006; Verona, Italy; 2006. p. 93-8.

[12] Li M, Badger JH, Chen X, Kwong S, Kearney P, Zhang H. An information-based sequence distance and its application to whole mitochondrial genome phylogeny. Bioinformatics 2001;17(2):149-54.

[13] Li M, Chen X, Li X, Ma B, Vitányi P. The Similarity Metric. The Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms; 2003; Baltimore, Maryland: Society for Industrial and Applied Mathematics; 2003. pp. 863-72.

[14] Li M, Vitányi P. An Introduction to Kolmogorov Complexity and its Applications. 2nd Edition ed. New York: Springer-Verlag 1997.

[15] Keogh E, Lonardi S, Ratanamahatana CA. Towards parameter-free data mining. Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining; 2004; Seattle, WA, USA: ACM Press; 2004. pp. 206-15.

[16] Krasnogor N, Pelta D. Measuring the similarity of protein structures by means of the universal similarity metric. Bioinformatics 2004;20(7):1015-21.

[17] Jain A, Murty M, Flynn P. Data Clustering: A Review. ACM Computing Surveys 1999;31(3):264-323.

Address for correspondence

Thomas Brox Røst
Norwegian University of Science and Technology
Department of Computer and Information Science
Sem Sælands vei 7-9
NO-7491 Trondheim, Norway
Email: brox@idi.ntnu.no

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

A Scale-Free Network View of the UMLS to Learn Terminology Translations

Chintan O. Patel, James J. Cimino

Department of Biomedical Informatics, Columbia University, New York, NY, USA

Abstract

The UMLS Metathesaurus belongs to the class of scale-free networks with few concept hubs possessing a large number of relationships. The hubs provide useful links between the concepts from disparate terminologies in the UMLS; however, they also exponentially increase the number of possible transitive cross-terminology paths.

Towards the goal of using machine learning to rank cross-terminology translations, we propose a traversal algorithm that exploits the scale-free property of the UMLS to reduce the number of candidate translations. We characterize the concept hubs into "informational" and "noisy" concept hubs and provide an automated method to detect them.

Using gold standard mappings from SNOMED-CT to ICD9CM, we found an average 20-fold reduction in the number of candidate mappings while achieving comparable recall and ranking results. A hub-driven traversal strategy provides a promising approach to generate high quality cross-terminology translations from the UMLS.

Keywords:

UMLS, scale-free networks, terminology, mapping, machine learning

Introduction

The Unified Medical Language System® (UMLS) [1] is a large compendium of biomedical terminologies that integrates synonymous *terms* from disparate source terminologies into a *concept*. As a result, the source-asserted relationships between the terms provide cross-terminology relationships across the concepts in the UMLS. Various informatics applications pertaining to terminology translation, [2, 3, 4] information retrieval [5] and knowledge discovery [6] perform transitive traversal of these relationships to find mappings, links and sub-networks across biomedical terminologies in the UMLS.

One of the critical challenges in performing such transitive traversal of the UMLS relationships is the exponential increase in the number of possible paths. For example, starting with the concept *Hemoglobin* in the UMLS, there are 868 direct relationships to other concepts, which in turn have 450,692 relationships to a different set of UMLS concepts. Identifying the relevant concepts of interest (such as *Anemia*, *Methemoglobinemia*, etc.) with high accuracy among all the candidate paths entails a significant computational burden.

We found that the enormous increase in the number of transitive paths can be attributed to the scale-free property [7] of the UMLS. A graph is said to have a scale-free property when a few nodes (or "hubs") have most of the edges and large number of nodes are each connected by only a small number of edges. An indicator of scale-free characteristics for a network is whether it obeys the *power-law*, which states that the probability p(k) of a given node connected to k other nodes is proportional to k^{-c}, where c is an arbitrary constant. On plotting the power-law for the connectivity distribution of the UMLS 2005AA Metathesaurus (when viewed as a network with concepts as nodes and relationships as edges), we found the constant c=3.0032, strongly indicating the presence of scale-free characteristics for the UMLS.

Another relevant class of network is the small world network wherein the transitive edge traversal distance between any two nodes in the network is very small. For any scale-free network with c=3 (where c is the power-law constant), the average shortest path distance between any two nodes of the network is close to *ln N/ln ln N*, where N is the number of nodes in the network [8]. For the UMLS Metathesaurus, the distance is 5.29 (N=1179179), hence, for a given concept in the UMLS, any other concept can be reached by traversing on average 5.29 transitive relationships.

It has been shown that many real-world networks, such as the World Wide Web, biological networks, and social networks, are also scale-free [7]. However, the meaning of hubs (the nodes with large number of edges) is different across all networks and has important implications in transitive traversal of edges across the hubs. Consider, for example, the p53 molecule, which is an important hub for pathways across regulatory networks. In contrast, for metabolic networks, the water molecule hub acts as noise and produces irrelevant pathways.

In this paper, we characterize the concepts that act as hubs in the UMLS and propose a method to computationally identify the "informational" and "noisy" hubs. We propose a hub-driven algorithm for transitive traversal of relationships in the UMLS to minimize the exponential growth and generate high quality transitive paths. We evaluate the proposed approach with gold-standard mappings from SNOMED-CT to ICD9CM.

Related work

Existing approaches for traversing the UMLS to find mappings, links, and sub-networks have largely relied on expert-designed heuristics that exploit the semantics of relation-

ships. Cimino et al. (1993) [2] presented an approach of using the relationships in the UMLS for translating ICD9 terms to MeSH. The *restrict to MeSH algorithm* [3] starts by traversing hierarchical relationships (parent, broader), followed by associative relationships, until a concept from MeSH is reached. A method [4] to map SNOMED-CT and ICD9CM using the UMLS resulted in recall of 42% and precision of 20%. In our previous work, [5] the exponential rise in the number of links was shown as the critical limitation for applying machine learning approaches to identify cross-terminology translations.

Datasets

UMLS

The Metathesaurus (Meta) of the Unified Medical Language System (UMLS) integrates synonymous terms from different terminologies into concepts. The source terminology relationships are abstracted into 11 top-level relationship types (REL) such as *Parent (PAR), Sibling (SIB) or Child (CHD)*. The REL can have a relationship attribute (RELA) that provides further granular meaning such as *clinically_associated_with, mapped_from etc*. The relationships are provided in the MRREL file of the UMLS. The UMLS 2005AA distribution was used for the experiments reported in this paper.

SNOMED-CT to ICD9CM mappings

The SNOMED-CT vocabulary in the UMLS provides mappings from SNOMED-CT to ICD9CM terms, for example, *Massive Hepatic Necrosis* (SNOMED-CT) to *Acute and subacute liver necrosis* (ICD9CM). These mappings exist as separate relationships in the UMLS with relationship attribute type *mapped_from and mapped_to*. We used 65,417 such concept mappings from the 2005AA distribution of UMLS as our gold-standard.

Approach

Connection definition

Two given terminologies in the UMLS are labeled as the source terminology and target terminology that provide the source concepts and target concepts (respectively). We define a *connection* as a set of transitive relationships between a source concept and a target concept connected by zero or more intermediate terminologies through the relationships in MRREL.

- *0-Step Connection* – The source concept and the target concept are identical.
- *1-Step Connection* – A direct relationship between a source concept and a different, unique target concept.
- *2-Step Connection* – A set of two transitive relationships, wherein the source concept is related to a second concept, which in turn is related to the target concept.

Hub identification

Several different metrics can be used to categorize a given node as a hub or non-hub such as the *degree* (number of edges for a given node), *betweenness centrality* (how often a given node is encountered in shortest paths across the network) and the *"authority" measure* (as determined by algorithms such as the PageRank or HITS) [9]. Given that our goal is to spot the concepts responsible for linking a large number of disparate concepts, the node degree provides a desirable metric for identifying hub concepts in the Meta.

The degree for a given concept in Meta is calculated based on the number of relationships present in MRREL to different unique concepts (i.e. if there are multiple relationships between two concepts, it is considered as a single relationship edge). The concepts are then sorted in descending order based on concept degree to rank the highly connected concepts first. A threshold variable, k (*Hub cutoff*), is used to label the top k concepts as the hubs.

Characterizing concept hubs in meta

In our analysis of the hub concepts in the Meta, we observed a systematic semantic pattern in the type of hubs:

1. *"Noisy" Concept Hubs*: The set of concepts that do not generate meaningful transitive connections across them. Several of these concepts are found at the top of the hub list (see Figure 1a). We can further classify these into:
   a) *"Property" Concept Hubs*: The hub concepts that are used as "attributes" or "properties", for example, *metabolic aspects,* which relates cells, organs and diseases with metabolic biochemical changes

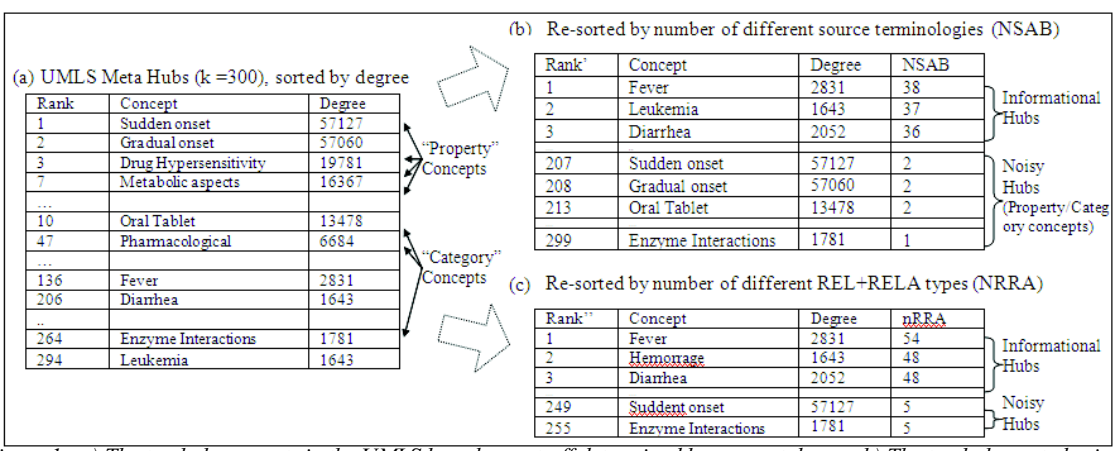


Figure 1 - a) The top hub concepts in the UMLS based on cut-off determined by concept degree. b) The top hubs sorted using the number of source terminologies for a given concept. c) The top hubs sorted using the number of unique relationship types

or *immunologic,* which provides immunological attributes to different concepts.

b) *"Category" Concept Hubs*: The hub concepts that are used as "containers" to group other concepts, for example, *Oral Tablets* (Methotrexate, Nitoman, etc.), or *Microbiological* to group organs, animals, etc used for microbiologic studies.

2. *"Informational" Concept Hubs*: In the concept hub list, we also found several concept hubs that are indispensable for generating useful cross-terminology connections across the Meta such as *Fever, Diarrhea, Edema, Hypertension, etc.* We were able to filter these "informational" concept hubs by re-sorting the hub list based on the number of source terminologies contributing to the Meta hub concept (Figure 1b). Another approach that produced similar results was sorting based on the number of different REL and RELA types associated with the concept e.g. *Fever* has 54 different types of relationships such as *RO_associated_with,RB_inverse_isa, RB_mapped_from* and so on (Figure 1c).

By using a threshold variable, *t (NSAB cutoff)*, the number of different source terminologies, we separate informational hubs from noisy hubs. We consider informational hubs as having number of source terminologies equal to or greater than *t* and the rest are labeled as noisy hubs. The noisy hubs are thrown away during the transitive traversal and the informational hubs are retained. The informational hubs nevertheless have significantly high degree and hence can possibly lead to explosion in number of possible connections. In the next section, we propose a strategy that deals with this explosion problem without eliminating any asserted relationships from the Meta.

Concept hub decomposition

To limit the exponential growth in the number of possible connections generated via hub concepts, we propose an approach to decompose the hubs. The core idea is to traverse relationships only from a single terminology when passing through a hub. Formally, if we have a connection of type $Concept_1 – Relationship_1 – Concept_2$ *(labeled as Hub)* $– Relationship_2 – Concept_3$ then the source terminology contributing the $Relationship_1$ should be same as the source terminology for $Relationship_2$ (Figure 2). The hub decomposition leads to a significant reduction in the number of possible connections, consider if we have a hub concept with *n* relationships all from different source terminologies, a naïve traversal would produce total *n*(n-1)* paths however with proposed approach we generate only *n* paths. An interesting aspect of the proposed approach is that informational hubs generally tend to have a large number of source terminologies, (Figure 1b) which further helps in reducing the number of connections.

Hub driven connection generating algorithm (CoGen)

We summarize here the connection generating algorithm (CoGen) that uses the aforementioned hub-based traversal strategies to generate high-quality connections from the Meta.

Algorithm: CoGen (SCUI, TSAB, n, k, t)

Input: *SCUI: Source Concepts, TSAB: Target Terminology, n: length of connection (n-step), k: Hub cutoff, t: NSAB cutoff*

Output: *Candidate Connections*

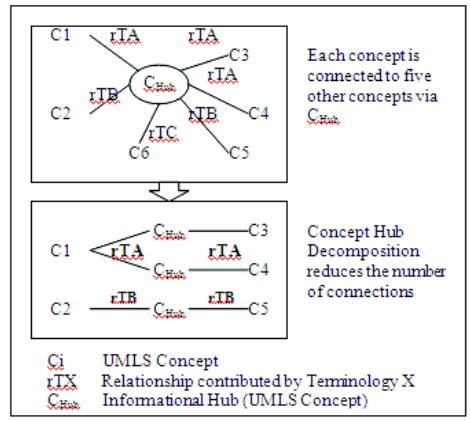


Figure 2 - An illustration showing how decomposing hub concepts reduces the number of connections by traversing only single source asserted terminology relationships via informational hubs

1. **[Clean Meta]**: *Remove all the relationships (REL) of type SIB (sibling), AQ (allowed qualifier) and QB (qualified by).*
2. **[Identify Hubs]**: *Sort the concepts in the cleaned Meta based on their degree. Label top k concepts as hubs.*
3. **[Differentiate Hubs]**: *Re-sort the hubs based on the number of source terminologies (or different relationship types). Label top t concepts as informational hubs and the rest as noisy hubs.*
4. **[Traverse Meta]**: *Generate Candidate Connections by traversing all n-step transitive connections from SCUI to concepts in TSAB, if encounter a concept of type*

   a. Noisy hub: break the traversal.

   b. Informational hub: use Concept Hub Decomposition.

Note that we added a pre-processing step to clean the Meta by removing all relationships of type SIB (sibling), AQ (allowed qualifier) and QB (qualified by) that do not generate meaningful connections in the transitive sense.

Learning features and training set

Towards our goal of using machine learning to identify and rank terminology translations from all possible connections, we describe here strategies for choosing learning features, developing training sets, and ranking the connections (refer to [5] for more details). The Semantic Network and Meta provide broad semantic categories for all concepts (Semantic Types, TUI) and relationships (REL, RELA). A connection can be abstracted into these broad semantic categories, which then creates learning tuples of the form $<REL_1, RELA_1, I_1TUI, REL_2, RELA_2$*:Decision Class*> for 2-step paths where $REL(A)_i$ indicates i^{th} step relationship, I_1TUI is the Semantic Type of intermediate concept and *Decision Class* can be *Positive* or *Negative*.

The positive examples are generated from the gold-standard terminology translations. However to generate negative examples we randomly sample 'strict' N-step connections i.e., the connections where the corresponding source and target concepts do not have any connection that is less than

N steps. The underlying rationale is that when the given concepts are related only by a 'strict' N-step connection, it implies they are relatively "unrelated" or "distant" as compared to the concepts related by less than N-step connection [5]. Once the connections are classified (using a supervised machine learning algorithm), we rank them based on the weighted difference of the number of positive and negative connections between the concepts. For example, given all 2-step connections between *Thyroid_Gland - Hypothyroidism*, 193 connections are classified as positive and 23 connections are classified negative, the connection rank is given by w_p*193-w_n*23, where w_p and w_n represent the average of confidence weight (for positive and negative connections respectively) obtained from a given classifier such as the likelihood of testing instance in Naïve Bayes.

Methods

1. The gold-standard mappings (SNOMED-CT to ICD9CM) were used to create the training sets. We randomly selected 1000 gold-standard mappings as positive training instances and 1000 'strict' 2-step connections as negative instances. The 2-step training connections were generated using the cleaned Meta (without SIB, AQ and QB).
2. Using the proposed CoGen algorithm, all possible 2 step connections were generated for 100 randomly selected source SNOMED-CT concepts to all possible candidate concepts in ICD9CM. The source concepts that were used for generating training sets were not included. Different set of connections for same source concepts were generated by varying each of the following thresholds
   a) The top k hub threshold *(Hub cutoff)* with values k=1000, k=5000 and k=9000
   b) The informational hub threshold *(NSAB cutoff)* with values t=5, t=10 and t=∞ (all hubs are considered as noisy)
3. Using simple transitive traversal (without scale-free view) over cleaned Meta, a different set of 2-step connections to ICD9CM were generated for the same 100 source SNOMED-CT concepts.
4. The training connections were used to learn a Naïve Bayes classifier. The test (candidate) connections were evaluated using the classifier and the likelihood was recorded for each connection. The connection rank was calculated for the classified connections.
5. The following measures were used to evaluate the results of the proposed approach.
   a) Recall: Number of source concepts, for which the correct (gold-standard mapped) target concept was *retrieved*.
   b) Top Rank: The number of correct target concepts ranked in the first position (top1), within top 5 and within top 10.

Results

For the randomly chosen 100 source SNOMED-CT concepts, 31 had the same target concept (0-Step Connection) and were excluded from further analysis. Considering the number of candidate connections generated (Figure 3a), the simple transitive traversal produced 315,329 connections, which was about 21.1 times more than the number of connections generated using CoGen (average 14,967.44). Subsequent analysis of connections produced by simple transitive traversal revealed 11,120 unique candidate ICD9CM concepts that represent 58.2 % of all ICD9CM concepts in the Meta. Within CoGen results, the increase in Hub cutoff parameter led to an overall reduction in the number of candidate connections and the reduction was more prominent for t=∞ (about 80% decrease from k=1000 to 9000), implying the importance of hub concepts in keeping the network connected. On removing all hubs (i.e. t=∞), 4,309 average connections were generated (62.3 connections per source concept) as compared to average 22,343 connections for t=5 (323.8 connections per source concept).

The 69 source SNOMED-CT concepts under analysis had gold standard mappings to 72 ICD9CM concepts. The simple transitive traversal retrieved a marginally higher number (57) of gold-standard ICD9CM concepts as compared to the

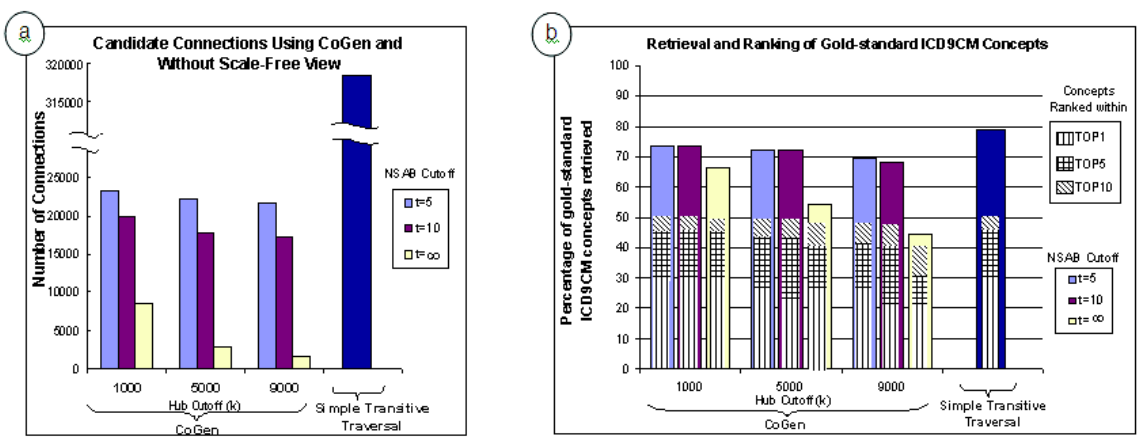


Figure 3 - a) the number of candidate 2-step connections generated for 69 source SNOMED-CT concepts to target ICD9CM using CoGen (with different parameters, k, t) and using simple transitive traversal. b) Percentage of total gold-standard target ICD9CM concepts (72) retrieved and the corresponding ranking after applying connection learning

CoGen results (54) (Figure 3b). The machine learning training set gave 10-fold cross-validation AUC (area under ROC curve) score of 0.97 for Naïve Bayes. No significant difference was observed in the ranking of gold-standard targets, despite of the large reduction in the number of candidate connections, indicating the high specificity of hub-based approach towards irrelevant translations. On average, 25.2% of time the gold-standard target ICD9CM concept was ranked in the top position and 48.9% of time in top 10 positions.

Discussion

Existing research on the UMLS has explored the semantic and linguistic aspects of the Meta at a greater depth, however, little analysis has been done to understand the large-scale structure and network of the Meta. Our results indicate that the hub concepts play a critical role in the Meta for applications that perform a transitive traversal of the relationships. We observed a 20-fold reduction in the number of candidate connections generated by proposed hub-based algorithm when compared to simple transitive traversal while the gold-standard connection ranking produced similar results; implying a significant decrease in the computational effort to traverse and rank the candidate connections. An argument can be made that all the connections corresponding to gold-standard mappings do not generally traverse through hubs. However, when we excluded all hubs ($t=\infty$) in the candidate connections, there was a significant decrease in recall for k=5000 and k=9000 and a marginal decrease for k=1000 (Figure 3b). This finding reflects the essential characteristic of all scale-free networks, where the few top hubs form the backbone of the network and removing them can cause the network to break apart.

Interestingly, the machine learning based ranking of connections was found to be independent of the size of the candidate set and the recall. This suggests that our machine learning approach is resilient to noisy connections. Secondly, it implies an inherent limitation in learning and detecting specific types of gold-standard connections. A possible solution would be to use different training sets to mine different types of connections.

Choosing an optimal value of cutoff parameters for the CoGen algorithm requires further analysis. Consider the following gold-standard connection,

Aircraft accidents – PAR - accidents (hub) – RN - Accident to powered aircraft, other and unspecified, injuring other person.

The hub concept *accidents* was removed when the hub cutoff changed from k=1000 to k=5000. Similarly, here is an example where change in NSAB cutoff led to disconnection of the gold-standard target

Acrokeratosis – RO – Hyperkeratosis (hub) – associated_with – Acquired Keratoderma

Hyperkeratosis is an informational hub at t=5 but becomes a noisy hub at t=10. Our experiments indicate that as more concepts are considered as hubs, (increasing k) a higher variation in the size of candidate connections is obtained for different values of NSAB cutoff parameter. We found that the optimal values of candidate size and ranking occur at around t=5.

In comparing our approach to the method by Fung et al. [4] over the same gold-standard dataset, our approach would obtain a higher recall given that we fetch all possible n-step connections for all relationship types and intermediate terminologies, instead of terminating the traversal when the first target terminology concept is found. Furthermore, our approach produces a ranked list of connections, thereby making it difficult to perform an exact comparison with the results obtained by Fung et al.

Our future work includes validating the approach on LOINC to CPT mappings and investigating the CoGen parameter settings for other terminological applications pertaining to knowledge discovery and information retrieval. Coupling other network properties of the UMLS (such as clustering coefficient, cohesion, and betweenness) to its semantic aspects might result in further improvements to our approach.

Conclusions

The inherent scale-free topology of the UMLS Metathesaurus provides a powerful feature to reduce the number of candidate cross-terminology connections without sacrificing the recall and ranking performance. Our experiments on gold-standard mappings from SNOMED-CT to ICD9CM showed a 20-fold decrease in the number of candidate translations by using a scale-free network view of the Metathesaurus. Generalizing the approach for terminology applications (other than translation) requires further investigation into optimal parameter settings of the proposed algorithm.

Acknowledgements

This work was supported in part by the NLM grant R01LM07593.

References

[1] Lindberg DA, Humphreys BL, McCray AT. The Unified Medical Language System. Methods Inf Med. 1993 Aug;32(4):281-91.

[2] Cimino JJ, Johnson S, Peng P, Aguirre A. From ICD9-CM to MeSH using the UMLS: a how-to guide. Proc.Annu.Symp.Comput.Appl.Med.Care. 1993:730-734

[3] Bodenreider O, Nelson S, Hole W, Chang H. Beyond synonymy: exploiting the UMLS semantics in mapping vocabularies. Proc.AMIA.Symp. 1998:815-819.

[4] Fung KW, Bodenreider O. Utilizing the UMLS for semantic mapping between terminologies. Proc AMIA Annu Symp 2005:266-270.

[5] Patel CO, Cimino JJ. Mining Cross-Terminology links in the UMLS. Proc AMIA Annu Symp 2006:624-628.

[6] Butte AJ, Kohane IS. Creation and implications of a phenome-genome network. Nat Biotechnol. 2006 Jan;24(1):55-62.

[7] Barabasi A, Albert R. Emergence of scaling in random networks. Science 1999;286(5439):509-512.

[8] Cohen R, Havlin S. Scale-free networks are ultrasmall. Phys Rev Lett. 2003 Feb 7;90(5):058701.

[9] Wu AY, Garland M, Han J. Mining scale-free networks using geodesic clustering. Proc of KDD 2004:719-724.

Address for correspondence

Chintan O. Patel
Email: chintan.patel@dbmi.columbia.edu

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Assigning Categorical Information to Japanese Medical Terms Using MeSH and MEDLINE

Yuzo Onogi[a], MD, PhD

[a] *Clinical Bioinformatics Research Unit, Graduate School of Medicine, the University of Tokyo, Japan*

Abstract

*This paper reports on the assigning of MeSH (Medical Subject Headings) categories to Japanese terms in an English-Japanese dictionary using the titles and abstracts of articles indexed in MEDLINE. In a previous study, 30,000 of 80,000 terms in the dictionary were mapped to MeSH terms by normalized comparison. It was reasoned that if the remaining dictionary terms appeared in MEDLINE-indexed articles that are indexed using MeSH terms, then relevancies between the dictionary terms and MeSH terms could be calculated, and thus MeSH categories assigned. This study compares two approaches for calculating the weight matrix. One is the TF*IDF method and the other uses the inner product of two weight matrices. About 20,000 additional dictionary terms were identified in MEDLINE-indexed articles published between 2000 and 2004. The precision and recall of these algorithms were evaluated separately for MeSH terms and non-MeSH terms. Unfortunately, the precision and recall of the algorithms was not good, but this method will help with manual assignment of MeSH categories to dictionary terms.*

Keywords:

medical subject headings, MEDLINE, algorithms, classification, Japanese medical terms

Introduction

Clinical documents are processed and stored in electronic format by many providers in Japan and elsewhere, and this trend will certainly increase in the future. Some clinical data is stored as numbers and codes, for example laboratory examination results and disease names or categories, but the remainder is stored as natural text, for example data relating to a patient's present illness, clinical findings, and radiology reports. If we want to make use of this accumulated information or drive any decision-support system based on the clinical documents, then natural text processing is indispensable, which is very difficult for Japanese-language documents.

Japanese is an agglutinative language, meaning that there are no separators such as spaces between words. There are computer tools for morphological analysis [1] that can segment a whole sentence into parts of speech, and can derive syntactic information. If these tools are used for medical texts, it is necessary for medical terms to be added to the vocabulary, because the above-mentioned tools were originally made for general, non-technical, purposes. However, it is difficult to proceed further with deep parsing because there are very few resources on semantics for Japanese medical terms. Here we define semantic information as relating to the category or categories that a term belongs to. For example, gallbladder (in Japanese) is the name of an anatomical organ, and gallbladder calculus is the name of a clinical finding.

To obtain such categorical information, medical taxonomies are useful, because terms are located within hierarchies of categories, although the categories vary between taxonomies. One such taxonomy is the Japanese medical thesaurus published by the Japan Medical Abstract Society (JAMAS), in which terms are mapped to MeSH (Medical Subject Headings) terms. Furthermore, Japanese MeSH terms have been released by the National Library of Medicine (NLM) since 2002. We verified the validity of the mapping between the JAMAS thesaurus and MeSH terms, so we were able to obtain categorical information for about 20,000 concepts and about 50,000 Japanese medical terms [2].

An English-Japanese dictionary published by the Japan Medical Society (JAMS) comprises about 80,000 entries, and is authored by medical subcommittees within Japan. This dictionary is an authority for standardized medical terminology in Japan. To determine what proportion of this dictionary is covered by MeSH, we compared terms in the dictionary with those in the Unified Medical Language System (UMLS) [3-5] by normalization (norm and lvg programs in the Lexical tools of the UMLS) and found that about 30,000 terms can be mapped to MeSH terms, but that the remaining 50,000 terms cannot be mapped to any existing controlled vocabularies in UMLS, and therefore we cannot obtain any categorical information for these 50,000 terms [6]. We reasoned that if these unmatched terms appeared in the titles and abstracts of articles indexed by MEDLINE, then we could calculate the most relevant MeSH category for the term, because each article is indexed using MeSH terms. Thus, the objective of this study was to assign categorical information to Japanese terms that appear in an English-Japanese medical dictionary using MEDLINE and MeSH.

Methods

The basic premise of this study is that if terms in the above-mentioned English-Japanese dictionary appear in the titles and abstracts of articles indexed by MEDLINE, and that if each article is indexed by MeSH, then a list of relevancies between dictionary terms and MeSH categories could be calculated, so the most precise categorical information for a term in the dictionary could be obtained by selecting the largest (most important) weight category for the term in the list of MeSH categories.

Materials used in this study were as follows: the 2005AA version of the UMLS published by the NLM; the 2005 version of the MEDLINE records leased from the MEDLARS Management Section of the NLM; and the English-Japanese medical dictionary published by the Japan Medical Association, which includes about 80,000 entries in English.

Terms in the dictionary were compared one at a time with titles and abstracts of articles indexed by MEDLINE, with both converted to lower-case letters, for articles published between 2000 and 2004 (a total of 5 years), and the frequencies (number of appearances of each term per article) were obtained. The comparison of terms was performed directly; that is, neither stemming nor inflection conversion was carried out.

Relevancies between dictionary terms and articles were obtained by the TF*IDF weighting method [7] as follows:

$$w_{ij} = tf_{ij} \times idf_i = \frac{freq_{ij}}{\mathrm{argmax}_l \, freq_{lj}} \times \log\frac{N}{n_i} \quad (1)$$

where i represents a term; and j represents a document; w is a weight value for term i in document j; tf_{ij} is a "term frequency," that is, the frequency of term i in document j, and it is normalized against the number of times the most frequent term l appears in the document j; idf_i is the "inverse document frequency for term i, that is, the reciprocal number of documents including term i (n_i), which is also normalized against the total number of documents, N.

Similarly, relevancies between articles and MeSH terms (main headings) were calculated. Here we used both major and minor MeSH terms indexed to an article.

We obtained two weight matrices: one comprised dictionary terms versus articles, and the other comprised articles versus MeSH terms. The inner products of these matrices gave us a weight matrix of relevancies between dictionary terms and MeSH terms as follows:

$$\begin{bmatrix} entry\ terms \\ articles \end{bmatrix} \bullet \begin{bmatrix} articles \\ MeSH\ terms \end{bmatrix} = \begin{bmatrix} entry\ terms \\ MeSH\ terms \end{bmatrix} \quad (2)$$

This method is hereafter referred to as the "proposed method." In order to evaluate the performance of this method, we calculated another weight matrix between dictionary terms and MeSH terms directly from all combinations of dictionary terms and MeSH terms for an article (referred to as the "baseline method) and compared the results.

Evaluations were performed by calculating precision and recall for both algorithms. We constructed two kinds of test sets (gold standards), one consisting of MeSH terms (main headings or synonyms by themselves) existing in the dictionary (9770 terms in total), and the other consisting of non-MeSH terms randomly chosen from the dictionary and categorized manually (100 terms) by a researcher other than the one who developed the algorithms. For each dictionary term, we obtained a list of categories and corresponding weight values in descending order, which were smaller if the category was less relevant. Then precision and recall for rank n was defined as follows:

$$precision_n = \frac{N(retrieved_{rank \le n} \Lambda correct)}{N(retrieved_{rank \le n})} \quad (3)$$

$$recall_n = \frac{N(retrieved_{rank \le n} \Lambda correct)}{N(test\ set)} \quad (4)$$

where N indicates the number of terms. ROC (Receiver Operating Characteristic Curve) can be drawn using a series of precision and recall values for various n values.

This algorithm should not be dependent on the features of the dictionary terms, but if the frequencies were calculated from MEDLINE-indexed articles only, then there might be no difference between the performances of the MeSH and non-MeSH terms. We wanted to investigate this possibility further.

Results

Identification of dictionary terms in MEDLINE

The total number of articles indexed in MEDLINE between 2000 and 2004 (a 5-year period) was about 2.8 million. The number of dictionary terms found in the titles and abstracts of these articles was 49,384 (62% of terms in the dictionary). Of these, 22,296 (28%) were non-MeSH terms; that is, we assigned for the first time some categorical information to these dictionary terms (Table 1). In addition, we compared the dictionary terms with the titles and abstracts of articles indexed in MEDLINE between 1995 and 2004 (a 10-year period), and found that 53,670 terms appeared, of which 25,227 (31%) were non-MeSH terms. This shows that the number of dictionary terms appearing in MEDLINE seems to reach a plateau at around a 5-year period.

We were able to calculate two kinds of weight matrices (dictionary terms versus articles and articles versus MeSH terms) for the 10-year period, but it was impossible to calculate the inner products of the matrices because of the limitations of our computer resources.

Table 1 - Number of dictionary terms in the titles and abstracts of articles indexed in MEDLINE over 5-year and 10-year periods.

	No. of terms (%)
Total no. of terms in the dictionary	80,131 (100)
No. of matched with MeSH terms	33,487 (42)
Five-year period	
No. appearing in MEDLINE	49,384 (62)
No. non-MeSH terms	22,296 (28)
No. unassigned terms	24,348 (30)
Ten-year period	
No. appearing in MEDLINE	53,670 (67)
No. non-MeSH terms	25,227 (31)
No. unassigned terms	21,417 (27)

Evaluation of categorization algorithms

The precision and recall results for the baseline and proposed methods are shown in Figure 1. This shows that the proposed method has better precision and recall (0.83 and 0.78, respectively) for the first rank result in both test sets compared with the baseline method (0.71 and 0.71, respectively). However, the performance results for non-MeSH terms were poor, with the proposed method showing better precision and recall (0.51 and 0.56, respectively) than the baseline method (0.26 and 0.26 respectively).

Table 2 shows a sample result of the categorization of non-MeSH terms in the dictionary using the proposed method. Japanese terms and their corresponding English terms are presented, and the English terms are categorized by using the proposed method to the "Assigned categories. When this category correctly matched the given Answer category, then the symbol in the "Correct column is O, otherwise it is X.

Discussion

Matching of dictionary terms with MEDLINE

Finding the dictionary terms in the titles and abstracts of articles indexed in MEDLINE was partially successful, because we were able to find about 50,000 terms in the present study. However, about half of these were MeSH terms, so we were only able to find 23,000 non-MeSH terms. The number of terms found in MEDLINE using articles indexed during a 10-year period was not significantly greater than that from a 5-year period, and so the number seems to reach a plateau after 5 years. This finding implies that we should use resources other than articles to find the dictionary terms. For the dictionary terms that didnt appear in MEDLINE, we attempted to find the reasons why this had occurred. Three main reasons were apparent. Firstly, we found that most were certainly valid medical terms, but were seldom used in recent articles. They comprised such terms as names of bacteria, infectious diseases, physiological tests or manipulations, and signs and symptoms. Secondly, there were many Latin terms, which also seldom appear in recent articles. Based

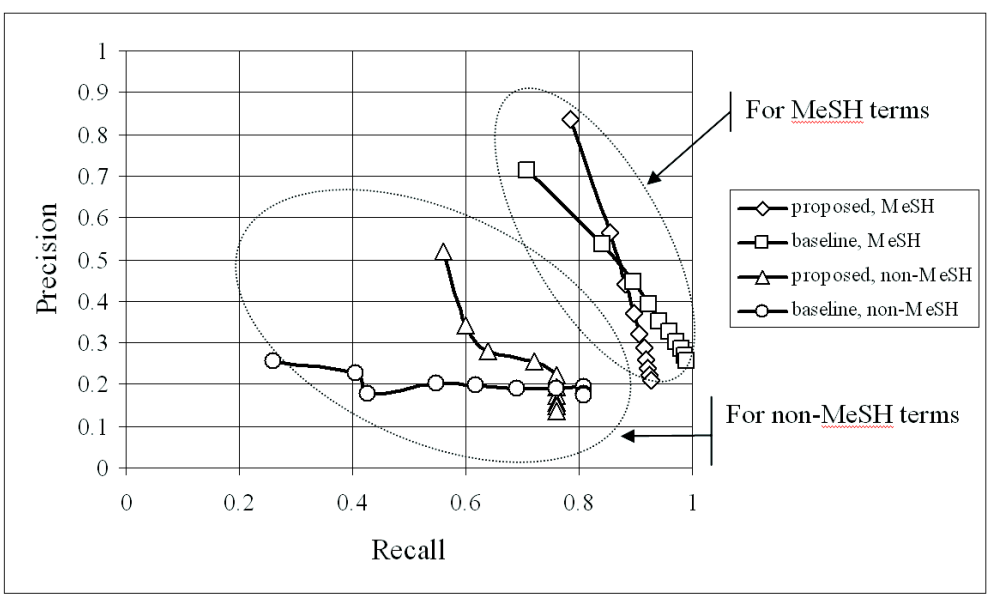


Figure 1- Result of ROC study of the baseline and proposed methods for assigning categories to dictionary terms, applied to MeSH and non-MeSH terms.

Japanese term	Dictionary term	Answer	Assigned	Correct
黒内障性の	amaurotic	C	C	○
救急車サービス	ambulance service	N	J	×
銀親和性	argentaffinity	A	B	×
自己免疫性胆管炎	autoimmune cholangitis	C	C	○
境界潤滑	boundary lubrication	H	H	○
接触角	contact angle	H	H	○
細胞学的分類学	cytotaxonomy	A	B	×
乾燥剤	drying agent	D	G	×
試験的	exploratory	E	F	×
努力呼気流量率	forced expiratory flow	E	E	○
成長生長抑制	growth inhibition	C	G	×
恒常性調節	homeostatic control	G	G	○
左室憩室	left ventricular diverticulum	C	C	○
膜性骨	membrane bone	A	D	×
臓器向性の	organotropic	G	B	×
前重合体	prepolymer	D	D	○
心身医学	psychosomatics	F	F	○
改床法	rebasing	E	N	×
呼吸性ニューロン	respiratory neuron	A	A	○
角	seta	A	A	○
単純癌	simple carcinoma	C	D	×

Table 2- Examples of categorization for dictionary terms (non-MeSH terms).

on these analyses, it may be better to use older articles in MEDLINE to efficiently find these old terms. Thirdly, we found some terms that may be in common use in English written by Japanese-speakers, but which may not be commonly used by authors from other countries. Next we checked the validity of terms not found in MEDLINE using MerckSource (http://www.mercksource.com/) to determine whether they had corresponding definitions or whether other term was suggested as a substitution, because the system could not find any corresponding terms and supposed to be misspelled. We found approximately 4,000 terms that were specific to Japan (e.g. relating to Japanese legislation) but invalid terms comprised a majority of the unassigned terms. If these invalid terms can be removed, then more terms could be matched with MEDLINE in the future.

We compared the dictionary terms and MEDLINE text directly in the present study, and this may have reduced the number of terms matched. Stemming (or canonical normalization in lvg) may be effective to increase the number of matches, but may be ineffective to calculate relevancies because false matches will increase when using stemming.

We used only titles and abstracts of articles indexed by MEDLINE in the present study, but there is a possibility we could increase the number of matches if we could use the entire text of articles, because the more terms that are compared, the more matches will be found not only with respect to number but also with respect to variations.

Categorization algorithm

The proposed method using the inner products of two weight matrixes (dictionary terms versus articles and articles versus MeSH terms) seems to perform better than the baseline method (dictionary terms versus MeSH terms directly) when the category that has the most significant weight value for each term is selected. However, the baseline method was slightly better when using third and lower ranks. For non-MeSH terms, the proposed method performed better than the baseline for all ranks, although precision and recall were both about 0.5 for the first ranked weight value. This result was not good, with only half of the non-MeSH terms being categorized accurately. The purpose of this study was to assign categorical information to Japanese dictionary terms, and the results so far show that automatically categorizing terms with this algorithm fails to assign terms correctly. However, this does not imply that the output of the proposed method is useless, rather we think that this assignment will help in the manual determination of which category a term belongs to (because only the first-ranked category was chosen in the proposed method). For this purpose, the proposed method, using the inner products of two weight matrices, seems to be suitable.

We expected that there would be no difference in the performance of these algorithms for MeSH terms and non-MeSH terms but, as shown in Figure 1, there were actually differences. We checked the non-MeSH terms and found that the average document frequency was about half that of the MeSH terms, but that the standard deviations were

almost the same. This shows that the non-MeSH terms do not frequently appear, and that the TF*IDF weight value mainly depends on document frequency. However, some non-MeSH terms have very large document frequencies (because we did not eliminate stop words and non-MeSH terms contain stop words such as At - astatine), causing the poor result, which was different from the results for MeSH terms.

The nature of the relevancy data should also be discussed. We obtained relevancies between dictionary terms and MeSH terms, but actually this implies collocation or co-occurrence, rather than the two having a categorization or subsumption relationship as we expected. For example, the term hypertension may have stronger relevancy to the term angiotensin-converting enzyme (drug category) than it does to vascular diseases (disease category) as categorized in MeSH. Furthermore, we cannot identify the meaning of each relevancy based on the TF*IDF method only. However, as is shown by the precision for MeSH terms, subsumption might happen to be the dominant relationship in the relevancies determined in the present study.

Conclusion

In order to obtain categorical information for Japanese terms in a English-Japanese medical dictionary, we calculated weight matrixes between English dictionary terms and MeSH terms with the TF*IDF method using MEDLINE title and abstract data for articles published between 2000 and 2004. To calculate the matrix, we used the inner products of two weight matrices. We found 22,000 non-MeSH dictionary terms and assigned corresponding MeSH categories. Although precision and recall was not good, the results are still useful for our purposes.

Acknowledgements

This research was supported in part by funding for a project entitled “Application of Japanese medical thesaurus linked to UMLS Metathesaurus” funded by Japan’s Ministry of Health, Labor and Welfare.

References

[1] Kurohashi S, Nagao M, Japanese Morphological Analysis System JUMAN version 3.5. Department of Informatics, Kyoto University. (in Japanese); 1998.

[2] Onogi Y, et al. Mapping Japanese Medical Terms to UMLS Metathesaurus. Proceedings of the 11th World Congress on Medical Informatics, 2004: pp. 406-410.

[3] Lindberg D, Humphreys B, McCray A. The Unified Medical Language System. Methods Inf Med 1993: 32(4); pp. 281-91.

[4] Campbell KE, Oliver DE, et al. Representing Thoughts, Words, and Things in the UMLS. JAMIA. 1998: 5; pp. 421-31.

[5] Cimino JJ. Auditing the Unified Medical Language System with Semantic Methods. JAMIA. 1998: 5; pp. 42-51.

[6] Onogi Y. Knowledge support system for medical information retrieval in Japanese using UMLS Metathesaurus. Proceedings/Japan Congress of Medical Informatics 2004; pp. 386-7

[7] Salton, G. and Buckley, C. Term-weighting approaches in automatic text retrieval. Information Processing & Management 1988: 24(5); pp. 513–523

Address for correspondence

Yuzo Onogi, MD, PhD:
Clinical Bioinformatics Unit,
University of Tokyo Hospital,
7-3-1 Hongo Bunkyo Tokyo,
113-8655 Japan
Email: yonogi@hcc.h.u-tokyo.ac.jp

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

PharmARTS: Terminology Web Services for Drug Safety Data Coding and Retrieval

Iulian Alecu[a], Cédric Bousquet[a,b], Patrice Degoulet[a,b], Marie-Christine Jaulent[a]

[a] *INSERM, UMRS_872 Eq20,Paris,F-75006 France; UnivParis Descartes, Paris, F-75006 France*
[b] *APHP,Hôpital Georges Pompidou,Département d'Informatique Hospitali_re,Paris,F-75015 France*

Abstract:

MedDRA and WHO-ART are the terminologies used to encode drug safety reports. The standardisation achieved with these terminologies facilitates: 1) The sharing of safety databases; 2) Data mining for the continuous reassessment of benefit-risk ratio at national or international level or in the pharmaceutical industry. There is some debate about the capacityof these terminologies for retrieving case reports related to similar medical conditions. We have developed a resource that allows grouping similar medical conditions more effectively than WHO-ART and MedDRA. We describe here a software tool facilitating the use of this terminological resource thanks to an RDF framework with support for RDF Schema inferencing and querying. This tool eases coding and data retrieval in drug safety.

Keywords:

semantic web, RDF(S), drug safety, graphic navigation

Introduction

The detection of new adverse drug reactions is essential to protect patients from potential harm due to medication. The rapid and effective review of drug safety reports requires 1) the accurate coding of case reports [1] and 2) efficient grouping of similar case reports by the terminology used [2].

Drug safety is an international public health issue. Drug safety reports describe and code observations of adverse reactions after drug administration. These reports are stored in national and international databases. Medical experts must review these reports manually or with the aid of statistical tools if they wish to detect or to confirm a problem related to a particular drug. The databases are huge (3.8 million reports in the WHO international database), but adverse reactions are usually rare events and are therefore rarely reported. The grouping of cases is thus of prime importance for the early detection of problems.

International data sharing is only possible if consistent standardisation is applied to the coding used for drug safety case reports. The coding terminologies used for drug safety are WHO-ART[1] and MedDRA[2], which are used for the indexing, sharing, analysis and reporting of data during clinical trials and for drug surveillance after market release. In the last few years, the focus on drug safety terminology has increasingly shifted from coding to data retrieval and analysis for risk assessment and safety signal detection (data mining).

The main problem faced during statistical analyses based on drug safety terminologies is currently the constraints imposed by regulatory activities. Regulatory authorities require the use of a terminological structure oriented according to surveillance requirements for major public health problems, such as congenital disorders, cancers and cardiac disorders [3]. The WHO-ART and MedDRA terminologies also have a hierarchical structure with restricted multiple inheritances, to avoid double counting of the same case.

There is some debate about the relevance of the structure of these terminologies in data mining [4]. The terminologies currently used for drug safety report coding are not optimal for data retrieval and case review. Indeed, they lack the ability to group cases efficiently when reviewing specific medical issues, particularly those not covered by the requirements of the regulatory authorities.

Structurally, these grouping capabilities are represented by *grouping terms* (e.g. acid base imbalance disorders) and *links* within the terminologies (e.g. duodenal ulcer – is a kind of – digestive system disorder). Although 11% of the reports sent to the FDA in the first three quarters of 2004 were reports of abnormal laboratory results, no link is made between abnormal laboratory results terms and diseases in MedDRA [3]. This results in lower sensitivity for queries concerning medical disorders (e.g. searches for hepatitis-related reports should, but do not always, identify reports of abnormal hepatic enzyme levels).

This is problematic in signal detection, necessitating searches for groupings of appropriate terms related to a medical condition, resulting in delays and variability in the detection of drug-related problems.

SNOMED-CT [5] is emerging as a standard for data coding in clinical domains. It represents signs, symptoms,

1 World Health Organisation – Adverse Drug Reaction Terminology
2 Medical Dictionary for Drug Regulatory Activities

diseases and laboratory examination results. The structure of SNOMED-CT appears to be more appropriate than those of WHO-ART and MedDRA for grouping tasks. We previously extracted the structure embedding the terms used in drug safety from SNOMED-CT [6]. One of the advantages of SNOMED-CT is that this resource embeds within its structure various semantic views of each term and provides detailed definitions for each of its terms. We made use of this asset in the development of our resource.

We decided to use OWL[3] as the representation framework for this resource. All the contents of WHO-ART are included (5267 terms) in this resource, which also includes grouping terms from SNOMED-CT. This resource is semantically rich and provides good results in terms of grouping WHO-ART terms describing related medical conditions. However, OWL software tools are often disappointing in routine use because of problems with scalability, rapidity and technological maturity.

Objective

Our aim was to make the OWL resource operational for routine use in coding and data retrieval from drug safety databases. We aimed to build a browser tool, making it possible to use this resource for intelligent safety data retrieval and coding. This work is part of a large French project with the overall aim of developing a knowledge-based environment to assist drug safety specialists in their work (i.e. a safety database and bibliographical searches, coding). The main problem here is dealing with the size of the resource. The technologies we chose to use are derived from the semantic web. In our previous attempts to use OWL to represent medical terms we have faced the problem of tools not being fast or powerful enough to deal with all of the almost 95,000 terms in our resource.

OWL builds on RDF(S)[4], adding more vocabulary for the description of properties and classes: relations between classes (e.g. disjointedness), cardinality (e.g. "exactly one"), equality, richer typing of properties, characteristics of properties (e.g. symmetry) and enumerated classes.

We believe that OWL is not indispensable for the use of this terminology, but that this formalism is necessary for its development. The use of RDF(S) standard and tools should make it possible to develop an operational and accessible resource.

We first downgraded the resource from OWL to RDFS formalism. This was done by transforming all OWL restrictions into RDFS triples, resulting in half a million such triples. We then built a browser based on, with three main functions: 1) The composition of queries generating groupings of similar terms (WHO-ART, MedDRA and SNOMED-CT), 2) Graphical navigation within the structure of the resource, 3) Case retrieval.

Background

It may be difficult to learn how to use large terminology systems with complex intertwined structures and to navigate within them. Expandable trees, like those from the browser supplied with MedDRA, work well for some tasks, but are of limited value for displaying terminologies allowing multiple contexts. It is also easy to run out of screen space and to lose context if trying to expand and focus on several nodes at once. This is also true with our improved resource. Most browsers for medical terminologies have not taken such requirements into account. The main reason for this is that medical terminologies were mostly used for coding purposes in most commercial software tools, and this task demands little context orientation.

Advanced navigation searches are used in medical terminologies [7], but there are currently no specific researches or implementations in the domain of drug safety. Advanced navigation requires specific resources and tuning for each specific use (i.e. the medical domain) for appropriate adaptation, reducing noise levels and maximising the chances of obtaining an answer to queries.

RDF frameworks with support for RDF Schema inferencing [8] and querying are mature enough for use in this context. Unlike OWL, RDFS does not allow expressive description logics definitions but can be used to represent knowledge as graphs.

RDQL[5] is the standard query language for RDFS, making it possible to describe the patterns of graphs. The answers to queries are the parts of the RDFS graph matching the patterns described in the query. For example, if RDFS was used to represent the genealogical tree of a family using only the “is_child_of” relationship and we wanted to identify all the nephews of given individual called Tom, then a simple query like:

```
Select ?nephews                                          (1)
    from
    (<tom> is_child_of ?father_of_Tom ),
    (?brother_of_Tom is_child_of ?father_of_Tom),
    (?nephews is_child_of ?brother_of_Tom )
```

should give the right answer.

Materials

The resource

The resource is a network of terms. It contains terms from WHO-ART, MedDRA and the parts of Snomed-CT [5] relevant to drug safety. The links are associative relations extracted from Snomed-CT (i.e. “bladder neoplasm” has localisation “Bladder structure”), and “is a kind of” relations used to indicate taxonomic relationship (“renal failure” is a “renal disease”), based on original structure of all the terminologies included. Mapping relationships also connect synonymous terms from different terminologies (e.g “Bladder neurogenic” from WHO-ART is a synonym

3 Web Ontology Language http://www.w3.org/2004/OWL/

4 Resource Description Framework http://www.w3.org/RDF/

5 RDF Query Language http://www.w3.org/Submission/RDQL/

of “Neurogenic bladder” from Snomed-CT). The links and grouping terms (e.g. “renal diseases”) make it possible to group terms in the terminology. This structural information was obtained from Snomed-CT [6].

Open-source technologies

Sesame[6] is an open-source RDF framework with partial inference capabilities, with support for RDF Schema storage, inferencing and querying. The authors of this framework used it to extend RDQL. The new query language, SeRQL, combines the features of other triple-based query languages and adds some of its own. Some of SeRQL’s most important features are: graph transformation, RDF Schema support, XML Schema datatype support, expressive path expression syntax and optional path matching.

Another interesting feature of Sesame is its ability to perform “construct queries”. These queries make it possible to develop persistent inference rules. For example, if we wish to add a new relationship "is_nephew_of" to the example cited above, we need only create a query similar to (1), replacing Select with Construct and specifying that all nodes of the genealogic tree corresponding to the given pattern will be related by the "is_nephew_of" relationship. We have used Sesame as the container of our resource.

The Graphviz layout programs take descriptions of graphs in simple text language, and transform them into diagrams in several useful formats. Graphviz has many useful features for concrete diagrams, such as options for colours, fonts, tabular node layouts, line styles, hyperlinks, and custom shapes. We decided to use Graphviz for graphical representation as it is robust and is appropriate for graph and web representations.

Methods

The work was performed in two steps. We first downgraded the resource from OWL to RDFS formalism. We then built a browser based on Sesame technology,

All the elements from the OWL resource (terms and links) are represented in RDFS formalism. This makes it straighforward to transform queries like “are all WHO-ART terms linked directly or not to renal diseases from SnomedCT” or “are all diseases located in bladder structure” into RDF query language, These queries allow logical compositions (and, or, not), providing greater flexibility in query composition. WHO-ART does not contain terms like “Bladder_structure” (referred to as grouping terms) or associative relations such as “has location” or “has morphology”. By relating the terms in this way, we have greatly increased flexibility and the number of groupings. The resource contains some 500 thousand RDF(S) triplets and 95,000 terms.

The flow of information in PharmARTS[7] is represented in Figure 1. The user connects to PharmARTS via the Internet. The interface is a web browser that interacts with QueryWriter module. During navigation, QueryWriter composes and sends queries to Sesame. The data returned by Sesame is processed by ResultPrint for on-screen display. When graphical navigation is requested by the user, DOTModule translates the data into Graphviz language and ResultPrint displays the image returned by Graphviz.

Once the right terminological cluster has been found, the user can send a request to a drug safety database to which he or she has access and can thus obtain the drug safety reports corresponding to the chosen cluster.

The user interface is a web browser that sends a request to the QueryWriter module. This module transforms the request into Se-RQL, to query the Sesame knowledge base.

The response from Sesame is processed by the ResultPrint module, which displays the result either on a test form or in graphical form. When graphical navigation is activated, the ResultPrint module sends a request to DOTmodule, which translates the information to be printed into dot language. Graphviz processes the dot representation and creates an image file, which is sent by ResultPrint across the Internet to the user.

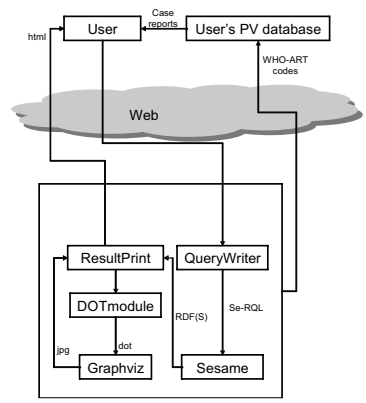


Figure 1 Functional architecture of PharmARTS

Results: The PharmARTS tool

The system was developed in Java and Java Server Pages and runs on a Tomcat server.

Two use cases are envisaged:

1. Data retrieval: The user wishes to obtain drug safety reports related to a specific medical condition.
2. Coding: The user wishes to obtain the most appropriate coding term for a specific drug safety case and navigates through different contexts of related terms.

The browser has four different panels (A to D). In figure 2, panel A includes the controls for free text searches in the terms from the three terminologies. In panel B, the terms resulting from the query are shown. The M: prefix indicates MedDRA, the S: prefix indicates SNOMED-CT and the W: prefix indicates WHO-ART. There are 166 terms

6 http://www.openrdf.org/

7 PharmARTS stands for Pharmaceutical Adverse Reaction Terminology Services

including the string “stomach”. The user can click on the terms displayed in panel B.

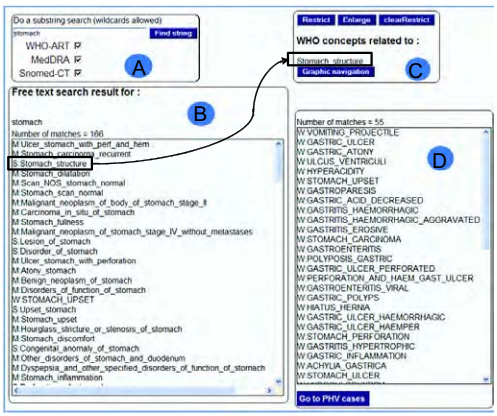


Figure 2 General presentation of the graphical user interface (text mode).

If the user clicks on a term t from panel B, all terms from WHO-ART related to term t are shown in panel D. For example, for “stomach structure” from Snomed-CT, 55 WHO-ART terms are found. There is also a button at the bottom of panel D allowing the user to display all the drug safety case reports coded with at least one of the terms listed in this panel.

The grouping terms are listed in panel C. In our example, “stomach structure”, query refining and graphical navigation are also proposed.

The semantic browser makes it possible to restrict the results of a first free-text search by the results of a second free-text search, using the “Restrict” button in panel C. This operation is actually a logical intersection. For example, in Figure 3, all WHO-ART terms related to haemorrhage in panel D AND located in the stomach structure are shown. Unions can also be created, in a similar manner, using the “Enlarge” button.

The WHO-ART terms related to both the "Haemorrhage" and "Stomach_structure" SNOMED-CT terms can be displayed in the graphical window (figure 4). In graphical mode, it is possible to navigate in the structure of the resource. Once a term of interest is selected, its definition is displayed. A colour code is used to differentiate between the different types of relationship, for example black arrows are hierarchical relationships and red dotted are anatomical localizations. Users can enlarge or restrain their clusters by selecting the ancestors/descendants of the grouping terms. For focusing, a special feature displays the descendants of the grouping terms, enabling the user to select the descendant desired.

Figure 3 - Restraining the "Stomach structure"-related WHO-ART term set to terms relating to both "Stomach structure" AND "Haemorrhage".

The formal definition of WHO-ART terms can be obtained by clicking on the WHO-ART term in question. Figure 5 shows the formal definition of the “Mallory-Weiss syndrome” WHO-ART term. The upper part of the screen shows the WHO-ART ancestors of this term. The middle part of the screen shows synonymous SNOMED-CT and MedDRA terms. Formal definitions of the WHO-ART term are displayed at the bottom of the screen. These formal definitions consist of projections on the morphological, topographic and causal agent axes of SNOMED-CT. For instance the “Mallory-Weiss syndrome” by being linked topographically with “cardia structure”, will be in the same cluster as “achalasia chardiae”, “gastroesophageal reflux” and “cardiospasm”.

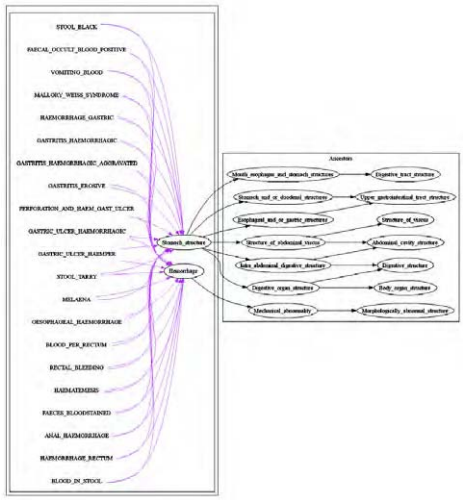


Figure 4 - General presentation of graphic navigator

mode. All nodes on the diagram are interactive.

This tool has been used for evaluation purposes at the WHO Uppsala Monitoring Centre for drug international monitoring. Despite the fact that no formal evaluation protocol was used, some qualitative ad-hoc evaluation has been done. The evaluation involved first restricting a general query (like "stomach"), by navigation. Users then tried to refine the query to obtain the appropriate set of terms. In a second evaluation step, the user tried to retrieve cases coded with the previous set of terms.

This evaluation showed also that 45 minutes of training was sufficient for effective use of the system. The results of these experiments remain qualitative, but the three people at the WHO centre who tested PharmARTS were satisfied with the case retrieval and other functions of this tool. The collaboration for a formal evaluation has been decided with the Uppsala Monitoring centre.

Discussion and conclusion

The semantic browser is a graphical interface communicating with the RDF server. It uses the resource for two main purposes: (1) To facilitate the building of topic-oriented semantic clusters of terms for querying the pharmacovigilance database and (2) To assist with the coding of documents (i.e. case reports) with WHO-ART. This tool provides a graphical visualisation of the context of a coding term defined in the resource. Navigation through the semantic context also helps the user to choose the most appropriate term corresponding to the description of the drug safety report and therefore assist coding procedure. This tool is developed in a web environment, allowing collaborative work for case coding and case retrieval. Drug safety specialists could work together on a cluster related to a specific medical condition.

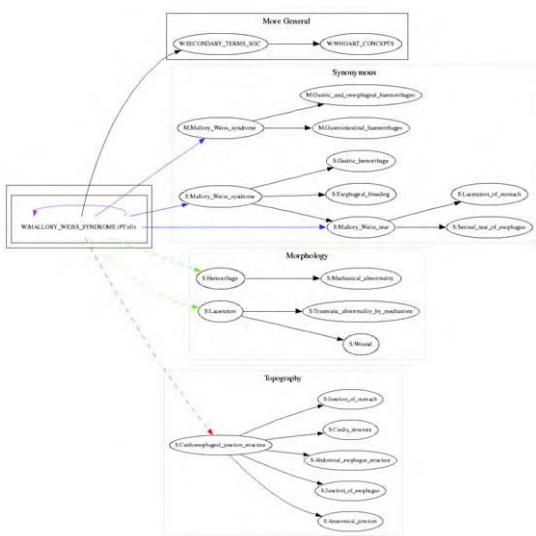


Figure 5 Modelling example of the "Malory Weiss syndrome" term.

Some areas require further development. The ontological resource must be completed. At the moment, only 86% of WHO-ART terms are completely linked to contexts (formal definitions according to SNOMED axes). MedDRA terms have been added for transcoding purposes but are not linked to Snomed-CT contexts. We have carried out feasibility studies for adding SNOMED-CT contexts to MedDRA using OWL, but the main difficulty was related to the number of terms (75,000 terms). The work described here overcomes this size limitation and MedDRA will shortly be added to the resource.

The development features of updating and quality checking must also be considered. The technologies already in use include such features and our future work will involve adapting these features to our needs.

RDF has limited expressivity (i.e. negation) and a low capacity for logical consistency check. It is therefore important to develop and to maintain the resource in OWL but to use RDF to exploit it.

Drug safety databases hold sensitive information and are therefore not made public. The development of internal data management software is required to control access to the private pharmacovigilance database. Users can retrieve cases from the drug safety databases in two ways: using an internal tool to access the database and searching based on the terms from the cluster supplied by the browser or developing an interface using the cluster of terms for automatic querying of the database, followed by display of the results.

Evaluations of Sesame [9] have shown that it performs very well with up to 7 million triples. The entire Unified Medical Language System (2005AC) could be transcribed into 25.5 million triples and Snomed CT, into 3 million triplets. This tool is therefore robust for large queries or intensive use.

Drug safety specialists currently carry out routine manual searches to identify relevant groupings of terms. An evaluation protocol has been developed with the Uppsala Monitoring Centre and will be implemented next year. This protocol will review term groupings for known drug safety signals and will try to determine whether the signals would have been detected earlier with the groupings provided by the terminological server.

In the near future, this tool will be integrated into a complete environment, enabling drug safety specialists to use different resources (bibliographical, report database, drug summary of product characteristics, etc.) and data mining tools (signal detection).

References

[1] Coulter DM. Signal generation in the New Zealand Intensive Medicines Monitoring Programme: a combined clinical and statistical approach. Drug Saf. 2002;25(6): 433-9.

[2] Brown EG. Effects of coding dictionary on signal generation: a consideration of use of MedDRA compared with WHO-ART.Drug Saf. 2002;25(6):445-52.

[3] Zhao-Wong A, Rump E, Moraleda T, Mozzicato P, Revelle P, "Proposed Terminology Changes to Facilitate the Analysis of MedDRA-coded Data" Drug Information J. 2006:40:291-303

[4] Bousquet C, Lagier G, Lillo-Le Louet A, Le Beller C, Venot A, Jaulent MC. Appraisal of the MedDRA conceptual structure for describing and grouping adverse drug reactions. Drug Saf. 2005;28(1):19-34.

[5] Spackman KA, Campbell KE. Compositional concept representation using SNOMED: towards further convergence of clinical terminologies. Proc AMIA Symp. 1998;740-4

[6] Alecu I, Bousquet C, Mougin F, Jaulent MC, Mapping of the WHO-ART terminology on Snomed CT to improve grouping of related adverse drug reactions., Stud Health Technol Inform. 2006;124:833-8.

[7] Sundvall E, Nystrom M, Petersson H, Ahlfeldt H. Interactive visualization and navigation of complex terminology systems, exemplified by SNOMED CT. Stud Health Technol Inform. 2006;124:851-6.

[8] Broekstra J, Kampman A, Inferencing and Truth Maintenance in RDF Schema: Exploring a naive practical approach, Second International Semantic Web Conference PSSS 2003.

[9] Broekstra J, Kampman A, van Harmelen F, A Generic Architecture for Storing and Querying RDF and RDF Schema Source Lecture Notes In Computer Science; Vol. 2342 Proceedings of the First International Semantic Web Conference on The Semantic Web 2002:54-68.

Address for correspondence

Email: Iulian.Alecu@spim.jussieu.fr

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Machine Learning Approach for Automatic Quality Criteria Detection of Health Web Pages

Arnaud Gaudinat, Natalia Grabar, Célia Boyer

Health on the Net Foundation, Geneva 14, Switzerland

Abstract

The number of medical websites is constantly growing [1]. Owing to the open nature of the Web, the reliability of information available on the Web is uneven. Internet users are overwhelmed by the quantity of information available on the Web. The situation is even more critical in the medical area, as the content proposed by health websites can have a direct impact on the users' well being. One way to control the reliability of health websites is to assess their quality and to make this assessment available to users. The HON Foundation has defined a set of eight ethical principles. HON's experts are working in order to manually define whether a given website complies with s the required principles. As the number of medical websites is constantly growing, manual expertise becomes insufficient and automatic systems should be used in order to help medical experts. In this paper we present the design and the evaluation of an automatic system conceived for the categorisation of medical and health documents according to he HONcode ethical principles. A first evaluation shows promising results. Currently the system shows 0.78 micro precision and 0.73 F-measure, with 0.06 errors.

Keywords:

medical web, patient safety, quality of health information, natural language processing

Introduction

The Web proposes an ever-increasing number of medical and health sites, and makes it possible to access quickly and easily several billion of pages devoted to health information. Regarding the overwhelming quantity of information and its uneven quality, how can users, especially patients, judge the information on the Internet? Over the last ten years, the Health On the Net (HON[1]) has responded to the risks and dangers represented by the ever increasing mass of online health and medical information by protecting online citizens by establishing a third party accreditation program, the HON Code of Conduct (HONcode www.hon.ch/HONcode) [2]. This program is now a *defacto* standard adopted worldwide by health web publishers all over the world.

The HONcode principles are a set of quality criteria, i.e *transparency, authority, complementarity, updating of information, authorship, reference, justifiability, sponsorship, advertising*. Each principle is clearly defined by a guideline. The websites that candidate for the accreditation must clearly disclose an explanation of how they respect the eight principles. For instance, the *privacy* principle means that *Confidentiality of data relating to individual patients and visitors to a medical/health website, including their identity, is respected by this website. The website owners undertake to honour or exceed the legal requirements of medical/health information privacy that apply in the country and state where the website and mirror sites are located.* The website must clearly state this (and other principles) in order to be accredited, as the following excerpt demonstrates : *"We respect and are committed to protecting your privacy. 1. We do not monitor individual usage of the website. 2. We collect site usage statistics from our server logs. This data helps us to manage and plan resource updates, and will be used as part of the evaluation of the site."*

The HON's network of experts is working on the inspection of websites that candidate for the HONcode accreditation in order to identify if the site clearly indicates the information required by the HONcode principles. This approach guarantees a high quality review of the websites and consequently allows to certify reliable medical and health pages. However, to remain efficient while the growth of online documents is accelerating, manual compliance reviewing needs to be complemented and systematically executed by automated means.

An interesting but limited study has been presented [3] that aims at filtering the results by using quality criteria. According to the study the system uses regular expressions and heuristics such as the presence of the HONcode seal or not. An automated method, based on the matching of regular expressions, has been proposed [4]. In our work, we assume that machine learning methods will provide more complete and reliable results. Moreover, such methods allow us to detect and modelise linguistic factors which are hardly managed by human experts. Two kinds of machine learning methods can be applied for the automatic identification of quality principles: supervised and unsupervised ones. Supervised methods are based on a learning step, while unsupervised methods are driven by internal proper-

1 Health On Net: *http://www.healthonnet.org*

Selected for best paper award.

ties of the processed data. In the first case expected categories are known, while in the last case they emerge from the data. We have chosen to apply supervised learning methods as they allow us to better characterise and constrain expected categories related to the eight HONcode criteria. Indeed, categorisation methods seem to be helpful in automatic systems working with textual documents, when detecting hostile messages [5] or racist content [6] or when filtering spams [7]. The objective of this paper is to propose an automatic system for the detection of websites which disclose ethical HONcode principles. This paper first describes the material used, and then the proposed methods. We finally present and discuss the results obtained and draw a conclusion.

Material

As far as the automatic text categorisation is concerned, a key component of any system is a knowledge base with positive examples. In our work, the learning dataset is composed of over 5,000 HONcode accredited websites from 72 countries and represents over 1,200,000 web pages indexed in Google.

This unique database is the result of a long experience in the field of health website accreditation. A site is accredited by HON when the site not only complies with the HONcode principles but also demonstrates how each principle is implemented in the site. To make the training dataset even more relevant for the learning process and text categorisation, human experts were asked to extract paragraphs demonstrating the respect of the principle as well as the URL. Thus, we obtain separate datasets related to current principles: *authority, complementarity, privacy, attribution, justifiability, authorship, sponsorship, advertising.* Table l indicates the number of documents available for each principle in four languages (English, French, Spanish and Italian) for the purpose of this study. Please note that we have divided the principle *attribution* in *reference* and *date* elements. As a result, we obtain nine learning datasets: eight HONcode principles and one *date* item.

The data extracted from the English material is the most complete one since the number of accredited websites in English plays a more important role as shown in Table 1. The figures of learning data in table 1 indicate also that, in all the languages, the justifiability principle has a lower number of statements. Although it is difficult to define the optimal size of training data, we are aware that learning on sets with a smaller number of data will show worse results as compared to those where the data is larger and more complete.

Table 1 - Size of the learning data (in number of extracts)

Principle	English	French	Spanish	Italian
Authority	1685	188	123	230
Complementary	1738	182	119	190
Privacy	1561	128	106	187
Reference	1039	112	71	128
Date	2069	232	196	297
Justifiability	323	25	17	28
Authorship	1813	177	120	201
Sponsorship	1473	163	101	163
Advertising	1030	103	86	142

Methods

Automatic categorisation methods consider documents as vectors within a vectorspace. The dimension of this space depends on the number of units (often words) of the whole collection of documents, and the size of each vector corresponds to the frequency of a given unit in a given document. In our application, we aim at categorising sentences and not documents. Indeed, sentences are more suitable for our purpose: (1) a statement about principles can be located in one or more sentences, and (2) the information contained in each sentence is expected to be more homogenous than the information contained in the whole paragraph. The segmentation of paragraphs into sentences is performed by regular expressions based on punctuation marks and html tags.

The features tested within the learning process are the following: (1) with or without use of stopwords (prepositions, determinants, etc.); (2) with or without application of Porter stemming algorithm [8] in order to lexically normalise words, *treating => treat*; (3) learning unit set up to the word combination (n-grams of 1 to 4 words); (4) learning unit set up to the word cooccurencies within sentences, or bag of words.

The unit weight within the sentences is defined by three elements [9,10]: term frequency, inverse document frequency and length normalisation.

The used machine learning algorithms are those proposed by our learning framework [11]: Naive Bayes (NB), Support Vector Machine (SVM), k-Nearest Neighbors (kNN), and Decision Tree (DT). Different combinations of features and categorisation algorithms have been applied to the English data which is the most large one. Features' combinations which showed the most satisfying results have been applied to other languages.

The feature selection aims at reducing vectorspace dimension by selecting the most discriminant features, and thus obtaining more relevant results [12]. We performed feature selection with the document frequency (DF) criterion, which favors units distributed in the largest number of sentences. DF is quick and efficient, and usually allows to reduce about 80% of features [13] without to loose performance. In this study 20% of the features are kept.

The learning and test sets are composed of 90% and 10% of the available documents, respectively.

The evaluation is performed with the following measures in their micro and macro versions: precision, recall and F-measure. Macro precision (*maP*) is representative for the distribution of elements in each category (principle), and micro precision *miP* in each sentence.

Results

Table 2 - Presentation of obtained results

Lang	Lem	Meth.	weight	miR	miP	miF1	Err
ENG	w1	NB	nnn	0.81	0.65	0.72	0.07
ENG	s1	NB	nnn	0.82	0.63	0.71	0.07
ENG	w1	NB	ann	0.81	0.65	0.72	0.07
ENG	w1	NB	ntn	0.79	0.64	0.71	0.07
ENG	w1	NB	nnc	0.77	0.61	0.68	0.08
ENG	w1	NB	atn	0.80	0.64	0.71	0.07
ENG	w1	NB	atc	0.77	0.61	0.68	0.08
ENG	w1	NB	lnn	0.81	0.65	0.72	0.07
ENG	w1	NB	ltn	0.79	0.64	0.71	0.07
ENG	cooc	NB	nnn	0.75	0.73	0.74	0.05
ENG	cooc	NB	atn	0.75	0.73	0.74	0.06
ENG	cooc	NB	atc	0.68	0.65	0.66	0.07
ENG	cooc	NB	ann	0.74	0.70	0.72	0.06
ENG	w2	NB	nnn	0.78	0.70	0.74	0.06
ENG	w2	NB	atn	0.78	0.71	0.74	0.06
ENG	w2	NB	ann	0.77	0.70	0.73	0.06
ENG	w3	NB	nnn	0.76	0.71	0.73	0.06
ENG	w4	NB	nnn	0.75	0.71	0.73	0.06
ENG	**w1**	**SVM**	**nnn**	**0.69**	**0.78**	**0.73**	**0.06**
ENG	cooc	SVM	nnn	0.69	0.72	0.70	0.06
ENG	cooc	SVM	ann	0.76	0.63	0.69	0.07
ENG	w1	KNN	nnn	0.45	0.84	0.59	0.07
FRE	w1	NB	nnn	0.81	0.61	0.70	0.08
FRE	cooc	SVM	nnn	0.64	0.75	0.69	0.06
FRE	w1	SVM	nnn	0.65	0.80	0.72	0.06
SPA	w1	NB	nnn	0.67	0.5	0.58	0.10
SPA	cooc	SVM	nnn	0.50	0.57	0.53	0.10
ITA	w1	NB	nnn	0.81	0.60	0.70	0.08
ITA	cooc	SVM	nnn	0.63	0.76	0.69	0.06

Table 2 shows global results obtained with different algorithms and features. First column *Lang* indicates the processed language and the following three columns indicate the setting up of a learning system : *w* = the segmentation of sentences into units (single word *w1*, bigram *w2*, three-gram *w3* ... w_n, cooccurence *cooc* and stemmed single word *s1*); *meth* = the used learning algorithm (Naive Bayes *NB*, Support Vector Machine *SVM,* k-Nearest Neighbors *kNN* and Decision Tree *DT*); *weight* = the weighting of units (first character indicates term frequency: natural *n*, logarithmic *l* or augmented *a*; second character indicates if inverse document frequency is taken into account *t* or not *n*; third character indicates if document length is normalised with cosine *c* or not *n*). The following columns indicate evaluation figures for *precision, recall, F-measure* in their *micro* versions, and *error rates*. The precision figures are included in the interval between 0.59 and 0.78.

Table 3 presents the contingency between precision and recall figures for all nine sets of the considered data. The rows of this table represent the assigned class and the columns represent the correct class. For instance, the cell at row 1 (authority) and column 6 (authorship) is the percentage of assigned authority sentences while the correct are authorship sentences. It allows to observe how successful the categorisation of data is according to the principles during the test step. Best results are observed when the contingency between precision and recall is high, i.e. *privacy* principle with the contingency 0.92/0.90.

Discussion and perspectives

We consider precision figures, which correspond to the percentage of correct categorisations among all the results, to be the most relevant ones as far as the judgement on the performance of the system in the reviewers' daily use is concerned. Furthermore, we consider that micro precision *miP* is more suitable to be taken into account as it corresponds to the precision with which a sentence is assigned to into a given category. According to this, SVM algorithm with unique word *w1* as processed unit and *nnn* weighting shows the best results (tab. 2): 0.78 micro precision. The recall of this setting is 0.69 which is one of the lowest figures, while F-measure is one of the highest (0.73).The error rate of this setting is one of the lowest (0.06). We can thus expect that applying SVM algorithm with such a setting would give results which show a relevance that is closer to the human categorisation.

All the combinations of features (some of them are presented in the first four columns of table 2) have been tested with English data but, globally speaking, the experiments show no significant differences. According to the indicated reasons, *nnn* setting up of *SVM* algorithm, which shows the most interesting results, has been applied to datasets in other languages. Its application to other language datasets, ...

Table 3 – Precision/Recall contingency table about quality criteria (SVM with unique word on English, bold row from table 2)

P/R		*Correct*								
		Authority	*Complem.*	*Privacy*	*Reference*	*Justif.*	*Author.*	*Sponsor.*	*Advert.*	*Date*
Assigned	*Authority*	**0.64/0.72**	0.05/0.05	0.01/0.01	0.19/0.34	0.01/0.09	0.04/0.13	0.04/0.09	0.01/0.01	0.00/0.01
	Complem.	0.05/0.05	**0.80/0.82**	0.05/0.03	0.01/0.02	0.06/0.44	0.00/0.00	0.03/0.05	0.00/0.01	0.00/0.00
	Privacy	0.02/0.03	0.02/0.04	**0.92/0.90**	0.00/0.01	0.00/0.03	0.01/0.02	0.01/0.02	0.02/0.06	0.00/0.00
	Reference	0.24/0.13	0.03/0.02	0.03/0.01	**0.64/0.57**	0.02/0.08	0.01/0.01	0.02/0.02	0.00/0.00	0.01/0.02
	Justif.	0.06/0.01	0.32/0.03	0.06/0.00	0.06/0.01	**0.45/0.33**	0.02/0.01	0.00/0.00	0.00/0.00	0.00/0.00
	Author.	0.06/0.02	0.02/0.01	0.08/0.02	0.02/0.01	0.00/0.00	**0.81/0.81**	0.00/0.00	0.01/0.00	0.00/0.00
	Sponsor.	0.05/0.03	0.04/0.02	0.02/0.01	0.01/0.01	0.00/0.02	0.02/0.02	**0.69/0.69**	0.16/0.17	0.00/0.00
	Advert.	0.01/0.01	0.02/0.01	0.05/0.01	0.00/0.00	0.00/0.02	0.00/0.00	0.13/0.12	**0.77/0.73**	0.00/0.00
	Date	0.00/0.00	0.01/0.00	0.01/0.00	0.06/0.03	0.00/0.00	0.00/0.00	0.01/0.01	0.01/0.01	**0.90/0.98**

The analysis of the contingency figures from table 3 indicates that the principle the better recognized is *privacy*. Indeed, its precision/recall contingency figures are the highest: 0.92/0.90. This means that on the lexical level, which provides basic data for the categorisation system, these principle statements are formulated with a specific lexicon. For instance, among units with the highest frequencies we can find *identity, personal, respected, individual, confidentiality* or *privacy*. The principle which is the most difficult to recognise is *justifiability as* the precision/recall contingency of only 0.45/0.33 demonstrates. Moreover, it appears to be ambiguous with the *complementarity* principle. Other couples of ambiguous principles are *reference/authority* and *advertising/sponsorship*. As far as the *justifiability* and *reference* principles are concerned, they are sub-populated and do not represent a learning set that is large enough. As for the *advertising/sponsorship* couple, the main reason for the confusion between the two is that, regarding the lexical level, there exist similarities. For instance, both principles mention *funding, acceptance and maintenance*. Furthermore, statements on these principles can be located in the same paragraphs.

As expected, due to the small number of documents referring to the two principles searched for, *justifiability* and *reference*, they could be hardly processed by the system. In this regard, other methods should be tested. For instance, similarity measures between documents [10] as inspired by the information retrieval field. Thus, the similarity can be computed directly with the principle definitions, which exist in various languages, and the small size of the reference data would not represent a limitation.

Furthermore, the role of `url` analyses can be important as they often convey indications about principles [14]. For instance, when a webpage is named *privacy.html* or *policy.html*, this offers a direct indication about the nature of the processed pages and the content that is to be expected. The combination of these different approaches and clues with the machine learning system that is currently used, offers new opportunities.

Currently we aim at the categorisation of sentences even though entire documents should be processed. Indeed, information on quality principles can be distributed among different sentences within a document or even among different pages of a website. Moreover, this information can be found on all the pages and the categorisation system must deal with this. Taking into account whole documents and websites, the categorisation process becomes even more difficult and should decrease the performances of the system. But first evaluations showed that our learning system acquires a database that is necessary for the categorisation of entire pages and websites. Even if a page can contain statements on several principles, the sentences usually deal with only one principle. However, if it happens, the consideration of the whole page will arise the ambiguity.

As discussed, our system categorises sentences according to principles, but it is not sensitive enough to detect in which kind of context, positive or negative, the statement occurs. For instance, a webpage can indicate that privacy policy is not respected on the site, while the system would just detect that this sentence is concerned with the *privacy* principle. This categorisation is correct but managing the nuances that occur with the main information is even more important. The detection of such details remains a challenging perspective [15] and would give interesting complementary indications and weighting of the categorisation resultsAs a next step of our current work, we will deal with this problem.

Another limitation is related to the fact that the statements on the principles can hardly be verified. Web publishers can declare to respect the privacy principle while unofficially they sell information on users. Notice that, in this regard, HON proposes a complaint solution within which three parties (user, web publisher and HON) can anonymously communicate and resolve such situations.

Currently, the evaluation of feature selection is based on the document frequency criterion. It could be interesting to use the Mutual information or the Chi2 criteria for this purpose. We suppose this would enhance the efficiency of the feature selection and improve the categorisation results.

In order to detect webpages which comply with the HONcode principles, we have used a database of positive examples as a training set taken from medical and health pages. But negative examples can also be used if we want to detect pages and sites which do not meet these principles. Each page or site could thus be weighted according to its positive and negative scores, and its global judgement further computed.

The ethical principles of the HONcode are currently translated into 32 languages and the accreditation process is being adopted all over the world. The System which has currently been equiped with for four languages (English, French, Spanish and Italian) can be adapted and applied to other languages. Furthermore, the difficulties related to the quality and transparency of the information on the web are not only reserved to the medical area. Our system can as well be equiped with data from other areas as well. In the medical domain, this system can be tested with other quality principles, which can be different from the ethical ones.

Conclusion

We have presented our work by designing an automatic system for the detection of qualitative and transparent HONcode principles in medical and health documents on the web. These principles respect the EU recommendations[2]. The System is based on machine learning methods for document categorisation. First evaluations show promising results: SVM algorithm with simple words as processed units and *nnn* weighting features shows 0.78 micro precision and one of the highest F-measure (0.73).

2 http://europa.eu.int/information_society/eeurope/ehealth/quality/draft_guidelines/index_en.htm

The error rate of this setting is one of the lowest (0.06). These results seem to confirm the relevance of our approach to the categorisation of webpages in accordance with the ethical principles of the HONcode. We outlined several perspectives which, we believe, will improve our system.

Additionally a manual evaluation is nevertheless needed. A comparison of the pages detected by our automatic system and those already manually categorised by experts, would convey indications about the suitability and reliability of the accreditation of documents that are automatically computed according to the HONcode principles.

Acknowledgments

This work has been realised as part of the PIPS (Personalised Information Platform for Life & Health) project funded by the European Commission programme Information society technology, contract number 507019.

Bibliography

[1] Fox S. Online Health Search 2006. Pew Internet & American Life Project, 2006.

[2] Boyer C, Baujard V, and Scherrer J. HONcode: a standard to improve the quality of medical/health information on the internet and HON's 5th survey on the use of internet for medical and health purposes. In: *6th Internet World Congress for Biomedical Sciences* (INABIS 2000), 1999.

[3] Price SL, Hersh WR. Filtering web pages for quality indicators: an empirical approach to finding high quality consumer health information on the world wide web. *Proc AMIA Symp* 1999:911-5.

[4] Wang Y. and Liu Z. Automatic detecting indicators for quality of health information on the Web. IJMI, In Press,

[5] Spertus E. Smokey: automatic recognition of hostile messages. 1997.

[6] Vinot R, Grabar N, and Valette M. Application d'algorithmes de classification automatique pour la détection des contenus racistes sur l'internet. In: *TALN*, 2003.

[7] Carreras X and Màrquez L. Boosting trees for anti-spam email filtering. In: Proceedings of RANLP-01, *4th International Conference on Recent Advances in Natural Language Processing*, Tzigov Chark, BG. 2001.

[8] Porter M. An algorithm for suffix stripping. *Program* 1980;14(3):130-7.

[9] Salton G. Developments in automatic text retrieval. *Science* 1991;253:974-9.

[10] Singhal A, Salton G, Mitra M, and Buckley C. Document length normalization. *Information Processing & Management* 1996;32(5):619-33.

[11] Williams K and Calvo RA. A framework for text categorization. In: *7th Australian document computing symposium*, 2002.

[12] Koller D and Sahami M. Toward optimal feature selection. In: *International Conference on Machine Learning*, 1996:284-92.

[13] Yang Y and Liu X. Re-examination of text categorisation methods. In: *Proc of 22nd Annual International SIGIR*, Berkley. 1999:42-9.

[14] Craswell N and Hawking D. Overview of the trec-2004 web track. In: Voorhees EM and Buckland LP, eds, *13th Text Retrieval Conference* (TREC 2004), Gaithersburg, Maryland, USA. 2004.

[15] Chapman W, Bridewell W, Hanbury P, Cooper G, and Buchanan B. Evaluation of negation phrases in narrative clinical reports. In: *Annual Symposium of the American Medical Informatics Association (AMIA),* Washington. 2001.

Address for correspondence

Arnaud Gaudinat
Health on the Net Foundation, HUG/DIM
24, rue Micheli-du-Crest
1211 Geneva 14 Switzerland
tel: +41 22 372 62 50
fax: +41 22 372 88 85
email: arnaud.gaudinat@healthonnet.org

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Using Discourse Analysis to Improve Text Categorization in MEDLINE

Patrick Ruch[a], Antoine Geissbühler[a], Julien Gobeill, Frederic Lisacek[c], Imad Tbahriti[ab], Anne-Lise Veuthey[b], Alan R. Aronson[d]

[a] *Medical Informatics Service, University and Hospital of Geneva, Geneva, Switzerland*
[b] *Swiss-Prot, Swiss Institute of Bioinformatics, Geneva, Switzerland*
[c] *Protein Informatics Group, Swiss Institute of Bioinformatics, Geneva, Switzerland*
[d] *Lister Hill Center, National Library of Medicine, Bethesda, MD, USA*

Abstract

PROBLEM: Automatic keyword assignment has been largely studied in medical informatics in the context of the MEDLINE database, both for helping search in MEDLINE and in order to provide an indicative "gist" of the content of an article. Automatic assignment of Medical Subject Headings (MeSH), which is formally an automatic text categorization task, has been proposed using different methods or combination of methods, including machine learning (naïve Bayes, neural networks…), linguistically-motivated methods (syntactic parsing, semantic tagging, or information retrieval. METHODS: In the present study, we propose to evaluate the impact of the argumentative structures of scientific articles to improve the categorization effectiveness of a categorizer, which combines linguistically-motivated and information retrieval methods. Our argumentative categorizer, which uses representation levels inherited from the field of discourse analysis, is able to classify sentences of an abstract in four classes: PURPOSE; METHODS; RESULTS and CONCLUSION. For the evaluation, the OHSUMED collection, a sample of MEDLINE, is used as a benchmark. For each abstract in the collection, the result of the argumentative classifier, i.e. the labeling of each sentence with an argumentative class, is used to modify the original ranking of the MeSH categorizer. RESULTS: The most effective combination (+2%, $p<0.003$) strongly overweights the METHODS section and moderately the RESULTS and CONCLUSION section. CONCLUSION: Although modest, the improvement brought by argumentative features for text categorization confirms that discourse analysis methods could benefit text mining in scientific digital libraries.

Keywords:

text mining; abstracting andiIndexing; information storage and retrieval; natural language processing; libraries, medical; artificial intelligence.

Introduction

Systems for text mining are becoming increasingly important in biomedicine because of the exponential growth of written information in this domain. The mass of scientific literature needs to be filtered and categorized to provide for the most efficient use of the data. The problem of accessing this increasing volume of data demands the development of systems that can extract pertinent information from unstructured texts, hence the importance of key words extraction, as well as key sentence extraction. While the former task has been largely addressed in text categorization studies [1], the latter has been more rarely studied. In this report, we propose to merge the two ideas to improve the first task. We intend to use sentence extraction and sentence ranking methods to improve text categorization in MEDLINE based on Medical Subject Headings (MeSH). Selecting a set of appropriate sentences likely to improve MeSH assignment is a complex task because, more than key words, the importance of a sentence is dependent on the point of view of every reader. However, as for key words, which can be more or less comprehensively listed in a controlled vocabulary, we believe it is possible to propose task-specific criteria, which can help to define such sentences. Our model is based on the implicit discourse-level information found in scientific reports; in particular, our model uses the argumentative sections as they often appear – sometimes explicitly but usually implicitly – in abstracts. The argumentative categories we are experimenting to rank the sentences originate from professional guidelines (ANSI Z39.14-1979). Indeed, articles in experimental sciences tend to respect strict argumentative patterns with at least four sections: purpose-methods-results-conclusion. These four moves –leaving aside minor variation of labels- are reported to be very stable across different scientific genres and experimental sciences in general (chemistry, anthropology, computer sciences, linguistics…) [4]. They are also confirmed in biomedical abstracts and articles [5][6][29]. Following recent trends, which show that argumentative criteria can be useful to improve various text mining and information retrieval applications such as related articles search [7], blind feedback [27] for ad hoc search, or feature and passage selection for automatic indexing in full-text articles [30], our objective is to evaluate the benefit of using discourse analysis representation levels for keyword assignment in MEDLINE.

Background: argumentative categorization

In this section, we present a set of background methods and already reported results, which are useful to understand the rationales supporting our approach. Our argumentative categorizer is formally defined as a mono class classifier: for each piece of text the system will have to decide whether it is a PURPOSE, or a METHOD, or a RESULT, or a CONCLUSION. Sentences are natural candidate segments for such a classification [8] because they are more self-contained than phrases; however anaphoric phenomena may demand larger segments.

Summarization has a related task

Modern summarization systems use annotated corpora in order to acquire appropriate knowledge; based on textual features summarization tools are able to conduct general summarization tasks. Thus, Kupiec et al. [10] report that in abstracts produced by professionals, 80% of sentences are also found in the source document. In such systems, the complex abstracting task is recast into a more modest sentence selection problem. To do so, experts identify a set of relevant sentences from large corpora; these sentences are then used to train the learning system. Complementary to machine learning approaches, Teufel et al. in [23], who design a task very similar to our one, combine a set of manually crafted triggered expressions, such as *finally, we have shown that, we conclude that*… to classify sentences into seven argumentative classes.

Basic classifiers, features selection and weighting

Choosing *a priori* an appropriate classifier for a given task is a fairly difficult task; therefore, empirical comparisons are often necessary. Among state-of-the-art classifiers for text categorization, such as k-NN [11], SVM [12][28], neural networks [13], and rule-induction systems [14], Bayesian classifiers [15] show a linear complexity, while most top performing algorithms have quadratic complexity; therefore they are often more adapted for rapid application development and exploratory studies [16] [24]. The basic features in text categorization are usually word-based. Possible variants are stems, which often implies stop-word removal, and/or sequences of stems, such as bi- or trigrams of stems. Examples of stemming algorithms are provided in Table 1.

Table 1: String Normalization Methods

Token	Lovins	Porter	S-stemmer
genetic	genet	genet	genetic
genetically	genet	genetically	genetically
genetics	genet	genet	genetic
gene	gene	gene	gene
genes	gene	gene	gene
homogeneous	gene	homogen	homogeneous
plaid	plai	plai	plai
play	plai	plai	plai

Our preliminary studies confirmed that elaborate string normalization and stop-word removal strategies such as Porter and Lovins did not outperform simpler approaches, such as plural normalization, which process morphological variations of plural forms (-s, -ies). This strategy appears sufficient to help the classifiers to generalize without removing interesting features, such as verb tense (as suggested in [4]), which is usually removed by stemming (cf. Table 1).

The last step concerns feature weighting. Indeed naive Bayes classifiers combine the log-likelihood of each feature in order to select the most probable category in the category space; however the real frequency observed in training corpora can follow different refinement and smoothing processes, known as feature weighting [11].

> PURPOSE: While gemcitabine (GEM) is widely accepted for the treatment of advanced pancreatic cancer, capecitabine (CAP) has shown single agent activity and promising efficacy in combination with GEM. […] METHODS: Patients had advanced pancreatic adenocarcinoma, no prior systemic chemotherapy other than that given concurrently with radiation therapy, at lease one measurable disease, and adequate organ functions. […] RESULTS: The objective RR among 45 patients was 40.0% (95% CI; 25.1-54.9), including 1CR (2.2%). The median TTP and OS were 5.4 months (95% CI; 1.8-9.0) and 10.4 months (95% CI; 6.2-14.5), respectively. […] CONCLUSIONS: The combination of GEM with dose escalated 14-day CAP is well tolerated and offers encouraging activity in the treatment of advanced pancreatic cancer. […]

Figure 1 - Partial example of an explicitly structured abstracts in MEDLINE.

We tested 3 weighting methods: tf-idf (term frequency-inverse document frequency), chi-2, and df-thresholding (only features appearing frequenty in each class are selected). Our conclusion is that chi-2 and df-thresholding perform similarly, while tf-idf weighting should be avoided. Indeed, tf-idf weighting is appropriate for weighting content-bearing features, while argumentative content is supported by functional words. Three types of features are linearly combined to get a final probability ranking per class: stems, stem bigrams and stem trigrams. As in [15], length normalization of sentences as been applied in order to overcome biases introduced by too long or too short sentences.

Training and test data

In text classification tasks, two types of strategies are competing: expert-driven and data-driven approaches. While the former, which rely on a domain expert, are often time and labour-intensive, the latter are directly dependent on the availability of large training sets. Fortunately, training data for our task can be acquired in a cheap way. Most abstracts in MEDLINE are unstructured (i.e. provided without explicit argumentative markers, such as METHODS, PURPOSES…); but fortunately, a significant fraction of these abstracts contain explicit argumentative

markers. Using PubMed and its Boolean query interface, we collected a set of 12000 MEDLINE citations containing strings such as "PURPOSE:", "METHODS:", "RESULTS", "CONCLUSION:". This fully automatic data collection process introduces some argumentative noise since some of the explicit markers gather additional argumentative content. Thus, explicit markers such as "BACKGROUND AND PURPOSES:" were also collected as pure "PURPOSES:" markers by this simple method. The collection was then split into two sets:

- set A (10800 abstracts) was used for training and validation purposed,
- set B (1200) was used for the final assessment.

In addition to sets A and B, we also collected a smaller set (C) of marker-free abstracts (100 items). Then, two human annotators were asked to manually annotate this set, using the four selected argumentative classes. In contrast with the automatically acquired sets, here we do not assume that argumentative segments and sentences are overlapping items. As shown in Figure 2, some sentences in set C receive more than one label, because they may express two different argumentative moves in the same sentence. In such cases, we do not attempt to identify segment boundaries (as explored in [18] and [19]) and instead ask the system to provide any of the relevant classes. The interannotator agreement on the C set for argumentative segments is 0.81, when measured by kappa statistics, which indicates that agreement is good.

As mentioned above, our goal is to extract conclusion sentences, but because the information is available in our training data, this binary task has been modified into a four-class problem: {PURPOSE, METHODS, RESULTS, CONCLUSION}. We expect that working with more classes will help the system to discriminate between classes that have been reported to be lexically similar [4][21], such as PURPOSE and CONCLUSION. In the data sets used for the evaluation (B and C), explicit argumentative markers have been removed.

<CONCLUSION> Skin surface proteolytic activity in the living animal was detected </CONCLUSION> <METHODS> by a sensitive, non-invasive methodol. Developed in our laboratory</ METHODS>. <METHODS>A non-leaky well was constructed on the shaved back of an anesthetized guinea pig. The well contained the reaction mixture including the substrate 125I-S-carboxymethylated <GPN> insulin B-chain</GPN> (<GPN>ICMI</ GPN>)</METHODS>. <RESULTS>The proteolytic activity was shown to be time-dependent. The activity was strongly inhibited by <ASP_GPN>pepstatin A</ASP_GPN>, indicating the involvement of aspartic proteinase(s) such as <ASP_GPN>cathepsin D</ASP_GPN> and/or <ASP_GPN>E</ASP_GPN>. Pretreatment of the skin with propylene glycol blocked the proteolyticity</RESULTS>.

<CONCLUSION>The present study demonstrates the presence of proteolytic activity located on skin surface<CONCLUSION> <METHODS>using a unique, non-invasive method for in situ proteinase detn. In the living animal</METHODS>.

Figure 2 - Example of an automatically structured abstracts. Four argumentative classes are annotated with XML tags; Gene and Protein Names (GPN and ASP_GPN), are also annotated.

Combining positions of segments

Optionally, we also investigated the sentence position's impact on the classification effectiveness through assigning a relative position to each sentence. Thus, if there were ten sentences in an abstract: the first sentence has a relative position of 0.1, while the sentence in position 5 receives a relative position of 0.5, and the last sentence has a relative position of 1. The following distributional heuristics are encoded in a distributional model: 1) if a sentence has a relative score strictly inferior to 0.4 and is classified as CONCLUSION, then its class becomes PURPOSE; 2) if a sentence has a relative score strictly superior to 0.6 and is classified as PURPOSE, then its class is rewritten as CONCLUSION.

Categorization effectiveness

Table 2 indicates the categorization effectiveness of our argumentative categorizer. In Table 2, we evaluate the effect of positional information on the categorizer. We also evaluate the performance of the system on explicitly and non-explicitly structured abstracts. Finally, with an F-score (i.e the harmonic mean, with recall and precision having the same importance; cf [22]) above 85%, recall and precision measures are competitive with the trigger-based approach proposed by Teufel et al [23] (F-score ~ 68%), and the SVM learner used in McKnight and Srinivasan [25] (F-score ~ 80%). While recall exhibits excellent levels of performance, precision could still be improved. Thus, expected conclusion segments are well classified, but we observe that some non-conclusion sentences are also classified as conclusion (false positives).

Without sentence positions				
	PURP	METH	RESU	CONC
PURP	80.65%	0%	3.23%	16%
METH	8%	78%	8%	6%
RESU	18.58%	5.31%	52.21%	23.89%
CONC	18.18%	0%	2.27%	79.55%

With sentence positions				
	PURP	METH	RESU	CONC
PURP	93.55%	0%	3.23%	3%
METH	8%	78%	8%	6%
RESU	12.43%	5.31%	74.25%	13.01%
CONC	2.27%	0%	2.27%	95.45%

Table 2 - Confusion matrices expressing the classification effectiveness of the argumentative categorizer, with and without using positional information.

Materials and methods

To conduct our key word assignment experiments and evaluations, we used the OHSUGEN collection [25], which contains more than 300,000 MEDLINE records. From these records, we randomly selected 500 citations with abstracts and keywords to tune our system and 1000 citations with abstracts and keywords to provide the final results. All experiments were conducted with the MeSH 2000 version, which roughly corresponds to the time period covered by the collection. Our baseline system, which was originally developed for the BioCreative challenge 2003 to perform text categorization tasks with Gene Ontology categories, combine a vector space ranker and a pattern matcher. Both stems (Porter) and linguistically-motivated features were used by the system [35]. The system achieved competitive results in the context of the BioCretative challenge; cf. [34] for a comparative presentation and [35] for a comprehensive presentation and evaluation.

In our MeSH categorizer, the ranking was based on the categorization status value returned by the system. Thus, the best candidate categories (i.e. the relevance estimate) obtained the highest score, which directly expresses a similarity between a category and the input abstract. To overweight a given argumentative section, we simply modified the term frequency of the feature in the abstract. Thus, if we wanted to emphasize features appearing in the METHODS section, then every feature in a sentence classified as METHODS by the argumentative categorizer received a boosting factor. The fine-tuning of the optimal model, which includes the calculus of the optimal boosting factor, was based on direct search. We worked with integer values, ranging from 1 to 7 and we strove to maximize the mean average precision of our system, which is the best metric to express the full ordering skill of the system.

Results

Figure 3 provides the expected list of Medical Subject Headings for the abstract in Figure 1.

Assigned MeSH:

Adult; Cystic Fibrosis/genetics; DNA/analysis*; Genotype; HLA-DQ Antigens/genetics; Humans; Research Support, Non-U.S. Gov't; Reverse Transcriptase Polymerase Chain Reaction; Sequence Analysis, DNA/methods*; Spectrum Analysis, Raman

Top-12 categories proposed by the categorizer:
12. dna mutational analysis
11. dna
10. genetics
9. fibrosis [not in NP index]
8. alleles
7. sequence analysis, dna
6. genotype
5. mutation
4. fluorescence
3. oligonucleotides
2. polymerase chain reaction
1. cystic fibrosis

Figure 3: Expected and predicted Medical Subject Headings for abstracts in Figure 1 (PMID: 12404725). Major headings are listed with a star. We do not separate between major and minor headings. Qualifiers are ignored in our benchmark.

We observe that several headings, which relates to age, or population-related groups (*Adult*), and to grants (*Research Support, Non-U.S. Gov't*) cannot be inferred from the abstract; therefore, as is well known, both the average precision and the recall of MeSH categorizers are generally low. In contrast, some categories, in particular major headings, which are more important for annotators, can be relevantly inferred from abstracts. Table 3 provides results of the final evaluation. In particular, we observe that the best combination emphasizes the PURPOSE section (x5 times) and the METHODS and RESULTS sections (each with a multiplicative factor of 2). The overall resulting improvement (+2%) is statistically significant, although quite modest ($p<0.003$) [32]. We also see that another simpler yet quite effective combination can be obtained by boosting just the PURPOSE section (x3 times). It is also interesting to observe that the best combination regarding the mean average precision (MAP), i.e. 0.217, is not the best one for precision at high ranks, as expressed by the Precision at recall = 0. In contrast, the best precision at high ranks (0.93) is achieved by trading recall for precision. Thus, in that setting, only 3064 relevant categories are proposed while 3068 were proposed with the baseline system.

Table 3 - Results of the argumentative boosting on the effectiveness of the MeSH categorizer (C=CONCLUSION, P=PURPOSE; M=METHODS, R=RESULTS, MAP= Mean average precision).

Parameters Kc=C,P,M,R	Relevant Retrieved	Prec. at Rec.=0	MAP
Kc=1,5,2,2	3068	0.927 (+0.8%)	**0.217 (+2%)**
Kc=1,3,1,1	3064	0.93 (+1.1%)	0.216 (+1.4%)
Kc=1,1,1,1 (Baseline)	3068	0.92 (100%)	0.213 (100%)

Conclusion

Our results (precision at high ranks ~ 93%) suggest that argumentative contents as available in abstracts stored in digital libraries are helpful for text categorization tasks, such as automatic assignment of Medical Subject Headings (MeSH). The reported improvement is statistically significant. Although modest (+2%), the improvement confirms that discourse analysis methods are useful for a growing number of text mining applications. Indeed, while attempts to apply linguistically-motivated approaches based on syntactic tools (part-of-speech tagging, shallow or deep parsing) to information retrieval and text categorization were rather inconclusive, methods inherited from discourse analysis could provide a more scalable and dependable improvement. Finally, it would be interesting to evaluate the benefit of argumentative features using more elaborate approaches such as those working with novelty detection [33], full-text articles [30], or with more advanced learners [31].

Acknowledgements

The first author has been supported by a visiting faculty grant at the National Library of Medicine of the National Institute of Health (Bethesda, MD). The Oak Ridge Institute for Science and Education (ORISE), managed and operated by Oak Ridge Associated Universities (ORAU), provided the grant through an interagency agreement with the Department of Energy. The author would like to thank the LHC/NCBI team, in particular Anantha Bangalore, Olivier Bodenreider, May Cheh, Dina Demner, Susan Humphrey, Cliff Gay, Will Roger, Larry Smith, Lorry Tanabe, and John Wilbur. The study was also funded by the EAGL project (SNSF 3252B0-.105755).

References

[1] F Sebastiani, Machine learning in automated text categorization. *ACM Computing Surveys* 34(1): 1-47, 2002.

[2] A Aronson, O Bodenreider, H Chang, S Humphrey, J Mork, S Nelson, T Rindflesch, W Wilbur, The NLM Indexing Initiative. *Proc AMIA Symp*. 2000;:17-21.

[3] P Ruch, R Baud, and A Geissbühler. Learning-free Text Categorization*, AIME* 2003, LNCS 2780.

[4] C Orasan, Patterns in scientific abstracts, in Proceedings of Corpus Linguistics Conference, Lancaster University, Lancaster, UK, pp. 433 - 443, 2001.

[5] J Swales, *Genre Analysis: English in academic and research settings*, Cambridge University Press, 1990.

[6] F Salanger-Meyer, Discoursal movements in medical English abstracts and their linguistic exponents: a genre analysis study, *INTERFACE: Journal of Applied Linguistics* 4(2), 1990, pp. 107 - 124

[7] Tbahriti, C Chichester, F Lisacek and P Ruch. Using Argumentation to Retrieve Articles with Similar Citations: an Inquiry into Improving Related Articles Search in the MEDLINE Digital Library. *Int J Med Inform*. 2006 Jun;75(6):488-95

[8] Ruch, C Chichester, G Cohen, G Coray, F Ehrler, H Ghorbel, H Müller, V Pallotta. Report on the TREC 2003 Experiment: Genomic Track, *TREC* 2003.

[9] J Kupiec, J Pedersen and F Chen: A Trainable Document Summarizer, *SIGIR* 1995, p. 55-60.

[10] Y Yang, An evaluation of statistical approaches to text categorization. *Journal of Information Retrieval* 1 (1999) p. 67-88.

[11] S Dumais, J Platt, D Sahami: Inductive learning algorithms and representations for text categorization. In: CIKM, ACM (1998) p. 148-155.

[12] K Ming Adam Chai, Hai Leong Chieu, Hwee Tou Ng, Bayesian online classifiers for text classification and filtering, SIGIR 2002, p. 89-96.

[13] D Beeferman, A Berger, and J Lafferty, Statistical models for text segmentation. *Machine Learning*, (34):177-210, 1999. Special Issue on Natural Language Learning (C. Cardie and R. Mooney, eds).

[14] D Lewis and M Ringuette, M.: A comparison of two learning algorithms for text categorization. In: *SDAIR*. 1994, p. 81-93

[15] P Domingos and M Pazzani, On the Optimality of the Simple Bayesian Classifier under Zero-One Loss, Machine Learning, 29 (2-3), 1997, p. 103-130.

[16] Y Yang, J Pedersen, A Comparative Study on Feature Selection in Text Categorization, *ICML*, 14th International Conference on Machine Learning, 1997, p. 114-121.

[17] K Nigam, J Lafferty and A McCallum. Using maximum entropy for text classification. In IJCAI Workshop on Machine Learning for Information Filtering, pages 61-67, 1999.

[18] M Hearst, Multi-Paragraph Segmentation of Expository Text. Proceedings of the 32nd Meeting of the Association for Computational Linguistics, 1994, p. 79-94.

[19] J Reynar and A Ratnaparkhi. A maximum entropy approach to identifying sentence boundaries. *ANLP*, pages 16-19, 1997. ACL

[20] S Teufel and M Moens, Argumentative classification of extracted sentences as a first step towards flexible abstracting. Mani, M. Maybury (eds), *Advances in automatic text summarization*, MIT, 1999.

[21] C van Rijsbergen, Information Retrieval. Buttersworth, 1979.

[22] S Teufel and M Moens, Sentence Extraction and rhetorical classification for flexible abstracts, AAAI Spring Symposium on Intelligent Text summarization, 1998, p. 89-97.

[23] P Ruch, L Perret and J Savoy. Feature Combination for Extraction of Gene Functions. ECIR 2005. Springer LNCS, 112-126.

[24] L McKnight and P Srinivasan. Categorization of Sentence Types in Medical Abstracts. *AMIA* 2003.

[25] P Ruch, R Baud, P Bouillon, and G Robert. Minimal Commitment and Full Lexical Disambiguation: Balancing Rules and Hidden Markov Models. CoNLL-2000, pages 111-115.

[26] W Hersh. Report on the TREC 2004 Genomics track. SIGIR Forum (2005)

[27] P Ruch, I Tbahriti, J Gobeill, AR Aronson: Argumentative Feedback: A Linguistically-Motivated Term Expansion for Information Retrieval. *ACL* 2006
[28] Y Mizuta and N Collier: Zone Identification in Biology Articles as a Basis for Information Extraction, In Porceedings of JNLPBA (NLPBA/BioNLP), 28-29 August, Geneva, Switzerland.
[29] F Couto, M Silva and P Coutinho, P. FIGO: Findings GO Terms in Unstructured Text (2004). BioCreative Notebook Papers.
[30] C Gay, M Kayaalp, A Aronson. Semi-Automatic Indexing of Full Text Biomedical Articles. AMIA 2005 (2005).
[31] M Ruiz and P Srinivasan. Hierarchical text categorization using neural networks. Information Retrieval, 5(1), pp. 87-118, 2002.
[32] J Zobel. How reliable are large-scale information retrieval experiments? ACM-SIGIR\/, pages 307—314, 1998.
[33] F Lisacek, C Chichester, A Kaplan, S Agnes. Discovering paradigm shift patterns in biomedical abstracts: Application to neurodegenerative diseases. *SMBM* 2005.
[34] F Ehrler, A Jimeno Yepes, A Geissbühler and P Ruch. Data-poor Categorization and Passage Retrieval for Gene Ontology Annotation in Swiss-Prot, *BMC Bioinformatics*, Special Issue on BioCreative: A Critical Assessment of Text Mining Methods in Molecular Biology, vol. 6 (suppl. 1), 2005.
[35] P Ruch: Automatic assignment of biomedical categories: toward a generic approach. *Bioinformatics* 22(6): 658-664 (2006).

Address for correspondence

Patrick Ruch, University Hospitals of Geneva
24 Micheli du Crest,
CH-1211 Geneva,
Switzerland
Email: patrick.ruch@sim.hcuge.ch,
tel: +41 22 372 61 64

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

A Comparison of Impact Factor, Clinical Query Filters, and Pattern Recognition Query Filters in Terms of Sensitivity to Topic

Lawrence D. Fu, MS[a], Lily Wang, PhD[b], Yindalon Aphinyanagphongs, MS[a], Constantin F. Aliferis, MD, PhD[a]

[a] *Department of Biomedical Informatics, Vanderbilt University, Nashville, TN USA*
[b] *Department of Biostatistics, Vanderbilt University, Nashville, TN USA*

Abstract

Evaluating journal quality and finding high-quality articles in the biomedical literature are challenging information retrieval tasks. The most widely used method for journal evaluation is impact factor, while novel approaches for finding articles are PubMed's clinical query filters and machine learning-based filter models. The related literature has focused on the average behavior of these methods over all topics. The present study evaluates the variability of these approaches for different topics. We find that impact factor and clinical query filters are unstable for different topics while a topic-specific impact factor and machine learning-based filter models appear more robust. Thus when using the less stable methods for a specific topic, researchers should realize that their performance may diverge from expected average performance. Better yet, the more stable methods should be preferred whenever applicable.

Introduction

The size of the biomedical literature makes it increasingly challenging for researchers to find high-quality articles. Since manually monitoring the literature in its entirety is impractical, quality assessment methods have been developed to aid researchers. Journal quality is typically measured with impact factor [1], which is a citation-based method. A methodological and content criteria-based approach for identifying high-quality articles is the PubMed clinical query filter [2]. Machine learning methods such as polynomial support vector machine (SVM) models [3] have been recently introduced as pattern recognition query filters for identifying high-quality articles.

Previous work showed that machine learning methods outperform citation-based methods for finding articles [4]. On the other hand, citation-based methods are more flexible since they are applicable for multiple uses. Compared to clinical query filters, machine learning models have distinct advantages. They have superior performance, are automatically generated, and allow users to specify a desired sensitivity or specificity. On the other hand, clinical query filters are easier to understand and more suited for use with standard PubMed interfaces.

While prior work studied the average behavior of these methods over all topics in the biomedical literature, the research in this paper examines a subtle but important property: *stability over topics*. A method with excellent average performance may fail in a focused domain. Suppose we have a set of articles about three topics (A, B, and C) where 90% of the articles relate to topic A and the remaining articles are divided equally among B and C. If a method has a sensitivity of 1 for topic A and .1 for the other topics, overall performance is .91. Researchers interested only in topics B or C would experience much worse than expected performance. The present work examines the extent of this performance variability for the three methods. If this phenomenon is prevalent, these tools should be modified accordingly.

Methods

Journal impact factor

The journal impact factor evaluates journal impact regardless of publication size or frequency [1,5,6]. It affects journal readership and helps researchers determine to which journal they submit their work. Essentially, it is the average number of citations received per article published in the journal. It is defined for a year y as the quotient of two terms [1]:

$$\text{Impact Factor} = \frac{\text{Number of citations in year } y \text{ to journal items published in years } (y-1) \text{ and } (y-2)}{\text{Number of journal articles published in years } (y-1) \text{ and } (y-2)} \quad (1)$$

The numerator is the number of citations received in a given year to journal items published in the previous two years. The denominator is the number of journal articles from the previous two years. Items in the numerator include articles, editorials, and letters to the editor, while the denominator consists only of articles [7]. For example, the impact factor of the New England Journal of Medicine (NEJM) for 2004 is the number of citations in 2004 to its published items from 2002 and 2003 divided by the number of articles from 2002 and 2003.

Topic-specific impact factor

Prior work considered the impact factor of topics irrespective of journal by computing the number of citations received by articles in a topic area (e.g. asbestos) [8,9].

However, this metric does not assess journals. To study the sensitivity of impact factor to the distribution of topics *in a journal*, we need a formula that isolates the contribution of a specific topic (arbitrarily defined) to the overall impact factor. We calculate a topic-specific impact factor (TIF) for a journal in year y by considering only publications related to a given topic:

$$\text{TIF} = \frac{\text{Number of citations in year } y \text{ to items published in years } (y-1) \text{ and } (y-2) \text{ that were relevant to topic}}{\text{Number of journal articles published in years } (y-1) \text{ and } (y-2) \text{ that were relevant to topic}} \quad (2)$$

For example, the numerator of the cardiology-specific impact factor of NEJM in 2004 is the number of citations in 2004 to cardiology-related items published in NEJM from 2002 and 2003. The denominator is the number of cardiology-related articles. Determining topic relevance is topic-specific. For example, we can consider an item relevant to cardiology if its MEDLINE record contains the MeSH term "Cardiology", a related topic such as "Cardiovascular Diseases" that is specified in the "See Also" field of the MeSH record, or a term residing in a sub-tree of these terms [10]. When we specify topics, the topics do not need to be exclusive or cover all items for the adjustment to be meaningful.

Topic-mix adjusted impact factor

Impact factor can be adjusted for a mix of topics with a weighted average of the topic-specific impact factors. We define the topic-mix adjusted impact factor for k topics as:

$$\text{Topic-mix adjusted impact factor} = \sum_{i=1}^{k} w_i \times TIF_i \quad (3)$$

TIF_i is the topic-specific impact factor of topic i, and w_i is a weight proportional to the importance of topic i normalized such that the sum of all weights equals one and each weight is between 0 and 1. For example, a researcher interested in gastroenterology twice as much as hematology would weight the topic-specific impact factors of gastroenterology and hematology by 2/3 and 1/3 respectively. If all topics are weighted equally, the topic-mix adjusted impact factor is the arithmetic mean of the topic-specific impact factors for all topics.

Analysis for journal methods

When computing topic-specific impact factors, we do not have p-values or confidence intervals since they are population totals and not point estimates. We calculated the absolute differences of impact factor to topic-specific impact factor to analyze variability. We determined the minimum, median, maximum, and interquartile ranges of the differences to assess the skewness and spread of the values. Interquartile range measures dispersion and is the difference of the third and first quartiles. We would expect little discrepancy between the methods if citations are evenly distributed over topics.

A Bland-Altman plot [11] was used to determine whether topic-specific impact factors are significantly different from impact factors. This plot shows whether a new measurement method agrees with another method by plotting the measurement differences against their mean and illustrating any dependence between the values. We considered the correlation coefficient, but it is not an appropriate method [11].

We also wanted to ensure that the variation was not randomly caused by smaller sample sizes independently of topic. By definition, journal impact factor is calculated on a larger number of publications than the topic-specific impact factor. To determine whether the difference between the two measures is associated with sample size, we computed the regression coefficients of the following regression model:

$$\text{Diff(TIF, IF)} = \beta_0 + \beta_1 * (\text{sample size difference}) + \beta_2 * \text{topic} + \beta_3 * \text{year} + \beta_4 * \text{journal} \quad (4)$$

where Diff(TIF, IF) is the difference between topic-specific impact factor and impact factor, and "sample size difference" is the difference between the number of articles used in each calculation. The "topic", "year", and "journal" variables are categorical variables representing different values for the topic, year, and journal. They were included in the model to account for any possible confounding effects.

Clinical query filters for articles

The clinical query filters were originally designed by Haynes and colleagues [12] and are the most widely available method for identifying high-quality articles through PubMed [2]. These filters are semi-manually constructed Boolean queries of terms in the MeSH headings, publication type, or text of the MEDLINE record. All articles that match a given combination of terms are returned. Performance is measured by sensitivity and specificity. Filters are defined for diagnosis, etiology, prognosis, and treatment with queries optimized for sensitivity and specificity [13]. For example, the specificity-optimized filter for therapy is: (randomized controlled trial [Publication Type] OR (randomized [Title/Abstract] AND controlled [Title/Abstract] AND trial [Title/Abstract])). This query returns all articles with publication type "randomized controlled trial" or with all three words in the title or abstract.

SVM models for articles

Machine learning methods provide another approach to identifying high-quality articles. In previous research, polynomial support vector machine models [14] had superior performance compared to the clinical query filters [3]. These models preprocess fields and text from MEDLINE records for use as features during learning. A kernel function maps the input space to a "feature" space where a hyperplane is calculated to separate the classes of data. We used the models learned from a previous study [3] which includes further details about the learning procedure. Performance is measured by area under the receiver operating curve (AUC).

Analysis for article methods

We compared the absolute differences between the overall performance when ignoring topic and performance for

each topic. For clinical query filters, the metrics were sensitivity and specificity, and for the SVM models, the

Table 1 - Topic-specific impact factors for general topics and journal impact factor in 2004 and 2003

	Journal	Topic-specific Impact Factors for General Topics								Impact Factor
		Cardiology	Endocrinology	Gastroenterology	Hematology	Medical Oncology	Nephrology	Pulmonary Disease	Rheumatology	
2004	AIM	16.07	13.85	16.92	7.94	12.49	23.17	12.66	15.4	13.11
	AJM	4.09	3.44	2.73	6.38	3.95	4.31	3.1	6.29	4.18
	BMJ	7.55	6.48	7.37	5.73	5.57	2.37	7.94	8.77	7.04
	JAMA	42.18	28.27	60.55	13.87	35.58	20.32	36.47	13.4	24.83
	Lancet	33.8	47.7	18.86	11.98	23.16	14.3	27.41	52.5	21.71
	NEJM	37.46	54.31	37.68	33.71	44.8	27.93	37.97	24.08	38.57
2003	AIM	14.37	19.83	12.73	10.63	12.14	23.06	13.21	14.5	12.43
	AJM	4.21	5.82	2.43	4.3	3.98	5.33	3.44	5.82	4.4
	BMJ	7.95	6.84	4.98	5.57	5.76	4.00	5.37	12.25	7.21
	JAMA	38.12	28.24	70.00	13.38	39.27	18.94	30.13	12.8	21.46
	Lancet	24.42	34.33	17.91	8.34	17.78	14.61	14.12	17.94	18.32
	NEJM	38.05	55.78	33.66	28.78	40.46	39.51	22.42	45.33	34.84

metric was AUC. We also computed the minimum, median, maximum, and interquartile ranges of these differences. Second, we performed Wilcoxon signed rank tests which test the difference between paired measurements. They compare repeated measurements after an experimental manipulation to determine if a value has changed. The null hypothesis is that the difference is zero. A p-value less than .05 means that the difference is significantly different from zero, which implies that the method does not retain performance for individual topics.

Corpus construction and topic identification

We analyzed the corpus from the previous work comparing clinical query filters and SVM models [3]. The ACP Journal Club [15] was the gold standard. It is a meta-publication where experts review the best journals in internal medicine on a monthly basis to identify high-quality articles for categories such as diagnosis, etiology, prognosis, and treatment. All MEDLINE articles from the ACP Journal club during the study period were positive cases or considered high-quality. The remaining articles from the journals during the same period were negative cases and not considered high-quality. For the treatment and etiology categories, there were 15,786 MEDLINE records from July 1998 to August 1999. For prognosis and diagnosis, there were 34,938 MEDLINE records from July 1998 to August 2000. The longer timeline provided a sufficient number of positive cases. Articles were converted into a format suitable for the learning methods by extracting and encoding terms from the abstract, title, MeSH terms, and publication type.

Table 2 - The minimum, median, maximum, and interquartile ranges for the absolute differences between impact factor and topic-specific impact factor in 2004

Topic	Min	Median	Max	IQR
Cardiology	0.09	2.04	17.35	11.58
Endocrinology	0.56	2.09	25.99	15
Gastroenterology	0.33	2.15	35.72	2.92
Hematology	1.31	5.02	10.96	7.53
Medical Oncology	0.23	1.46	10.75	5.61
Nephrology	0.13	6.04	10.64	5.55
Pulmonary Disease	0.45	0.99	11.64	5.1
Rheumatology	1.73	6.86	30.79	12.38

We randomly selected 18 MeSH terms covering a range of topics. The topics were: Bone Diseases, Cardiovascular Diseases, Cysts, Diabetes Mellitus, Endocrine System Diseases, Gastroenteritis, Gastrointestinal Diseases, Heart Diseases, Hematologic Diseases, Hernia, Infection, Kidney Diseases, Lung Neoplasms, Myocardial Infarction, Muscular Diseases, Neoplasms, Respiratory Tract Diseases, and Rheumatic Diseases. Articles were relevant to a topic if its MEDLINE record contained the MeSH term or a term residing in a sub-tree.

Results

Variability of journal ranking for different topics

We measured the variability of impact factor over topics and journals by adjusting it for topic with equation (2). We chose eight general topics of internal medicine according to the MeSH vocabulary and a set of narrowly-defined subtopics randomly selected from Gastroenterology. The 6 journals were Annals of Internal Medicine (AIM), American Journal of Medicine (AJM), British Medical Journal (BMJ), Journal of the American Medical Association (JAMA), Lancet, and New England Journal of Medicine (NEJM). For each journal and topic, we retrieved all MEDLINE records with related MeSH terms and obtained citation counts and journal impact factors from the ISI Web of Knowledge [16] in July 2006.

Our results showed that rankings based on impact factor and topic-specific impact factor were not equivalent. When comparing journals for a given topic, the higher impact journal did not always have the higher topic-specific impact factor, as shown in Table 1. For example, NEJM had a higher impact factor than JAMA but had a lower cardiology-specific impact factor. Of the 120 comparisons among the 15 journal pairs and 8 topics, there were 10 reversals (8.33% of the comparisons, 95% confidence interval 3.39% to 13.28%). There were 3 extremecases where a journal impact factor was 1.5 times greater than another journal while the other journal's topic-specific impact factor was 1.5 times greater. The topics were nephrology (AJM, BMJ), gastroenterology (NEJM, JAMA), and rheumatology (Lancet, NEJM).

Table 3 - The minimum, medium, maximum and interquartile ranges for the absolute differences between overall and topic-specific sensitivity/specificity. The specificity-optimized filter for diagnosis did not return any articles

		Sensitivity				Specificity			
Optimized for	Category	Min	Median	Max	IQR	Min	Median	Max	IQR
Sensitivity	Diagnosis	0.02	0.02	0.15	0.0013	0.015	0.087	0.23	0.097
	Etiology	0.028	0.07	0.07	0	0.00047	0.059	0.22	0.10
	Prognosis	0.031	0.1	0.57	0.15	0.0029	0.053	0.18	0.042
	Treatment	0.0035	0.01	0.026	0.0025	0.0027	0.030	0.17	0.053
Specificity	Diagnosis	-	-	-	-	-	-	-	-
	Etiology	0.16	0.34	0.49	0.28	0.0066	0.13	0.31	0.086
	Prognosis	0.11	0.24	0.52	0.33	0.030	0.099	0.22	0.035
	Treatment	0.034	0.053	0.07	0.023	0.00037	0.048	0.13	0.033

The absolute differences between the two measures as well as the Bland-Altman plot further support that the methods are not equivalent. Table 2 shows the minimum, median, maximum, and interquartile ranges of these differences over all journals for each topic. There was much instability since the maximum differences ranged from about 10 to 35. In Figure 1, the Bland-Altman plot showed that the difference in impact factor and topic-specific impact factor depended on their values, and the divergence increased as the values increased. Also, the difference did not depend on specialty since all topics showed some difference. If the methods were in agreement, all values would lie between horizontal lines at -22.17 and 17.7, which is the range of two standard deviations from the mean difference of -2.24. Three values fall outside this range.

The observations for the eight general topics from internal medicine were also evident for gastroenterology subtopics (data not shown due to space restrictions). As with the general topics, there were a number of ranking reversals. The variation increased for more specialized topics and was most pronounced in the 3 highest impact journals. JAMA had the greatest variability with a maximum topic-specific impact factor that was over 13 times larger than its minimum. For increasingly specialized topics, the overall impact factor became less meaningful. In 2004, JAMA had an impact factor of 24.83, gastroenterology-specific impact factor of 60.55, and topic-specific impact factors for gastroenterology-based subtopics ranging from 6 to 80.07. These data support the idea that researchers studying a specific disease should not rely on overall impact factor for journal evaluation.

To ensure that variation was not a random occurrence unique to a single year, we performed additional experiments. First, we replicated the experiments for 2003 and found consistent results, as shown in Table 1. Many of the relative rankings of the journals were retained, while some of the same reversals existed. Also, ranges of topic-specific impact factors were comparable. Next, we verified that variation was not randomly caused by sampling with the regression model in equation (4). We found that β_1, the regression coefficient for sample size difference, was .0021 and not significantly different from zero (p-value = .6062). Thus, the difference between topic-specific impact factor and impact factor did not appear associated with differences in sample size.

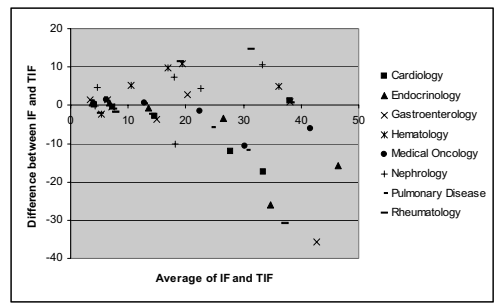


Figure 1 - Bland-altman plot for the differences between impact factor and topic-specific impact factor

For an example of a topic-mix adjusted impact factor, we used the 2004 data and a topic mix where cardiology is weighted three times more than pulmonary disease. JAMA had a topic-mix adjusted impact factor of 40.75 while NEJM was 37.59. In this case, JAMA had a higher cardiology-specific impact factor, while NEJM had a higher pulmonary disease-specific impact factor. Due to the emphasis on cardiology in this example, JAMA had a higher topic-mix adjusted impact factor despite the fact that NEJM had a higher overall impact factor. This example shows that the unadjusted impact factor may not be the best guide in evaluating journals for topic mixes.

Finding high-quality articles for different topics

Performance of the clinical query filters differed for specific topics. Articles were relevant to a topic if its MEDLINE record contained related MeSH terms. The gold standard for high quality articles was inclusion in the ACP Journal Club. Table 3 summarizes the differences between the overall sensitivity/specificity and the observed values. For some categories, there was considerable variability. For example, the sensitivity-optimized prognosis filter had a median difference of 0.1, maximum difference of 0.57, and an interquartile range of 0.15 for sensitivity. These values are relatively large since sensitivity ranges from 0 to 1. The Wilcoxon signed rank tests suggested that performance was unstable for most categories. The p-values were less than .05 for all cases except sensitivity with the sensitivity-optimized diagnosis filter and both values for the sensitivity-optimized prognosis filter.

The SVM models demonstrated more stable results over topics as shown in Table 4. Since AUC values are not sensitivity or specificity values, they cannot be compared directly

with the Haynes' filters results. However, AUC values also range from 0 to 1. Differences are much smaller since the largest interquartile range is .065, and the largest maximum difference is 0.13. The Wilcoxon tests for the SVM models showed that all categories except for diagnosis did not differ significantly from the overall AUC values. These results imply that the SVM models are less sensitive to topic or are more stable for specific topics. One important observation for the diagnosis category is that it had few positive documents. A number of the topics had no positive documents, and most of the topics had fewer than 4 positive cases out of several hundred or thousand articles. With more positive cases, the diagnosis results may be consistent with the results for the other categories.

Table 4 - The minimum, median, maximum, and interquartile ranges for the absolute differences between AUC values

Category	Min.	Median	Max.	IQR
Diagnosis	0.0083	0.038	0.04	0.012
Etiology	0.0027	0.028	0.13	0.05
Prognosis	0.0041	0.045	0.10	0.065
Treatment	0.00054	0.0040	0.041	0.0078

Discussion

Previous research studied impact factor, clinical query filters, and SVM-based models as quality evaluation tools for journals and articles. Performance was measured as an average over all topics. The present study builds on prior work by analyzing the stability of these methods for specific topics. Our results demonstrated that impact factor and clinical query filters are sensitive to topic and vary widely for different subjects. Researchers should realize that average performance cannot be expected for focused searches. Approaches that adjust for topic or are insensitive to topic should be used, if available. The topic-specific impact factor and SVM models are two such options.

Impact factor and clinical query filters are unstable over topics since they are query-independent methods built separately from the learning task. Citation-based metrics (i.e. impact factor) suffer since a citation is not necessarily an endorsement related to the topic of interest. It can acknowledge prior work, identify methodology, correct or criticize, or disclaim the work of others [17]. The reason for the citation may not be relevant to the query topic and distort topic-specific rankings.

The topic-sensitivity of clinical query filters may be due to their manual creation. Experts choose terms that reflect their expertise. Since research areas use different jargon and vocabulary, the coverage of terms may not be exhaustive, and some topics may lack adequate consideration. On the other hand, SVM models automatically learn terms for all topics in the corpus. The machine learning methods should not perform poorly for some topics if they are included in the corpus.

This work's findings have practical implications. Relying on topic-sensitive methods can provide misleading conclusions. For example, researchers interested in gastrointestinal diseases would believe that NEJM is the best journal to read according to impact factor. However, JAMA may be a better choice since it has higher topic-specific impact factors for gastroenterology and gastrointestinal diseases. The variability could result in queries of lower than expected sensitivity or specificity when searching for articles. Taken together, these consequences represent the potential for a habitually flawed literature evaluation. Researchers' work may not be reaching as large an audience as possible, and articles may frequently be misjudged with respect to quality.

We found that the variability problem was not as extensive as the example in the introduction. However, it is still present and should be considered when using the studied methods. This work was the first step in characterizing the application of these approaches for specific domains. Aphinyanaphongs and colleagues showed that it is naive to believe that citation-based metrics can describe all clinical uses [4], and this work shows that it is unrealistic to expect impact factor and clinical query filters to exhibit average performance for all clinical contexts.

Acknowledgments

This work was supported by training grants from the National Library of Medicine (T15 LM 007 450-03, LM007948-02). The authors thank Drs. Cindy Gadd, Nunzia Giuse, and Daniel Masys for their helpful comments.

References

[1] Garfield E. The history and meaning of the journal impact factor. JAMA. 2006 Jan 4; 295(1):90-3.

[2] PubMed Clinical Queries. htttp://www.ncbi.nlm.nih.gov/entrez/query/static/clinical.shtml (accessed Nov. 2006).

[3] Aphinyanaphongs, Y, Tsamardinos I, Statnikov A, Hardin D, Aliferis CF. Text Categorization Models for High Quality Article Retrieval in Internal Medicine. J Amer Med Inform Assoc. 2005;15(2):207-216.

[4] Aphinyanaphongs Y, Statnikov A, Aliferis CF. A Comparison of Citation Metrics to Machine Learning Filters for the Identification of High Quality MEDLINE Documents. J Amer Med Inform Assoc. 2006;13:446-455.

[5] Garfield E. Citation analysis as a tool in journal evaluation. Science. 1972; 178:471-479.

[6] Glanzel W, Moed HF. Journal impact measures in bibliometric research. Scientometrics. 2002; 53(2):171-193.

[7] The ISI Impact Factor: Thomson Scientific.htttp://scientific.thomson.com/free/essays/journalcitationreports/impactfactor/ (accessed November 2006).

[8] Takahashi K, Aw TC, Koh D. An alternative to journal-based impact factors. Occup Med. 1999 Jan; 49(1):57-9.

[9] Uehara M, Takahashi K, Hoshuyama T, Tanaka C. A proposal for topic-based impact factors and their application to occupational health literature. J Occup Health. 2003 Jul; 45(4):248-53.

[10] MeSH Browser: National Library of Medicine. htttp://www.nlm.nih.gov/mesh/MBrowser.html (accessed Nov 2006).

[11] Bland JM, Altman DG. Statistical Methods for Assessing Agreement Between Two Methods of Clinical Measurement. Lancet. 1986;i:307-310.

[12] Haynes RB, Wilczynski N, McKibbon KA, Walker CJ, Sinclair JC. Developing optimal search strategies for detecting clinically sound studies in MEDLINE. J Am Med Inform Assoc. 1994;1:447-458.

[13] PubMed Clinical Queries Table. http://www.ncbi.nlm.nih.gov/entrez/query/static/clinicaltable.html (accessed November 2006).

[14] Burges C. A tutorial on support vector machines for pattern recognition. Data Mining Knowl Discov. 1998;2:121-167.

[15] ACP Journal Club.http://www.acpjc.org (accessed Nov 2006).

[16] ISI Web of Knowledge: Thomson Scientific. http://www.isiknowledge.com (accessed November 2006).

[17] Garfield E, (ed). Can citation indexing be automated? Washington, DC: National bureau of Standards; 1965.

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

A Method for Defining a Journal Subset for a Clinical Discipline using the Bibliographies of Systematic Reviews

Nancy L. Wilczynski, MSc*, Amit X. Garg, MD, PhD†, R. Brian Haynes, MD, PhD*, for the Nephrology Hedges Team

Health Information Research Unit, McMaster University, Hamilton, Ontario, Canada, †Division of Nephrology, University of Western Ontario, London, Ontario, Canada

Abstract

Background: Searching for best evidence for clinical decisions in large biomedical databases is problematic because advances in health care practice that are ready for application are but a very dilute constituent in a much larger pool of biomedical literature. Sensitive search strategies have been developed to help alleviate this problem but search precision is still generally low. If "virtual journal subsets" that are likely to include all relevant articles can be defined for clinical discipline areas or disease content areas this will likely improve search precision.

Objective: *To determine whether studies cited in systematic literature reviews can define a journal subset for a given clinical discipline.*

Design: *Survey of the primary studies included in systematic reviews that are relevant to the clinical discipline of nephrology.*

Methods: *Four data sources were searched to identify systematic reviews relevant to clinical nephrology: the Cochrane Database of Systematic Reviews, McMaster PLUS (Premium LiteratUre Service), MEDLINE, and the Renal Health Library. Three research assistants recorded data pertinent to each of the included primary studies.*

Results: *195 systematic reviews relevant to nephrology were defined and the 2,779 unique original articles they cited were concentrated in 466 journals, with 90% of the articles in 217 titles. This journal subset can be stored online and used when searching the large biomedical databases such as MEDLINE.*

Conclusion: *The bibliographies of systematic reviews can be used to define a journal subset for a clinical discipline area.*

Keywords:

information storage and retrieval, nephrology, medical informatics

Introduction

With the increasing generation of research evidence of high relevance to public health and patient care, ways to manage the knowledge reported in the medical literature are of paramount importance. This is true for the public (who must be able to determine how to best avoid health problems), patients (who must often negotiate for best care), clinicians (who must offer their patients "current best care" and be in a position to provide patients with evidence concerning the many competing claims patients can encounter), managers (who must provide an efficient organizational infrastructure for the best care that is affordable), and policy makers (who must decide on the optimal allocation of scarce resources to support best care). Research to address many of these issues is first widely accessible in the biomedical and health services journal literature. Access to original research and summaries of this evidence is widely and freely available throughout the world through MEDLINE and its PubMed interface, as well as through commercial vendors, including Ovid, Dialog, Aires, and others.

Unfortunately, the task for end-users of accessing, appraising, understanding and applying this literature is difficult mainly because advances in health practice that are ready for application are in very low concentration in a much larger pool of biomedical literature, the vast majority of which is preliminary "scientist to scientist" communication. Thus, the "signal-to-noise" ("ready-to-not-ready") ratio for clinically relevant and ready studies is exceedingly low. To help alleviate this problem we developed search strategies (hedges) for use when searching in MEDLINE and several other electronic databases (e.g., EMBASE). These search strategies were developed to improve the retrieval of clinically relevant and scientifically sound study reports and achieved very high sensitivity (>90%) and specificity (>90%) but with typically low precision (<50%). The search strategies that we developed are available for use on the Clinical Queries page of PubMed (http://www.ncbi.nlm.nih.gov/entrez/query/static/clinical.shtml) and on the Limits screen in Ovid (Ovid (http://gateway.ut.ovid.com/gw1/ovidweb.cgi).

To enhance precision while retaining sensitivity we [1, 2] and others [3, 4] posited that subsets of journals could be defined for given clinical content areas that would contain all or most of the relevant documents that are of sufficient quality to warrant clinical attention. Such journal subsets could be easily implemented as "stored searches" for any number of disciplines, given the computational power of

modern information technology. Rather than searching through the over 5,000 journals that are indexed in MEDLINE the search would be restricted to the journal subset. We hypothesized that using these journal subsets, with a much higher concentration of the relevant documents, would dramatically increase the precision of searching without losing highly relevant reports.

Thus, our preliminary work, in a restricted sample of journals, showed that for even a broad clinical discipline such as internal medicine, 90% of the scientifically valid and clinically important content is captured in 18 journals and 100% in 33 journals [1]. The journal subsets for each of the disciplines of general medical practice/primary care, nursing and mental health are somewhat broader, with just over 30 journals capturing 90% of the content and 40-46 journals capturing 100% [1]. The methodology used to derive the journal subsets for these broad discipline areas involved determining which journals contributed content to 4 evidence-based secondary journals: *ACP Journal Club, Evidence-Based Medicine, Evidence-Based Nursing, and Evidence-Based Mental Health* [1].

In our subsequent work on nephrology we used an alternative method to deriving journal subsets. We hypothesized that journal subsets could be developed for any discipline using bibliographies of systematic review articles because systematic reviews of the medical literature include exhaustive searches for high quality clinical articles (primarily concerning treatment of disease conditions). Using the bibliographies of systematic reviews to derive journal subsets has at least 2 advantages over our previous method. First, it would be possible to use this method for any clinical content area because it is not restricted to locating evidence-based secondary journals on a given clinical content area. Second, this method utilizes a much larger journal set in deriving the subset. Using this approach we found that 49% of nephrology content is captured in 20 journals and the remaining 51% is published across 446 journals [2]. A very different spread of content than observed for the broad clinical discipline areas using an entirely different approach.

Some findings of this research were previously reported [2]. In this paper we describe the detailed methods (not previously reported) of generating a journal subset using the bibliographies of systematic review articles using the clinical discipline area of nephrology as an example.

Methods

The bibliographies of systematic review articles about the management of renal disorders were used to generate a journal subset for the clinical discipline area of nephrology. More specifically, the journals that published the primary studies included in the systematic reviews were tallied and rank ordered to determine where the majority of literature is published for nephrology. To identify potentially relevant systematic reviews 4 databases were searched: the Cochrane Database of Systematic Reviews (CDSR), McMaster Premium LiteratUre Service (PLUS), MEDLINE, and the Renal Health Library. A review was defined as systematic if it included a methods section that outlined the data sources used to identify the primary studies, and the inclusion/exclusion criteria were specified for determining which primary studies were included in the review.

The first database we searched was the CDSR, Issue 3, 2005. CDSR houses systematic reviews that are produced by review groups who use exhaustive database and hand searching techniques and explicitly defined standards of clinical appraisal [5]. We used reviews prepared by the Cochrane Renal Group within CDSR to define the "universe" of reviews for nephrology for each of the following potentially relevant renal subtopics: acute renal failure, diagnostic tests, drugs and the kidney, chronic kidney disease, end-stage kidney disease, general nephrology, process of care, transplantation, urinary tract infection, and urology. All completed reviews (not protocols) listed for each relevant subtopic were included in this study. The subtopics "urinary tract infection" and "urology" were deemed not relevant to the clinical practice area of nephrology by two practicing nephrologists, one of the authors (Amit Garg) and a colleague (Robert Yang). Additionally, the subtropics "diagnostic tests" and "process of care" had no completed reviews.

Data were extracted by 3 research assistants working independently. For each of the relevant Cochrane systematic reviews we initially recorded on an Excel spreadsheet the following data: title of the Cochrane review, renal subtopic group, and the complete citation of all included primary studies. Subsequently, for each of the included primary studies we recorded: the PubMed ID (if indexed in MEDLINE), whether the citation was for a full-text article, whether the citation was for an English language study, whether the included study was a randomized controlled trial, and the year the included primary study was published. The journal titles of included primary studies were tallied and rank ordered by frequency to determine which journals published content relevant to the clinical discipline area of nephrology.

The CDSR contains mainly reviews about treatment. Given that studies researching questions relating to diagnosis, prognosis, and etiology would also be of interest to clinicians working in nephrology we decided to search additional databases.

The second database was McMaster PLUS. McMaster PLUS powers bmjupdates+ (http://bmjupdates.mcmaster.ca/index.asp) which provides access to current best evidence from research, tailored to the specific clinician's health care interests, to support evidence-based clinical decisions. McMaster PLUS contains systematic reviews and primary studies identified from over 130 clinical journals since January 2003. Primary studies and review articles that met specific criteria for scientific merit are categorized for various clinical practice areas including nephrology. The PLUS database was queried for all treatment, diagnostic, prognostic, and etiologic systematic reviews disciplined for Internal Medicine and Nephrology using the advanced searching interface in bmjupdates+.

Cochrane reviews were excluded. Two practicing nephrologists (Amit Garg and Robert Yang) independently assessed each review for having content relevant to nephrology. The definition used to determine relevant nephrology content was based on a list of medical criteria that was being operationalized for a subsequent study. Data collection for each of the primary studies included in the relevant systematic reviews proceeded in the same fashion as outlined for the CDSR.

To further expand the number of systematic reviews in the areas of diagnosis, prognosis, and etiology a third database was searched, MEDLINE using the Ovid interface. On February 2nd, 2006, MEDLINE was queried for the following types of articles: treatment, diagnosis, prognosis, and etiology using the search strategy shown in Table 1. Reviews identified via the previous 2 searches were discarded. All citations retrieved were reviewed independently by 2 nephrologists who determined if the content was relevant to nephrology.

Table 1 - Search strategy used in MEDLINE

#	Search History	Results
1	exp Kidney Diseases/	279165
2	exp Kidney Failure, Acute/	22481
3	exp Diabetic Nephropathies/	12202
4	exp Kidney Failure, Chronic/	52835
5	exp Proteinuria/	20317
6	exp Glomerulonephritis/	30063
7	exp Kidney Calculi/	11560
8	exp Kidney, Artificial/	2802
9	exp Renal Dialysis/	64851
10	exp Peritoneal Dialysis/	16820
11	exp Renal Replacement Therapy/	118097
12	exp Kidney Transplantation/	55939
13	exp Purpura, Schoenlein-Henoch/	2242
14	*Kidney/de	10503
15	1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14	364595
16	15 not Kidney Neoplasms/	328686
17	limit 16 to (humans and english language & abstracts)	123476
18	17 not "cochrane database of systematic reviews".jn.	123419
19	limit 18 to ("reviews (specificity)" and yr="2001 - 2005")	349*

* 349 articles were retrieved, including 113 nephrology reviews not previously identified, 25 reviews already identified through the previous searches, 72 reviews not relevant to nephrology, and 139 articles that were not systematic reviews.

Subsequently, 3 research assistants assessed the relevant reviews for methodological rigor. Data collection for each of the primary studies included in the relevant systematic reviews proceeded in the same fashion as outlined for the previous 2 databases.

The fourth and final database searched was the Renal Health Library (http://www.cochrane-renal.org/renal-healthlibrary.php), an electronic database that is regularly updated by the Cochrane Renal Group and contains non-Cochrane meta-analyses (among many other types of articles). Two nephrologists independently assessed each non-Cochrane meta-analysis to determine if the content was relevant to nephrology, Data collection proceeded in the same way as outlined for the previous 3 databases.

Results

The results are shown in Table 2. A total of 195 unique systematic reviews relevant to clinical nephrology were identified. In these 195 systematic reviews, 3,260 primary studies were included of which 2,779 were unique, published in 765 journal titles of which 466 were unique.

Table 2 - Number of Unique Systematic Reviews, Primary Studies and Journal Titles Generated by Each Data Source

Data Source	Unique Systematic Reviews	Primary Studies	Journal Titles
CDSR	30	513	102
McMaster PLUS	30	726	211
MEDLINE	113	1696	354
Renal Health Library	22	325	98
Total	195	3,260 **Unique** 2,779	765 **Unique** 466

The systematic reviews investigated questions relating to treatment (n=160, 1986 primary studies), diagnosis (n=8, 182 primary studies), prognosis (n=12, 381 primary studies), etiology (n=14, 222 primary studies), and economics (n=1, 8 primary studies).

Ninety percent of nephrology content was captured in 217 journals; 75% was captured in 81 journals. Horizon estimation using Poisson regression was used to estimate how many journals were missed with this 4 search strategy approach. It was estimated that 20 journals were missed for a completeness of 96%.

Discussion

This paper outlines the methods used to define a journal subset for a clinical discipline, nephrology, using the included studies cited in systematic reviews. As indicated

in the introduction, this method led to the derivation of a journal subset with greater dispersion than when content contained in evidence-based secondary journals is used.

At least 2 factors may explain this difference. First, the articles contributing content to the 4 evidence-based journals must meet all of several basic criteria for scientific merit as outlined at http://www.acpjc.org/shared/purpose_and_procedure.htm. Thus, when using the evidence-based journal method to defining a journal subset the resulting subset is based on content of very high quality. Systematic reviews, however, often include studies of lesser merit (for questions of therapy nonrandomized trials and before-after studies may have been included). Thus, this method shows a broader dispersion of the literature including content that is relevant to the clinical topic but not necessarily ready for clinical attention.

Second, the journals contributing content to the 4 evidence-based journals have been accessed for their ability to produce high-quality content. Thus, when using this method a restricted set of high-quality journals is used to derive the subset. When using the method of included studies cited in systematic reviews the pool of potential contributing journals is much larger.

The method used to define a journal subset for a given clinical content area has implications because the resulting subset can be very different. If journal subsets are stored on-line and used when searching in large biomedical databases such as MEDLINE the methods used to derive the subset must be explicitly stated. Various types of journal subsets for the same discipline could be made available, for example, those that were derived to find high-quality content for direct clinical application and those derived to find all content for a given clinical content area. Then, depending on the purpose of the search, the appropriate journal subset could be chosen. For example, if a clinician is searching MEDLINE to find an answer to a patient care question the search could be restricted to a journal subset derived to find high-quality content (those studies that are ready for clinical application) in the disease area in question. On other hand, if a researcher is conducting a systematic review the search could be restricted to a journal subset derived to find all content for a given content area. Thus, both journal subsets serve a purpose.

Regardless of the journal subset chosen searching in MEDLINE, using a "virtual journal subset" will improve the precision of the search in terms of absolute numbers. Rather than searching through the entire MEDLINE database that contains references from over 5,000 journals, the search will be conducted using only 466 journals (in the case of the nephrology example). This will substantively decrease the number of articles that will need to be sorted through in order to find those that are on target, and the concentration of these articles will be much higher than in the entire database, resulting in increased search precision.

Conclusion

The bibliographies of systematic reviews can be used to define a journal subset for a clinical discipline area. The journal titles can be used to create a "virtual journal subset" that can be stored on-line and used when searching the large biomedical databases such as MEDLINE.

Acknowledgments

This research was funded by the National Library of Medicine, USA, the Canadian Institutes of Heath Research and the Kidney Foundation of Canada. The Nephrology Hedges Team includes Amit Garg, Brian Haynes, Ann McKibbon, Nancy Wilczynski, Arthur Iansavitchous, Monika Kastner, Leslie Walters, Robert Yang, Faisal Rehman, and Charlie Goldsmith.

References

[1] McKibbon KA, Wilczynski NL, Haynes RB. What do evidence-based secondary journals tell us about the publication of clinically important articles in primary healthcare journals? *BMC Med* 2004;2:33.

[2] Garg AX, Iansavichus AV, Kastner M, Walters LA, Wilczynski N, McKibbon KA, Yang RC, Rehman F, Haynes RB. Lost in publication: Half of all renal practice evidence is published in non-renal journals. *Kidney Int* 2006;70:1995-2005.

[3] Birken CS, Parkin PC. In which journals will paediatricians find the best evidence for clinical practice? *Pediatrics* 1999;103:941-7.

[4] Crawley-Low J. Bibliometric analysis of the *American Journal of Veterinary Research* to produce a list of core veterinary medicine journals. *J Med Libr Assoc* 2006;94:430-4.

[5] Bero L, Rennie D. The Cochrane Collaboration. Preparing, maintaining, and disseminating systematic reviews of the effects of health care. *JAMA* 1995;274:1935-8.

Address for correspondence

Nancy Wilczynski, McMaster University, 1200 Main Street West, HSC 3H7, Hamilton, Ontario, Canada, L8N 3Z5
e-mail address: wilczyn@mcmaster.

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Aequus Communis Sententia: Defining Levels of Interoperability

Peter L. Elkin, MD[a], David Froehling, MD[a], Brent A. Bauer, MD[a], Dietlind Wahner-Roedler, MD[a], S. Trent Rosenbloom, MD[b], Kent Bailey, MD[a], Steven H. Brown, MD[b]

[a] *Mayo Clinic College of Medicine* [b] *Vanderbilt University*

Abstract

Interoperability is a common understanding of the meaning of data between a sending and receiving computer system [1]. The level of interoperability required varies with application needs. The specification of data in enough detail to create a common shared meaning between organizations is a complex task as systems work within organizational and human factors contexts as well as having specific technical requirements. Aequus communis sententia translates from Latin to the "level of common meaning." In this manuscript, we define an Ontology of Interoperability. The scale asks reviewers of a specification to define it's level in terms of Syntactic, Semantic and Pragmatic Interoperability. We tested the scale by having five medical Informaticians rate a set of ANSI standard specifications and we report the inter-rater variability of the interoperability rating scheme. We learned that some elements of the scale presented more difficulty for our reviewers and based on our findings we present a final version of the interoperability scale in our discussion. Our interoperability rating ontology has high inter-rater reliability and is a relatively simple mechanism for comparing the levels of interoperability afforded by different specifications or the same specification over multiple versions.

Keywords:

interoperability, standards

Introduction

Interoperability is one of the main goals of standardization for health information technology (IT) [2]. Although many Informaticians appreciate what would define an ideal state where we have perfect interoperability such that the receiver appreciates the information in the same context and meaning as the sender of the data. Levels of interoperability that fall short of this ideal have to date been only minimally defined. We believe that there are occasions when standards have been touted as supplying interoperability without actually accomplishing this lofty goal [3]. We contend that it would be useful to reviewers and users of standard specifications to know what level of interoperability is afforded by the specification.

Users of specifications, we believe, should be informed of what level of interoperability is afforded by any particular specification. Implementers of specifications could compare the functional requirements for interoperability for their applications with the level of interoperability afforded by any specification that they are contemplating implementing to ensure that their needs are well served by their choice of which standard to implement. This is implementation protective if the standard being considered would if implemented fall short of providing the level of interoperability needed by the functional requirements of their application. This method is cost protective if the standard contains significant unnecessary machinery to ensure a level of interoperability well beyond the functional requirements of the application in question.

As new versions of specifications are developed or released standards development organizations (SDOs) and implementers of specifications could be informed regarding how the changes in the new version of the standard effects the level of interoperability provided by each version of the standard.

Syntax, semantics and pragmatics

In 1938, Charles Morris published his seminal work dividing interoperability into three components [4]. Syntactic Interoperability deals with interoperable structures. Semantic Interoperability deals with the interoperability of a common shared meaning. Pragmatic Interoperability deals with the external constraints on the system. This last category takes into account the level of granularity needed for common understanding and the complexity or difficulty required to achieve a certain level of interoperability. Although Morris was referring to the Pragmatic Philosophers, we have extended this in our scale to address the practical side of standards development and implementation.

Levels of Interoperability

Many authors have discussed the benefits of Interoperability. Some like Hammond et al in 1998, have differentiated partial from full or true interoperability [5].

Interoperability		
	Syntactic Interoperability	
	a	Headings (e.g. Section of the clinical record)
	b	Select Fields Are Delimited, but relationships between fields are not specified
	c	4 plus data types are fixed and reliable
	d	5 plus numbers are broken out along with values (e.g. Blood pressure and values are diastolic and systolic values)
	e	6 plus Hierarchical structure of data without non hierarchical relationships between fields (e.g. XML structures)
	f	7 plus non-hierarchical relationships can be specified
	Semantic Interoperability	
	1	Free Text
	2	Free Text with fixed data types
	3	Codification of data by local codes
	4	Codification of data by nationally standard aggregate codes
	5	Codification of data by nationally standard detailed coding system allowing only pre-coordinated concepts
	6	Codification of data by nationally standard detailed coding system allowing both atomic and pre-coordinated concepts
	7	Codification of data by nationally standard detailed coding system allowing post-coordination (based on formal logic)
	8	Model based knowledge representation with local codes
	9	Model based knowledge representation with nationally standard aggregate codes
	10	Model based knowledge representation with nationally standard granular pre-coordinated codes
	11	Model based knowledge representation with nationally standard detailed coding system allowing both atomic and pre-coordinated concepts
	12	Model based knowledge representation coordinated semantically nationally standard detailed coding system allowing post-coordination (based on formal first order logic)
	13	Model based knowledge representation coordinated semantically nationally standard detailed coding system allowing post-coordination with support for context (based on formal higher order logic)
	Pragmatic Interoperability	
	α	Currently available and easily implemented
	β	Currently available but with barriers to implementation
	γ	Could be built and implemented within one year without barriers to implementation
		Could be built and implemented within one year with barriers to implementation
	ϵ	Could be built and implemented within three years without barriers to implementation
		Could be built and implemented within three years with barriers to implementation
	$\hat{\epsilon}$	Could be built and implemented within ten years
	λ	Would take longer than ten years to achieve
	μ	Not practically achievable
		Not possibly achievable

Table 1: Interoperability scale

The scale has three components. Syntactic interoperability ranges from simple headings to fixed and formatted hierarchically organized fields with possible structural links indicating non-hierarchical relationships between concepts. Semantic interoperability has as its simplest incarnation, free text, to the formal representation of knowledge using higher order logics which can fully support context. The authors believe that context is integral to how clinicians process information and will be important to achieving adequate computer based clinical decision support, which is one of the important goals of standards that deliver interoperable data. Pragmatic Interoperability speaks to the practical application of the standard. Here the levels go from currently available and easily implementable to impossible to achieve, advancing using a combination of how difficult it is to achieve and how long it is estimated to take to reach the goal of being practically implemented.

Methods

In order to evaluate the inter-rater reliability of the scale we had five Informatics professionals independently evaluate three standard specifications. The specifications chosen were Health Level 7 (HL7) v2.5, HL7 v3.0 [6] and the American Society of Testing and Materials (ASTM) Continuity of Care Record (CCR) specification [7]. These standards were chosen as they represent important national and international efforts that have taken somewhat different approaches to similar clinical problems (i.e. the representation of clinical laboratory data). Each Informatician rated each of these three standards by the three components of the interoperability scale.

For each scale, we simulated 50,000 "deals" of 15 cards into 3 hands of 5 each. The 15 cards dealt out were determined by the overall distribution of each scale. We then had the computer count the number of "pairs" in each hand and summed this over the 3 hands (3 of a kind counts as 3 pairs, 4 of a kind counts as 6 pairs, and 2 of a kind is obviously just 1 pair). In doing so we generated for each "deck", what the expected total pair count from a random deal would be. In addition to calculating where our observed pair count fell in the tail of the distribution (this is a one-sided testing situation), we also calculated the mean and SD of the pair count.

Results

The five reviewers showed excellent inter-rater agreement. The agreement with respect to Syntax and Pragmatics was greater than with Semantics. The agreement in the semantic category could be improved by conflating categories 5 and 6 and also conflating categories 10 and 11, as reviewers had particular difficulty making this distinction.

Although not directly studied, comments from the reviewers recommended simplification of the Pragmatic scale leading to the final recommended scale given below

HL7 v2.5	R1	R2	R3	R4	R5
Semantics	5	5	3	5	6
Syntax	d	d	d	d	d
Pragmatics	alpha	alpha	beta	alpha	alpha

Table 2 - Review data for HL7 v2.5

HL7 v3.0	R1	R2	R3	R4	R5
Semantics	11	11	4	10	11
Syntax	e	e	e	e	e
Pragmatics	delta	delta	beta	beta	beta

Table 3 - Review data for HL7 v3.0

ASTM CCR	R1	R2	R3	R4	R5
Semantics	6	6	3	5	6
Syntax	e	e	d	e	e
Pragmatics	beta	beta	beta	alpha	beta

Table 4 - Review data for the ASTM CCR.

Statistical analysis:

The agreement among the reviewers was analyzed as follows. For each scale the overall distribution of 15 responses was assumed fixed, and a probabilistic calculation was performed for how likely these 15 responses, when grouped by chance into 3 subsets (1 for each test case) would contain as many or more reviewers agreeing in a pairwise fashion. For each scale, the maximum number of pairwise agreements over the 3 cases was 10*3 = 30.

Case 1: Semantics, no grouping; Distribution of scale:

3	4	5	6	10	11
2	1	4	4	1	3

Number of pairwise matches = 9 (3 for each case).

The probability of at least 9 pairwise matches occurring in all, given this distribution, assuming ***random*** clustering of

responses, was p = 0.026. The expected number of pairwise matches was 4.6 with an SD of 1.7.

Case 2: Semantics, grouping 5/6, 10/11; Distribution of scale:

3	4	5/6	10/11
2	1	8	4

In this case, there were 18 pairwise matches in all (6 in each case). The chance of at least 18 pairwise matches occurring under random clustering, was P = 0.0042. The expected number of matches under random assortment was 10 with an SD of 1.9.

Case 3: Syntax; Distribution of scale:

d	e
6	9

The number of pairwise matches was 26 (out of a possible maximum of 30).

The chance of this good an agreement occurring by chance was p = 0.006. The expected number of pairwise matches in this case was 14.6 with an SD of 2.4

Case 4: Pragmatics; Distribution of scale:

Alpha	Beta	Epsilon
5	8	2

Number of pairwise agreements was 16 (6+4+6)

The probability by chance of obtaining at least 16 pairwise agreements was 0.054. The expected number of agreements was 11.2 with an SD of 2.2. Thus in this case only, we cannot rule out chance as accounting for the amount of agreement, if we use the 0.05 significance level.

Discussion

Here we present a revised (Table #5, below) scale for indicating the level of interoperability provided by any health informatics standard. The authors hope that ratings using this scale will help consumers of health informatics standard to better understand the level of interoperability provided by any particular specification. Further we believe that the use of this scale will help these same consumers, who are faced with the choice of which standards to implement, to compare the relative levels of syntactic, semantic and pragmatic interoperability provided by each of the specifications under review

Interoperability		**Table #5**
	Syntactic Interoperability	
	a	Headings (e.g. Section of the clinical record)
	b	Select Fields Are Delimited
	c	b plus data types are fixed and reliable
	d	c plus numbers are broken out along with values (e.g. Blood pressure and values are diastolic and systolic values)
	e	d plus Hierarchical structure of data without non hierarchical relationships between fields (e.g. XML structures)
	f	e plus non-hierarchical relationships can be specified
	Semantic Interoperability	
	1	Free Text
	2	Free Text with fixed data types
	3	Codification of data by local codes
	4	Codification of data by nationally standard aggregate codes
	5	Codification of data by nationally standard detailed coding system allowing both atomic and pre-coordinated concepts
	6	Codification of data by nationally standard detailed coding system allowing post-coordination (based on formal logic)
	7	Model based knowledge representation with local codes
	8	Model based knowledge representation with nationally standard aggregate codes
	9	Model based knowledge representation with nationally standard detailed coding system allowing both atomic and pre-coordinated concepts
	10	Model based knowledge representation coordinated semantically nationally standard detailed coding system allowing post-coordination (based on formal first order logic)

	11	Model based knowledge representation coordinated semantically nationally standard detailed coding system allowing post-coordination with support for context (based on formal higher order logic)
	Pragmatic Interoperability	
	α	Currently available and easily implemented
	β	Currently available but with barriers to implementation
	γ	Barriers could be overcome within one year
		Barriers could be overcome within three years
	ε	Barriers could be overcome within ten years
	ζ	Would take longer than ten years to achieve
	μ	Not practically achievable
		Not possibly achievable

Conclusions

Interoperability is essential for information about patients to be shipped from one computer to another in a reliable and computable manner. The information once transferred should be adequate to drive the local clinical decision support software, thereby helping to improve patient safety and optimize patient outcomes. We have presented a scale with good inter-rater agreement which can help implementers of healthcare standards to better understand the level of interoperability provided by standard specifications that they are considering implementing. This transparency, we believe, will help mitigate the risk of choosing a healthcare standard and in that regard will fuel adoption of standards in health IT solutions.

Acknowledgements

This work has been funded in part by grants from the National Library of Medicine LM06918 and LM008576. This work has also been funded in part by grants from the Centers for Disease Control PH000022 and HK00014.

References

[1] Brailer, D. J. (2005). "Interoperability: the key to the future health care system." Health Aff (Millwood) Suppl Web Exclusives: W5-19-W5-21.

[2] Hammond, W. E. (1994). "The role of standards in creating a health information infrastructure." Int J Biomed Comput 34(1-4): 29-44.

[3] Norgall, T. (2005). "Interoperability--a key infrastructure requirement for personalised health services." Stud Health Technol Inform 117: 125-33.

[4] Charles W Morris, Foundations of the Theory of Signs. International Encyclopedia of Unified Science Volumes 1 and 2 foundations of the unity of science, volume 1 number 2, Editor in chief, Otto Neurath University of Chicago Press 1938, Ninth impression 1957

[5] Hammond, W. E., D. L. Pollard, et al. (1998). "Managing healthcare: a view of tomorrow." Medinfo 9 Pt 1: 26-30.

[6] www.hl7.org

[7] http://www.astm.org/cgi-bin/SoftCart.exe/COMMIT/COMMITTEE/E31.htm?L+mystore+xynh7717

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

What's in a code?
Towards a Formal Account of the Relation of Ontologies and Coding Systems

Alan L Rector MD PhD

School of Computer Science, University of Manchester, Manchester, England M13 9PL

Abstract

Terminologies are increasingly based on "ontologies" developed in description logics and related languages such as the new Web Ontology Language, OWL. The use of description logic has been expected to reduce ambiguity and make it easier determine logical equivalence, deal with negation, and specify EHRs. However, this promise has not been fully realised: in part because early description logics were relatively inexpressive, in part, because the relation between coding systems, EHRs, and ontologies expressed in description logics has not been fully understood. This paper presents a unifying approach using the expressive formalisms available in the latest version of OWL, OWL 1.1.

Keywords:

knowledge representation, terminology, ontology, electronic health records, OWL

Introduction

Coding systems, such as SNOMED-CT [1] and the NCI Thesaurus [2] are increasingly being developed using "ontologies" represented in description logics or languages based on them such as OWL. Other groups, such as the Open Biomedical Ontologies (OBO) consortium in the basic biological sciences, are developing what they overtly describe as "ontologies" [1], many of which are implemented in OWL.

A major benefit of using description logics and ontologies to represent coding systems is purported to be the ability to infer logical equivalence between sets of codes and to classify codes automatically. However, that promise is still far from being realised routinely in practice. We suggest here that one reasons for the difficulty is that the relationship between "ontologies" and coding systems has not been clarified. We put forward here a procedure that clearly distinguishes between the ontology as a logical representation about the world and the code as a data structure, and show how different questions can be answered by each.

Throughout this paper we shall use OWL 1.1[2] as our representation language in the simplified Manchester syntax described in [3]. Exemplar ontologies are available on the Web.[3]

1 http://obo.sourceforge.net/

2 http://owl1_1.cs.manchester.ac.uk/

Requirements

The first question for this paper is: "What should a code represent?" Our basic requirements are to be able to express:

1. Individual "pre-coordinated" codes – *e.g.* the code for "head injury".
2. Clinical complexes common in many coding systems such as "head injury with/without intracranial bleed"[4].
3. Syndromes in two senses: a) well defined invariant combinations of conditions – *e.g.* tetralogy of Fallot – and ill defined variable combinations of symptoms – *e.g.* chronic fatigue syndrome.
4. Composite "post-coordinated" code expressions (what HL7 refers to as "code phrases").
5. The logical equivalence, or not, of alternative combinations of codes – *e.g.* tof identify "intracranial bleed" whether it occurs singly or as part of the complex "head injury with intracranial bleed".
6. Negation – definitely not having a condition and "absence" in the sense of negative findings such as an absent pedal pulse.
7. Formation of arbitrary "value sets" – sets of codes for use in particular situations.
8. The clinical dialogue.

Finally, we require that these requirements be met within a uniform logical framework for what constitutes a code, so that an inference engine or "classifier" can infer the subsumption hierarchy according to well specified semantics.

Framework: Data structures and ontologies

In a separate paper [4] we have argued that we need to consider information models at two levels:

- *Representations of the world* (or our conceptualisation of it) – "ontologies" and formal logical statements about patients, their disease and treatments, etc. The criteria for correctness is prediction of observations of the world.

3 http://www.cs.man.ac.uk/~rector/ontologies/whats-in-a-code/

4 In this paper we have sometimes substituted "bleed" for "haemorrhage" to conserve space in figures and formal expressions.

- *Models of data structures* – "information models" – which we use to specify which data structures are valid. The criteria for adequacy is that it sufficiently constrains data structures that those produced by one system can be correctly processed by another.

We argue that the task for information systems is: a) to begin with a representation of our understanding of the patient's situation in the world – the level of the ontology, b) to transform it into valid data structures – messages or EHR fragments – for storage and/or transfer to other systems, and then c) to re-interpret these data structures to derive representations of statements about the world – again at the level of the ontology Furthermore, we wish to perform these transformations with no, or only well defined, loss of information. These transformations are shown diagrammatically in Figure 1.

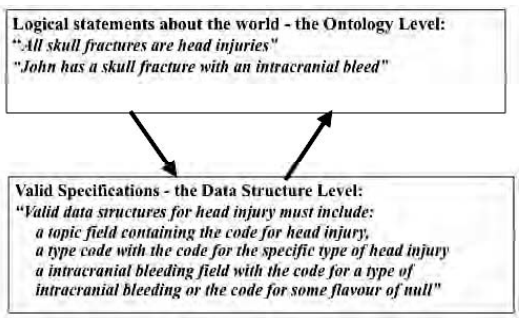


Figure 1 – Relation of ontology about the world and specification of valid data structures

Codes, data structures, and ontologies.

We argue that, although they may be derived from ontologies, "codes" are themselves data structures. An important reason to develop ontologies is to support coding systems, but the ontologies and the coding systems are distinct. Coding systems derived from ontologies may be thought of as "meta-models" of the underlying ontology – *i.e.* as models of the representation of the ontology in which each individual in the coding system represents the representation of a class in a particular formalisation of an ontology.[5] In hierarchical coding systems, the hierarchical relation – which we will here term "has_sub/is_sub_of" – reflects that subsumption (superclass-subclass) relationship in the underlying ontology.

The relationship between the ontology and the coding system derived from it is shown in Figure 2: each individual (dot) in the coding system represents a class (oval) in the ontology from which it is derived. The individuals (dots) in the ontology represent cases of the condition, or more precisely patient situations including those conditions.

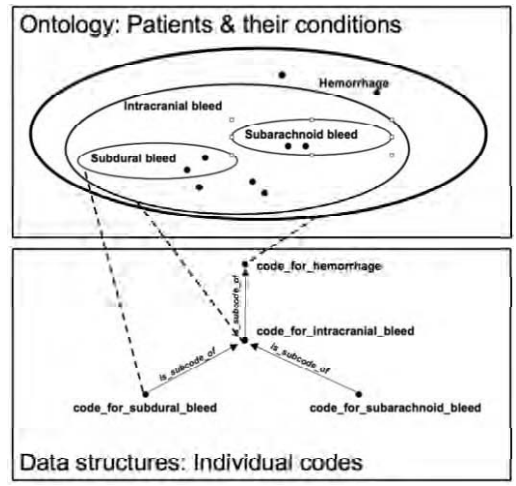


Figure 2 – Relation of ontology to codes. Each class in ontology corresponds to an individual code.

Note that the classes in the ontology and the codes in the coding system have very different characteristics. The ontology is "open". There are an indefinite number of possible subclasses of cases in the ontology – and an indefinite number of different ways of classifying the world. By contrast, the coding system is "closed". It consists of enumerated lists of codes, in some cases with enumerated lists of possible qualifiers. For example, in the above schema, there might be many descriptions for intracranial bleeds expressible in the ontology, but the only two for which we have codes in this limited coding system are "subarachnoid bleed" and "subdural bleed".

Ontological interpretation of codes

In discussing the ontological interpretation of codes, it is useful to deal with two cases which, following SNOMED, we shall term "findings" and "observables". "Findings" are things that apply to only some patients – *e.g.* diseases, injuries, symptoms etc. The existence of a "finding" is significant information, regardless of its detailed description. "Observables" are characteristics of all patients but with different values or states – values measured by laboratory tests, examinations, or other means. For example, only some patients have head injuries. The presence of a head injury is information. By contrast, all patients have a serum potassium concentration; the information is in the *value* of that concentration.

Case 1: Findings and clinical situations

An "ontology" in the narrow sense, represents the fundamental entities in a domain and the relations between them. In clinical medicine, these entities include anatomical structures, pathophysiological processes, organisms, cells, cognitive processes, etc. Authors such as Smith [5] advocate confining the word "ontology" to this narrow scope.

However, a major function of clinical knowledge is how these basic phenomena are organised into more complex entities with clinical significance – "macrocytic anaemia", "head injury without intracranial bleeding", "grade II stage

5 This is very roughly the relation of the SNOMED-CT "distribution form" to the underlying description logic form

1 carcinoma of the breast", etc. Clinical coding systems typically reflect this higher level of organisation. To ascribe a code to a patient is to say that the patient has (or in some cases does not have) the complex conditions indicated.

OpenGALEN used the term "ClinicalSituation" to define classes of such complexes [6] . They correspond roughly to what SNOMED-CT has termed "Context Dependent Entities" and it has recently rechristened "situations"[6] and are related to what Smith terms "Spans" [5].

Figure 3 gives examples of a simple representation of pathophysiological entities and classes of clinical situations in OWL. Note that many of the classes of clinical situation contain only a single condition. "Wrapping" single conditions in "Situations" in this way may seem redundant. However, it provides the uniform structure needed to support uniform automatic classification as required. To each of these classes of situations there corresponds a code, hi_s_code, sfx_s_code, etc. organised in a hierarchy that mirrors the subsumption hierarchy for the classes of situations as indicated in Figure 2.

HI_S = Situation THAT includes SOME Head_injury.
SFx_S = Situation THAT includes some Skull_fracture.
…
ICB_S = Situation THAT includes SOME Intracranial_bleed.
SDB_S = Situation THAT includes SOME Subdural_bleed.
HI_ICB_S = Situation THAT includes SOME Head_injury
AND
includes SOME Intracranial_bleed
…
Not_ICB_S =
Situation THAT not (includes SOME Intracranial_bleed).
HI_Not_ICB_S =
Situation THAT includes SOME Head_injury AND
NOT (includes SOME Intracranial_bleed).
…

Figure 3 – Example definitions of "situations" in OWL (Abbreviated names correspond to labels in Fig 5)

Case 2: Observables, codes and values

"Observables" are qualities of patients that are present in all patients and whose values or states are determined by observation – often by means of laboratory tests or physical examination. Typically, observables are represented by a "code-value pair". However, as often noted, an observable plus its value – *e.g.* "<Serum potassium, elevated>" – can be equivalent to a finding – "elevated serum potassium". We therefore propose an interpretation in the ontology that makes this equivalence apparent following the example in Figure 4:

Whereas for findings a single code is interpreted in the ontology as a "Situation", for observables, it is usually a code-value pair, *e.g.* "<serum_potassium_code, elevated_code>", that is interpreted as a "Situation". However, not uncommonly there is also a code assigned to the entire Situation, especially when a qualitative symbolic value such as "elevated" is involved, *e.g.* in Figure 4, the class of situations involving elevated serum potassium, ESP_S, and the corresponding code esp_s_code. To determine if the code-value pair and single code are equivalent merely requires interpreting each according to their meanings in the ontology and then using the reasoner to determine if the two meanings are equivalent.

Ontological Level:
ESP_S = Situation THAT includes SOME
(Serum_potassium THAT has_state VALUE elevated)
PQ5_S = Situation THAT includes SOME
(Serum_potassium THAT has_quantity VALUE
[5.1 mMolPerL]
Corresponding coded representation:
For code-value pairs: <serum_potassium_code, elevated_code>
For Situations: pq5_s_code, esp_s_code

Figure 4 – Ontological representation and corresponding codes for example observable

Consequences: Addressing the requirements

Of the requirements in the introduction, the proposed framework meets requirements 1-3 directly. As indicated in the examples in Figure 3, pre-coordinated codes, complexes, and syndromes are treated uniformly. The extension to qualifiers and post-coordinated codes (requirement 4) is straightforward and omitted for reasons of space. Requirements 5-8 are more subtle and are discussed below.

Classification and equivalence (requirement 5)

Is a patient who is assigned the code-value pair "<serum_potassium_code, elevated_code>" the same as a patient assigned the single finding code "esp_s_code"? Is a patient who is assigned separately the codes for "Situation THAT includes SOME Head_injury" (hi_s_code) and "Situation THAT includes SOME Intracranial_bleed" (icb_s_code) separately equivalent to a patient that has assigned the single code for "Situation that includes some Head_injury AND includes SOME Intracranial_bleed" (hi_icb_s_code)?

In the proposed framework, all such questions are answered by re-interpreting the codes as representations in the ontology and then comparing these representations, using an appropriate classifier where necessary. In the case of the equivalence of a code value pair and the corresponding finding code, the answer is obvious, since the interpretations as expressions in the ontology are identical – see Figure 4.

In the case of comparing several separate findings of single conditions with a single finding of those conditions combined, the answer follows naturally from the notion of a "clinical situation". If each patient can have only one "situation" at any one time (perhaps as observed by a given observer), then we need only form the conjunction of the criteria and compare the result. This can be done manually for simple lists, or the inference engine can be used for more complicated cases.

6 Kent Spackman, Personal communication, 2006.

For example, to determine if patients with "head injury" and "no intracranial bleed" coded separately are equivalent to patients with the single code for "head injury without intracranial bleed", first interpret the codes to describe a patient or class of patients at a particular time:

(Patient THAT at_time SOME Time_point)
has SOME
(Situation THAT includes SOME Head_injury) AND
has SOME
(Situation THAT NOT (includes SOME Intracranial_bleed).

Because a patient can have only one situation at one time, the classifier will recognise that the second Situation is redundant and find that this is logically equivalent to:

Patient at_time SOME Time_point)
has SOME
(Situation THAT includes SOME Head_injury AND
NOT includes SOME Intracranial_bleed).

The axiom that each patient at a given time can have only one situation is captured by the generic axiom in OWL:

(Patient at_time SOME Time_point) has MAX 1
Situation

Note that this axiom holds at the level of the ontology but not at the level of codes. It is not true that a patient can be ascribed only a single code in an EHR or message. It is true that there can be only one clinical situation for a given patient at a given time (as determined by a single observer).

Dealing with negation (requirement 6)

Negation of a finding – "not any"

Many codes represent classes of situations that involve a patient *not* have certain findings. Examples in Figure 3 include "No intracranial bleed" (Not_ICB_S) or "Head injury without intracranial bleed" (HI_Not_ICB_S). The classes of findings in these cases would best be defined by analogy with "patients who do not have *any* intracranial bleed".

Correct classification of negation manually without formal inference is difficult. The result of applying the classifier to an extended set of definitions based on Figure 3 is shown in Figure 4. This achieves the correct results automatically. For example, "No intracranial bleed" (Not_ICB_S) is a kind of "No subdural bleed" (Not_SDB_S) rather than vice versa. This is an example of the rule that negation inverts the kind-of hierarchy. If "B is a kind of A" – *i.e.* "all Bs are As" – then "NOT A is a kind of NOT B" – *i.e.* "all non-As are non-Bs".

However, note that the negation in Figure 3 applies to the entire criterion "includes SOME Intracranial_bleed" rather than to "Intracranial_bleed" itself. To say "Situation THAT not includes SOME Intracranial_bleed" means "the situation does not include *any* intracranial bleed", as intended. By contrast, to say "Situation THAT includes SOME NOT Intracranial_bleed" is to say that "the situation includes *something that is not* an intracranial bleed" – a different statement altogether.

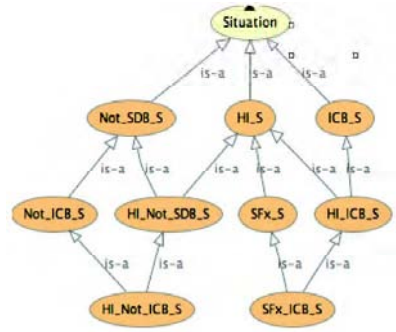


Figure 5 – Fragment of logical classification of complexes in Figure 3as determined by the OWL classifier (abbreviations follow pattern of Fig 3).

Negative findings and negative observables – "some not"

In most cases negation follows the example of intracranial bleeding above. However, there is a group of what are often referred to as "negative findings" that cause confusion.

The classic example is "absence of pedal pulse". Such rubrics are fundamentally ambiguous. Does it mean "Absence of any pedal pulse" or "Absence of some pedal pulse". In the first case, we would expect it to imply the absence of all pedal pulses – *e.g.* "absence of dorsalis pedis pulse", "absence of posterior tibial pulse", etc. We would therefore expect the hierarchy to be inverted as for types of intracranial bleeding. "Absence of any pedis pulse" would therefore fall under "absence of dorsalis pedis pulse" in the hierarchy.

However, in the second case, if we mean "Absence of some pedal pulse", we would expect the reverse. The "absence of the dorsalis pedis pulse" is certainly an example of the "absence of *some* pedal pulse", so we would expect "absence of dorsalis pedis pulse" to fall under "absence of some pedal pulse". The hierarchy would not be inverted.

The cleanest way to deal with this case is to use the ontological interpretations as shown in Figure 6. The first, for the absence of *any pedal pulse,* is analogous to the usual negation of situations in Figure 5. The second, for the absence of *some* pedal pulse, deals with the special case of "negative findings".

Absence_of_any_pedal_pulse = Situation THAT NOT (includes SOME Pedal_pulse).
Absence_of_some_pedal_pulse = Situation THAT excludes SOME Pedal_pulse
Preferred property axiom: excludes = NOT includes
OWL 1.1 property axiom: DISJOINT excludes includes

Figure 6 – Absence of any vs Absence of some

The scope of the negation in the second case is critical. In the definition of "Absence_of_some_pedal_pulse" the negation is included in the definition of the property, "excludes" and affects just that property, whereas in the definition of "Absence_of_any_pedal_pulse" it negates the entire restriction "includes SOME Pedal_pulse".[7]

Unfortunately, no current description logic implements the constructor for negating properties, although it is known to be tractable [7]. OWL 1.1 implements a slightly weaker construct, disjoint properties, as shown in the "*OWL 1.1 property axiom*", which gives the same results in except in cases involving double negation.[8]

Theoretically, the OWL 1.1 solution is the best approximation that can be implemented currently, but because it has just become available, there is little experience with it in practice. An alternative construct, available in most formalisms with which there is more experience, is to regard "having a pedal pulse", "having a dorsalis pedis pulse" etc. as "observables" with possible values "detectable" and "NOT detectable" as shown in Figure 7. This leads to correct classification but is arguably less faithful to the intended meaning.

```
Absence_of_some_pedal_pulse =
 Situation THAT includes SOME
  (Having_pedal_pulse THAT has_state SOME (NOT
Detectable))
```

Figure 7 – Alternative representation for negative findings

Value sets and faithfulness to the clinical dialogue (requirements 7 and 8)

Whereas equivalence of meaning can only be addressed by re-interpreting the information structures, including codes, into the ontology, specifying valid value sets and a faithful representation of the actual clinical statements made can only be made in terms data structures and codes themselves.

Value sets are closed lists of codes or tightly specified code phrases. While we can talk about requiring "only the code for head injury and none of its subcodes", at the ontological level, we cannot talk about the "all cases of head injury" without including all cases of all kinds of head injury. The process of specifying code sets and binding code sets to health records is discussed in detail in a separate paper [4].

Similarly, on the one hand, we want the *meaning* for a patient of the two statements – "This patient has a head injury" and "this patient has an intracranial bleed" – to be the same as the meaning for a patient of a single statement combining the two conditions. On the other hand, for purposes of clinical responsibility and documenting the clinical dialogue, we want to keep track separately of each statement made about the patient in the form in which it was made. For example we may want to record who made which statement, when, why, etc. This can only be done at the level of the codes and data structures where the statements are distinct, not at the level of the ontology where their meanings are indistinguishable.

7 In standard predicate logic notation, the difference is between ∃y. PP(y) & includes(s,y) and ∃y . PP(y) & includes(s,y).

8 The difference between "includes = NOT excludes" and "DISJOINT includes, excludes" is that in the case of true negation, we can infer that something not included is excluded, where in the case of mere disjointness we cannot.

Conclusion

This paper presents two key ideas: a) That codes should be regarded as individual data structures which can be interpreted in terms of meanings in a separate ontology about patients and their conditions, and b) that the ontology should be structured in two layers: a layer of kernel concepts – conditions, anatomy, etc – and a layer of "clinical situations" which describe classes of patient states at a particular times as viewed by particular observers. It suggests that both codes for "findings" and code-value pairs for "observables" should be interpreted as representing classes of situations in the ontology. It shows how these notions can be used to provide a unified framework for dealing with the questions of equivalence of the meaning of negation. An account of patients' situations, the conditions for validity of data structures conveying information on those situations, and the dialogue about those patients' care requires taking account of both the level of the ontology and the level of data structures and codes.

Acknowledgements

This work supported in part by the UK Department of Health "Connecting for Health" programme, the UK MRC CLEF and CLEF-Services projects (G0100852), the JISC and UK EPSRC projects CO-ODE and HyOntUse (GR/S44686/1) and the EU Funded Semantic Mining Network of Excellence. The HL7 Terminfo working group stimulated and contributed to many of the ideas presented here.

References

[1] Stearns M, Price C, Spackman K, Wang A. SNOMED clinical terms: overview of the development process and project status. In: *AMIA-2001*. Henley & Belfus; 2001. p. 662-666.

[2] Golbeck J, Fragoso G, Hartel F, Hendler J, Oberthaler J, Parsia B. The National Cancer Institute's thesaurus and ontology. J Web Semantics 2003;1(1):75-80.

[3] Horridge M, Drummond N, Goodwin J, Rector A, Stevens R, Wang H. The Manchester OWL syntax. In: Cuenca Grau B, Hitzler P, Shankey C, Wallace E, editors. *OWL: Experiences and Directions (OWLED 06)*; 2006; Athens, Georgia: CEUR; 2006. p. http://sunsite.informatik.rwth-aachen.de/Publications/CEUR-WS//Vol-216/submission_9.pdf.

[4] Rector A, Qamar R, Marley T. Binding ontologies & coding systems to electronic health records and messages. In: Bodenreider O, editor. *Formal Biomedical Knowledge Representation (KR-MED 2006)*; 2006; Baltimore: CEUR; 2006. p. 11-19.

[5] Smith B. The logic of biological classification and the foundations of biomedical ontology. In: Westerstahl D, editor. 10th In *Int Conf on Logic Methodology and Philosophy of Science; 2004*; Oviedo Spain: Elsevier-North-Holland; 2004.

[6] Rector AL, Rogers JE. Ontological and practical issues in using a description logic to represent medical concept systems: Experience from GALEN. In: Barahona P, Bry F, Franconi E, Henze N, Sattler U, eds. *Reasoning Web*. Heidelberg: Springer-Verlag; 2006. p. 197-231.

[7] Lutz C, Sattler U. Mary likes all Cats. In: International Workshop in Description Logics (DL2000); 2000; 2000. p. 213-226.

Address for correspondence

Alan Rector, School of Computer Science
Email: rector@cs.manchester.ac.uk

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

A Road from Health Care Classifications and Coding Systems to Biomedical Ontology: the CEN Categorial Structure for Terminologies of Human Anatomy: Catanat

JM Rodrigues[a], C Rosse[b], M Fogelberg [c], A Kumar[a] , B Trombert Paviot[a]

[a] *University of Saint Etienne, Department of Public Health and Medical Informatics, France*
[b] *University of Washington, Department of Medical Education and Biomedical Informatics, USA*
[c] *University of Linköping, Department of Biomedical Engineering, Sweden, convenor of CEN TC 251 WG 2*

Abstract

There is an increasing need for updated and harmonised health care classifications and coding systems to allow international comparisons and cooperation for instance for population based WHO indicators, Electronic Health Record safety, trans border migration of population, case mix and procedure payment. It is not feasible to propose a standardisation of the linguistic expressiveness of different health care professionals and of the citizens. Natural language expressions show inconsistencies and ambiguities as assessed by biomedical ontology driven tools. In order to build a road towards standardisation the European Standard Body CEN has stated that it was not possible to convince the different European member states using different national languages to agree on a reference clinical terminology as Snomed-CT or to standardise a detailed language independent biomedical ontology. It has developed since 1990 an approach named categorial structure as a step standardising only the terminologies model structure. The categorial structure for terminologies of human anatomy currently in the phase of final approval is presented as a methodology for bridging between classifications and coding systems and biomedical ontology on the way to semantic interoperability between different languages, legacies and goals.

Keywords:

classification, coding system, terminology, ontology, standards, categorial structure, anatomy

Introduction

Interoperability – the fact that data produced in one Information and Communication Technology (ICT) system is readily useable in another ICT system – has by now matured to provide reliable operational platforms for technical and functional aspects. In healthcare or for e-Health, content interoperability now emerges as a top challenge under the name of semantic interoperability.

If we defined semantic interoperability as ìthe ability for information shared by systems to be understood at the level of formally defined entities, so that the receiving system can process the information effectively and safelyî [1] it is easy to understand that it is a major issue for most health care organisations and countries.

There is an increasing need to address national and international comparisons and cooperation across the most advanced and less developed organizations or countries for instance for population based WHO indicators, Electronic Health Record safety , trans border migration of population, case mix and procedure payment et.

Unfortunately, clinical terminological systems, classifications and coding systems have been developed by independent, divergent and uncoordinated approaches which have produced non reusable systems on overlapping fields for different needs. For some decades several broad pre-coordinated or compositional systems have been proposed to users targeting different goals. The most well known are the UMLS (Unified Medical Language System) [2], SNOMED international [3], Read Clinical Classification Version 3 [4], LOINC [5] for clinical laboratories, DICOM SDM [6] for imaging, SNOMED CT [7] and Convergent Medical Terminology (CMT) [8].

On the other hand, most of developed countries have kept on maintaining, updating and modifying their own coding systems for procedures, as well as national adaptations of ICD, in order to manage and to fund their health care delivery. The most significant efforts were done in Australia with ACHI (Australian Classification of Health Interventions) or ICD10-AM [9], in Canada with the Canadian Classification of Health Interventions (CCI) [10] developed by the Canadian Institute for Health Information (CIHI) and in France with CCAM (Classification Commune des Actes Médicaux) [11][12].

Natural language expressions show inconsistencies and ambiguities as it can be assessed by biomedical ontology driven tools [13] which have been enabled by advances in computer science, artificial intelligence, linguistics and philosophy. These are based on terminology server architecture making available knowledge bases representing multi-hierarchies of concepts associated by semantics subsuming the logical meaning of their relations and automated language generation. These knowledge repre-

sentations are called *biomedical ontologies* [14] [15], – a term coming from metaphysics – but are in fact formal logic using mathematical expressions.

These representations have been supported by specific software tools and natural language processing techniques, including the linkage of lexicons from different national languages [16] to the knowledge representation. The most important achievements are GALEN (Generalised Architecture for Languages, Encyclopaedias and Nomenclatures in Medicine) [17], FMA (Foundational Model of Anatomy) [18] [19]. For biomedical terminologies and OBO (Open Biomedical Ontology) which provides an interesting training process [20].

On the other hand, the standardisation in health informatics started in the US with the HL7 user group and was formalised by 1990 in Europe with CEN (ComitÈ EuropÈen de Normalisation) TC 251 and internationally in 1998 with ISO (International Standard Organisation) TC 215. The process aimed to the electronic health record of the patients (EHR) has opened the way to divergent initiatives encompassing information models (HL7 Reference Information Model RIM, CEN Continuity of care, et.) messages, architecture (HL7 Common Document Architecture CDA, CEN EHRcom, et.), security, context (Templates or Archetypes, et.) and the forest of ìpragmaticî or ìreenceî clinical terminologies. Finally WHO has initiated the revision of ICD which will take advantage of such achievements to enhance the interoperability of products of the family of international classifications (ICD, ICF and classification of procedures) by the year 2011.fer

We address the standard approach for biomedical terminology elaborated and developed by the European Standard Body CEN TC 251 WG2 (Comité Européen de Normalisation Technical Committee 251 Working Group 2) named categorial structure in part 2 and the last standard in the state of final approval on terminologies of human anatomy [21] in part 3 .We finally discuss the role of this standard approach on the road from bridging between classifications and coding systems of biomedical terminology to a more complete semantic interoperability based on a shared biomedical ontology.

CEN Categorial Structure Standard approach

Rationale

As a prerequisite it was clearly stated at the beginning of the CEN standard process that it was not realistic to impose within Europe a standard terminology to health care professionals even when saying it is a reference terminology or a pragmatic terminology. The 2 top arguments supporting this point were that European countries are speaking different natural languages and that different health care professionals within the same natural language are not using the same meanings for the same terminology statements. The increasing role of citizens in health care has hampered this point for they currently use natural language different statements to mean the same things.

The CEN categorial structure standard approach was based on advances in computer sciences, artificial intelligence, linguistic and philosophy allowing to define a new approach to support unambiguous exchanges of meaning through computers named biomedical ontology [14-15].

We have proposed in a FP 6 European Union seminar [22] to define biomedical ontology as: a representation of some pre-existing domain of reality which

1. reflects the properties of the entities within its domain in such a way that it obtains a systematic or according to rule correlation between reality and the representation itself and the non ambiguous comparison between different formal representations of the reality
2. is intelligible to a health care domain expert,
3. is formalised in a way that allows it to support automatic information processing.

CEN has considered not possible to standardise the clinical knowledge in an only one biomedical ontology which could support one reference clinical terminology or different types of clinical terminology due to the volume of terms ,the quickly evolving size and the international and not only European production and use of clinical terminologies.

Definition

As a consequence the CEN categorial structure was defined within some linguistic variations [23-25], as a minimal set of health care domain constraints to represent a biomedical terminology in a precise health care domain with a precise goal to communicate safely.

It is a definition of a minimal semantic structure describing the main properties of the different artefacts used as terminology (controlled vocabularies, nomenclatures, coding systems and classifications), a model of knowledge restricted to:

1. a list of semantic categories
2. the goal of the categorial structure
3. the list of semantic links between semantic categories authorised with their associated semantic categories and
4. the minimal constraints allowing the generation and the validation of well formed terminological phrases.

This methodology has been applied by CEN for the coding systems and classifications of surgical procedures [24], clinical laboratory [26] and by ISO for nursing diagnostics and procedures [27]. The standardisation processes are on going for medical devices, medicinal products, continuity of care and anatomy.

The categorial structure for terminologies of human anatomy: Catanat

To conform with the CEN standard terminologies of human anatomy we shall describe the 4 elements of an anatomical categorial structure.

The Anatomical Categories

The anatomical categories are defined as types of anatomical entity shared by all the individual instances in existence in the present, past and future.

For example, the anatomical category liver is instantiated by this liver and all individual livers in existence in the present, past and future.

Anatomical categories may be more or less general. Where one anatomical category is subsumed by another, the "is_a" relation is asserted to obtain between the more specific and the general anatomical entity [21].

To conform with the CEN standard terminologies of human anatomy we shall use as root one of the 19 anatomical categories defined in [21] and adapted from the FMA [18-19] and the hierarchy defined in [21] and adapted from the FMA [18-19] as well.

The following 19 anatomical categories are defined within their hierarchy in [21]: Physical anatomical entity, Immaterial physical anatomical entity, Anatomical space, Anatomical surface, Anatomical line, Anatomical point, Material physical anatomical entity, Body substance, Anatomical structure, Cell, Organ, Cardinal organ part, Portion of tissue, Cardinal body part, Body region, Organ systems, Anatomical cluster, Anatomical set, and Anatomical junction.

We give here some examples from [21].

Physical anatomical entity:

anatomical entity which has a spatial dimension e.g. organ, surface, apex of the orbit.

Immaterial physical anatomical entity:

a physical anatomical entity which has no mass e.g. diaphragmatic surface of left ventricle.

Anatomical space:

immaterial physical anatomical entity which has spatial dimension of value 3 e.g. thoracic cavity.

Anatomical line:

immaterial physical anatomical entity which has spatial dimension of value 1 e.g. inferior margin of liver.

Anatomical point:

immaterial physical anatomical entity which has spatial dimension of value 0 .Apex of heart.

Material physical anatomical entity:

physical anatomical entity which has a mass e.g. liver, cell nucleus.

Body substance:

material physical anatomical entity which has no inherent shape e.g. portion of blood.

Anatomical structure:

material physical anatomical entity which has inherent shape and is generated by coordinated expression of the organism's own structural genes e.g. thorax, tibia.

Organ

anatomical structure, which consists of a maximal collection of cardinal organ parts so connected to one another that together they constitute a self-contained unit of macroscopic anatomy, morphologically distinct from other such units e.g. heart, urinary bladder.

Cardinal organ part

anatomical structure, which consists of two or more portions of tissue, spatially related to one another in patterns determined by coordinated gene expression; together with other contiguous cardinal organ parts it constitutes an organ e.g. upper lobe of right lung, shaft of humerus.

Cardinal body part

anatomical structure, which has as its parts the most complete set of diverse subclasses of organ and cardinal_organ part spatially associated with either the skull, a segment of the vertebral column or a complete set of bones of the appendicular skeleton; it is partially surrounded by skin and forms a distinct morphological subdivision of the body; together all cardinal body parts constitute the body e.g. head, neck, trunk, upper limb.

Body region

sub volume of a cardinal body part demarcated by at least one fiat boundary e.g. femoral triangle, epigastrium.

Organ systems

anatomical structure which consists of organs predominantly of the same anatomical category, which are interconnected by zones of continuity e.g. alimentary system musculoskeletal system. Organ systems shall be distinguished from body systems based on function as endocrine or immune system.

Anatomical cluster:

anatomical structure which consists of a heterogeneous set of organ parts grouped together in a predetermined manner, but which do not constitute the whole or a subdivision of either a body part or an organ system e.g. joint, adnexa of uterus, renal pedicle..

Anatomical set

material anatomical entity which consist of the maximum number of discontinuous members of the same class e.g. set of cranial nerves, thoracic viscera.

Anatomical junction

anatomical structure in which two or more anatomical structures are in physical continuity with one another or intermingle their component parts.

Precise goal of the anatomical categorial structure

The goal of each anatomical terminology used in the terminology systems of health care and biomedical science shall be defined by the users to explicit the scope and the content of the categorial structure. This point must at least mention the situations and applications for which the categorial structure is intended and a statement on the limits of use, for example controlled vocabulary production for clinicians or comparison with another terminological system for coding centres.

List of the representations of anatomical relations between anatomical categories authorised by anatomical domain constraints

The representations of anatomical relations which shall be used when necessary to be conformant with the CEN standard are the 8 following anatomical relations plus the Is_a

anatomical relation defined in [21] and adapted from the FMA [18] [19].

has_part

Anatomical relation which holds between each anatomical entity of one to three dimensions in category A and some anatomical entity of the same dimension in category B such that if A has_part B there is a complement C which together with B accounts for the whole of A; for example, Stomach has_part fundus. Together with body and pyloric antrum, fundus accounts for the whole of Stomach.

A contained_in B

Anatomical relation which holds between each anatomical entity in category A contained in some anatomical entity in category B. The former is a body substance or an anatomical structure; the latter is an anatomical space. For example Urinary bladder contained_in pelvic cavity. Contained_in does not imply part_of. Although cavity of urinary bladder is part_of urinary bladder, urine part_of urinary bladder is an invalid assertion.

A adjacent_to B

Anatomical relation which holds between each anatomical entity in category A is adjacent to some entity in category B. Two anatomical entities of the same dimension are adjacent when they are spatially proximate, share no boundary or parts, and are separated by no further anatomical entities of the same dimension. For example Spleen adjacent_to stomach, Posterior surface of kidney adjacent_to anterior surface of quadratus lumborum

Adjacent_to which is adirectional may be qualified with the aid of qualitative anatomical coordinates such as anterior_to, posterior_to, superior_to, inferior_to, medial_to, lateral_to depending upon the value of the trajectory relationship; These qualifiers are directional and asymmetric.

A continuous_with B

Anatomical relation which holds between each anatomical entity in category A and some entity in category B when there is no bona fide boundary (real physical discontinuity) between the related entities and their parts. For example Arterial trunk continuous_with branch of arterial trunk

A attached_to B

Anatomical relation which holds between each anatomical entity in category A and some entity in category B when some of the parts of the entity in category A are continuous with some entity in category B across a portion of their maximal boundary which the related entities share. For example each patellar ligament is attached_to the patella at a narrow area along the lower margin of the latter and also to the tuberosity of the tibia, or the circumference of the tympanic membrane is attached_to bones of the skull forming the external auditory meatus.

has_dimension

Anatomical relation which relates an anatomical entity to the number of its spatial dimension; for example wall of stomach has_dimension 3

has_shape

Anatomical relation which relates an anatomical entity to its three dimensional shape. For example oesophagus has_shape hollow cylinder

has_boundary

Anatomical relation which relates categories of anatomical entities of one to three dimensions to categories of immaterial physical anatomical entities of one dimension lower, called bounding anatomical entities. Such bounding entities delimit anatomical entities of one, two or three dimensions from one another. A boundary may be bona fide or fiat. A bona fide boundary of an anatomical structure is a real physical discontinuity. A fiat boundary is a virtual plane or line such as those that demarcate the oesophagus from the stomach. For example cavity of stomach has_boundary internal surface of stomach.

The practical application of boundary information is critical for example to processes of automatic image segmentation and to the analysis of volumetric datasets.

List of minimal anatomical domain constraints

The list shall contain the different anatomical relations and the different related anatomical categories which are valid and necessary for the precise goal of an anatomical categorial structure.

As a summary to be conformant to the CEN standard any biomedical terminology using a terminology of human anatomy shall explicit its anatomical categorial structure which shall provide the information described in 1, 2, and 3 within part **22** and shall be conformant to the following three rules:

1. An anatomical categorial structure shall have as root node any anatomical entity from one anatomical category listed in this CEN European standard [21].
2. An anatomical categorial structure shall make precise the level of granularity of the classes of anatomical categories used as described in this CEN European standard [21].
3. An anatomical categorial structure shall use when necessary the anatomical relations only as described in this CEN European standard [21].

Discussion

- The first issue to address on the road from divergent initiatives on health care classifications and coding systems to semantic interoperability is the role of standard.

There are at least 3 different ways to address the standard issues.

The first way can be called the way of the Family of classifications used since the nineteenth century by the International Classification of Diseases (ICD) and now by the WHO Family of International Classifications (FIC). This is the rationale behind the ICD eleventh revision and the International Health Terminology (IHT) Standard Development Organisation (SDO) [28] currently under creation and based on Snomed CT.

Unfortunately this model faces a lot of drawbacks Let us stress the most important ones.

First the size of knowledge is increasing so quickly with biomedical research and technology developments that maintenance and updating is huge and never quality assured. It is a bottomless permanent work based on divergent individual and time related expert opinions when there is no open and transparent access to a commonly agreed and shared knowledge representation.

On the other hand the multilingualism is a major constraint and multi-translation a very important effort not only across national languages but as well across different specialists and researchers using one type or another of biomedical terminology for different situations.

The second way can be called the biomedical formal ontology and the biomedical ontology tools coordinated with natural language processing and web based tools[17,19-20]. This way provides a new perspective to semantic interoperability by de-multiplying the workload organised by disseminated social computing.

Unfortunately this perspective is at its beginning and can only fulfil the needs step by step in specific areas [13,20].

Between these 2 trends the international standardisation organisations have the responsibility to support or to build the road to increase semantic interoperability as defined in [1].

This has been the choice of the CEN European standard body within Technical Committee 251 and ISO Technical Committee 215 for information model, network infrastructure, architecture and messages, security and terminology. For terminology CEN has proposed a third way between the Family of classifications and the biomedical formal ontology named the Categorial Structure:

The Categorial structures as presented in part 2 cannot be considered separately from the other standards of this framework.

- The second issue is how this type of standard can work.

The categorial structure CEN standard for terminologies of human anatomy shall be used for any biomedical terminology using anatomical terms under the framework of interoperable communications between different ICT or e-Health systems including an Electronic Health Record (EHR).

It must be communicated by the e-Health sender system to the e-Health receiver system. The template sent shall contain the 4 parts of the categorial structure as developed earlier and more precisely in [21].

The intention is not to insure a full semantic interoperability defined in [1] as if Users of System B are able to use information acquired automatically from System A with equivalent meaning to its local data.

The intention is to reach an intermediate interoperability defined in [1] as if System A is able to send information automatically to system B, but System B (and Users B) can only interpret the meaning from the perspective of users A and System A.

As a consequence any biomedical terminology using the FMA (model and terms) is conformant to the CEN standard but any biomedical terminology based on a categorial structure defined as in [21] by a subset of FMA will be as well. This choice faces the criticism that the anatomical relation representations are not exhaustive and that a biomedical terminology may need to use additional relations not part of the standard.

Of course the alternative should be to use the whole FMA (model and terms) as the standard.

The Foundational Model of Anatomy (FMA), initially developed as an enhancement of the anatomical content of UMLS, is the biomedical domain ontology of the concepts and relationships that pertain to the structural organization of the human body. It encompasses the material objects from the molecular to the macroscopic levels that constitute the body and associates with them non-material entities (spaces, surfaces, lines, and points) required for describing structural relationships. The disciplined modelling approach employed for the development of the FMA relies on a set of declared principles, high level schemes, Aristotelian definitions and a frame-based authoring environment. The FMA is applied as the reference ontology in biomedical informatics for correlating different views of anatomy, aligning existing and emerging ontologies in bioinformatics and providing a structure-based template for representing biological functions.

On one hand CEN is not considering to be entitled by its charter of multinational and multilingual European organisation to dictate the terms to be used by health care professionals, administrators and citizens within Europe. On the other hand this new categorial structure is more detailed, precise and prescriptive than the previous ones [23-27] and for the first period of the standard (5 years) it has been decided that the incentive to share the same anatomical terminology by an increase use of at least a part of FMA was sufficient for the time being.

- The third issue is the relation between the CEN standard categorial structures and a biomedical ontology as defined in [22].

They fulfil the point 2 and partly the point 1 of the biomedical ontology definition by giving the categories and the relations between the categories as we have shown earlier for Anatomy but they do not allow a systematic (i.e. according to the precise rules of description logic) correlation between representations and between the representation and the reality. Finally they are not formalised enough to support automatic information processing. For these reasons they must be considered as a milestone on the road to more complete semantic interoperability.

- The last issue is how this categorial structure approach can facilitate the road towards formal ontology developments and applications.

The first categorial structures were defined for surgical interventions and clinical laboratory measurements which are occurrent events

The principles, anatomical categories, part relations and other structural relations adopted for the construction of the categorial structure for terminologies of human anatomy have been largely derived from the Foundational Model of Anatomy (FMA) [18]. This categorial structure [21] is a sensible move for it is a continuant present in most of biomedical and clinical terminologies for instance in surgical procedures terminologies [13]. It is intended to train more and more people to base their biomedical terminologies on a well formed biomedical ontology for Anatomy to decrease inconsistencies and ambiguity.

On the other hand this minimal structure separated from terminology artefacts gives a common framework to analyse them, their comparison, alignment, their gaps and inconsistencies. It is proposed as a methodology to demonstrate the need of a common worldwide full set of biomedical ontologies necessary to achieve a full semantic interoperability.

Acknowledgements

We wish to thank our partners in the CEN TC251 WG2 and terminology and biomedical ontology experts who commented and proposed useful modifications to Catanat and namely Alan Rector from University of Manchester and Barry Smith from the University at Buffalo.

References

[1] SemanticHEALTH Deliverable 1.1. Conceptual Framework for e-Health Interoperability. European Union Commission; Brussels 2006.

[2] McCray AT, Nelson SJ. The representation of meaning in the UMLS. Methods Inf Med 1995; 34(1-2):193-201.

[3] Iggulden P,Price C. SNOMED Clinical Terms ñ Looking forward to the next phase Br J Healthcare Comput Info Manage 2001; 18(10):20-22.

[4] Read Clinical Classification (RCC, Read Codes). Updated by NHS Centre for Coding and Classification. Crown copyright, 1990.

[5] Logical Observation Identifiers Names and Codes(LOINC). See :http://www.loinc.org/

[6] DICOM see http://www.xray.hmc.psu.edu/dicom/dicom_home.html

[7] SNOMED Clinical Terms. College of American Pathologists.see http://www.snomed.org/

[8] Dolin RH. Kaiser Permanente's Convergent Medical Terminology. Testimony to the National Committee on Vital and Health Statistics, Subcommittee on Standards and Security. May 22, 2003. http://ncvhs.hhs.gov/030522sstr.htm]

[9] National Centre for Classification in Health see http://www3.fhs.usyd.edu.au

[10] See the Canadian Classification of Health Interventions (CCI) home page at: http://secure.cihi.ca/cihiweb/dispPage.jsp

[11] Agence Technique de líInformation HospitaliËre sehttp://www.sante.atih.gov.fr e

[12] Trombert-Paviot B, Rodrigues J-M, Rogers J, Baud R, van der Haring E, Rassinoux A-M, Abrial V, Clavel L, Idir H. GALEN: a third-generation terminology tool to support a multipurpose national coding system for surgical procedures. International Journal of Medical Informatics 2000; 58ñ9: 71ñ85.

[13] Rodrigues J-M, Rector A, Zanstra P, Baud R, Innes K, Rogers J, Rassinoux A-M, Schulz S, Trombert Paviot B, ten Napel H, Clavel L, vander Haring E, Mateus C. An ontology driven collaborative development for biomedical terminologies: from the French CCAM to the Australian ICHI coding system. In: Hasman A et al. (Eds.) MIE 2006 Proc. IOS Press, 2006; pp. 863-8.

[14] Gruber T. A translation approach to portable ontology specifications. Knowledge Acquisition 1993; 5(2):199-220.

[15] Smith B, Ceusters W. Towards industrial strength philosophy: how analytical ontology can help medical informatics. Interdisciplinary Science Reviews 2003; 28:106-11.

[16] Baud RH, Rodrigues J-M, Wagner JC, Rassinoux A-M, Lovis C, Rush P, Trombert-Paviot B, Scherrer J-R. Validation of Concept Representation Using Natural Language Generation. In: Masys DR (Ed.). AMIA Annu Symp Proc. 1997: 841.

[17] Rector A. Compositional Models of Medical Concepts: Towards Re-usable Application-Independent Medical Terminologies. In: Barahona P and Christensen JP (Eds.). Knowledge and Decisions in Health Telematics. Amsterdam: IOS Press, 1994; pp. 109-14.

[18] UWDA; University of Washington Digital Anatomy .A reference ontology for biomedical informatics: the Foundational Model of Anatomy (FMA).see http://fma.biostr.Washington.edu

[19] Rosse C, Mejino JVL. 2003.A reference ontology for biomedical informatics: the Foundational Model of Anatomy. J Biomed Inform. 36:478-500.

[20] Smith B ,Ceusters W ,Klugges B ,Kohler J ,Kumar A ,Limax j ,Mungall C J ,Neuhaus F ,Rector A ,Rosse C. Relations in biomedical ontologies. Genome biology.2005,6:R46

[21] CEN prEN 15261:2007; Health informatics - Categorial structures for terminologies of human anatomy.

[22] Semantic Interoperability seminar .Brussels February 2005

[23] CEN/TC251/ENV12264:1995 Medical informatics-Categorial structures of concepts systems. Further EN 12264:2005 Health informatics - Categorial structures of concepts systems.

[24] CEN AFNOR EN NF 1828:2002.Health informaticsñ Categorial Structure for classifications and coding systems of surgical procedures.

[25] CEN AFNOR.EN NF 12264:2005 Health informatics-Categorial structures of concepts systems.

[26] CEN EN 1614:2006 Representation of dedicated kinds of property in laboratory medicine.

[27] EN ISO 18104 2003.Integration of a Reference Terminology Model for Nursing.

[28] Severs M .The new organisation for Snomed CT, support and development .Semantic mining conference on Snomed CT Copenhagen October 2006

Address for correspondence

JM Rodrigues, CHU de St Etienne, SSPIM, Hôpital St Jean Bonnefonds, 42 650 St Jean Bonnefonds, France
rodrigues@univ-st-etienne.fr.

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

The Nodes Focusing Tool for Clinical Course Data of Hypergraph Structure in the Ontological Framework CSX Output from POMR-based EMR system

Yasuyuki Hirose[a], Ryuichi Yamamoto[b], Shinichiro Ueda[c]

[a] *Medical Informatics, University of the Ryukyus, Okinawa, Japan*
[b] *Interfaculty Initiative in Information Studies, The University of Tokyo, Japan*
[c] *Clinical Pharmacology, University of the Ryukyus, Japan*

Abstract

Knowledge management is one of the significant issues in this century. In medical informatics, the concept of ontology is an important current topic. In addition, there is great interest in knowledge acquisition from clinical data. When doctors use clinical information systems (CIS), their operation implicitly represents, to some extent, their thinking processes with clear reasons and goals according to their intent. If we can capture such resources in an appropriate representation, considerable empirical knowledge could be utilized in various research fields. With this prospect, the authors built an experimental CIS and a nodes-focusing tool for abstracting a summary of the clinical course being examined. The information model is based on a thinking process model, and data are represented in the ontological framework CSX. Then we showed the knowledge-abstracting procedures with this tool. Some domain experts and intern students showed great interest. The authors thereby concluded that such an empirical knowledge acquisition environment is useful for research, education, and so on.

Keywords:

intention, decision-making, clinical course, visualization, knowledge acquisition and representation, ontology

Introduction

As clinical information systems (CIS) extend their functionalities to support the practices of physicians and nurses in a more comprehensive and direct manner, the systems should be designed based on a profound understanding of clinical workflows. We think well-designed CIS can play a considerable part in the clinical practitioners' thinking process behind these workflows, and therefore CIS can provide a significant resource for empirical knowledge about clinical thinking process, by recording the *reasons, goals and intentions of medical interventions* in a semantically cohered data structure during clinical courses.

Knowledge can be represented in a net or hypergraph structure. Each node corresponds to each concept or piece of information, and each edge shows the meaning of the inter-relationship between nodes. As a clinical process ontology, this hypergraph must have enough capability to represent the facts or facets of the facts at point-of-care to acquire clinical empirical knowledge.

In general, a CIS requires several information models: at least one for a transaction and another for clinical document representation, and they may be formulated as health IT standards [1, 2]. Now we propose another model for the thinking process. Those three refer to the same data in the same repository in a CIS to contain them in a semantically related structure of each model. In this sense, the structure of the information model is expressed as semantic inter-relationships among semantically defined data containers according to its model perspective. The differences of models depend on their perspectives, which then leads to different structures.

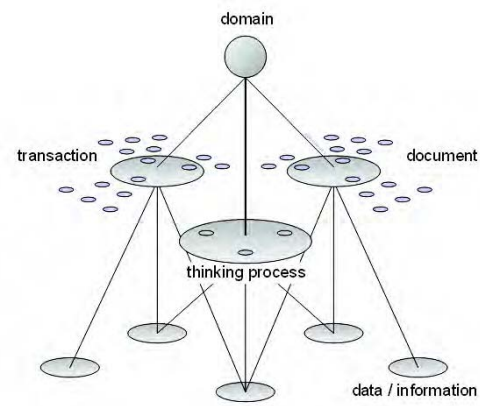


Figure 1 – Perspective variations in a domain

Of course a container may contain other containers when a structure is complex and the hierarchy is deep. An element in a container may be identified with a code of other terminology systems.

Smith et al. have asserted the significance of perspective in methodology axes, and the possibility of the representation of in a single framework that holds plural ontologies [3]. We agree because of our design policy that was based on 'perspective' in the sense of oriental philosophy [4], and because of our experience in research supported by grants from the Research on Health Technology Assessment,

Ministry of Health, Labor and Welfare, Japan from 2000 [5, 6].

To support plural ontologies in one system, we decided to utilize an ontological framework called CSX that we originally developed as a metamodeling environment [7]. Then we developed an experimental CIS based on the model of the clinical thinking process to capture clinical empirical knowledge. We also developed a node-focusing tool for abstracting a kind of partial graph of the clinical course being examined.

Methods and designs

The clinical thinking process model

Model outline

A clinical course model should have two aspects: the clinical course itself and the health care provider's thinking process at point-of-care. We had already developed the latter, as a *decision* process model, along with the problem-transition model known as a problem-oriented system (POS) or the problem-oriented medical record (POMR) [8]. Generally speaking, this model is an *intention* realizing process model; therefore, each *reason* in the thinking process and the *goals* of interventions are clearly delimited. One cycle of the thinking process in a clinical session (i.e., encounter) is illustrated in the following:

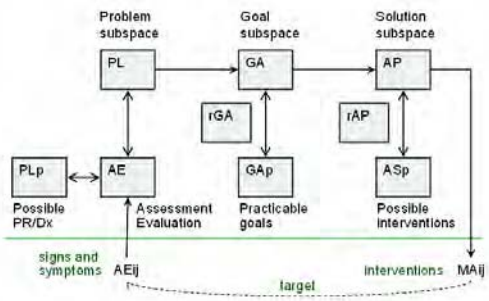


Figure 2 – The clinical thinking process model

The decision process is placed in three spaces: problem, goal and solution. Each space contains elements that represent an 'assorted step' in the thinking process. Each is a container that holds child objects (e.g., each problem or disease (PRi) in 'ProblemList' (PL), and each assessment and evaluation (AEi) in 'AssessmentEvaluation' (AE)). AEi may contain a physician's inference and/or reasoning for diagnoses, with referring atomic data of signs and symptoms. These are called S/O in SOAP in POMR. The Goal (GA) is decided due to PRi or PL with reasons of each goal item (GAi), and those reasons are held in reason of GA (rGA). The action plan (AP) consists of some protocols and/or particles of interventions (ASi/MAi). They are adopted to gain GA with reasons held in reason in rAP.

Then, any clinical course is composed of a series of clinical sessions, where each phase connects to another with a screw structure or spiral binding through PL as an axis.

In this research, we carried out the implementation by focusing mainly on problem list and problem transitions, and on the problem-intervention relation.

Some requirements for a representation framework

The representation framework should have the capability (i) to represent the hypergraph structure and to support n-arity, (ii) to discriminate between semantic roles and semantic relations', and (iii) to represent perspective for handling plural ontologies. At the same time, these are the reasons for adopting the ontological framework CSX.

Those three are coupled to each other in knowledge operation and system functions. A complicated semantic graph becomes more complicated because there are multiple interrelationships among nodes with plural perspectives of plural ontologies. Consequently, the usual semantic link does not work well when finding an appropriate path to a target node.

One has to take a notice of the degree or valence of the node, i.e., number of arcs of a node, with respect to data processing, because they may become too much tasking to control a 'path walk in a knowledge graph. Simply stated, one needs a strategy to find out correct 'semantically grouped links' within an appropriate processing time. This is not clearly known as yet because, in some sense, each application module in CIS knows' all arcs and their target information objects. However, it is quite another story to find out the exact path to the target node from the starting node that has many arcs, and some arcs may belong to a different 'semantic group' i.e., semantic relation.

For example, there are elements in AE or PL or 'signs and symptoms' referred to and semantic roles of semantic relations among them: [cardiac silhouette size in thorax] (reason) conclude (object) [cardiac hypertrophy], [decreased ejection fraction] (reason) [oliguria] (reason) conclude (object) [dysfunctional pump], [[cardiac hypertrophy], [dysfunctional pump]] (reason) define (object) [cardiac failure]. Here, [] is a node, () is an arc, and the underlined is the meaning or scope of the 'semantic relation. In this sample, a relation represents predicate and an arc represents a 'semantic role as a deep case that is required in each predicate.

Therefore, one solution is that the separation of 'semantic link into semantic relation and semantic role. The ontological framework CSX provides this capacity.

The ontological framework CSX

The outline of this framework is as follows:

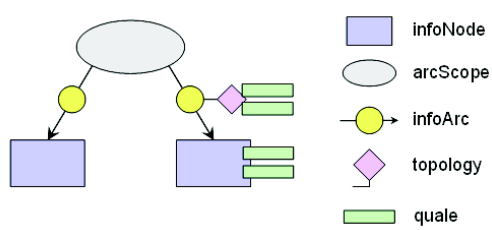


Figure 3 – The ontological framework CSX

O=<P, S, N, vr>, P{dm, sg}, S=<A..A, sg, id>, A=<N, N, T, sg>, N{C, Q, sg, id}(t), C{cs, ci, vr}, T{dr, cd, Q}, Q{dn, ms, ut, eq}, where O:ontology, P:perspective, dm:domain, sg:designation, id:identifier, S(arcScope):semantic relation that integrates A and represents the predicate or copula, A(infoArc):rhetorical role or semantic role in a semantic relation, N(infoNode):information object that represents thing or event, C:code that identifies what a element is, T(topology):represents the topological relationship between N, Q(quale):represents quantity/quality, vr:version, t:timestamp, cs:code system that defines the combination of code/content /designation, ci:identifier in cs, dr:direction, cd:coordinate, dn:dimensionality, ms: measure, ut:unit, and eq:equivalence.

S, A and N have the attributes of category and family. Those two attributes hold the value that represents the position of meaning in the ontological category. Please refer to Figure 1 and the preliminary works [5, 9].

Here, ontology means 'a framework or methodology that manifests concepts/entities and relationships between/ among them in the target world'; and, the representation of relationship is weighted in this research. On the other hand, building a vocabulary/concept repository is beyond the scope of this paper; thus, positively referring vocabularies are defined in terminologies, especially in C held in N.

Implementation scope in this paper

System environment

To evaluate our approach, we implemented an experimental CIS with two edit/browse windows (Disease/Problem Transition and Clinical Chart) and four supporting entry tools (problem/diagnosis, prescription order, laboratory order, and operation and procedure). Edit/browse windows were designed to imitate Japanese unique regulated forms of No.1 and No.2 respectively. In this system, controlled vocabularies and identifiers in following code systems are used: ICD-10 based disease code system, JLAC10 (resembles LOINC) based laboratory code system, HOT based drug code system, and operation/procedure code system for reimbursement. The history of the clinical course is output as xml defined for CSX, which is used as an input file for the nodes-focusing tool.

Visualization of clinical course and focusing

This tool is designed to facilitate the browsing of the whole image of a graph that illustrates a clinical course, then to abstract the nodes and arcs that are focused on according to the end users purpose. With this tool, the model of a clinical course or empirical knowledge from a certain case might be clarified, although it has large numbers of encounters in a very long course; for example, only four essential PLs and its transition illustrated in Figure 4. Such a summary would be found through the trial and error process of comparing, masking and tracking.

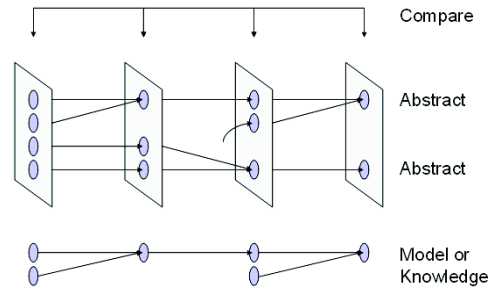


Figure 4 – Abstraction with nodes-focusing tool

Extracted data (or knowledge or pre-knowledge) is output as xml defined for CSX. The image of graph can also be output in JPG format.

Results

Disease/problem transition and relevant care information

The Disease/Problem transition window has a maximum of three panes. One pane view is for a usual Dx/PRi list, and the three-pane view (Figure 5) is for editing and browsing disease/problem transitions. In the three panes, an end-user can control the pane linking which one is connected or unconnected. The latter provides end-users with the capability to look discretely at problem lists.

The kinds of semantic roles in each PRi is automatically determined due to a partial graph between the PRi in the former PL and PRi in the latter PL: proceeded, diverged, converged, promoted, demoted, and terminated in semantic role A in CSX and transition or transmute in a semantic relation S in CSX.

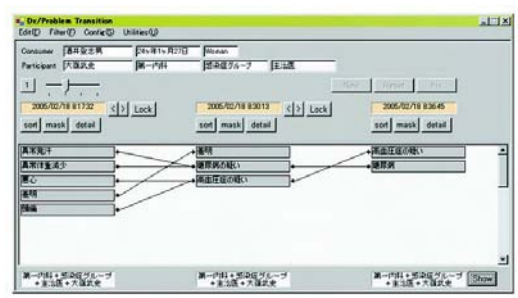


Figure 5 – Sample of the disease/problem transition window

Each PRi is selected in the problem/diagnosis entry tool by the end-user operation, often to be compiled with prefix and/or suffix. Then the PRi is inserted into the Disease/ Problem transition window when the problem/diagnosis selecting operation is fixed. Therefore, the PRi node has three codes in C held in N: main, prefix and suffix.

Each problem (again, including diagnosis) has the attributes ofpriority, activity and Basso Continuo. They are also editable in this window. Basso Continuo means the problem with the attribute influences other problems and their course, including the available interventions for those problems.

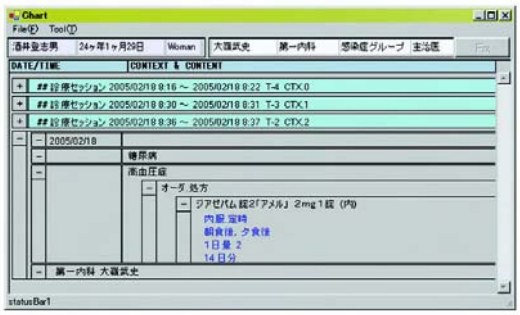


Figure 6 – Sample of the chart window

In the Chart window, interventions are semantically linked to a problem, as a 'reason or object in a sense of deep case in A; a 'cure or intervene in a semantic relation S. Of course, one can list various patterns in such a relation; for example, a single problem holds many interventions, a group of problems is linked to a set of interventions, and different problems hold the same intervention. All of them are implemented in the system, and the latter one is realized as an *alias* in the GUI. The application runs appropriate responses.

Highlighting significant clinical course

Figure 7 shows a sample of a bird's eye view of a clinical course in expressed in a graph.

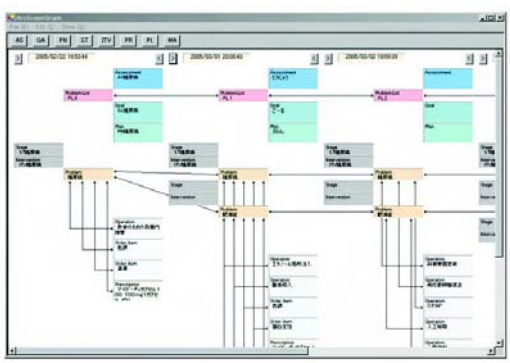


Figure 7 – Sample of a whole graph

The number of panes for each encounter is unlimited (or depends on the system environment). An end-user can control the pane linking in the same way as the Disease/ Problem transition window.

It is worth noting that the path tracking function is available. An end-user selects a start node, then selects the path walk policy, then clicks the extract button, and finally gets to the situation shown in Figure 8, for example.

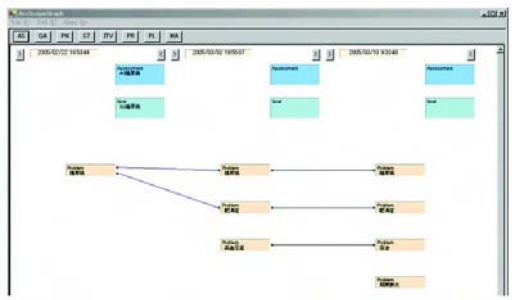


Figure 8 – Sample of a focused and abstracted

In this experimental implementation of the node-focusing tool, four kinds of tactics are independently available: start and end date, tracking direction (backward, forward, or both), code matching in problem (main code, whole codes, or any when linked), and selectable divergence walk (track whole path, track if targeted on Basso Continuo, or no track).

Discussion

Visualization and knowledge acquisition

This tool provides the visualization capability of a clinical course by focusing nodes and arcs on the various purposes of end-users. As it is enabled to trace the focused nodes and arcs through the crowd, it is obviously useful for case review, clinical research, clinical education, the creation of a new clinical path or the consideration of the revision of it, and variance analysis for financial management in a hospital. Using the visual tracing function, inexperienced residents or medical students could refer to experienced doctors' successful processes, which contributes to appropriate decision-making and therefore better outcomes. Hence, we believe this tool will help health care providers in certain cases, especially for the compromised patient.

Such advantages follow the output from the POMR system, a file is in xml format defined for CSX, and data represent the clinical course according to the ontological framework CSX.

Some functions might be insufficient but others can be added to this tool.

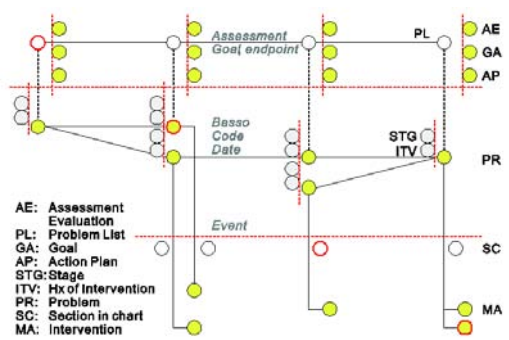


Figure 9 – Graph of clinical course data

In this framework, we used designations for entities and semantic relations that resemble or parallel ISO/CD TS 22789 in the implemented model, given the need for this experimental implementation.

Evaluation of the experimental implementation

Six domain experts (all were physicians and their subject was clinical trials) and a hundred of intern students showed great interest. When the authors showed medical students this tool, most immediately understood its idea. They also stated that they wanted to use this tool at clinical conferences, especially for complicated cases. Therefore, we believe that such an empirical knowledge acquisition tool is useful for clinical research and education.

Conclusion

We developed the experimental POMR based on the clinical thinking process model with the ontological representation framework CSX, which has the capability of handling plural ontologies, and the nodes-focusing tool for clinical course data in a hypergraph structure to capture and acquire or mine some portion of health care providers empirical knowledge at point-of-care, including their intent. Then some domain experts and intern students showed great interest. Therefore, the authors concluded that such empirical knowledge acquisition environment is useful for research, education and so on.

Acknowledgments

The authors are grateful for the support and suggestions by Mr. Seiichi Yamada, Mr. Tatsuya Yonamine, Mr. Takeshi Ohmine, Mr. Satoshi Yamamoto, Mr. Shigeru Bito, Mr. Ei Murakami, Dr. Kenichi Yajima, Dr. Mitsugu Kanda, Dr. Noriaki Morimoto, Dr. Taishi Nago, and Prof. Kentaro Inui. This work was supported by grants from the Research on Health Technology Assessment of Ministry of Health, Labor and Welfare, Japan H17-RHTA-043 from 2005 to 2007, H15-RHTA-050 from 2003 to 2005, and H12-RHTA-009 from 2000 to 2003 in a previous work.

References

[1] HL7 version 2.x, and 3. At www.hl7.org, 2006.
[2] OpenEHR. At www.openehr.org, 2006.
[3] Grenon P, Smith B, Goldberg L. Biodynamic Ontology: Applying BFO in the Biomedical Domain. In: Pisanelli DM, ed. Ontologies in Medicine. Technology and Informatics 102: IOS Press, 2004; pp.20-38.
[4] The Prajna Paramita Hrdaya Sutra.
[5] Hirose Y. Methodology for clinical information sharing (H12-RHTA-009). ISBN 4902408066, 2003.
[6] Hirose Y. Information model in EMR for the representation of problem transition and problem-intervention relationship (H15-RHTA-050). ISBN 4902408090, 2005.
[7] ISO/IEC JTC1 SC32 WG2. ISO/IEC 11179-1:2004.
[8] Weed LL. Medical Records, Medical Education and Patient Care. 2nd ed. Cleveland: The Press of Case Western Reserve University, 1970.
[9] Hirose Y. Tiny and Compact Meta Meta-information Model. MEDINFO 2004; 1640.
[10] Rogers JE. Quality Assurance of Medical Ontologies. Methods Inf Med 2006; 45: 267-274.

Address for correspondence

Medical Informatics,
Hospital, University of the Ryukyus
Uehara 207, Nishihara, Nakagami, Okinawa 903-0215, JAPAN
hirose@hosp.u-ryukyu.ac.jp

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Integrating Descriptive, Functional and Cooperative Aspects in a Domain Ontology to Computerize Home Care Charts

N. Bricon-Souf [a], S Ngom [a], S Hamek [b]

[a] *CERIM, EA 2694, Université de Lille 2, France*
[b] *CNRS UMR 8530 -LAMIH-PERCOTEC, Université de Valenciennes et du Hainaut-Cambrésis, France*

Abstract

To coordinate themselves, home care (HC) professionals use artifacts to keep a mutual understanding on their common activity: lightly structured charts written in natural language. Instead of trying to define a record to capture them, we want to focus on efficient indexing of this information. The use of a Domain Ontology was proposed. This paper explains how we built and implemented it. Three complementary aspects of the HC charts were analyzed (i) functional aspects performed with precise analyses of actual charts; (ii) interoperability aspects with the use of some HL7-RIM standardized descriptions; and (iii) cooperative aspects, with the integration of a cooperation model. We proposed a Domain Ontology to represent the concepts and the relations used in the charts. The implementation was done with Protégé; the ontology was built in OWL-DL. The IAnnotate application helps us to index the HC chart with the domain ontology. The next step of the work will use the ontology to reason about the different items of information contained in the charts so that contextual use of them should be envisaged.

Keywords:

medical informatics, cooperation, quality of care, ontology

Introduction

- Health professionals, who work together in order to care their patients, very often have to coordinate themselves asynchronously. They use some artifacts in order to keep a mutual understanding of the situation such as the knowledge of the patient's state or the knowledge of the performed medical tasks. In particular, documents such as Home Care Charts or Annotations are shared. They are readable by different individuals at different times, and they allow the elaboration of an awareness of the different activities performed by the caring team. These documents are highly complex: information is often given in natural language, they are not very structured or even not structured at all, they mix all kinds of information (medical, legal, logistic and so on) and they are consulted by different people assuming different roles for the care (nurse, practitioner, family, patient, physiotherapist and so on). Moreover, this information has a double status for the cooperation: (i) it is used at an operational level (coordination of actions); (ii) it is used at a more abstract level of planning (synchronization of the knowledge that everyone has on the current care process). It also builds the common ground indispensable for efficiency, good cooperation and coordination. The medical domain is highly concerned with the computerization of the information. The use of electronic records helps people to store, retrieve, display and exchange many items of complex information. Such electronics records have shown their strength and are actually used in everyday situations of care, but they have also shown some weaknesses [1].

In the context of our studies of Home Care Situations, one of our purposes is to understand how to represent and how to use the information contained in the Home Care Charts. Instead of trying to define an efficient record for such information, we thought it could be interesting to treat this information according to the "Semantic Web" paradigm and to focus on efficient indexing of the information. The items of knowledge contained in the Home Care Chart should:

- be displayed in an efficient and "business oriented" way;
- support the cooperation of the health care professionals by improving the mutual awareness they have on the care; and
- develop new exchanges between Home Care organization and other medical Information Systems.

These three dimensions should be captured by three interconnected descriptions of the complementary aspects of the domain: functional, interoperability-oriented and cooperation-oriented.

The following definition of ontology is proposed by Sowa [2]: "The subject of *ontology* is the study of the *categories* of things that exist or may exist in some domain. The product of such a study, called *ontology*, is a catalog of the types of things that are assumed to exist in a domain of interest ***D*** from the perspective of a person who uses a language ***L*** for the purpose of talking about ***D***. The types in the ontology represent the *predicates, word senses*, or *concept and relation types* of the language ***L*** when used to discuss topics in the domain ***D***."

Indexing the Home Care Charts items by using domain ontology, as proposed in other areas of research [3], could improve the understanding of the written home care charts while keeping a flexible way of expression for the health care professionals.

If the long-term perspective of this work is to propose an efficient way of indexing the home care charts in order to focus on useful information according to the context of use or to the user's identity, the first steps of this work are to build a domain ontology able to answer the problem of indexation of the home care charts as well as to validate the idea that this ontology could help to reason on the items to focus on.

The objective of this paper is to relate the work we have done to analyze the home care charts and to explain the ontology we have then proposed to support such indexation. A brief description of the use of the ontology will also be presented.

Materials and methods

Materials

The study was performed:

- From raw and real information: we used 16 home care charts as a corpus, each one containing all the information exchanged between nurses throughout the entire care period (from 1 month to 6 months). These charts come from the SANTELYS association that is a rather big association (200 employees), with an active and well-organized branch dedicated to Home Care. Patient's anonymity has been preserved. Home care charts have been analyzed by a psychologist involved in this project, and then described. These descriptions have been confronted with the opinion of some of the nurses involved in such a care.
- From medical standards: in the medical field, a lot of work has been done to define standards in order to increase the interoperability of the health care information systems. In particular, HL7 was created "to provide (global) standards for the exchange, management and integration of data that supports clinical patient care and the management, delivery and evaluation of healthcare services."[4]. Moreover, a formal model, the Reference Information Model (RIM) "specifies the grammar of HL7 messages and, pecifically, the basic building blocks of the language and their permitted relationships" [5].

These materials will help us improve our understanding and representation of: (i) functional aspects of the charts, (ii) interoperability and (iii) cooperation.

Functional aspects of the Home Care Chart

Thanks to the activity analysis performed by the psychologists working on this project [6, 7]; we had rich information about how to combine meaningful units in the domain of the home care. A typology of the home care information, according to a functional point of view, has been performed.

Interoperability

As one of our purposes was to assume interoperability between the Home Care System and other Medical Information Systems, it was then really interesting to integrate the RIM standardized description to our model for Home Care Charts indexation. We used the pre-existing RIM ontologies made by Bhavna Orgun [8] and Samson Tu [9] to propose an ontology answering to the reference ontology of the Health care domain aspects.

Cooperation

Hoc's architecture of collaboration [10] which proposes three levels of collaboration (action, plan and meta-collaboration) can anchor a conceptual framework for an ontology of cooperation. Two levels of synchronization can be identified: (i) cognitive synchronization (plan and metacooperation), which aims to build a mutual knowledge context, to elaborate and maintain a shared representation structure; (ii) temporal and operative synchronization (action level) which facilitates the coordination by the implementation of two functions: it ensures the task allocation between partners and it ensures the train of actions, their simultaneity, their starting or their stopping. Such refined descriptions actually allow the psychologists to perform their study of the asynchronous cooperation with the home care chart. We decided to integrate such knowledge in our work.

Methods

Smith and Ceusters [11] criticizes HL7/RIM that presents some weakness while mixing different aspects in the same standard. These authors recommend to distinguish between the 'Reference Ontology of the Healthcare Domain' (needed to provide a compact and coherent high-level framework in terms of which the lower-level types captured in vocabularies like SNOMED CT could be coherently organized) and the 'Model of Healthcare Information' (needed to specify how information about the entities that instantiate the mentioned types can be combined into meaningful units and used for further processing). In [12] Thieu et al. suggest a way to combine two different ontologies to represent both morphological and diagnostic concepts to define one cell, in [4] Isaac combines ontologies to describe audiovisual documents.

Our hypothesis was then that it could be possible to combine the functional, the descriptive and the cooperative aspects of the Home Care Charts and to propose a useful ontology for indexing them.

Healthcare professionals' cooperation context study and frameworks' elaboration:

A systemic approach [13] which proposed the following steps: systemic exploration, qualitative representation, dynamic representation and treatment of the created framework, revealed helpful to elaborate our frameworks. Each step was discussed with the psychologists:

- *Systemic exploration; representing the objects of the world of the homecare process.* Taxonomy of the home care charts was proposed so that any observations contained in one of the 16 despoiled home care charts

could be classifiable in one of the elements of the taxonomy.

- *Qualitative representation; representing the taxonomy of the home care charts in terms of flows and hierarchical organization*. Based on the previous mapping, a model was refined to visualize the interactions between the main components of the ystem, and the different information flows between different levels of organization.
- *Dynamic representation; superposition of the HL7/ RIM and of the qualitative representation.* During this step we determined how the elements found in the taxonomy of the home care charts could match the entities of HL7/RIM: obviously, some fundamentals belong to these different representations.
- *Treatment of the created framework; domain ontology building.* To propose an ontology for indexing the home care charts, the model of HL7's RIM does not appear to be sufficient enough to describe the home-care process from a functional viewpoint as only the subclasses of the qualitative representation were visible. Truncating the representation of the home care charts' taxonomy to integrate it into the RIM could generate some problems too. For example, one could find it incomprehensible that a person is considered as an "Entity", such as a place or a material. To merge the two models, we represented the home healthcare process from a descriptive sight (according to the HL7's RIM model), from a functional sight and a cooperation sight (according to the qualitative representation, the latest including the home care information model and the home care cooperation model).

Results

The qualitative model

Home Care was described through 6 main functional parts which are: Initial Prescription; General Evaluations and Actions; Organization of Care, Therapeutics; Non-Action; Patient Information. These parts can be refined with "is-A" relations, or can be linked with "isConsequenceOf" relations.

If the higher-level classes have no RIM correspondences, some subclasses of the qualitative model match some RIM classes. Some choices have been made in order to build the complete classification. For example, the class "Indication for care follow-up" in the functional model could match the "To-do List" in RIM representation. In fact, "To-do List" is a subclass of "Working List", defined by HL7 workers as: "A dynamic list of individual instances of Act which reflects the needs of an individual worker, team of workers, or an organization to view groups of Acts for clinical or administrative reasons. (…). – as: problem lists, goal lists, allergy lists, to-do lists, etc." Another example is "Visit stopping," which is considered as a "Control Act," defined as: "An act representing a change to the state of another class, a user event (e.g. query), or a system event (e.g. time-based occurrences). – *as:* Discharging a patient (Encounter from Active to Completed); Stopping a medication (Substance Administration from Active to Aborted); Sending an end-of-day summary (time-based event)."

The Home Care Domain Ontology

To represent the concepts and the links of our qualitative model, we decided to build Home Care Domain Ontology. Three sub-ontologies were used to represent the different aspects of the homecare process.

1. The descriptive ontology of the domain (Ontodescriptive).
   This ontology was adapted from existing RIM ontologies. [8][9]. It allows to characterize the home-care documents according to HL7/RIM concepts.
2. The functional ontology of the domain (Ontofonctionnelle).
   Drawn from the functional representation of the home care charts, it should give the professional a way to index communication units in a business oriented way.
3. The cooperative ontology (Ontocooperative).
   It introduces the concepts of the cooperation and supports a new way to characterize the activities according to a cooperative point of view.

To build the global ontology of the homecare process domain, we described the relationships between fundamentals in the different sub-ontologies.

- An object property called "hasRimDefinition" served to link some instances of the functional domain to their RIM equivalent so that: "Instance.functional_ontology" hasRimDefinition "Instance.descriptive_ontology". The cause-consequence and class-subclasses relationships have been kept, respectively thanks to the object properties "hasConsequence" and "isA".
- In the same way, we could see the superposition of the functional and the cooperative models: a property "hasCooperationType" will connect some instances of the functional ontology to the matching instances of the cooperative ontology:

For example, a "realisation of a physical or a behavioural evaluation" is linked with the "management of a common ground". A "demand for therapeutics change" is linked to a "creation of interference".

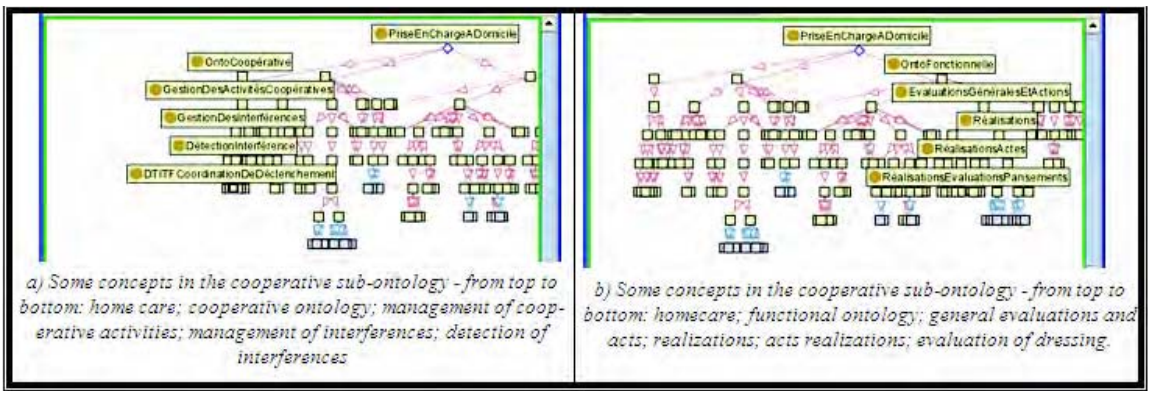


Figure 1 - Examples of the concepts used in the ontologies built from the communication unit analyses

The Indexation of the Home Care Charts

The coordination ontology as well as the functional ontology has been used to index the medical events. For example, when it is written on the chart "nurse care, morning and evening, every days including Sundays and public holidays:"

- it can be marked as an "initial prescription" in a functional representation;
- it deals with temporal and operative synchronisation and can be marked as a "creation of planning" in a cooperative point of view.

A sentence like "constipation since 3 days:"

- is a "reported event" in the functional representation;
- can be marked as "an elaboration of shared representation structure" in a cooperative representation.

Moreover, as the indexation is supported by ontology, we can use the subsumption links to increase the knowledge we have on the marked sentence; for example, "reported event" is an "evaluation".

We can have very flexible views on the information as we can choose to underline the information according to some of the descriptions we have on it.

Implementation

The ontology was built in OWL-DL, under the 3.1 version of the Protégé ontology editor [14]. Protégé plugins were used o have graphic representations such as in fig. 1,2. So as to make reasoning on the ontology, the RACER [15] inference engine, based on description logics was installed. This Reasoner was used to automatically classify the instances of the classes on regard to their properties, to compute the inferred ontology class hierarchy and to check for inconsistencies.

To perform the annotation of the home care chart, we chose to start from the iAannotate [16] application that uses an ontology to index content (file or web pages). It allows choosing a concept of the ontology to select part of the text and to index it with the selected concept. Once part of the text is annotated with an ontology concept, it is highlighted in a colour associated to this concept. Moreover, the indexations are shown or hidden by checking or not the checkboxes associated to each concept. (Cf figure 2).

IAnnotate can be used as a plug-in in Protégé. Some modifications have been applied to the iAnnotate program so that it could support multiple indexing for one text entity.

Figure 2 - Example of an annotated Home Care Chart where the underlined items are those that are underlined with cooperative concepts (here, the ticked checkboxes are those that describe cooperative concepts)

Discussion and conclusion

The production of Communication Notes (informal written sentences, annotations, and so on) is important to manage all the informal and unforeseeable information produced by the Health Care Professionals during care delivery [17]. Being able to use information for different purposes, according to different points of view or different contexts is one of the challenges we have to face in our information society. Thanks to free content, in the homecare setting, we want to keep the flexibility of the notes and to allow people to express all the information that is needed to support efficient collaboration. At the same time, we want to use some descriptions to highlight relevant information according to the collaborative nature of the activity. It could be interesting to apply the Semantic Web paradigm to the homecare chart and to perform efficient indexing of the information. But just how to build such indexing is less than obvious.

We have made different deep analyses of home care charts to feed our work and we are able to describe the main concepts and their relations in a functional or a cooperative approach. The RIM concepts can also feed our work. The first step of work was then to build a domain ontology that reorganises all the concepts associated to the homecare charts. Now, we have rich material to use as a basis for more extensive studies.

The second step of the work is then to have a tool that uses the ontology to index the homecare charts. At this stage of work, it was not realistic to perform the indexing automatically, so we decided to implement software with a user interface that allows selection of a concept from the ontology and to use it to index some text. By slightly modifying iAnnotate, we now have such a tool to index the charts.

A further step of the work will explore the use of the indexing.

A direct application is that the psychologists would use it to index the home care chart in a convenient way for their next studies. But obviously, the strength of this domain ontology representation is that we will now be able to reason on the knowledge we have. The functional descriptions we have on these charts, increased with the relationships between functional descriptions and cooperative descriptions, should help us to suggest cooperative views on the written items. For example, is it possible and useful to highlight all the items focusing on the planning concepts, when people organize their work? In the same way, we wanted to know how far we could use our system to provide automatically information from the home care chart to a homecare information system, using the relationships between functional and descriptive ontologies.

Acknowledgments

This work has been supported by the Region Nord - Pas de Calais in France (CPER).

References

[1] Hartswood M, Procter R, Rouncefield M, Slack R. Making a Case in Medical Work: Implications for Electronic Medical Record. Computer Supported Cooperative Work. 2003: 12: 241–266.
[2] Sowa J F.http://www.jfsowa.com/ontology
[3] Isaac A, Troncy R. Using several ontologies for describing audiovisual documents: a case study in the medical domain In workshop on multimedia and the semantic web, ESWC 2005.Heraklion Crete ,2005
[4] HL7, http://www.hl7.org/about/
[5] http://www.hl7.org.au/HL7-V3-Resrcs.htm
[6] Hamek S, Anceaux F, Souf N, Flahou S. La prise en charge des soins lors de la coopération asynchrone. Journées Francophones d'Informatique Médicale, Lille 12-13 mai 2005.
[7] Hamek S, Anceaux F, Pelayo S, Beuscart-Zéphir M-C, Rogalski J. Cooperation in healthcare - Theoretical and methodological issues. A study of two situations: Hospital and Homecare. In Marmaras N, Kontogiannis T, & Nathanael D (eds.), Proceedings of the annual conference of the european association of cognitive ergonomics - EACE'05 Athens, G: National Technical University of Athens/ErgoU, 2005, pp. 225-232.
[8] HL7-RIM Ontology as a Protégé project - cleaned up version. http://www.ics.mq.edu.au/~borgun/Software.html
[9] Health Level Seven (HL7) Data Types and Top-Level Reference Information Model (RIM) Classeshttp:// protege.stanford.edu/ontologies/HL7RIM/index.html
[10] Hoc J-M. Towards a cognitive approach to human-machine cooperation in dynamic situations. International Journal of Human-Computer Studies. 2001:54: 509-540.
[11] Smith B, Ceusters W. HL7 RIM: An Incoherent Standard, Stud Health Technol Inform. 2006; 124: 133–138.
[12] Thieu M,.Steichen O, Le Bozec C,.Zapletal E,.Jaulent M-C. « Mesures de similarité pour l'aide au consensus en anatomie pathologique ». IC 2004; 2004, 225-236.
[13] Donnadieu G, Karsky M. « La systémique : penser et agir dans la complexité ». Editions Liaisons. 2002. pp78-95.
[14] Protégé OWL - Ontology Editor for the Semantic Web http:/ /protege.stanford.edu/plugins/owl/
[15] RACER http://www.racer-systems.com
[16] http://www.dbmi.columbia.edu/~coop7001/iAnnotateTab/ iannotate.htm
[17] Bricon-Souf N, Bringay S, Anceaux F, Hamek S, Degardin N, Barry C, Charlet J. A study of the communication notes for two Asynchronous Collaborative Activities. Stud Health Technol Inform. 2006: 124:713-8.

Address for correspondence

Nathalie BRICON-SOUF ; CERIM – Faculté de Médecine-Université de Lille 2 - 1, place de Verdun -59 045 Lille Cedex France - nsouf@univ-lille2.fr
Equipe d'accueil CNRS 2694; Santé publique: épidémiologie et modélisation des maladies chroniques

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

The ICNP-BaT - A Multilingual Web-based Tool to Support the Collaborative Translation of the International Classification for Nursing Practice (ICNP)

Ulrich Schrader[a], Peter Tackenberg[b], Rudolf Widmer[c], Lucien Portenier[d], Peter König[e]

[a] *University of Applied Sciences, Frankfurt, Germany*
[b] *University Wuppertal, Germany*
[c] *SBK - IG PflegeInformatik, Switzerland*
[d] *SBK - ASI, Bern, Switzerland*
[e] *ICN Accredited Research & Development Centre, Freiburg i. Br., Germany*

Abstract

To ease and speed up the translation of the ICNP version 1 into the German language a web service was developed to support the collaborative work of all Austrian, Swiss, and German translators and subsequently of the evaluators of the resultant translation. The web service does help to support a modified Delphi technique. Since the web service is multilingual by design it can facilitate the translation of the ICNP into other languages as well. The process chosen can be adopted by other projects involved in translating terminologies.

Keywords:

ICNP, nursing terminology, translation, internet, web services, collaboration.

Introduction

The International Classification for Nursing Practice (ICNP) is a terminology system of the International Council of Nurses. It defines itself as "a unified nursing language system. It is a compositional terminology for nursing practice that facilitates the development of and the cross-mapping among local terms and existing terminologies" [1]. The current version 1 has been published in English in 2005 [2].

In order to be usable and as an important prerequisite to being an accepted nursing language it has to be translated. Building on the experience of the previous translation of the beta-version, which took almost three years to be translated and published [3], a different approach was sought to speed up the process.

As the ICNP version 1 is published in its original form in English it must be observed that the content has been collected internationally and the use of the English language reflects a more international style than just American or British English. A different problem arises as the ICNP version 1 has to be translated into the German language which is used not quite congruently also in Austria and Switzerland. Even if the translation is targeting high German, it also has to take into account the different uses of the German language. This makes the use of synonyms a necessity to accommodate for regional or national variations.

Translation efforts of the previous versions relied highly on documents being send around by email to the different translators, and also on moderators collecting and consolidating the different translations. This is a very time consuming process involving a large amount of work especially for the moderators putting out consolidated versions in each translation step.

For the translation of the ICNP version 1 a web-based service was designed to speed up this work. An additional challenge to be considered is that all work is performed by volunteers. So the different time frames each volunteer can donated have to be taken into account.

Methods

Concept

The basic idea of the concept is that at any time all ICNP terms and all their translations are visible to all translators. Any terms not yet translated can be seen by all translators at any time thus indicating areas still in need to be worked upon.

Since all existing translations of all translators are visible once they are made this offers a way to eliminate redundant work. If a translator agrees with an already existing translation by another person he does not have to translate the term again. Only if he wants to suggest a different translation it has to be entered into the system and is stored along with all other translations. It then will be available immediately for all other translators. Items hard to translate or with regional differences will show up by having multiple translations. Additionally it should be possible to have simple discussion threads for each translation in order to annotate them.

Because their is only one body of ICNP terms and their translations at any time the consolidating work of the moderator is not needed at that stage, resulting in a reduction of workload.

Once the translations of an ICNP term stabilizes, e. g. no new translations are offered, the term will be closed for transla-

tions and be made available for evaluation. Each evaluator will use a questionnaire to evaluate the quality of the translation. The evaluations are collected and analyzed centrally

Realization

Since the translation and evaluation work has to be performed by volunteers from home or from the office a web-based solution was proposed. The solution should require as little as possible on the user's side. There are two components to the complete system: a wiki to publish and discuss translation and evaluation issues as they arise, and to look up translation guidelines and a newly developed browser and translation tool for the ICNP (BaT). The wiki is based on the open source project PhpWiki realized using PHP as scripting language and a MySQL database. Both are usually available on a standard web server running under the Windows or Linux operat

ing system. The browser and translation tool was developed using the same tools.

In order to keep the requirements to just a browser on the user's side JavaScipt was not used. Only cookies have to be enabled for the authentication methods to work.

The BaT can handle several translation languages simultaneously. Each user is allowed to offer only translations in one language but he can view all the other languages. One example of the user interface is depicted in fig. 1. The original English version is on the left side, the corresponding German translation on the right. As shown on the very top of the window the user can select the languages of the interface and the translation displayed. Navigation to and through the ICNP terms is realized using a linear list with several options for filtering and sorting, a direct link to a specific code, and a text search option. The linear list can be easily printed and the printout be used as basis for the translation. Translations can then be easily entered by going directly to the term using its code.

Additionally, the ICNP terms can be browsed according to the hierarchical structures of the ICNP. Here each term is displayed with its parent, child, and sibling terms. To be able to see each term in its context is important for the correct translation of it. This way of navigation also favors users who prefer to translate specific topics like everything concerned with the term 'mobility'.

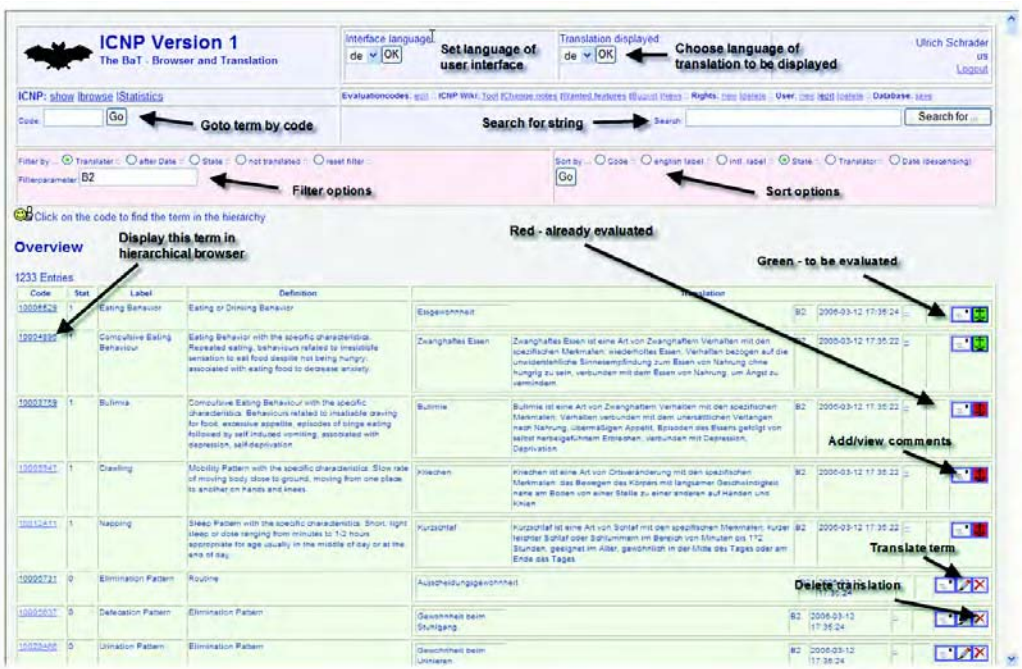


Figure 1 - Example of the user interface

Rights management

The management of access rights of the BaT is function based. Each function of the BaT has a specific right attached to it. When the functionality is invoked or made accessible, e.g. as a link or icon, it is checked whether the right is also assigned to the user. All rights can be assigned individually to each user, thus making it possible to create individual profiles.

Since the translation process is done collaborative and multilingual, care must be taken, that no one does undo the work of another translator. So if a user has the right to translate, he can translate only to one language assigned to him. All other languages are read only for this user. He can either add his own translation, or create a new translation building on an existing translation by another person. A regular user can only edit or delete his own translations as long as the ICNP term is open for translation. Only a user

with moderation rights can edit or delete the translations of others. Once the evaluation phase has started for an ICNP term it will be closed for translation meaning no translations of this term can be added, edited, or deleted.

Due to the multilingual design there are two types of user administrators necessary:

- The 'super' administrator can create regular user administrators responsible for each language.
- The regular user administrator can only create and administer users translating into the same language as is assigned to the regular user administrator. The same restriction applies to the creation of evaluators.

Guidelines of the translation process

The quality of the translation is dependent of the process chosen. The process again is a compromise between what could ideally be done and what is realizable considering limited resources. Given limited resources the web-based tools are used to make the most of them. A guideline for the translation into the German language has been defined and can be accessed by each translator in the wiki. The guidelines are based on the recommendations of the WHO used in the revision process of the ICF [4].

In 2005 the ICN announced the development of a guideline for the translation process. Once it will be available it will be incorporated into the guideline for the translation into the German language.

For the German translation there are currently four main points to consider:

1. Translation of the meaning of the term
2. Identification of a preferred term. A translator can suggest a translation to be a preferred term. The evaluation committee will make the final decision.
3. Optional translations. If a translator is uncertain about a possible translation this can be declared as an 'optional translation'. The BaT allows for multiple optional translations per ICNP term.
4. Synonyms. If there is more than one translation possible preserving the original meaning, the BaT permits the declaration of multiple synonyms per ICNP term. A translator can suggest a translation to be a synonym. Again the evaluation committee will make the final decision.

Since the status of each ICNP term can be set individually it is possible that the translation phase and the evaluation phase are organized to run in parallel.

Rules for the translation phase of each term

Every person translating a term should also do a first self-evaluation of his translation. Along with the translation he should rate his translation as being a:

1. Preferred term,
2. Synonym,
3. Optional translation,
4. or undefined.

Also each translation should be self-evaluated by the translator according to the following criteria:

1. Translation has the same meaning as the original ICNP-term.
2. Translation has a narrower meaning as the original ICNP-term.
3. Translation has a broader meaning as the original ICNP-term.
4. Translating requires more than one translation due to regional differences.
5. Differences between the meaning of two or more original ICNP-terms are lost during the translation process.
6. ICNP-term can be translated, but the meaning is different in the target language.
7. ICNP-term can not be translated. The target language does not have such a concept.

Rules for the evaluation phase of each term

It is planned to nominate seven experts to the evaluation committee. Once the translation of an ICNP-term seems to stabilize it will be closed for translation in the BaT tool. The experts then are evaluating each translation using the BaT tool. For each translation they have to fill out a form to evaluate whether the translation:

1. is a preferred term,
2. is a synomym,
3. requires further translation work,
4. should be considered to be deleted,
5. or the ICNP-term can not be translated.

All translations with an at least 66% vote for preferred term are declared preferred term and finalized. The same applies for the synonyms.

The experts also evaluate each translation according to the same seven criteria of the self-evaluation. This evaluation is used in the processing of the translations requiring further work.

After evaluation of the translations of an ICNP term the results will be made public using the BaT by highlighting preferred terms, synonyms, and translations to be deleted. The ICNP terms with preferred terms and synonyms will be finalized. No further translation or evaluation work will be possible in this stage.

All translations considered for deletion will be open for comments for a while. If there is no reason given for keeping, they will be deleted.

All terms needing further translation work to be done will be opened for translation again and the circle begins. ICNP terms that are impossible to translate will be sent to the ICN to be considered in further releases of the ICNP.

A web-based two-phase Delphi technique

Taking a closer look at the defined process for getting consensus in the translation and evaluation phases similarities to a Delphi technique are appearing. There are actually two teams of experts at work in the different phases. The ques-

tionnaires used in the Delphi technique are generated by the web server.

In the first phase - the translation phase - the experts are the translators. The initial questionnaire is the list of all ICNP terms to translate. But instead of having several rounds of distinct questionnaires that are summed, analyzed, and send back as feed back to the translators, they get a dynamic instrument e.g. a list of translations each time they log into the BaT. They reach consensus once the translations are stabilizing. This means no translator finds it necessary to add another or modify an existing translation. Since at this phase the feedback offered to the translators consists of all the translations for a given ICNP term, this list (questionnaire) is generated automatically by the web server. During this phase there is not need for a moderator to consolidate the translations as they are collected by the application.

In the second phase – the evaluation phase – the experts are the evaluators. Because the form (questionnaire) generated by the BaT tool when evaluating a translation does not contain any open questions, it again can be automatically summed, analyzed and fed back to the evaluators as statistics of the current state of evaluation of a translation. Again the application is doing most of the work.

The first and second phase continue repeatedly until consensus is found for most translations.

The time consuming and expensive procedure of a Delphi technique can be sped up by using a web server to generate and distribute the questionnaires and feedbacks to the experts. The only intervention that has to be done is located at the interface between the two phases in order to decide if the translation of a specific ICNP term has stabilized enough so that the evaluation phase for this term can follow.

Results/experiences

The translation of the ICNP into the German language started in May 2006. About two dozen nurses from Austria, Switzerland, and Germany are working on the translation. A first preliminary translation will be available at the end of December 2006.

The translation is done completely using the BaT tool and the wiki. Both systems have been running without any problems and seem to be well accepted for the process. There seems to be very little discussion necessary outside of the tools. The instructions to use the system are restricted to a few pages in the wiki, but it seems that the tools used are quite self-explanatory.

There seems to be interest in using the system for the translation of the ICNP into other languages as well.

Discussion

So far the system has been used for only one target language. Although it is designed and tested for a multilingual use supporting several translations at the same time this still has to be proven in routine work. But there has been some interest to integrate other languages as well so that might happen in the near future. Up to today the wiki is not really integrated with the BaT. While there are links used to navigate from one application to the other, both are still using a separate user administration requiring each user to login twice which should not be necessary.

Up to today demand for user support has been minimal, although the system is used currently by about two dozen nurses for translation purposes. This could be an indicator that the philosophy of following standard web browsing techniques might pay off in terms of ease of use and low support requirements.

The evaluation module of the BaT has not yet been tested in routine application. This will start in the beginning of January 2007 after the preliminary translation into German is completed at the end of December. Depending on the upcoming guidelines for translation by the ICN some changes might have to be applied to the evaluation module.

Conclusion

Translation of any terminology is time and resource consuming. This can be eased by using a web-based approach. While the process of translating and evaluating being chosen is a good compromise between what should be done ideally and what can be realized given limited resources, principles like back translation should still be considered as a follow up if resources might be made available in the future. Since the BaT tool is already multilingual it can be used for other translations of the ICNP. The already existing infrastructure can be made available free of cost. The source code of the BaT will be made licensed under an open source license.

Acknowledgments

The coordination of the translation is supported by administrative and financial resources of the nursing associations DBfK, SPK and ÖGKV. But without the voluntary efforts donated by the members of the German-speaking ICNP Users group such a translation would be unfeasible.

References

[1] ICN. About ICNP. http://www.icn.ch/icnp_def.htm (visited 2006-12-03).
[2] ICN. ICNP Version 1.Geneva: ICN, 2005.
[3] ICN. ICNP - Internationale Klassifikation für die Pflegepraxis. Bern: Hans Huber, 2003.
[4] WHO. ICIDH-2 Beta-2 Field Trial Studies. Study One: Translation and linguistic evaluation. WHO, 1999. http://www.stakes.fi/oske/icf/study1.pdf (visited 2006-04-13).

Address for correspondence

Prof. Dr. Ulrich Schrader
Mühlgasse 33
D-61231 Bad Nauheim
Germany
Email: mail@ulrich-schrader.de
http://info.ulrich-schrader.de

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Ontology Based Modeling of Pandemic Simulation Scenarios

Henrik Eriksson[a], Magnus Morin[b], Johan Jenvald[b], Elin Gursky[c], Einar Holm [d], Toomas Timpka[a,e]

[a] *Dept. of Computer and Information Sci., Linköping University, Linköping, Sweden*
[b] *VSL Research Labs, Linköping, Sweden*
[c] *ANSER Analytic Services Inc., Arlington, USA*
[d] *Dept. of Social and Economic Geography, Umeå University, Umeå, Sweden*
[e] *Dept. of Social Medicine and Public Health, Linköping University, Linköping, Sweden*

Abstract

Computer-based simulation of influenza outbreaks in local communities can help researchers, epidemiologists, and decision makers better understand the impact of the community structure on the reproduction rate of disease, and the relative benefits of different types of prevention and interventions. The goal of scenario modeling is to develop a description of scenario components, such as the disease, the community, and interventions. An ontology-based representation of the scenario model together with a modeling tool, which is based on an extension to Protégé, assist scenario developers in formulating simulation specifications. This approach allows the exploration of new ideas by rapidly formulating and reconstructing scenarios from novel components.

Keywords:

epidemiological simulations, ontologies, public health informatics, influenza

Introduction

The recent outbreaks of avian influenza (H5N1) have put renewed emphasis on preparing for a human influenza pandemic. Of particular concern in this planning is the allocation of resources, such as vaccine and antiviral medications, which will likely become scarce during a pandemic. One component of pandemic preparedness is surveillance and disease monitoring. Currently, there are operational systems that allow the global monitoring of influenza [1]. A second component of preparedness is simulation and forecasting epidemiological outcomes. In such simulations, networks and the epidemiology of infectious diseases are fundamentally linked. The basis for early epidemiological models is population-wide random-mixing. In practice, however, each individual has a finite set of contacts to which they can pass infection, described as a *mixing network*. Knowledge of the structure of such networks allows models to be created that compute the epidemic dynamics at the population level from the individual-level behavior of infections [2]. Therefore, representation of mixing networks, and how these deviate from the random-mixing norm, can enhance the understanding and prediction of epidemic patterns and intervention measures. Another central problem with forecasting in the context of influenza pandemics is the ecological fallacy; that is, that results obtained at regional or national levels may not necessarily be congruent with local levels, because regional averages may mask diversity between smaller areas. [3,4].

Simulation environments for the prediction of influenza pandemics must support rapid development, must display limited complexity, and allow for flexibility in the underlying models. However, nontrivial changes in software environments currently used for epidemiological simulations often require reprogramming of the simulator, which is time-consuming and potentially error prone. Furthermore, it is interesting to simulate multiple scenarios, such as scenarios adapted to local contexts and variations of scenarios for sensitivity analysis and alternative interventions. Thus, the management of different scenarios and scenario versions, and their implementation in the simulator is an important task for simulator developers. Finally, it is important to perform simulations during an ongoing outbreak as more information about the disease and how it is spread becomes available, which requires fast turnaround times for modeling and simulation.

Our approach is to develop a framework for separating simulation modeling and execution that maintains simulator run-time performance while allowing flexible high-level modeling [5]. Modeling and implementation of models in simulators are different areas of expertise that should be reflected in the architecture. Ontology-based models allow the scenario developer to define the relevant concepts and relationships for the major parts of the model, such as the community, the disease, and interventions. We use an extension of the Protégé ontology editor [6] to model scenarios and their components, such as household and school structures of communities. Furthermore, we use a custom-developed simulation engine that takes input scenario specifications from the modeling tool and produces the simulation report. This approach combines the benefit of an advanced modeling tool with the performance of simulation engines designed for efficient execution.

Background

Mixing with a group of people increases the probability of infection from exposure to an infectious individual in the group. Because people are part of several *mixing groups*, their total probability of becoming infected is the combination of the risks contributed by each of the mixing groups. Figure 1 illustrates sample mixing groups common in most communities.

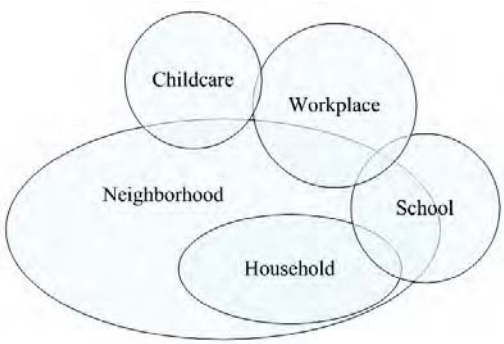


Figure 1 – Sample mixing groups

It is possible to simulate the interaction of people in mixing groups. Several researchers have used stochastic discrete-event simulation models to study outbreaks in simulated populations [7,8,9]. Typically, these approaches use mixing groups with defined transmission probabilities and contact rates, and they simulate the probability of transmission for every individual for the duration they are part of the group. The simulator maintains the dynamic state of each person with respect to infection and mixes people with others in their mixing groups based on the social context. Infectious persons in a mixing group contribute to the total probability that a susceptible person becomes infected.

The advantage of this type of simulation is that it allows us to study the outcome of different scenarios by varying the underlying model. Although the core of this simulation approach is relatively straightforward, there are several modeling issues that scenario developers must address. For example, the community model must reflect the age distribution of the population used to assign children and adults to schools and workplaces, respectively. In addition, the model must define the mixing groups used in the simulation and their properties.

Methods

The requirements of the modeling environment were formulated in terms of flexibility and turn-around-time for the modeling-and-simulation cycle. These requirements resulted in an architecture separating the modeling tool and the simulator engine, and the use of ontologies for modeling. Furthermore, the basis for the structure of the ontology model is the generalization of a previous version of the simulator using hard-coded models founded on relevant literature.

Results

The resulting modeling and simulation architecture approach comprises an ontology for scenario modeling, model formulation tools, and a simulation engine.

Scenario model

The purpose of the scenario model is to define a virtual situation to simulate, such as the spread of a disease in a certain community. Performing a simulator-based study usually involves creating multiple scenarios that alter the variable or variables under study. The scenario developer can prepare a series of scenarios that define the study by reusing (keeping constant) parts of the model while adjusting (varying) other parts. For example, a series of scenarios could simulate the impact of different countermeasures on the basic reproductive number, R_0. Alternatively, it would be possible to study the impact of differences in communities.

A scenario consists primarily of a community model and a disease model. In addition, a scenario can contain references to a model of the intervention policy and a model for mapping the community model to basic mixing groups.

Community model

Our approach is to inform models for simulation of pandemics by social geography by transforming Hägerstrand's concept of *pockets of local order* [10] to homogenous localized population mixing groups. A network structure is used to represent the mixing groups as nodes with contact frequencies between individuals within the nodes, as well as other properties of social pockets. Instead of a fixed arbitrary hierarchy of household groups and regions, we use a system of meeting place centered, floating, overlapping reference areas. As part of the modeling framework, we have implemented a generator that produces a randomized population that fits the household size and household member age distribution specified in the community model. The simulator engine uses this generated population to form the mixing groups during initialization.

Basically, there are two ways in which we can use detailed population data. One approach is to use as much of the detailed data as possible in the simulation; that is, to use information about households, such as residence location, workplace locations, and age of household members, and form mixing groups for the simulator from these data. An alternative approach is to extract the relevant features for the population in question, such as age distribution and household size distribution, and then generate a synthetic population with the same key distributions as the factual population for the simulation.

Disease model

The basic disease model, the susceptible-infectious-recovered (SIR) model, describes the number of individuals (or proportion of the population) that are susceptible to, infected with, and recovered from an infectious disease. Many of the details of infection progression are therefore neglected, as are differences in response among individuals. The SIR model is appropriate for infectious diseases

that confer lifelong immunity, such as influenza. Biomedically motivated modifications can be made to this basic model, such as involving the inclusion of more heterogeneities by further subdividing the SIR classification to reflect either more complex pathogen biology or greater structure within the host population. However, the number of contacts each individual has may be considerably smaller than the population size, and, in such circumstances, random mixing does not occur. Models that incorporate network structure avoid the random-mixing assumption by assigning to each individual a finite set of permanent contacts to whom they can transmit infection, and from whom they can be infected. Although in network and random-mixing models, individuals may have the same number of effective contacts per unit time and within a network, this set of contacts is fixed, whereas in random-mixing models, it is continually changing. Networks thus capture the permanence of interactions. The present disease model has therefore been customized to accommodate also network components in future developments of the SIR disease model.

Scenario modeling

Figure 2 shows the general workflow for scenario modeling. The scenario developer starts with a problem formulation and proceeds with conceptualization and definition of the concepts, properties, and relationships in the scenario ontology. Typically, the developer extends a basic ontology with scenario-specific concepts and instances. This basic ontology provides guidance for the developer. Moreover, the basic ontology can restrict relationships in the extended ontology. Because the scenario ontology is partitioned into community, disease, and intervention models, the developer can modularize the work and, perhaps, use different data sources for different parts, e.g. population databases for the community model, when extending the basic ontology. Once the modeling is completed, the developer can generate the intermediate specification format for use in the simulator engine. Because scenario ontologies extend the basic ontology, the subsequent generation and simulation steps can make assumptions about the concepts in the basic ontology, such as the base concept of mixing group.

Scenario reuse

An important aspect of simulation modeling is the potential to reuse models and model parts. For example, simulator users may want to keep the community and disease models constant while varying the intervention model. Alternatively, simulator users may want to vary the community models while keeping other models constant, for instance to simulate the impact of different community structures and community-oriented interventions on R_0.

Scenario developers can use this approach to build a library of scenarios and scenario components. The advantage of using this type of library approach is that it makes it easier to work with several types of local communities, which could correspond to actual cities, municipalities, and villages, in multiple scenarios. The library could contain past scenarios, i.e. previously simulated scenarios, and scenario components suitable for reuse. Furthermore, it is advantageous to archive scenarios from previous simulation runs together with results for future reference.

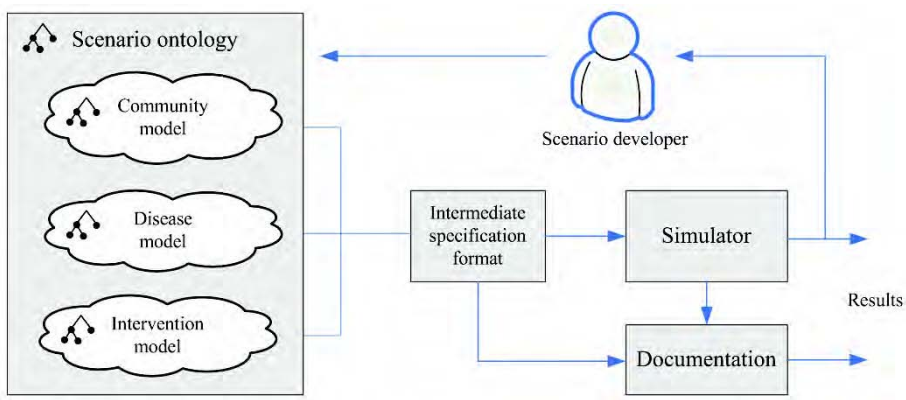


Figure 2 – General workflow for scenario modeling and simulation

Modeling support

Protégé is a widely-used ontology-development environment that provides several graphical tools for creating and editing ontologies [6]. We have extended Protégé with a plug-in for managing scenarios, and for generating specifications for the simulator engine. This extension adds a tab to the Protégé user interface, which allows users to create and define scenarios (see Figure 3). Using the scenario editor, developers can link models to the scenario, such as a community model representing the population used in the scenario.

Figure 5 shows the editing of a community specification in Protégé. This sample community instance consists mainly of the population, a household model with household size distribution, and a school structure model with the number of schools of each type (ranging from daycare centers to high schools).

The community model in Figure 4 corresponds to the city of Linköping, Sweden. It is possible to reuse this model in several scenarios as it represents the standard model for this city. Furthermore, is possible to construct a new scenario by replacing the standard Linköping model with another population, such as the household structure of

Washington DC, which has a high proportion of single-person households, or Utah, which has a low proportion of single-person households.

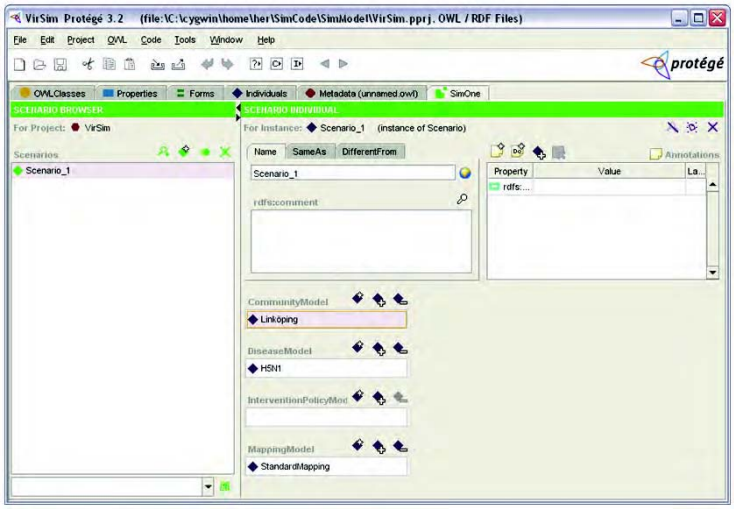


Figure 3 – Protégé tab extension for scenario editing

Figure 4 – Editing of community definition in Protégé

Architecture

Figure 5 shows a summary of the architecture for the modeling and simulation system. Protégé manages the ontology with the community, disease, and intervention models. The population generator, which is implemented as a Protégé extension in Java, expands the abstract community model to a set of generic mixing groups for use in the simulator engine. The intermediate specification format, which is XML-based, defines the mixing groups and their members as well as the disease and intervention model parts. The simulator then parses this intermediate format and instantiates the simulation. One of the advantages of the XML-based specification is that it makes it possible to transfer the specification to alternate simulation-execution architectures, such as an array of processors performing parallel simulations.

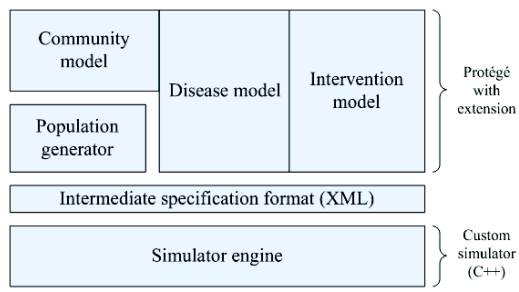


Figure 5 – Architectural layers

Discussion

Scenario modeling is an important part of the scientific process. In particular, modeling assists hypothesis formulation and simulation documentation. However, epidemiological simulation is limited by a modeling bottleneck in that it is difficult to develop and modify models for simulators. Additionally, epidemiological simulation often requires simulation experts and programmers to perform the modeling task. An important goal of our work is adaptation to local contexts to support simulation of local outbreaks. Ultimately, we want to enable local epidemiologists and decision makers to use simulation. Thus, the modeling approach must allow developers to construct models that reflect the structure of neighborhoods and interventions at the local level, informing decisions such as closing schools and workplaces. The systematic development of scenarios from components helps developers to construct case models rapidly, and to run them in the simulation engine.

In our modeling of H5N1 influenza outbreaks in local communities, we found that the ontological modeling helped structure the problem, reuse model components, and organize different scenarios and their corresponding simulation results. Furthermore, the use of a spatially explicit community model allows us to employ several alternative community structures for sensitivity analysis and for experimenting with moving an alternative population, such as one with a different household structure, into a community. In our continued work, we plan to extend the ontology to support the aggregation of scenario descriptions in larger simulation jobs that specify simulation series. In addition, we plan to generate documents, e.g. PDF documents that summarize the simulation model used and describe simulation results.

Summary and conclusions

Simulation of influenza outbreaks benefits from a clear separation of modeling and implementation. Ontologies provide a suitable representation scheme for such epidemiological models. It is useful to divide the scenario model into subparts for communities, diseases, and interventions and to support construction of new scenarios from these components. Protégé, combined with the scenario extension, provides a modeling environment that uses ontologies to define scenarios and scenario modules. Moreover, such ontologies, when populated, make up a scenario library with a collection of instances representing the scenarios. The modeling result is a specification for efficient execution in the simulation engine.

Acknowledgments

This work was supported in part by the Swedish Emergency Management Agency under contract 0700/2004 and in part by the Swedish Research Council under contract 2006-4433.

References

[1] Flahault A, Blanchon T, Dorleans Y, Toubiana L, Vibert JF, Valleron AJ. Virtual surveillance of communicable diseases: A 20-year experience in France. Stat Methods Med Res. 2006 Oct;15(5):413–21.

[2] Keeling MJ, Eames KT. Networks and epidemic models. J R Soc Interface. 2005 Sep 22;2(4):295–307.

[3] Meade MS, Earickson EJ. Medical geography (2ed.). New York: The Guildford Press, 2005, p.348, 439-40.

[4] Timpka T, Morin M, Jenvald J, Gursky E, Eriksson H. Dealing with ecological fallacy in preparations for influenza pandemics: Use of a flexible environment for adaptation of simulations to household structures in local contexts. Proc Medinfo 2007.

[5] Timpka T, Morin M, Jenvald J, Eriksson H, Gursky E. Towards a simulation environment for modeling of local influenza outbreaks. AMIA Annu Symp Proc. 2005; 729–33.

[6] Gennari JH, Musen MA, Fergerson RW, et al. The evolution of Protégé: An environment for knowledge-based systems development. Int J Hum Comp Stud 2003; 58(1):89–123.

[7] Longini IM, Nizam A, Xu S, et al. Containing pandemic influenza at the source. Science 309:1083–1087.

[8] Elveback LR, Fox JP, Ackerman E., et al. An influenza simulation model for immunization studies. Am J Epidem 1976; 103:152–165.

[9] Longini IM, Halloran ME, Nizam A, Yang, Y. Containing pandemic influenza with antiviral agents. Am J Epidem 2004; 159:623–633.

[10] Hägerstrand T. Time-geography: Focus on the corporeality of men, society and environment. In The Science and Praxis of Complexity. New York: United Nations University, 1985.

Address for correspondence

Henrik Eriksson, PhD.
Department of Computer and Information Science
Linköping University
SE-581 83 Linköping, Sweden
Email: her@ida.liu.se

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

An Ontology-based Model of Clinical Information

Thomas Beale[†a], Sam Heard MD[‡ab]

[†] *CTO,* [‡] *CEO Ocean Informatics Pty, Ltd*
[a] *Visiting Senior Research Fellow, University College London*
[b] *Adjunct Professor, University Central Queensland*

Abstract

In this paper we describe a model of clinical information designed to make health information systems properly interoperable and safely computable. The model is a response to a number of categories of requirements, ranging from the semantic to the performance of software at runtime. We argue that the starting point of a successful model must be an ontological analysis of the process of clinical care delivery, seen as a scientific problem-solving process. From this approach we develop a classification of types of clinical information called the Clinical Investigator Record (CIR) ontology.

Keywords:

EHR, health records, clinical process, information models, information systems, archetypes, problem-solving.

Introduction

Interoperability and computability of information in the healthcare sector promise a great deal in the global effort for better quality and more efficient healthcare, yet remain largely unachieved. Models of clinical information, including models of the Electronic Health Record (EHR) and variants (EMR, EPR, CPR etc), do not currently have a theoretical basis sufficiently strong to guarantee interoperability and computability. Any information model developed must also satisfy a myriad of other requirements when actually implemented including: computational efficiency and performance; economically viable implementation; maintainability; system scalability and extensibility; and the considerable privacy and security requirements of the health domain. A successful model of health information is thus like any other good model; it is an expression not only of the semantics of the domain, but also a response to the needs of economics of system construction and runtime performance.

Methods

Our methodological approach to developing a model of health information consisted of four steps. Firstly a conceptual model for understanding the creation and recording of information during the healthcare process was developed. This was used to develop an ontology of recorded health information. An information model suitable for software development was developed based on the process model and ontology as well as other requirements - this became the *open*EHR Entry model. Finally we validated the model in a number of ways, including testing it with real clinical statements and queries and showing how it works in implemented systems. This paper describes the first two steps.

Background

Systems and models whose design ignores the real world in which they operate will fail to work well (see e.g. [1], [2]). The challenge is to find a way of basing a system design on the reality in which it is intended to function. An ontological approach provides a formal basis. An 'ontology' can be understood as a 'description of reality' and may be subdivided into a description of *process* (e.g what occurs, what is done) anda description of *information* (e.g. what is thought, said, communicated, remembered).

Sowa [3] provides a top-level and very general classification of the world, which is a useful reference point for any ontology, while more health-specific ontologies include the Ontology of Biomedical Reality (OBR) [4] and the SNOMED-CT clinical terminology [5]. The CIR ontology describes information created in the process of healthcare delivery.

A model of clinical information that will perform at all levels must meet three broad categories of requirement:

- *semantic requirements*: accurately represent and convey intended meaning of its users;
- *functional requirements*: provide the functions required by its users when deployed; and
- *economic requirements*: enable economically viable software construction and maintenance into the future.

A process model for clinical healthcare

It is important that any model used as a basis for software be based on a generic and high-level conceptualisation of the care delivery process. Various attempts to do this have been made in the past. Weed's problem-oriented medical record (POMR) methodology [6] formally linked a particular model of the process of care (relying heavily on testing) to the information gathered during that care (leading to the well-known "SOAP" headings). Elstein's

'hypothetico-deductive' model of clinical reasoning [7] mainly accounts for the cognitive aspects of clinical care (cue recognition and evaluation) during diagnosis. The model makes a good case for clinical healthcare to be seen as an iterative, scientific problem-solving process. The model originally developed by Rector, Nolan and Kay in the PEN&PAD system [8] is not so much based on problem-solving, rather a faithful yet flexible ability to *record* clinical action, thinking and dialogue. They propose only a limited ontology of information, namely 'direct observations' and 'meta-observations', thus distinguishing 'facts' from 'opinions'.

The Danish 'general electronic patient journal' (G-EPJ) [9], shown in Figure 1, includes a conceptual model of the iterative problem-solving process and categories of information generated. This model represents the clinical investigative process in the form of the cycle with process steps (circles) and information arising (blocks). While it has proved too rigid in practice, it formalises both process and information based on a paradigm of rational problem-solving.

Various clinical modelling efforts dating from the RICHE [10] project to the present HL7 version 3 standard [11] have based their models on an 'act management' paradigm in which all aspects of clinical care are represented as Acts. This approach enables 'everything that is done' to be recorded, which has clear attractions for business process tracking and cost-recovery (based on costs associated with fine-grained acts). It also has applications in messaging, where most notifications are to do with acts being requested, or being carried out.

Useful principles

A proper understanding of the notion of 'recording' is required. The point of view taken here is that what is recorded in the health record is likely to be a small, selective choice of notes about real events, situations etc, intended for interpretation by other professionals. The implication is that any model of health information must capture the *cognitive* communication processes of health professionals wherein very partial information may be recorded, rather than some more general notion of 'comprehensive fact representation'.

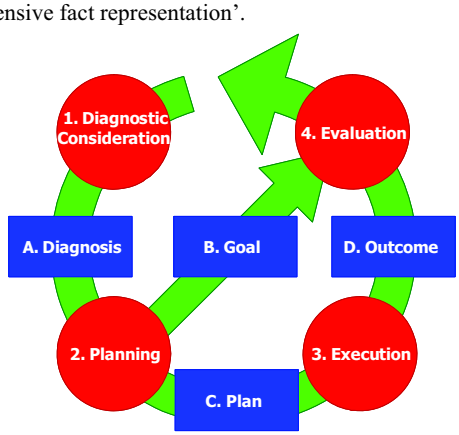
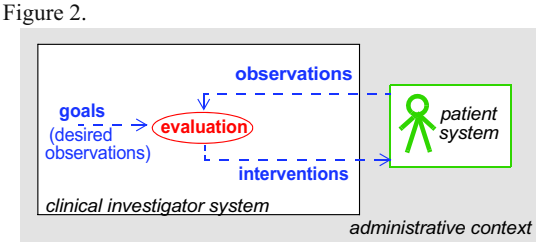


Figure 1 Danish G-EPJ Conceptual Model

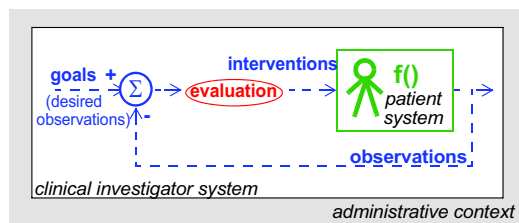
It is also important to avoid the danger of confusion of linguistic entities with real entities and phenomena. This often manifests in models that define types with names like 'problem', 'observation', 'plan' and 'result', without properly defining the meaning of these categories.

The clinical investigator model

The theory we describe in this paper was originally conceived during the GeHR (Australia) [13] (pp29-32) and *open*EHR [14] projects, and built on some of the earlier work described above. We model health care delivery as two kinds of process: a *clinical process,* corresponding to the interaction between a 'clinical investigator system' and a 'patient system', situated within a *business process*, which is owned by an 'administrative context'. The model can be illustrated in two equivalent ways, as shown in Figure 2.



a) problem-solving metaphor



b) control system metaphor

Figure 2 - The Clinical Investigator Model

In part a) of this figure, the clinical process is shown as a two-entity system: a subject of investigation - the 'patient system', and the investigating entity - the 'investigator system'. The former is defined as the subject of care (typically one person, but may be more) seen essentially as a biological or social system (depending on the perspective of the investigator), while the latter is defined as the totality of healthcare professionals and other agents who perform actions to do with care of the patient; this includes the patient in the role of self-carer or self-medicator, as well as any family members or other people. The investigator system's purpose is to use *observations* as a basis for *evaluation*, leading to *interventions*. The investigator system is driven by *goals*.

Part b) of the diagram shows exactly the same elements, but redrawn in a manner familiar to control system engineers and those familiar with systems theory. In this schema, goals (desired observations) are seen as the controlling input signal, with observations being the output, and the patient the filter, or 'transfer function'. The difference between the actual and desired observations is used

by the controller (the evaluation activity) to determine the actual (corrective) input signal to the patient, in the form of interventions. In acute care, the external control system realised by the clinical investigator(s) is attempting to perform the same kind of homeostatic function that the body itself does, and in some cases, it is performing *exactly* the same function, such as in the case of patients using substances such as insulin. The fact that the clinical process can be mapped directly onto the standard model of negative feedback is not only intellectually satisfying, but indicative that this particular conceptualisation of the process is likely to be correct.

An ontology of clinical information

Since our main goal here is to understand how to support healthcare delivery using information systems, we initially need to know what kinds of information might be created by the processes described above *for recording and use in the information system*. We can redraw the investigator system in order to more clearly show the types of information created during the care process, as shown in Figure 3. Five types of information are identified, as follows:

- *observation*: information created by an act of observation, measurement, questioning, or testing of the patient or related substance (tissue, urine etc.), including by the patient himself (e.g. taking own blood glucose measurement), in short, *the entire stream of information captured by the investigator, used to characterise the patient system*;
- *opinion*: inferences of the investigator using the personal and published knowledge base about what the observations mean, and what to do about them; includes all diagnoses, assessments, plans, goals;
- *instruction*: opinion-based instructions sufficiently detailed so as to be directly executable by investigator agents (people or machines), in order to effect a desired intervention (including obtaining a sample for further investigation, as in a biopsy);
- *action*: a record of intervention actions that have occurred, due to instructions or otherwise;
- *administrative event*: a record of a business event occurring within the administrative context, such as admission, booking, referral, discharge etc.

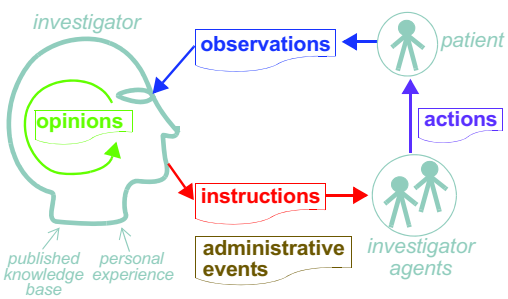


Figure 3 - Information Created by the Investigator

From these categories we construct an initial ontology, shown in Figure 4, the main purpose of which is to introduce the abstract categories 'recorded information' and 'care information', and to situate the five information types with respect to these.

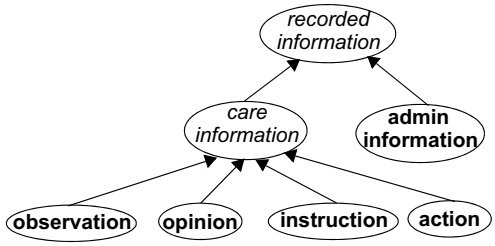


Figure 4 - Initial Clinical Information Ontology

Historical information: observation and action

We have already defined 'observations' as any information received by the clinical investigator used to characterise the patient state. It is axiomatic that observation information is 'in the past', since it expresses states or events that have already occurred. Observations are therefore *historical* in nature.

The Action information category is also historical in nature: a record of actions taken cannot be written until the actions have been completed (the documentation of actions 'to be done' is another category we call Instruction). Clinical actions such as administration of a medication or discontinuing a medication can all create Action instances. We can conclude, therefore, that the Observation and Action categories are subsumed by a category which we will simply call Historical (following Sowa's top-level categories). These categories can now be understood respectively as: a historical record of natural phenomena to do with the patient, and a historical record of things done to or for the patient.

Information in the cognitive 'present': opinions

Within the cognitive investigation process, clinical thinking, including any analysis, assessment or planning, is always 'in the present'. This is true of any analytical or decision-making activity: temporally it comes between evidence about the subject of study and actions to be performed in the future. When specific examples of clinical opinion are examined, the relationship with time may not always seem so clear. Consider a diagnosis; to properly characterise the problem, it will include time-related information such as the following:

- date of initial onset
- date clinically recognised
- for each occurrence or exacerbation:
  - date of last onset of occurrence
  - date of resolution
- date of resolution

Timings like this represent the 'continuant' or persistent relevance of this information and as such the Opinion category includes the idea of summary or review. As such the diagnosis is not itself an *observational* record of onset, exacerbation, resolution etc. (these events may have been documented as Observations when they occurred), but a

structured *summary* of dates and other items seen as important by the physician. The Opinion category of information is thoughts and analysis *about* previously recorded facts, not a historical record of the facts themselves. This is the same as items reported in the "news" section of a newspaper, and later analytical pieces, such as editorials and current affairs TV shows - the latter cannot help mentioning facts and dates from the former, but is not itself the contemporary record of the events in question. The Opinion category corresponds to Sowa's Description category.

Information about the future: instructions

We denote the last major category of information coming from the clinical care process as Instruction. Information in this category is in the form of orders, requests, and other directions to be performed by clinical investigator agents, so as to effect desired changes in the patient system, or else to gather new evidence for further consideration. Instruction time is therefore in the future, since the information speaks of events that have not yet occurred or are not completed. The Instruction category corresponds to Sowa's Script category.

We can describe two sub-categories of Instruction, distinguished on the basis of being for the purpose of investigation (to make further observations) or for intervention (aimed at changing the state of the patient system). We will call these categories Investigation-request and Intervention-request. As with the other categories, these names are to be understood as contractions for 'record of instruction for investigation' and 'record of instruction for intervention'. These may overlap when investigations have some treatment aspect; for example, removal of fluid around the lung will aid in the diagnosis as well as alleviate symptoms.

One of the key clarifications provided in the types described so far is between Instruction and Action. The former is a specification of what to do, while the latter is a record of what was done. A medication order may say to take 1 or 2 tablets every 4 hours, whereas an administration action will record the actual number taken. Models that are not clear about this distinction are likely to be problematic when connected to workflow systems. In simple terms, a series of Actions always represents a particular path taken through the set of possibilities described by an Instruction.

Accordingly, we arrive at an improved form of the ontology in which the three categories History, Opinion and Instruction are distinguished on the basis of the temporal relationship between the recorded information and the investigation process.

Observation and intervention sub-categories

We can make further improvements by analysing the categories so far in terms of their relationship to the two sides of the investigation process, i.e. 'observation' (characterising the patient system) and 'intervention' (changing it). We have included two subclassifications on this basis, namely Observation/Action and History and Investigation-request with Intervention-request below Instruction level.

Types of opinion

The top-level category *opinion* corresponds to inferences about observations, as well as about interventions, such as goals, and plans. Types of opinion have historically posed the most difficulties for models of health information, because they are so variable. The Opinion category corresponds to the notion of 'hypothesis' in general science. In health informatics, it corresponds to Weed's 'assessment' and (partially) 'plan' categories, to Van de Velde & Degoulet's [12] 'abstraction' category, and to Rector *et al's* [8] 'meta-observations'.

Here we propose two sub-categories, more directly related to the systems view of healthcare proposed above: *assessments*, i.e. opinions about what is happening in the Patient System, and *proposals*, i.e. opinions about what should be done by the Investigator system. Under this approach, we can say that the first category relates to *current* or *projected* states of affairs, and the second to a *desired* state of affairs.

The primary example of an Assessment in clinical medicine is the diagnosis. A diagnosis is the attachment of a label to a group of observed signs and symptoms, which designates it in the understanding of the Investigator as being a particular known phenomenon. A differential diagnosis, another category of assessment, allows for multiple possibilities, due to the lack of sufficient information or understanding to discount all but one.

Assessments include *quantification of risk* providing the basis for prevention or investigation. Such a notion includes the idea of *family history* being a risk arising from an unusual prevalence of a condition amongst relatives. *Prognosis*, the likely outcome corresponding to a current diagnosis, is a further example. In general science, risk assessment and prognosis correspond to *prediction*.

The second major sub-category of Opinion corresponds to opinions about the *desired* state of the patient system, and includes much of the creative thinking of the clinical investigator. Once a diagnosis, risk assessment or other evaluation about the patient's condition has been formulated, the clinical investigator determines what to do about it. Three key types of clinical thought at this point are *scenarios* (or what-if statements, *goals* and *plans*. A goal, such as weightloss and a target weight of 85kg, is a statement about *what* the desired state of the patient system should be, while a plan is a statement about *how* to get there.

In summary, the Opinion category is distinguished from the Observation category by representing inferences from evidence, rather than representing the evidence. Two investigators can form different interpretations of the same set of observations, but the observations themselves remain an objective picture of some aspect of the patient's situation, within the limits of the observational method itself. Similarly, two investigators can formulate different goals and plans based on the same observations, and even the same diagnosis.

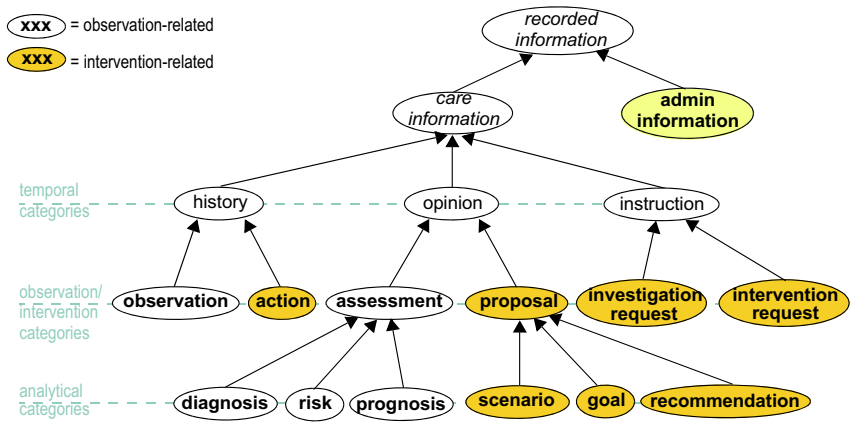


Figure 5 - The Clinical Investigator Record (CIR) Ontology

The clinical investigator record ontology

The ontology thus described is illustrated in Figure 5. Categories which correspond to the describing the patient system are shown using plain bubbles, while categories corresponding to interventions intended to change the system are shown using filled bubbles. We call this ontology the Clinical Investigator Record (CIR) ontology. The CIR ontology does not yet include any detailed subcategories beneath Admin information. Far less work has been done on this subject from an ontological point of view in health to date.

Conclusion

The Clinical Investigator Record Ontology provides the basis for the Entry classes in the *open*EHR reference model. The latter has proved to be semantically robust and a successful basis for defining clinical content models (known as 'archetypes') used in various kinds of health computing, as well as for EHR systems. The archetyping methodology has been used to create over 250 archetypes based on the *open*EHR Entry classes during the last few years, including in recent work (early 2007) in the UK NHS Connecting for Health programme.

Acknowledgements

The authors wish to thank the following for their collaboration and input over many years: Dipak Kalra, David Lloyd, Alain Maskens, Alan Rector.

References

[1] Heeks R, Mundy D, Salazar A. *Why Health Care Information Systems Succeed or Fail*. Information Systems for Public Sector Management - Working Paper Series Paper No. 9. 1999. Institute for Development Policy and Management.

[2] Scott J T, Rundall T G, Vogt T M, Hsu J. *Kaiser Permanente's experience of implementing an electronic medical record: a qualitative study.* BMJ 2005;331: 1313-1316.

[3] Sowa J. *Knowledge Representation: LogicalPhilosophical and Computational Foundations*. 2000 Brooks/Cole. (Top-level categories available at http://www.jfsowa.com/ontology/toplevel.htm).

[4] Rosse C, Kumar A, Mejino JLV, Cook DL, Detwiler LT, Smith B. *A Strategy for Improving and Integrating Biomedical Ontologies*. Proceedings of AMIA Symposium 2005, Washington DC, 639–643. (available at http://ontology.buffalo.edu/bio/OBR.pdf).

[5] SNOMED-CT. See http://www.snomed.org.

[6] Weed LL. Medical Records, Medical Education and Patient Care. *The Problem Oriented Medical Record as a Basic Tool*. Cleveland: Case Western Reserve University ress, 1968.

[7] Elstein AS, Shulman LS, Sprafka SA. *Medical problem solving: an analysis of clinical reasoning*. Cambridge, MA: Harvard University Press 1978.

[8] Rector A L, Nowlan W A, Kay S. *Foundations for an Electronic Medical Record*. The IMIA Yearbook of Medical Informatics 1992 (Eds. van Bemmel J, McRay A). Stuttgart Schattauer 1994.

[9] Bruun-Rasmussen M, Bernstein K, Vingtoft S, Nøhr C, Andersen SK. *Quality labelling and certification of Electronic Health Record Systems*. Studies in Health Technology and Informatics 2005; 116: p47-52.

[10] RICHE Consortium. RICHE ESPRIT Project. Final Report. Nov 30 1992.

[11] HL7 International. *Reference Information Model (RIM)*. See http://www.hl7.org.

[12] Van de Velde R, Degoulet P. *Clinical Information Systems: a Component-Based Approach*. 2003 Springer-Verlag New York.

[13] Beale T, Heard S. *The GEHR Object Model - Technical Requirements*. 2000. See http://www.openehr.org/gehr_australia/gehr_requirements.pdf.

[14] Beale T, Heard S. *Architectural Overview of openEHR*. 2006, The *open*EHR Foundation. See http://svn.openehr.org/specification/BRANCHES/Release-1.1-candidate/publishing/architecture/overview.pdf.

Address for correspondence

Thomas.Beale@OceanInformatics.biz.

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Creation of a Local Interface Terminology to SNOMED CT

Alejandro Lopez Osornio [a], Daniel Luna [a], Maria Laura Gambarte [a], Adrian Gomez [a], Guillermo Reynoso [b], Fernán González Bernaldo de Quirós [a]

[a] *Medical Informatics Department, Hospital Italiano de Buenos Aires, Argentina*
[b] *Conceptum, Buenos Aires, Argentina*

Abstract

This paper describes the steps followed in the creation of a local Interface Terminology to SNOMED CT (as reference terminology) with a strong focus on user acceptability. The resulting list of terms is used for clinical data input by physicians and nurses at the Hospital Italiano in Buenos Aires, Argentina. Description includes data model, mappings to SNOMED CT and classifications, subsets definitions and extensibility mechanisms. The Interface Terminology is currently used in the recording of diagnosis and procedures in inpatient discharge summaries and its coverage is improving from user feedback. Its current size is 24,800 concepts, 67% of them needed post-coordination for appropriate semantic representation, due to a very flexible policy that allows the use of any number of modifiers on concepts.

Keywords:

interface terminology, terminology services

Introduction

Interface terminologies (IfT) had been used for a long time in electronic medical records, and consisting in a list of local terms, familiar to the user, that are used for clinical data entry[1]. The name "interface" comes from its role as a link between user's vocabulary and a stricter, standardized, list of terms used to describe clinical data. Each term in the IfT is mapped to a term in any standard terminology used as a "reference". This strategy isolates the user from the complexity of standard classifications or nomenclatures used as reference terminologies, whose terms may not be appropriate for the setting, may have an arbitrary level of detail or may have too many rules for code selection.

SNOMED CT®[2] is currently regarded as the most advanced terminology system for storing clinical information, and is a government mandated standard in many countries[3-8]. SNOMED CT includes mappings to several standard classifications as ICD-9CM and ICD-10. In this paper we will review the approach of the Hospital Italiano of Buenos Aires for implementing an Interface Terminology mapped to SNOMED CT in a hospital wide Health Information System.

Background

The need of a local interface terminology for SNOMED CT

SNOMED CT includes IfT capabilities, with a list of different terms applied to each concept. The advantage of using a local IfT is to provide easier adaptation to local content and to create a permanent, list of terms, isolated from periodic SNOMED CT changes.

Localization involves two basic functions: hiding not desired content of SNOMED CT and including new terms and concepts needed in the local setting but not present in the standard terminology. These objectives can be achieved using two SNOMED CT functions, the "subsets" for restricting content and the "extensions" for adding new content.

Defining subsets for content restriction can define different scopes, using large subsets for extensive topics like "Problems List" and smaller ones for specific purposes like data input from templates, etc. The local IfT consists in all the subsets created for local use, including a mix of standard terms and local terms, created in an extension of SNOMED CT.

The main drawback of creating a local IfT is the additional work required for its development and maintenance. In our case, conforming to usability demands from our user base was a key point in our implementation, so the extra effort in creating and maintaining an IfT was deemed acceptable.

Setting

The Hospital Italiano of Buenos Aires is a 650-bed non-profit university hospital located in Buenos Aires, Argentina. More than 150,000 outpatient visits and 3,000 hospitalizations are registered every month. It is affiliated with a Health Maintenance Organization (Plan de Salud) that takes care of a population of 140,000 patients.

Since 1998, a full scale HIS has been gradually implemented, including ambulatory Electronic Medical Record (EMR), inpatient discharge summary, administrative systems, scheduling systems, inpatient tracking systems, pharmacy systems and complementary studies report and visualization. Several health informatics standards had been implemented, including HL7, CDA Version 2, ICD-9CM, DRG, ICD10, and ICPC.

Existing experiences

Several health systems created local IfTs for their electronic medical records, mapped to necessary billing classifications like ICD-9CM and other standard terminologies, some examples are shown in Table 1. Sometimes, IfTs were created from scratch and others were based on previously existing standards.

Table 1 - Examples of interface terminologies

Interface Terminology	Site
MED (Medical Entities Dictionary)[9]	Columbia University and the New York Presbyterian Hospital
VA terminology Lexicon[10]	U.S. Department of Veterans Affairs
Convergent Medical Terminology[11]	Kaiser Permanente
UNMC Lexicon[12]	University of Nebraska Medical Center
ICPC Plus[13]	Australia, General Practice
Mayo Master Sheet Index[3]	Mayo Clinic in Rochester, Minn.

Terminology Services had been described as a way to implement centralized terminology access in enterprise wide electronic health information systems[14, 15]. HL7 Common Terminology Services (CTS) specification proposes a full set of features that a terminology server should provide.

Recently, Rosenbloom et al. published a review about IfT features and challenges[1], which we used to adjust and rethink some parts of our project, and we are now applying some of these concepts to structure this paper.

Objectives

The key objective of our project was to build a local Interface Terminology that allows users to record clinical data choosing options from a list of familiar terms but storing information SNOMED CT -compatible.

The IfT should provide adequate coverage for our reality. Users should have the ability to refine terms, choosing a more specific option of a given term, and propose new terms to improve coverage. The system should provide the equivalence of a given local term in standard classifications.

A "Terminology Team" would be in charge of the maintenance of the IfT, the system should also provide the tools required by this team.

System description

Concepts are unique cognitive representations of objects or situations in the reality in people's minds. Concepts are represented in textual form by terms, also called descriptions in this setting. Terms included in an IfT are arranged in concepts, expressing synonymy, and concepts are linked to other semantically related concepts in hierarchies.

Scope of the local IfT covers most aspects of medicine in a setting like ours. For the purposes of this paper, we will describe the development of the "Diagnosis Subset", 'Procedures Subset", "Drugs Subset" and a very specific "Causes of liver failure Subset" as an example.

The creation of the local IfT included the steps described in the following sections using the "Diagnosis Subset" as the main example.

Definition of IfT data structure

Our IfT was designed to take full advantage of SNOMED CT semantic structure and description logics, so our data structure is also tied to the standard SNOMED CT structure. SNOMED CT provides a data structure[16] (Figure 1) that allows storing local data following simple guidelines; local data can be differentiated from original SNOMED CT standard data by a "namespace" identifier included as part of each concept, term or description identifier.

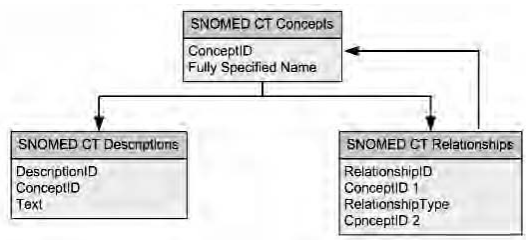


Figure 1 - Overview of SNOMED CT core data structure

This strategy is appropriate for small additions to SNOMED CT. In our case, we aimed to create an extensive local IfT, as independent as possible from the reference terminology. We decided to duplicate SNOMED CT data structure, and fill the empty model with our local data. The concept and description tables will contain our local concepts and descriptions, regardless if they are also present in standard SNOMED CT data or not. The relationship table will work as the link between the two models, linking from each local concept to standard SNOMED CT concepts (Figure 2).

Concepts that exist in both terminologies are linked with a "same as" relationship. New concepts in the local IfT are linked with a "is-a" relationship with the appropriate supertypes in SNOMED CT structure. The defining characteristics of the new concept, those that make it different from its super-type or father concept, are represented using other "attribute" relationships to standard SNOMED CT concepts, following official SNOMED CT modeling guidelines. The definition process using additional attributes is called post-coordination, which is done by experts from the Terminology Team based on user request of new terms. In this way, our local relationship table always links from local concepts to standard SNOMED CT concepts, providing semantic information to the local model.

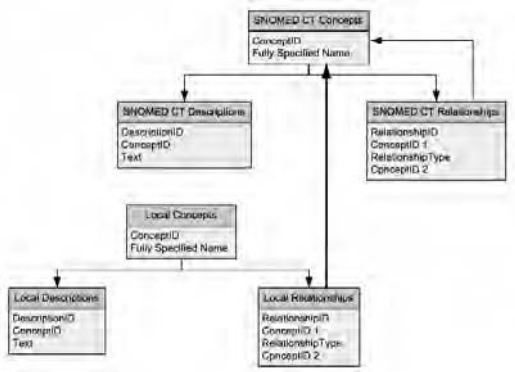


Figure 2 - Relationship between local IfT data structure and SNOMED CT core data structure

All concepts will be unique and each will have a unique "Fully Specified" description that fully describes the concept meaning, a preferred term for displaying to the user and a list of synonyms (descriptions).

Unique identifiers for concepts, descriptions and relationships are independent numbers, conforming to the SNOMED CT concept ID specification, using a local namespace. Applications use these identifiers to store data in the clinical repository and to retrieve subsets for structured entry in user interfaces.

Applications will code clinical data using the identifier for the description selected by the user, not the concept identifier. In this way, whenever the terminology team changes the relationship between a concept and some of its terms, pointing to another concept, there is no need to recode previously stored data in the clinical data repository.

We also decided that terms would be unique in our IfT, so each term is a synonym of only one concept. Ambiguous terms, defined as terms that possibly would relate to more than one concept were treated separately, as will be described later.

Subsets are implemented using SNOMED CT subset mechanism for enumeration of members (Figure 3).

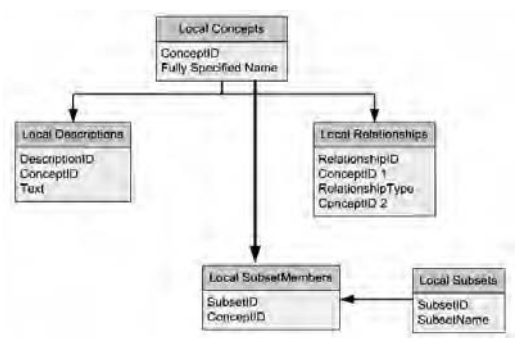


Figure 3 - Local Subsets data structure

Local terms collection and vonceptual aggregation

Some IfT's domains are expected to have more local variation and need to be more adjusted to our setting, such as terms for the diagnosis subset. On the other hand, some smaller issues, like drugs, have a much more limited number of textual variations and any standard list can be presented to the user as the IfT.

The Hospital Italiano previous vocabulary control system for the problems list consisted in physicians entering free text descriptions in the electronic medical record, which were later assigned classification codes by a group of coders. This process generated a backlog of more than 2 million free text phrases describing medical problems, entered by the professionals of the institution. We used this data for creating the terms included in the "diagnosis" subset of our IfT.

After applying a normalization process to these data, a list of more than 250,000 different short text phrases was created; about 110,000 descriptions appeared with 10 repetitions or more, accounting for more than 90% of the free text expressiveness variation of our setting. These 110,000 terms were manually revised, and synonyms were grouped together. We adopted the definition that synonyms are a group of terms that describe the same unique concept. After this process, the 110,000 terms were arranged in 24,800 concepts. We included these concept and synonym data into the "diagnosis" subset of our Interface Terminology. Only 33% of these concepts are direct links to SNOMED CT, a number of additional attributes where needed to represent the remaining 67%.

Terms for "Drugs" subset were imported from our Computer Physician Order Entry (CPOE) software databases.

For specific data-entry subsets like the "Cause of liver failure Subset" content was provided by specialists, describing the possible valid inputs.

Defining invalid terms, refinement and disambiguation rules

During manual revision of the free text entries, several terms where flagged as "Invalid" when their use was not appropriate in each sub-domain. Examples of invalid terms are uncompressible strings of text or terms of concepts not considered valid entries in the sub-domain, like entering procedures in a diagnosis field. In the implementation of the IfT, user interfaces should not allow terms included in this list to be added to the clinical record.

Some concepts were flagged as "Mandatory Refinement", when represented too general concepts and its clinical use would be improved with more detail. An example of that is the term "Diabetes Mellitus", which would be much more useful if refined to "Diabetes Type I", "Diabetes Type II", etc. In the implementation of the IfT, if the user selects one of these concepts he/she should be enforced to select a more specific concept, and a set of appropriate options should be presented, according to the semantic relationships in the local IfT.

Physicians were found to record data using less detail than they actually have obtained during a consultation or examination. Therefore, it must be possible to refine every concept to its more detailed sub-types, and these options should be presented to the user as non-mandatory alternatives.

Ambiguous terms cannot be assigned as valid descriptions of any concept and should be manually disambiguated by the users. These terms were stored separately and were related to the possible concepts in the Interface Terminology, so user interfaces should present the options to the users and force them to specify the concept they wish to use.

Adding semantic definitions through SNOMED CT mapping

Each local concept was modeled either as a direct equivalence to SNOMED CT or its meaning was described with a set of attribute relationships following SNOMED CT guidelines. The resulting relationships were also used for detecting synonymy, as there is a good chance that two terms that share the same definition refer to the same concept.

At this point, our Interface Terminology consisted only in a list of unrelated concepts in each subset, each of them with one or more textual descriptions. In our model all semantic information is derived from SNOMED CT. That means that it would be possible to understand the relationship between two terms in the Interface Terminology by following SNOMED CT semantic structure as can be interpreted from our data structure (Figure 2). This is equally possible for our post-coordinated terms, as they share the same semantic structure with original SNOMED CT concepts.

Definition of cross-maps to standard classifications

Mapping from our local IfT to standard classifications like ICD-9CM, ICD-10 or ICPC2 is provided through the SNOMED CT standard cross-maps mechanism.

Concepts included in the local IfT that are not present in the original SNOMED CT distribution, are mapped through its super-types, the more general, standard concepts used for represent their meaning.

This strategy was preferred instead of a direct linking in order to provide a full access to available standard cross-maps sets from SNOMED CT to classifications.

Defining strategies to mMaintain domain coverage

The coverage of the Interface Terminology can be described at two levels, conceptual level, and term level. The conceptual level refers to the proportion of the locally used medical concepts that are included in the terminology. Term level coverage refers to the proportion of the different text descriptions or terms used to describe those concepts which are included in the Interface Terminology.

The coverage is originally determined by the accuracy of the initial process of data recollection for inclusion in the Interface Terminology. Later, this coverage is maintained and adjusted by the expansion of the terminology, based on user experience.

According to its definition, the expansion process had to be based on direct user suggestions that would be manually evaluated by the Terminology Teams prior to its formal incorporation in the interface terminology. Software user interfaces would enforce user participation with suggestions and commentaries.

Definition of strategies for updating to new versions of SNOMED CT

The IfT should provide an adequate solution to the management of new SNOMED CT versions. New versions may carry ambiguity, including new standard concepts that were previously included in the local extension. In this case, the local IfT concept should map directly to the new SNOMED CT concept instead of the super type and the attributes used to model its meaning. Classification algorithms can automatically detect these cases and remap the concepts maintaining consistency.

Another problem is SNOMED retiring concepts or terms in new versions, in order to correct problems in its representation .When those concepts were used in the representation of the local IfT, concepts would be pointing to inactive codes in the standard. Current SNOMED CT mechanisms do not support automatic selection of new, valid, codes to replace retired ones. The IfT should provide a local valid list of concepts and terms whose representation is affected by changes in the new version of SNOMED CT to manually correct its representation.

Current status

Implementation of the local IfT

The local IfT was implemented in our health information system on June 2006 using a set of Terminology Services. The first area was the inpatient structured discharge summary input. Subsets for Diagnosis and Procedures are used in a user interface that allows text input and search of related terms. The user interface uses the rules for dealing with invalid, ambiguous and refinement rules.

During the initial 6-month deployment, the user proposal of terms is enforced and free-text input is an alternative available to the users. These inputs are redirected as new terms proposals.

The IfT administration software provides tools to support all the design objectives described before.

Coverage

Coverage is estimated by subset, the incidence of “new terms” proposal issued by users. We measured this during a week after 3 month of implementation; using the Diagnosis input interface new terms were proposed in 12% of the interactions by users, meaning than in 86% of the cases a good option was selected from the IfT. Using the Procedures input interface, proposal rate was 30% of the cases.

New terms proposals were evaluated by the terminology team, and accepted as new terms in the IfT in less than 40% of the cases for Diagnosis and 80% for Procedures. Reasons for rejections were the proposal of invalid terms or already existing terms.

Discussion

Our model performance was very satisfactory for our extensibility needs and the creation of local subsets. The duplication of data models with SNOMED CT adds complexity to data access, but once the basic mechanisms were created and tuned they did not impact on system performance or software development speed.

Our aim is to gradually enforce the user to only use terms included in the IfT, allowing suggestions but in a different process outside the data entry workflow. From the results is clear that the Diagnosis domain has achieved a much more mature status than Procedures. New improvements in search algorithms and contents are still needed until final implementation of a more restrictive interface.

The desired level of user expression freedom we wanted to support included describing laterality in findings and procedures. This lead to a combinatory "explosion" of thousands of new concepts to model outside SNOMED CT and an important increase in manual work by the terminology team. This is the cause of the 67% of post-coordination, which puts us on the need of a very strict quality assurance process over those more than 16,000 post-coordinated concepts. This will be the focus of a future paper.

Additionally, body parts are not always represented as lateralized concepts in SNOMED CT, for example different codes exist in SNOMED CT (1-2006) for left and right hand or breast, but not for great toe. We used a non-official "Laterality" attribute applied to the finding or procedure instead of modeling new concepts for body parts in SNOMED CT. In order to provide interoperability this can be easily solved at the moment of sharing data.

The procedures subset proved to be a much more complex domain than findings (diagnosis). Surgeons in our setting used a complex language, often combining different procedures in the same sentence, or adding details on the intention or result of the procedure. The rate of proposals of new terms by the user remains high in this domain; perhaps a structured, compositional approach to the user interface would be useful.

Analysis of user proposals was very important for introducing changes in the search algorithms, as in many cases the term already existed and was not found by the user.

In this experience we implemented post-coordination to SNOMED CT by expert authors of the Terminology Team. We are testing end-user interfaces to post-coordinate new concepts in specific settings, like the "Family History" subset. In this way, any user can create a new term in the IfT expressing that is a "Family History" of a given existing local concept, and selecting in which family member occurs. The proposed model is compared to existing "Family History" concepts. In case it has not previously appeared, a new concept is created, as well as the automatic creation of relationships to the corresponding standard SNOMED CT concepts.

A test of interoperability sharing post-coordinated data with other institutions with a similar SNOMED CT-based IfT is being tested and will be the subject of a future paper.

We are also testing a model for dynamic subset definitions, a set of rules that allows the definition of a subset in terms of their relationships with SNOMED CT concepts. In this way we can define in the local IfT a "Diabetes Subset" with all the terms related to the concept "Diabetes Mellitus" in SNOMED CT or any of its subtypes. Dynamic subsets are updated nightly in order to include recently added local concepts, and are the way to select data from the data repository with a clinical focus, including a wide range of terms, from the standard concept "Diabetes Mellitus Type II" to the very specific, post coordinated, local concept "Diabetes Mellitus Type II with mild diabetic neuropathy".

Future implementation of the system in the outpatient ambulatory setting and general restriction of free text input will prove how much work is required to maintain good domain coverage of the local IfT.

References

[1] Rosenbloom, S.T., et al., Interface terminologies: facilitating direct entry of clinical data into electronic health record systems. J Am Med Inform Assoc, 2006. 13(3): 277-88.

[2] Snomed, SNOMED Clinical Terms User Guide. 2005: College of American Pathologists.

[3] Elkin, P.L., et al., Evaluation of the content coverage of SNOMED CT: ability of SNOMED clinical terms to represent clinical problem lists. Mayo Clin Proc 2006; 81(6): 741-8.

[4] Elevitch, F.R., SNOMED CT: electronic health record enhances anesthesia patient safety. Aana J, 2005. 73(5): p. 361-6.

[5] Penz, J.F., et al., Evaluation of SNOMED coverage of Veterans Health Administration terms. Medinfo, 2004. 11(Pt 1): 540-4.

[6] Wasserman, H. and J. Wang, An applied evaluation of SNOMED CT as a clinical vocabulary for the computerized diagnosis and problem list. AMIA Annu Symp Proc, 2003: p. 699-703.

[7] Dykes, P.C., L.M. Currie, and J.J. Cimino, Adequacy of evolving national standardized terminologies for interdisciplinary coded concepts in an automated clinical pathway. J Biomed Inform, 2003. 36(4-5): p. 313-25.

[8] McClay, J.C. and J. Campbell, Improved coding of the primary reason for visit to the emergency department using SNOMED. Proc AMIA Symp, 2002: p. 499-503.

[9] Cimino, J.J., From data to knowledge through concept-oriented terminologies: experience with the Medical Entities Dictionary. J Am Med Inform Assoc, 2000. 7(3): p. 288-97.

[10] Lincoln, M.J., et al., U.S. Department of Veterans Affairs Enterprise Reference Terminology strategic overview. Medinfo, 2004. 11(Pt 1): p. 391-5.

[11] Dolin, R.H., et al., Kaiser Permanente's Convergent Medical Terminology. Medinfo, 2004. 11(Pt 1): p. 346-50.

[12] Campbell, J.R., Semantic features of an enterprise interface terminology for SNOMED RT. Medinfo, 2001. 10(Pt 1): p. 82-5.

[13] Britt, H., A new coding tool for computerised clinical systems in primary care--ICPC plus. Aust Fam Physician, 1997. 26 Suppl 2: p. S79-82.

[14] Chute, C.G., et al., Desiderata for a clinical terminology server. Proc AMIA Symp, 1999: p. 42-6.

[15] Rector, A.L., et al., A Terminology Server for medical language and medical information systems. Methods Inf Med, 1995. 34(1-2): p. 147-57.

[16] Snomed, SNOMED Clinical Terms Technical Reference Guide. 2005: College of American Pathologists.

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

A Feasibility Study on Clinical Templates for the National Health Service in Scotland

Derek Hoy[a], Nicholas R. Hardiker[b], Ian T. McNicoll[c], Phil Westwell[d]

[a] *Nursing, Midwifery and Community Health Research Centre, Glasgow Caledonian University, Glasgow, Scotland*
[b] *Salford Centre for Nursing, Midwifery & Collaborative Research, University of Salford, Salford, England*
[c] *McNicoll Medical Informatics, Glasgow, Scotland*
[d] *NHS Orkney*

Abstract

There is growing interest in the development of standards which structure information round discrete clinical concepts, in a way that supports system development and interoperability. Most notable are the openEHR Archetypes and Templates, and HL7 Templates.

A project is described which explored the potential for these to engage and support clinical practitioners in developing clinical information standards in an open and accessible way.

The project followed three phases: professional development, including networking for content development; library implementation, including modelling and processes for managing content and relating to information standards; and implementation in working systems in a manner that is professionally acceptable.

The project findings are encouraging although there remain some important issues to be explored in further work. The topic has now emerged as an important area of standards development, and a useful focus for international cooperation.

Keywords:

archetypes, clinical templates, nursing informatics

Introduction

At the heart of systems development is formalisation [1]: the process by which we describe objects in the world and actions that change them in ways that allow us to develop systems to help us achieve our goals. As our systems become more complex, interoperability becomes a key requirement and is dependant on shared formalisations.

This is the main driver behind current standardisation efforts: the goal of consistency and predictability across our data and our system processes, with the added goal of saving effort through re-use, and more controversially, raising quality.

Background to relevant standards development in NHS Scotland

During the 1990s, in the UK there was a large investment in terminology as an essential clinical information standard. There may be many reasons why this did not give the anticipated return on investment, but one explanation is that standardised terminology systems do not give enough support to developers and users who do not want to interact with terminology browsers, but with data items structured in ways that fit well with the context of the user at a particular point in time. [2]

The NHS Clinical Terms Project subsequently became part of SNOMED-CT, and NHS Scotland has adopted this as a national standard for clinical terminology. Other standards work has included the National Data Dictionary, and the National Clinical Datasets Development Programme. [3] However, none of this work meets the need identified above: context-specific domain models as the basis for standard components for building clinical information systems.

In the absence of standards in this area, there has been widespread ad hoc development within end-user applications such as GP systems, and, for example, in the national Scottish SCI XML messaging standards used for structured clinical communication. [4] Although the terminology concepts used are standardised, the contextual 'wrappers' around the terms are idiosyncratic.

A recent NHS Scotland project, the Generic Clinical System (GCS), has anticipated the issue of standard domain models applied to the user interface, and has done design work for a National Forms Library.

Archetypes and templates as international standards

The issues outlined above have motivated standards work in both openEHR and HL7.

An openEHR archetype is "*a computable expression of a domain content model in the form of structured constraint statements, based on some reference model.*" [5]

Archetypes are seen as a means of defining clinical knowledge in an explicit way, separating it out from the system

Selected for best paper award.

software that uses it. This has dual benefits of enhancing clinical ownership and making system development and maintenance easier.

An openEHR template is "*a directly, locally usable definition which composes archetypes into a larger structure logically corresponding to a screen form*." [5]

Templates have an important role in grouping and refining archetypes for specific local applications.

A HL7 Template is "*... an expression of a set of constraints on the RIM which is used to apply additional constraints to a portion of an instance of data which is expressed in terms of some other Static Model. Templates are used to further define and refine these existing models within a narrower and more focused scope*." [6]

The most obvious difference between the two approaches is that HL7 starts with a highly abstracted pattern, the Reference Information Model (RIM) that is then constrained by templates that define the local content of the model. In contrast, openEHR uses a small, generic and granular information model. Archetypes specify instances based on and built up from objects in the model. It is argued that this latter approach is closer to development methods such as object-oriented techniques where generic objects are specialised until they are fit for purpose.

HL7 was primarily intended to support the modelling of messaging between clinical systems whereas openEHR's design focus is on the clinical record itself, but both models intrude on the other's natural territory. The HL7 Clinical Document Architecture (CDA) [7] is designed to model elements of the clinical record, whilst the openEHR EHR Extract provides the basis for constructing clinical messages.

There have been interesting 'frank and open' exchanges on email discussion lists as some proponents point out differences, while those interested in solutions search for the similarities. [8, 9]

The feasibility study

As interest in this topic has grown in recent years, a project was commissioned by the Scottish Executive Health Department, funded by the Primary Care Division, and sponsored by the Community Nursing Network and the Chief Nurse for Scotland. The project started in November 2005, running until May 2007. The project was delivered by Glasgow Caledonian University and titled *'a National Library of Clinical Templates for Community Nursing in Scotland: a Feasibility Study'*.

Study aims

The project aimed to describe the options for developing and implementing a National Library of electronic Clinical Templates for Nursing in the Community in Scotland and evaluate the benefits to clinical care and secondary information users.

The objectives were to:

- identify the extent of commonality in the structure and content of community nursing records in Scotland;
- prototype and test a method of collecting record content, evidence-based, and other expert sources to develop and maintain templates;
- evaluate options for template types;
- prototype and evaluate an option for publication of templates, including information on sources and levels of authority of content;
- examine the use of template-driven data entry for community nurses to identify models of integration with clinical practice; and
- evaluate the implications of these models to produce implementation guidance.

The desired outcomes were to have:

- a clinically-owned reliable methodology for getting nationally standardised templates that support care delivery processes into electronic records and updating them; and
- a sustainable mechanism, the national library, for maintaining the templates.

This emphasis on clinical involvement was key in shaping the approach taken by the project.

Method

The project comprised three phases: *professional development, library management,* and *implementation*.

The *professional development* phase explored options for supporting clinical involvement in the process of development and maintenance of national standard clinical information tools.

Existing national groups of Tissue Viability Nurses, Continence Advisors and a Health and Well-being Network (concerning the physical health of people with serious mental health problems) were approached and cooperated over the development of national standard templates. In addition some standardised tools have been modelled, and content from the Dutch Care Information Model [10] included. A third source is template content from existing community information systems, included as example material.

The project has developed and tested on-line collaboration for working groups, using an approach based on the Open Source movement [11] at a public web site. [12] Working groups have been open for anyone to join. Joining means that members are notified by email of news or discussion items posted in the group. Anyone can visit the group page and view resources or templates, and comment on the work. Comments can be entered against individual template items to make the process easier.

Changes to templates are controlled by a group of *core developers*, and membership is restricted to those approved by the group itself.

The *library management* phase explored the development outputs to produce clinical domain models, candidate templates/archetypes, and prototype tools and architectures for maintenance and electronic publishing.

One use case explored collaboration and re-use. The Health and Wellbeing working group wanted to work on a general physical health screening tool. Their examples included content that would be commonly used in other primary care settings, including sections on smoking and alcohol use. The project invited experts on these topics to set up their own working groups, which would produce the component templates that the Health and Wellbeing group would re-use.

Owing to uncertainty over standards for archetypes, the project used simple XML mark-up, with a schema that supported metadata, individual items, and groups, which could be nested to an arbitrary level.

The *implementation* phase included evaluation of user experience with form-based systems at existing sites, for 'professional' acceptability. It also considered issues around importing schema from the library and issues of localised form development and updating.

A survey of expert practitioners is to be completed between January and May 2007, and results will be available for presentation at Medinfo 2007.

Emerging model

It seems that the more our user groups were able to engage in developing tools that were close to what they would use in their work, the easier they found their task. Very few users expressed any interest in the underlying technical detail of archetype definition and management.

The project experience therefore suggests that there is a line of demarcation between the clinical domain with forms and templates, and the technical domain of archetypes and other information standards that underpin them.

The project has therefore proposed that template development be continued with strong clinical ownership, while archetypes (or whatever approach is taken to defining standardised components) are best developed and managed by NHS national information standards bodies. This should also help with implementation of SNOMED-CT in a consistent way.

Discussion

By focussing on processes and methods to support clinical ownership and development of templates, these benefits are anticipated:

- immediate needs of national and local systems development for standardized forms will be met;
- a larger corpus of clinically acceptable templates will be available for 'technical' standards development;
- professional issues around clinical standards can be better addressed and awareness raised through clinical communities;
- re-use of templates will be increased if local developers know of a national source; and
- once the underlying standards work (archetypes etc) is sufficiently mature, there will already be a means for this to be used as the basis for further template development.

Given this emerging model, the openEHR model was a good fit, distinguishing as it does between templates and archetypes. However, given the developmental nature of the standards at the time, it was too early to select one over the other. The project therefore made an effort to include content developed using HL7 Templates, from the Netherlands.

The project was limited in its exploration of issues round *local customisation*. Attempts to promote reuse in software engineering, notably in object oriented programming, have resulted in a raft of techniques to enable reusable objects to behave in a slightly modified way. [13] Templates will require the application of similar techniques.

While national standard templates can carry identifiers when implanted on local systems, there are many 'soft' organisational issues around change control. Should local changes be fed back, as requests for change? A refactoring approach to local changes might be the best approach, i.e. an approach that contains the mechanisms for local changes to be 'pushed up' to achieve more national visibility if it is agreed that the local change should be used by others.

There is clearly a requirement for national knowledge of 'who is using what, and which version' and this could be done either by a nominated change control group, or collaboratively using a similar approach to the one used in this project.

The NHS already has a content management system, the technology that underpins the National Data Dictionary, and this would be suitable for containing templates to be used nationally.

The project did not investigate the effects on clinician/patient interaction due to structured data entry becoming part of the encounter. This is a major issue in clinical domains like nursing, where this will be new to many practitioners. However, we do have a large body of experience to draw on from General Practice medicine, where it has been well-researched [14, 15, 16] and guidance is available. [17] General practitioner training now specifically addresses the issues raised but the integration of decision support into the clinical setting remains problematic due to the competing cognitive demands on the practitioner. [18]

Mention was made earlier of the over-expectation in the 1990s that clinical terminology systems would 'enable' clinical system improvements, an expectation that was largely unrealised. It is important not to 'over-sell' the potential benefits of clinical templates. They do not remove the requirement for careful requirements analysis, and imaginative and sensitive system design. There is a doomsday scenario where a national *library* with a large number of forms would result in *systems* with large num-

bers of forms, which only increase the burden of data entry on clinical practitioners.

Successful implementation of a large national library would depend on template management. In common with openEHR archetypes, the project found that while templates may be simple in conception, we should not underestimate their complexity in representation. As templates are essentially combinatorial in nature it is important that template fragments can easily be reused across templates. Thus effective indexing becomes key, i.e. the ability easily to locate relevant templates or template fragments within a large repository.

A variety of approaches to representation and indexing were trialled in the pilot project including: an openEHR approach using the Ocean Informatics Archetype Editor [19]; and full representation in the Web Ontology Language (OWL) [20] using the Protégé ontology modelling environment. [21]

Each approach brought its own merits, for example the ability within OWL to specify and enforce the basic form of templates. However, no one approach emerged as a standard. Indeed, much simpler XML mark-up and metadata (as described previously) met the needs of this project. This represents a shift in emphasis but is consistent with the Open Source movement in providing relatively simple solutions to global problems (rather than vice versa). However, it is not yet known how such a simple approach might scale up.

Conclusion

The project has attracted strong interest from key stakeholders:

- *clinical groups* who see national templates as a means of supporting best practice;
- *clinical informatics staff*, who see the benefits of reuse, rapid development, and interoperability; and
- *national information standards bodies*, who require clinically acceptable standardized content, consistency for national information requirements, and a means of supporting implementation of SNOMED-CT.

Given the encouraging findings of the project, next steps should increase and widen clinical engagement and template development, most likely by involving Allied Health Professionals and Primary Care teams.

Finally, contact has been established with international colleagues working in this area, in particular the Netherlands [10], Australia [22] and developers in the USA, and there is agreement between interested parties to cooperate.

Acknowledgments

The authors acknowledge the support of the Scottish Executive Health Department Primary Care Division, the Community Nursing Network, NHS Scotland Information Services Division, and Dr William Goossen of the Dutch Care Information Model Project.

References

[1] Blois MS. Information and Medicine: The Nature of Medical Descriptions. Berkeley, California: University of California Press, 1984.

[2] Rector AL. Clinical Terminology: Why is it so hard? Methods of Information in Medicine 1999;38(4):239-252

[3] National Clinical Dataset Development Programme [homepage on the internet]. Edinburgh: ISD. Available at http://www.clinicaldatasets.scot.nhs.uk. Accessed: 2006 Dec 4.

[4] Scottish Care Information (SCI). [SCI Gateway XML Schemas]. Edinburgh; NHS National Services Scotland. Available at http://www.sci.scot.nhs.uk/products/gateway/gate_down_xml.htm. Accessed: 2006 Dec 4.

[5] Beale T, Heard S, editors. Archetype Definitions and Prinicples. London: The openEHR foundation; revision 0.6. Available at http://svn.openehr.org/specification/TRUNK/publishing/architecture/am/archetype_principles.pdf. Accessed: 2006 Dec 4.

[6] HL7. [Templates Specification]. Health Level Seven, Inc. Version 3 Standard, 2006. Available at: http://www.hl7.org/v3ballot/html/infrastructure/templates/templates.htm#Definition. Accessed: 2006 Dec 4.

[7] Dolin RH, Alschuler L, Boyer S, Beebe C, Behlen FM, Biron PV, et al. HL7 Clinical Document Architecture, Release 2. J Am Med Inform Assoc. 2005 October 12, 2005:M1888.

[8] Beale T. Unified EHR standards--is convergence possible? Stud Health Technol Inform. 2002;87:78-97.

[9] Blobel B. Advanced and secure architectural EHR approaches. Int J Med Inform. 2006 Mar-Apr;75(3-4):185-90.

[10] Zorg Informatie Model [homepage on the Internet]. Amsterdam: Portavita BV. Available at: http://www.zorginformatiemodel.nl. Accessed: 2006 Dec 4.

[11] Open Source Initiative (OSI) [homepage on the Internet]. Open Source Initiative. Available at: http://www.opensource.org. Accessed: 2006 Dec 4.

[12] ClinicalTemplates.org [homepage on the Internet]. Glasgow: Glasgow Caledonian University. Available at: http://www.clintemplate.org. Accessed: 2006 Dec 4.

[13] Fowler M. Analysis Patterns: Reusable Object Models. London: Addison-Wesley; 1997.

[14] Theadom A, de Lusignan S, Wilson E, Chan T. Using three-channel video to evaluate the impact of the use of the computer on the patient-centredness of the general practice consultation. Inform Prim Care. 2003;11(3):149-56.

[15] de Lusignan S, Wells SE, Hague NJ, Thiru K. Managers see the problems associated with coding clinical data as a technical issue whilst clinicians also see cultural barriers. Methods Inf Med. 2003;42(4):416-22.

[16] Booth N, Robinson P, Kohannejad J. Identification of high-quality consultation practice in primary care: the effects of computer use on doctor-patient rapport. Inform Prim Care. 2004;12(2):75-83.

[17] Liaw ST, Marty JJ. Learning to consult with computers. Med Educ. 2001 Jul;35(7):645-51.

[18] Gabbay J, le May A. Evidence based guidelines or collectively constructed "mindlines?" Ethnographic study of knowledge management in primary care. BMJ. 2004;329(7473):1013.

[19] Ocean Informatics. [Ocean Archetype Editor] Adelaide: Ocean Informatics. Available at: http://www.oceaninformatics.biz/archetype_editor/ArchetypeEditor_download.html. Accessed: 2006 Dec 4.
[20] W3C. [OWL Web Ontology Language Overview] World Wide Web Consortium. Available at: http://www.w3.org/TR/owl-features. Accessed: 2006 Dec 4.
[21] Protégé [What is Protégé?] Stanford Medical Informatics. Available at: http://protege.stanford.edu/overview. Accessed: 2006 Dec 4.
[22] Extensia Solutions [Health Records for Nursing and Midwifery]. Extensia Solutions. Available at: https://portal.dstc.edu.au/nursing. Accessed: 2006 Dec 4.

Address for correspondence

Derek Hoy,
68 Brunstane Road, Edinburgh EH15 2QR, UK.
d.hoy@gcal.ac.uk

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Keeping Up with Changing Source System Terms in a Local Health Information Infrastructure: Running to Stand Still

Daniel J. Vreeman, PT, DPT[a]

[a] *Regenstrief Institute, Inc and Indiana University, United States*

Abstract

Keeping up with changes in source system terms in a local health information infrastructure requires substantial effort. I developed a program to assist us that returns candidate mappings based on string similarities between newly encountered source test names, existing source test names, and our master dictionary term names. I evaluated this program's performance in identifying correct mappings through a retrospective study of term mappings to our master dictionary from four radiology systems. For source terms created after the initial system integration, the semi-automated mapping program identified correct mappings for 76.3% of terms from all systems. Overall, the program correctly identified mappings for 45.6% of all terms by exact string match to an existing term. The program identified correct mappings for 36.9% of the terms without an exact string match by string comparison to existing source terms, and for 54.4% of the remaining unmapped terms by string comparison directly to master dictionary terms. Because managing vocabulary mappings is resource-intensive, accurate automated tools can help reduce the effort required for ongoing health information exchange among disparate systems.

Keywords:

terminology; medical record systems, computerized; interinstitutional relations

Introduction

Interoperable health information exchange is hindered by the myriad internal, idiosyncratic conventions for identifying identical data in separate electronic systems. A comprehensive health information exchange must coalesce all of the various sources that produce health data in order to provide clinicians with information when and where they need it. The Indiana Network for Patient Care (INPC) [1] is an early example of a local health information infrastructure (LHII) that has been operating in central Indiana for more than eight years.

The INPC includes data from all five major hospital systems (fifteen different hospitals and more than a hundred clinics), the state and county public health departments, Indiana Medicaid, and RxHub. The federated INPC repository now contains more than a billion clinical observations, including laboratory results, text reports, radiology images, and EKG tracings. As a neutral, third party convener and data "Switzerland", the Regenstrief Institute receives all of the clinical message streams from participating systems. Last year approximately one hundred source systems sent HL7 clinical result messages to us in this collaboration. We accomplish the task of integrating data elements from all of these contributing systems by mapping the internal codes from the source systems to a common master dictionary that is based on vocabulary standards such as LOINC®.[2]

Mapping the local observation codes from all of these source systems requires substantial manual effort, domain expertise, and may be inconsistent because of the variability inherent in human review. Our centralized approach is effective at consolidating the expertise and tools needed for the task. Although the mapping effort for a given system is typically largest during the initial system integration, each source system continues to evolve after that initial period. INPC participants retain control over their institution's naming conventions and local term dictionaries, adding, deleting, and changing terms to meet their needs. Ideally, everyone would follow good terminology principles to allow "graceful evolution" [3] and would give the terminologists who manage the common dictionary advanced notice of upcoming changes. In practice, we have found the work of surveillance for and managing these source system changes to be a challenging aspect of operating an LHII.

To help address these challenges, we develop automated tools to assist us wherever possible. A recent focus in the INPC has been on integrating reports from six radiology centers throughout the network. By integrating these systems and distributing their content throughout the network, we hope to combat the common problem of radiology reports being unavailable when and where they are needed.[4] We previously developed and evaluated [5,6] programs to help map local radiology terms to LOINC®. While these tools are valuable in the initial mapping period, we have observed that many of the term changes occurring after this initial phase are the result of inventing new terms for the same test that are intended to distinguish among the various facilities performing that test. Because these new terms are often variants of an existing test term, their string names are often the same as (or very similar to) a term we have already mapped, but their test codes are different.

Approximate string comparators compute a measure of similarity between two strings. Although approximate string comparators have long been used in record linkage,[7-10] there is a paucity of literature about their performance in term mapping. I previously conducted a short term pilot study [11] of one institution's codes, which suggested that approximate string comparators could be valuable in managing these maintenance mappings. Based on these initial findings, I developed a revised program to help semi-automatically map new observation codes from our message processor exception logs to the INPC's master dictionary. Here I present a retrospective analysis of its performance in maintenance mapping test terms from four radiology centers in the INPC.

Materials and methods

Data sources

To evaluate the performance of the mapping program, I extracted local radiology terms from four systems participating in the INPC. The goal of my analysis was to simulate performance with maintenance mapping using the historical data contained in the INPC master dictionary. For each institution, I extracted local terms that were created after the initial system mapping. System A was initially mapped in March 2001, system B in October 2003, and system C and D were both mapped in December 2005.

In practice, as terms from each system are mapped to the master dictionary, we encounter some tests for which no suitable master dictionary term exists and therefore a new master dictionary term must be created. To simulate this retrospectively, I excluded local terms that were mapped to a master dictionary term created after the record of the local term was created. This process excluded 561 terms from system A, 318 terms from system B, and no terms from either system C or D. The extract thus included a total of 3,243 terms, of which 1,100 were from system A, 1,552 terms were from system B, 243 terms were from system C, and 348 terms were from system D. Because all of these terms have already been mapped into the master dictionary, the existing mappings served as the reference standard.

Maintenance mapping process

I developed a maintenance mapping program in Perl that is designed to semi-automate the process of finding master dictionary matches for unmapped local codes. In practice, these codes are caught by our HL7 message exception processor because they are not present in our dictionary mapping tables, and then stored in an exception log table. For this study, I modified the program to operate on the file of extracted terms rather than the log table. Because of the common feature of radiology systems inventing multiple codes for one test to distinguish among facilities, I designed the program to leverage the existing term mappings assigned during the initial system integration. The program identifies potential master dictionary matches in three phases.

Phase 1 - Exact string matches

For radiology tests from the same institution, we assume that if the string names are identical, there is a very low likelihood that the tests are actually different. Thus, we feel relatively confident in automatically assigning a master dictionary mapping based on identical strings. Therefore, the mapping program first looks for existing local terms from the same system that are mapped to the master dictionary and whose name ***exactly matches*** the string name of the unmapped code.

Phases 2 and 3 - Approximate string comparators

In the second phase, the program identifies candidate matches for all of the terms without an exact string match to an existing local term. It does this by calculating approximate string comparator scores between the new term name and existing local term names that have already been mapped to the master dictionary. The comparators we employed all generate a score between 0 and 1, with 0 being little similarity and 1 being an exact match between the strings. Because different string comparators exhibit varying sensitivity and specificity [7] related to the kinds of mismatches between strings, the program returns candidates based on the root mean square (RMS) of three different comparators.

The ***longest common substring*** algorithm generates its score by iteratively locating and deleting the longest common substring between the two strings.[12] The ***Levenshtein edit distance*** [13] between two strings is given by the minimum number of insertions, deletions, or single character substitutions needed to transform one string into the other. In this implementation, the program transforms the edit distance metric into the 0 to 1 range using a previously described method.[7] The third string comparator used is the ***Ukkonen similarity function*** [14], which is similar to the Levenshtein edit distance, but includes the operation of allowing arbitrary length substitutions. The program implements the Ukkonen function with the freely available Perl module String::Similarity (http://cpan.org) written by Marc Lehmann.

The mapping program generates its list of candidates from existing local terms with an RMS similarity score of 0.75 or greater. Candidates are ranked by descending RMS score, and the top three candidates are returned for human review.

In the third and final phase of the mapping process, the program returns candidate mappings directly to master dictionary terms for local codes that could not be mapped based on exact string matches or string similarity scores with existing local terms. Because of differences in naming conventions, master dictionary terms are much more likely to be dissimilar to local term names than are terms from the same system. Therefore, the candidate mappings in this phase are based on the less stringent RMS similarity score of 0.65 or greater. The program returns the top three candidates for human review. Figure 1 shows a Microsoft Access form for reviewing candidate maps and their similarity scores.

Figure 1- Form for reviewing candidates and their string comparator scores

Table 1 – Mapping program identification of correct matches by phase and radiology system

	Correct Matches by System				
Mapping Phase	**A** n=1100	**B** n=1552	**C** n=243	**D** n=348	**Overall** n=3243
Exact string match to an existing local term % (n)	12.5 (137)	68 (1054)	0 (0)	82.5 (287)	45.6 (1478)
Match to a top 3 candidate local term by string comparator score % (n)	14.6 (141)	76.8 (380)	41.2 (100)	49.2 (30)	36.9 (651)
Match to a top 3 candidate master dictionary term by string comparator score % (n)	34.3 (282)	18.5 (22)	23.1 (33)	25.8 (8)	54.4 (345)
Total of all phases % (n)	50.9 (560)	93.8 (1456)	54.7 (133)	93.4 (325)	76.3 (2474)

Measures

To assess the performance of the mapping program, I recorded the number of correct matches that could be identified in each sequential phase: exact match to an existing local term, match to a local term returned by the string comparators, and match to a master dictionary term by the string comparators. Correct matches were ascertained by comparing the program's results with the reference standard.

Results

Overall, the three steps of the mapping program identified correct matches for 76.3% (2474/3243) of terms from all systems. The mapping performance at each step in the process for the four radiology systems is given in Table 1. Correct matches were identified by exact string match to an existing mapped local term for 45.6% (1478/3243) of the terms. These matches represent the subset of terms that we felt could be automatically assigned a master dictionary mapping.

Of the terms that did not have an exact match to an existing mapped local term, the program returned at least one candidate mapped local term for 47.5% (839/1765) based on RMS string comparator scores exceeding the threshold. A local term with the correct master dictionary mapping was present in the top three ranking for 77.6% (651/839) of the terms with candidates. The top ranked existing local term had the correct master dictionary mapping in 84.0% (547/651) of the terms for which the correct mapping was present in the list of candidates. In summary, the string comparator-generated candidate list contained the local term with the correct mapping for 36.9% (651/1765) of the terms that did not have an exact match to an existing local term.

Of the terms without an exact match or a string comparator candidate match to an existing mapped local term, the mapping program returned at least one candidate master dictionary term for 56.9% (634/1115). The correct master dictionary term was present in the top three ranking for 54.4% (345/634) of the terms with candidates. The top ranked master dictionary term was the correct term in 81.7% (282/345) of the terms for which the correct master dictionary term was present in the list of candidates. In summary, the string comparator-generated list of master dictionary terms contained the correct mapping for 30.9% (345/1115) of the terms that did not have a match from the first two phases.

Discussion

The semi-automated process of the mapping program identified the correct master dictionary term mapping for two-thirds of the terms in this extract from four radiology

systems. Nearly half of the terms encountered since the initial system mappings could be automatically assigned a master dictionary code because of an exact string match with an existing term from that source system. More than half of the remaining terms could be assigned a map by a human reviewer examining at most six ranked candidate terms based on approximate string similarity scores (three existing mapped local terms and three master dictionary terms). For terms that could be assigned a map in this way, the likelihood of the correct mapping coming from the top ranked code is very high (82-84%).

This study was designed to evaluate the performance of the mapping program's unique, tiered approach to identifying candidate matches. Using the first phase of exact string matches to automatically assign a map without human review could greatly reduce the manual effort of maintenance mapping. Likewise, human review of a short, high probability, ranked candidate list is more efficient than our current term-by-term searching mechanism, even with database searching tools. Anecdotally, a domain expert can review these candidates and determine if a correct mapping exists in less than a minute per term. Because the volume of maintenance mapping is relatively low and there are relatively few pair-wise comparisons, the computational costs of using approximate string comparators is not prohibitive of this approach.

The method of bootstrapping existing term mappings from the initial system integration may not be as successful in domains other than radiology, because test naming conventions differ across domains. Although automatically assigning a master dictionary mapping based on an exact string match is defensible and pragmatic, it is not without error. Comparing the mappings that could be assigned by exact string match with actual historical mappings assigned by our largely manual process identified discrepancies in 4.7% (70/1478) of the terms. These discrepancies may indicate errors in either the initial system mapping or our current manual maintenance mapping procedures. Nevertheless, it highlights how fully automating this phase of the mapping process could propagate errors over time.

The success in identifying correct master dictionary matches varied widely among the four institutions examined in this study, both in total, and by phase. For example, 82.5% of the new terms from system D could be mapped based on exact string matches with existing terms, whereas none of the new terms from system C could be mapped in this manner. This variable success highlights the non-uniformity in naming conventions and terminology practices among radiology systems. Internal variation within a system's term names as well as that system's likelihood of assigning multiple codes for the same test both influence mapping success in phases 1 and 2. Mapping success in phase 3 is linked to the similarity of local system naming conventions to the master dictionary conventions. For example, some systems tended to use very few abbreviations, e.g. "CT Upper Extremity Angiography Without then W Contrast". In comparison, abbreviations and acronyms are commonly employed in master dictionary term names, e.g. "Extremity Up CTA WWO Contr". These discrepancies hinder identification by string comparison.

The maintenance mapping program evaluated in this study used three different approximate string comparators to assess string similarity, but inclusion in the candidate rankings was based only on the RMS combined metric. It is possible that the individual comparators have meaningfully different characteristics that were masked by aggregating them into one score. I intend to evaluate the relative individual performance of these comparators in future work. Furthermore, our RMS thresholds for phases 2 and 3 were arbitrarily set rather than by comprehensive empiric analysis. Moreover, it is possible that other string comparator functions[8], phonetic compression algorithms[15,16], alone or in combination with blocking schemes[17] or probabilistic methods of obtaining an overall agreement score[18] may perform better at the task of term mapping than the current approach. Other terminology mapping approaches, such as ontology-based integration techniques [19] may also be valuable in maintenance mapping from source systems.

Conclusion

This study demonstrates the utility of a simple program in maintaining the mappings from radiology source systems in an LHII. By leveraging the mappings generated during the initial system implementation and employing approximate string comparators to identify high likelihood candidate matches for focused human review, the program can reduce the manual effort of managing these changes. Nearly half of the terms encountered in this retrospective analysis could be automatically mapped with high accuracy. Half of the remaining terms could be mapped by human review of a short list of ranked candidates. Because the program so often includes the correct match first in its rankings, domain experts can review the candidates and choose the correct mapping more quickly than our current term-by-term searching methods. Because vocabulary mapping is a resource-intensive step in managing an LHII, automated tools that help us keep up with changes in source terminologies will reduce the effort required for ongoing health information exchange among disparate systems.

Acknowledgments

The author thanks Shaun Grannis, MD for providing Perl implementations of the longest common substring and Levenshtein edit distance algorithms, and the Regenstrief Institute for supporting this work.

References

[1] McDonald CJ, Overhage JM, Barnes M, Schadow G, Blevins L, Dexter PR, Mamlin B, and the INPC Management Committee. The Indiana network for patient care: a working local health information infrastructure. Health Affairs. 2005;24(5):1214-20.

[2] McDonald CJ, Huff SM, Suico JG, Hill G, Leavelle D, Aller R, Forrey A, Mercer K, DeMoor G, Hook J, Williams W, Case J, Maloney P. LOINC, a universal standard for

identifying laboratory observations: a 5-year update. Clin Chem. 2003;49(4):624-33.
[3] Cimino JJ. Desiderata for controlled medical vocabularies in the twenty-first century. Methods Inf Medicine. 1998; 37(4-5):394-403.
[4] Smith CP, Araya-Guerra R, Bublitz C, Parnes B, Dickinson LM, Van Vorst R, Westfall JM, Pace WD. Missing clinical information during primary care visits. JAMA. 2005;293(5):565-71.
[5] Vreeman DJ, McDonald CJ. Automated mapping of local radiology system terms to LOINC. AMIA Annu Symp Proc. 2005;769-773.
[6] Vreeman DJ, McDonald CJ. A comparison of Intelligent Mapper and document similarity scores for mapping local radiology terms to LOINC. AMIA Annu Symp Proc. 2006;809-813.
[7] Grannis SJ, Overhage JM, McDonald CJ. Real world performance of approximate string comparators for use in patient matching. Medinfo. 2004;11:43-7.
[8] Porter EH, Winkler WE. Approximate string comparison and its effect on an advanced record linkage system. Washington DC, US Bureau of the Census. 1997.
[9] Winkler WE. The state of record linkage and current research problems. Technical Report RR/1999/04. Statistical Research Report Series. Washington DC, US Bureau of the Census. 1999.
[10] Navarro G, Raffinot M. Flexible pattern matching in strings. Association of Computing Machinery Computing Surveys. 2001;33:31-88.
[11] Vreeman DJ. Maintaining mappings from source systems in a local health information infrastructure. AMIA Annu Symp Proc. 2005;1131.
[12] Sidelli R, Friedman C. Validating patient names in an integrated clinical information system. In: *Symposium on Computer Applications in Medical Care.* 1991; Washington, DC. p 588-592.
[13] Levenshtein VI. Binary codes capable of correcting deletions, insertions, and reversals. Soviet Physics Doklady. 1966;10(8):707-10.
[14] Ukkonen E. Algorithms for approximate string matching. Information and Control. 1985;64:100-18.
[15] Kunth D. The art of computer programming, Volume 3/ Sorting and searching. 2nd ed: Addison-Wesley Publishing Company. 1998.
[16] Lynch B, Arends W. Selection of a surname encoding procedure for the Statistical Reporting Service record linkage system. Washington DC, US Department of Agriculture. 1977.
[17] Sauleau EA, Paumier JP, Buemi A. Medical record linkage in health information systems by approximate string matching and clustering. BMC Med Inform Decis Mak. 2005;5:32.
[18] Felligi I, Sunter A. A theory for record linkage. J Am Stat Assn. 1969;328(64):1183-1210.
[19] Lee Y, Supekar K, Geller J. Ontology integration: Experience with medical terminologies. Comput Biol Med. 2006;36(7-8):803-919.

Address for correspondence

Daniel J. Vreeman, PT, DPT
1050 Wishard Blvd. RG-5
Indianapolis, IN 46202
dvreeman@regenstrief.org

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

The Role of Local Terminologies in Electronic Health Records. The HEGP Experience

Christel Daniel-Le Bozec[a,b], Olivier Steichen[a], Thierry Dart[b], Marie-Christine Jaulent[a]

[a]INSERM, UMR_S 872, eq.20 Paris, F-75006 France; Université Paris Descartes, Paris, F-75006 France;
[b]APHP, Hôpital Européen Georges Pompidou, Département d'Informatique Hospitalière, Paris, F-75015 France.

Abstract

Despite decades of work, there is no universally accepted standard medical terminology and no generally usable terminological tools have yet emerged. The local dictionary of concepts of the Georges Pompidou European Hospital (HEGP) is a Terminological System (TS) designed to support clinical data entry. It covers 93 data entry forms and contains definitions and synonyms of more than 5000 concepts, sometimes linked to reference terminologies such as ICD-10.In this article, we evaluate to which extend SNOMED CT could fully replace or rather be mapped to the local terminology system. We first describe the local dictionary of concepts of HEGP according to some published TS characterization framework. Then we discuss the specific role that a local terminology system plays with regards to reference terminologies.

Keywords:

terminology; controlled vocabulary; medical records systems, computerized

Introduction

The need for terminological systems (TS) to support clinical statements in Electronic Health Records (EHR) has been widely recognized [1]. A terminology system can range from simple code-name-hierarchies, e.g. ICD and MeSH to knowledge-based ontologies such as FMA [2], SNOMED-CT [3] or Gene Ontology [4]. Beside this, terminology browsers, editors and servers have been developed to facilitate creation, maintenance and use of terminological systems [5]. Nowadays, no universally accepted standard medical terminology and no generally usable tools have yet emerged.

By contrast, consensus on requirements for terminological systems and servers to be useful in EHR is emerging [6]. Evaluation methodologies are already available to describe and compare TS in an objective and reproducible manner [7-9]. For Cornet et al., a TS has to be characterized as a terminology (providing a list of terms) and possibly also as one or more among: a thesaurus (alphabetically ordered terms with synonyms), a classification (arrangement of concepts using is-a relationships), a vocabulary (providing free-text or formal definitions of concepts), a nomenclature (providing composition rules) and a coding system. Cornet *et al.* have applied their evaluation framework to SNOMED CT used in combination with the CLUE browser [8].

In the Georges Pompidou European Hospital (HEGP), a local dictionary of concepts was designed to support structured data entry and to provide decision support based on the recorded clinical information. The local dictionary of concepts was populated with the concepts derived from the items of data entry forms. In the next future, we plan to use SNOMED as a reference terminology in order to ensure semantic interoperability. Our main concern is to investigate if SNOMED should fully replace or rather be mapped to the local terminology system.

We first describe the local dictionary of concepts of HEGP and the evaluation framework proposed by Cornet et al. Then we characterize the local dictionary of concepts according to this framework and compare it to SNOMED CT. We further discuss the specific role that a local terminology system may play with regard to reference terminologies.

Material and methods

Local dictionary of concepts of HEGP

The component-based hospital information system of the HEGP is provided by a consortium driven by Thales©. This company develops generic components and middleware tools for the integration of healthcare components into a multilayered architecture. A generic component stores reference terminologies used by various other components. These terminologies are: international reference terminologies (e.g. ICD-10), national reference terminologies (e.g. CCAM, the French national procedures nomenclature) and local reference terminologies (e.g. list of HEGP care units or healthcare professionals, list of pharmaceutical drugs available in the hospital, etc.).

One of the healthcare components provides EHR components including patient order entry. Clinical information can be entered as free text in Word® documents or as semi-structured documents using entry forms called "questionnaires".

A module allows users developing their own local dictionary of concepts. Another module allows users to create questionnaires. Local concepts directly derive from the items (attribute-value pairs) of questionnaires. In practice,

each question (attribute) is expressed in colloquial terms in the questionnaire and linked to one and only one concept of the local dictionary of concepts. If no matching concept is found in the dictionary, then a new concept is created by the user. For checklist questions, with a predefined list of possible values, each answer (value) is also linked to one and only one concept, preexisting in the dictionary or newly defined. It is also possible to link these predefined answers to concepts selected in one of the reference terminologies rather than to concepts of the local dictionary.

Since 2003, HEGP users derived the local dictionary of concepts from clinical items of 93 questionnaires. Among these questionnaires, 58 were created for physicians, 33 for other healthcare professionals and 2 questionnaires about past medical history and allergy were common to all healthcare professionals.

Characterization of the HEGP local dictionary of concepts

We analyzed the local dictionary of concepts of the HEGP (LDC-HEGP) according to the TS characterization framework proposed by Cornet *et al* [8]. This framework distinguishes between application-dependant features and application-independent features.

The only application-dependant feature is content coverage of the TS according to predefined purpose and scope.

Application-independent features are categorized along two axes. Along the first axis, features are categorized as being related to one of the six main types of TS: list of terms management for "basic" terminologies, indexes and synonyms management for thesauri, free-text or formal concept definitions for vocabularies, classification principles for classifications, composition rules for nomenclatures and coding process for coding systems. Along the second axis, features are categorized as being formalism-related characteristics, function-related characteristics and content-related characteristics.

This two-axial evaluation framework was applied by Cornet *et al.* to the July 2003 English version of SNOMED CT used in combination with the CLUE browser 5.5 [8]. We applied the same framework to the HEGP local dictionary of concept and its browser. Although the HEGP dictionary of concepts should mainly be seen as a thesaurus and a vocabulary, we applied the full set of criteria found in Cornet's evaluation framework in order to expand our view on the system. Content coverage of the local dictionary of concept was analyzed according to its predefined purpose and scope – i.e. enhancing structured data entry and decision support.

Results

Application-independent evaluation

Table 1 - Criteria related to "Management of a list of terms" (Terminology)

List of terms management		LDC HEGP	*SNOMED CT [5]*
Formalism	· Concepts and terms distinguished	· Yes	· *Yes*
	· Term length restriction	· Yes	· *None*
	· Character encoding	· n/a	· *UTF-8*
	· Concept obsoletion mechanism	· Yes, concept status flag and author tracked	· *Yes, concept status flag, with motivations*
Content Tool	· Convert code to text	· No	· *For concept and description IDs*
	· Lookup phrases for a string	· Yes	· *Yes*
	· Lookup phrases with wildcards	· No	· *Yes*
	· Inexact match	· No	· *No*
	· Code refinement	· No	· *No*
	· Keyword matching	· No	· *n/a*
	· Case-insensitive	· Yes	
	· Total nb of concepts	· 5426	· *352662*
	· Total nb of terms	· 7693	· *939705*
	· Covered areas	· See below	· *Cf [5]*

The local dictionary of concepts comprises 5426 concepts among which 2087 are from the questions (attributes) of the questionnaires and 3339 are from the predefined answers (values) of checklist questions (table 1). Concepts were organized according to SNOMED covered areas. Concepts distribution is the following: procedures: 44% (medical procedures (including lab and radiology): 27%, social procedures: 15%, nursing procedure: 1%, physiotherapy procedure: 1%), social context: 26%, subjective symptoms and findings: 11%, care management: 7%, diseases: 4%, healthcare providers: 2%, anatomy: 1%, other areas: 5%.

Table 1 shows that changes in name, synonyms or description of concepts are tracked but not their motivation.

Table 2 - Criteria related to "Indexing and synonyms" (Thesaurus)

Indexing and synonyms		LDC HEGP	*SNOMED CT [8]*
Formalism	• Terms indexed	• Yes	• *None*
	• Supports synonyms	• Yes	• *Yes*
	• Synonym representation	• Yes	• *Yes*
	• Multilingual representation	• No	• *Yes, language code*
	• Synonyms for fragments	• No	• *No*
	• Description obsoletion mechanism	• No	• *Yes, status flag, with motivations*
Tool	• Translation to other languages	• No	• *No*
Content	• Nb of synonyms by concepts (average)	• 0.61	• *2.66*
	• Langages	• French	• *UK&US English, German, Spanish*

As shown in table 2, each concept of the dictionary is characterized by a preferred term or expression. A term can be preferred for no more than one concept, since the system recognizes redundancy. Meaningless local identifiers are associated to concepts. The identifier assigned to a concept is not inextricably bound to a hierarchy position in the terminology, so that there is no need to change the code as we update our understanding. Some synonyms can optionally be stated for each concept. At the present time, there are only French terms and descriptions in the HEGP dictionary of concepts.

Table 3 - Criteria related to "Concept definition" (Vocabulary)

Concept definition	LDC HEGP	*SNOMED CT [8]*
• Free-text concept definition	• Yes	• *None*
• Formal concept definition	• No	• *Yes, Description Logics*
• Atomic concepts	• No	• *No*
• Explicitly defined relationships	• No	• *Yes*
• Retrieve definitions (properties of a concept)	• Yes (if available)	• *Yes*
• Subsumption testing	• No	• *No*
• Instance checking	• No	• *No*
• Detection of equivalent definition	• No	• *No*
• Query for concepts matching structural criteria	• No	• *n/a*
• Nb of vague concepts	• n/a	• *13151 (3,7%)*[1]
• Nb of ambiguous/ redundant concepts	• n/a	• *n/a*
• Relationships used	• no typed relationships in the hierarchy	• *993 attributes defined, 42 used*

As shown in table 3, each concept of the dictionary of concepts can be defined by a free text description but there is no formal representation of concepts and no explicit relationships between concepts.

Table 4 - Criteria related to "Classification" (Classification)

Classification		LDC HEGP	*SNOMED CT [8]*
Formalism	• Hierarchical relationships	• Yes	• *Yes*
	• Properties inheritance	• No	• *Yes*
	• Polyhierarchies	• Not used	• *Yes*
	• Hierarchical depth restriction	• None	• *None*
	• Hierarchical breadth restriction	• None	• *None*
	• Classification inferred according to concept definitions	• No	• *Yes, since DL based*[2]
T	• Retrieve descendants	• Yes	• *Yes*
C	• Nb of parents per concepts (average)	• One	• *1.3*

To ease the navigation within the dictionary of concepts, concepts are organized into a taxonomic hierarchy. As previously said, the first level of the hierarchy consists in twenty-two categories defined according to the SNOMED top-level classes. For the "Body structure", "Finding", "Disease" and "Procedure" categories, the second level of the hierarchy consists in a partition based on anatomy. For the "procedure" category, the third level of the hierarchy consists in a partition based on the techniques defined in French national procedures nomenclature (CCAM, e.g. computed radiography, computed tomography, etc).

Concepts related to clinical information about someone else than the patient (relatives or others (like organ donors)) are prefixed by 'NP-'standing for 'NonPatient' and categorized in a separate hierarchy. We don't use multiple inheritances for the local dictionary of concepts.

Table 5 - Criteria related "Composition rules" (Nomenclature)

Composition rules		LDC HEGP	*SNOMED CT [8]*
Formalism	• Composition	• Only pre-coordination	• *Yes*[3]
	• Composition formalism	• Not used	• *Refinability*[4]
	• Detection of equivalent definitions	• Not used	• *Yes (DL reasoners)*[5]
	• Composition change meaning or only specify more detail	• Not used	• *More detail only*
Tool	• Support in concept composition	• Not used	• *No*
	• Retrieve refinable relations	• Not used	• *Yes*
C	• Nb of refinable concepts	• n/a	• *153080 (43%)*

Pre-coordinated concepts of the local dictionary of concepts are not linked to the two or more atomic concepts they derive from. Concepts can not be post-coordinated during data entry through questionnaires.

Table 6 - Criteria related to "Codes" (Coding system)

Codes	LDC HEGP	*SNOMED CT [8]*
· Codes assigned	· Only for concepts related to values	· Yes
· Code generation mechanism	· Manual	· Number +partition id+ check digit
· Code length restriction	· n/a	· 18 positions
· Meaning of identifiers	· No	· No
· Limitation of taxonomic placements	· No	· No
· Cross coding	· No	· No
· All concepts coded	· No	· Yes
· Cross mappings	· No	· Yes[6]

Concepts related to values (checklist predefined answers) can optionally be associated either to one local code or linked to a coded concept of an external codification system. There is no complete cross mapping between the local dictionary of concepts and the reference terminologies.

Application-dependant evaluation

The domain coverage is evaluated in relation to the main purpose and scope of the HEGP dictionary of concepts: supporting structured data entry by health care providers. Due to the building process of the dictionary of concepts, its coverage for structured data entry is necessarily good. Furthermore 22,6% of the concepts are "reused", i.e. linked with questions in at least two different questionnaires. Reusing concepts make it possible to report values/answers from one questionnaire to another, even if the wordings of the questions are different. For example, if the question "Hypertension?" in one questionnaire and the question "Does the patient suffer from high blood pressure?" in another questionnaire are linked to the same concept, and if the answer was "Yes" in the first questionnaire, then it is possible to take this answer as default value for the second questionnaire. This process makes data entry easier and faster.

Another purpose of the local dictionary of concepts was to use clinical information to perform alerts for patient order entry. Due to the building process of the dictionary of concepts, its coverage is not sufficient. Only clinical information recorded through questionnaire is linked to concepts. Then, data not recorded through questionnaire (i.e. from inpatient and outpatient reports and other documents) cannot be processed by the rules engine based on medical concepts. Moreover, automatic process can only rely on clinical data entered as answers to checklist questions since data entered as answers to free text or numeric questions are not linked to concepts.

Discussion

We have characterized the LDC-HEGP according to the evaluation framework proposed by Cornet *et al.* in order to identify precisely the main differences between such a local terminology and reference terminologies like SNOMED CT that includes 120 times more concepts than the LDC. The results show that the reference terminology provides many features that are lacking in the LDC.

Evaluated as a vocabulary, LDC-HEGP provides only optional free-text description of the concepts while SNOMED CT provides formal definition of concepts based on Description Logic. In SNOMED CT, atomic concepts are distinguished and explicit relationships are defined.

With regards to the criteria related to classification, we haven't define strict "is-a" relationships within the hierarchy of concepts and we don't make use of multiple inheritance. The hierarchical organization of LDC-HEGP was mainly meant to ease the navigation process within the dictionary and not to perform aggregation of data.

Criteria related to "Composition rules" underline that concepts of the LDC cannot be further post-coordinated. This ensures that coordinated concepts have a valid medical meaning. However, searching in long lists of fully defined pre-coordinated concepts is tedious as well as the creation of numerous pre-coordinated concepts (for refinability, temporal duration, uncertainty or context information, etc). For example, for clinical information about someone else than the patient, we chose to duplicate and prefix the corresponding concepts. In SNOMED CT, concepts may be defined by composition and/or modified by a qualifier. Refinability can be optional or mandatory. Description Logic provides semantic normalization insuring automatic comparability of composite expression (e.g. of "Stenosis of coronary artery" and "Coronary artery stenosis" or of "Cholecystectomy" and "Excision of Gallbladder"). This could allow detection of equivalent definitions although no SNOMED browser actually has this functionality [8].

At last, even though the local dictionary of concepts could be used as a coding system, automation of Disease Related Group (DRG) coding relies on reference terminologies. Today, when DRG coding is intended predefined values of checklist questions are linked to ICD-10 or CCAM instead of being linked to a concept of the local dictionary.

The evaluation framework proposed by Cornet *et al.* increases the understanding of specific properties of any TS and allows comparison between TS according to prominent features. We noted that some criteria could be added to the evaluation framework., for example authorship and motivation as a criteria for "basic" terminologies (who did introduce the concept and on which grounds?); mapping to other TS as a quality criteria for thesauri; detection of vague or ambiguous concepts and the rate of concepts with formal and/or free-text definition as criteria for vocabularies.

We found it useful to combine in a single evaluation framework both intrinsic features (formalism related-characteristics and content-related characteristics), directly related to the terminology itself, and extrinsic features (function-related characteristics), more related to the terminology browser.

Terminology developers historically have focused on reference terminologies designed to provide sound representation of a given domain (entities and their relationships) and to support the storage, retrieval, and

classification of clinical data. Ongoing efforts aim to define the binding processes between reference terminologies and information model of EHR. But other types of terminologies may complement reference terminologies since they are not easily and directly usable by health care providers during routine clinical tasks [10]. Rector has suggested that TS should conform to conflicting requirements: being usable and flexible for humans and providing formal representation for computer programs [11].

Spackman *et al.* distinguished three different types of terminologies: interface terminologies supporting data entry, processing terminologies optimizing natural language processing and finally reference terminologies for computer storage, retrieval and analysis of clinical data [12]. Rosenbloom *et al.* reviewed existing definitions of interface terminologies, which have also been called colloquial terminologies, application terminologies or entry terminologies [13]. The LCD-HEGP may be considered as an interface terminology primarily intended to support data entry. It interfaces clinicians' own colloquial conceptualizations and more structured elements that can be shared or used by computers.

According to Rosenbloom, existing evaluation frameworks for medical terminologies, such as the one proposed by Cornet *et al.* may not be applicable to interface terminologies. He defines prominent features of interface terminologies: representation of relevant assertional medical knowledge, adequate support of synonymy, balance between pre- and post-coordination and mapping to terminologies having formal concept representation [13].

With regard to this last point, we think that a local terminology may be useful not only as an interface terminology but also as an integration locus for multiple terminologies into a thesaurus. Indeed, it is doubtful that single TS will ever support as diverse needs as clinical care, billing, and research. Therefore, terminologies mapping and integration is an unavoidable challenge.

Our perspectives are first to characterize the local dictionary of concepts focusing on criteria specific to interface terminologies and to improve this local terminology in order to efficiently "interface" SNOMED CT to support data entry for structured documentation. We plan to use automated methods for mapping the local interface terminology to large scale reference terminologies.

Acknowledgments

This work was supported by a grant from the Semantic Mining (NOE).

References

[1] Hammond WE. Call for a standard clinical vocabulary. J Am Med Inform Assoc. 1997 May-Jun;4(3):254-5.

[2] Rosse C, Mejino JL Jr. A reference ontology for biomedical informatics: the Foundational Model of Anatomy. J Biomed Inform. 2003 Dec;36(6):478-500.

[3] Spackman K. SNOMED RT and SNOMED CT. Promise of an international clinical terminology. MD computng, 17(6):29, 2000.

[4] Gene Ontology Consortium.The Gene Ontology (GO) project in 2006. Nucleic Acids Res. 2006 Jan 1;34(Database issue):D322-6..

[5] Cimino J.J. et al. Terminology Tools: State of the Art and Practical Lessons. Meth Inform Med 2001:298-307.

[6] Cimino, JJ. Desiderata for controlled medical vocabularies in the Twenty-First Century. Meth Inform Med 1998; 37:394-403.

[7] Elkin et al. Guideline and quality indicators for development, purchase and use of controlled health vocabularies. Int J Med Inform. 2002 Dec 18;68(1-3):175-86.

[8] Cornet R, de Keizer NF, Abu-Hanna A. A framework for characterizing terminological systems. Methods Inf Med. 2006;45(3):253-66.

[9] ISO/TC251 WG3. Standard specification for quality indicators for controlled health vocabularies; 2000 July. Report n° TS 17117.

[10] Ash JS, Berg M, Coiera E. Some unintended consequences of information technology in health care: the nature of patient care information system-related errors. J Am Med Inform Assoc. 2004 Mar-Apr;11(2):104-12.

[11] Rector AL. Clinical terminology: why is it so hard? Methods Inf Med. 1999 Dec;38(4-5):239-52.

[12] Spackman KA, Campbell KE, Cote RA. SNOMED RT: a reference terminology for health care.Proc AMIA Annu Fall Symp. 1997;640-4.

[13] Rosenbloom ST, Miller RA, Johnson KB. Interface terminologies: facilitating direct entry of clinical data into electronic health record systems. J Am Med Inform Assoc. 2006 May-Jun;13(3):277-88.

Address for correspondence

Christel Daniel – INSERM U729 – Faculté de Médecine Paris V– 15, rue de l'École de Médecine – 75006 Paris – France
Email: christel.daniel@spim.jussieu.fr

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Ontology-based Knowledge Base Model Construction-OntoKBCF

Xia Jing, Stephen Kay, Nicholas Hardiker, Tom Marley

SHIRE, IHSCR, University of Salford, UK

Abstract

Semantic web technologies are used in the construction of a bio-health knowledge base model, which, when coupled with an Electronic Health Record (EHR), is to be used by clinicians. Specifically, this ontology provides the basis for a domain knowledge resource that attempts to bridge biological and clinical information. The prototype is focused on a Cystic Fibrosis exemplar, and the content of the model includes: Cochrane reviews; a time-oriented description; gene therapy; and the most common cystic fibrosis gene mutations. The facts within the model range from nucleobase mutation and amino acid change to clinical phenotype. The knowledge is represented by layers from the micro level to the macro level. Here, emphasis is placed upon the details between levels (i.e., the vertical axis) and these are made available to bridge the knowledge from different levels. The description of gender, age, mutation and clinical manifestations are clues for matching points within an EHR system. OWL is the ontology representation language used and the output from Protégé-OWL is a XML-based file format, which facilitates further application and communication.

Keywords:

knowledge representation; OWL; cystic fibrosis; OntoKBCF

Introduction

The ultimate goal is to provide relevant bio-health information to doctors in a clinical environment. To move towards this goal, we construct an ontology-based knowledge base (KB) model (called OntoKBCF) to provide the required domain knowledge and then consider how to make it available to an electronic health record (EHR) system as a plug-in. Figure 1 provides a schematic overview of the layered knowledge base exploiting the metaphor of an EHR as a container drawing from clinical and bio-information.

The content produced for the model, depicted by the horizontal levels in figure 1, is merely sufficient to organize the relationships between the vertical levels for this research. This paper introduces the OntoKBCF construction and the results obtained thus far.

Semantic web technology is the major technology used in the OntoKBCF construction, recognising that this is the next evolution of the current World Wide Web. Simply put, the semantic web provides more "labels" for data so that machines can process the data more precisely and in more meaningful ways. The semantic web requires common formats and language; the former are useful for data sharing, reuse and exchange; the latter is used for expressing data meaningfully and relating it with real objects[1].

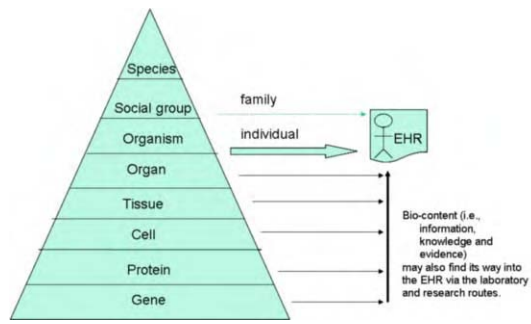


Figure 1 - The EHR as container of diverse and complex data

An ontology is an important part of the infrastructure for semantic web applications and shares similar characteristics with it. However, it is particularly useful for providing a "common understanding of a domain" and provides opportunity to improve knowledge management[2](p5). Considering these characteristics i.e., universal format (URI, uniform resource identifiers), expression capability (to represent data related with real objects), extensibility and standardization, we select the ontology-based KB to construct the applications knowledge source. Cystic Fibrosis is chosen as an exemplar.

Procedures, technology and tool

General procedures

The general procedures for OntoKBCF construction are outlined: (A): To understand doctors information needs; (B): To define OntoKBCFs border, axes and granularity; (C): To dissect knowledge resources into basic concepts; (D): To list relevant vocabularies; (E): To extract target concepts from UMLS (Unified Medical Language System)[3] or GO(Gene Ontology)[4]; (F): To organize and arrange concepts; (G): To create terms necessary for OntoKBCF ; (H): To discuss with colleagues and to iterate all the procedures until the concepts are stable. The OntoKBCF is constructed bottom-up. Most of the hierar-

chy of the basic concepts in OntoKBCF follow UMLS and GO when appropriate, and so do the semantic types in UMLS. However, we also add, adjust relationships and vocabularies if there are no appropriate choices. For example, we use "Gender_group" to connect "Population_group" and "Female"; but "Gender-group" is not an independent concept in UMLS. Likewise we create "Patient_CF" and all its subclasses and adjust the concept "CFTR gene" (cystic fibrosis transmembrane regulator) in UMLS to "Human CFTR gene". We justify these changes with respect to the OntoKBCF border, which is discussed in more detail in the "model construction" section of this paper.

Technologies and tool: OWL and Protégé-OWL

OWL (web ontology language) is used to explicitly represent vocabularies and their relationships in an ontology and this representation is directly usable by machine[5]. OWL has more semantic representation ability than XML (extensible markup language), RDF, and RDFS (resource description framework schema)[5] and OWL facilitate greater machine interoperability. OWL ontology can be encoded in different syntactic forms including RDF/XML, which is defined by XML syntax for RDF[6].

OWL Lite is an ontology language supporting a subset of the OWL language , specially designed to develop OWL tools; OWL DL (description logic) and the complete OWL language (OWL Full) support the same set of OWL language, however the former has desirable functions for reasoning system with more restrict constraints. We choose OWL Full as the ontology language in the OntoKBCF construction, as it provides the maximal RDF compatibility. OWL Full ontologies contain all the RDF content in a consistent manner, and OWL Full also assigns more meaning for certain RDF triples[7]. The OWL ontology uses the Protégé-OWL 3.2 platform and is encoded in XML. OWL statements make the OntoKBCF consistent, extensible and computer-readable.

OntoKBCF construction

OntoKBCFs border, main axes and granularity

The first thing in order to construct a KB model is to define its scope or border[8]. The research assumes that the OntoKBCF will be a plug-in to an EHR system, with the target end-users being clinicians (more specifically doctors). Initial scoping must therefore include the doctors information need.

There are several ways to obtain details of information need: e.g., survey, questionnaire, interview, observation, system log files analysis and so on. We utilized the published literature to set the initial scope. According to references [9-11], the most common clinical question is about treatment. Since we are trying to construct a knowledge base prototype to show proof of concept rather than produce a full commercial product, pragmatically we limited our search to a few reputable sources. In particular, we chose the Cochrane review topics as our starting clinical requirements.

There content of OntoKBCF comprises four parts: cystic fibrosis related Cochrane reviews [12]; time related cystic fibrosis description [13]; gene therapy of cystic fibrosis [14, 15]; the most common CFTR mutations and their characteristics [16, 17]. All the detailed content is available on request.

The two main axes developed in the OntoKBCF are time and problem orientation. With respect to Time the interest is in how cystic fibrosis may announce itself at different ages; with respect to problem-orientation the emphasis is from the Cochrane review on treatment, gene therapy and the most common CFTR mutations and their characteristics. We did not include the entire Cochrane reviews conclusions about cystic fibrosis, only some definite positive conclusions. Negative and uncertainty conclusions were excluded.

We only represent the definite knowledge facts about cystic fibrosis in the OntoKBCF-content i.e., the "what", without disease mechanism and reasons explanation i.e., "how" and "why". In the OntoKBCF the granularity of genetic level starts from nucleobase, which is the important component for the elementary unit (nucleotide) of RNA and DNA; the most of representation granularity in phenotype level starts from those concepts, such as diarrhea, nausea and coughing; other part of representation granularity in phenotype level depends on the border of the OntoKBCF, if the subclasses of the concepts are out of the border, we will not split the concepts any more. In OntoKBCF construction, the border and granularity are also the major criteria in considering an upper class concepts inclusion or exclusion.

Construction

We will explain construction of the OntoKBCF in detail by using the example knowledge fact (1) and biological concepts organization (2). The example given is from the time-oriented cystic fibrosis description. We organize the biological concepts since we have been unable to find an agreed reference hierarchy, and therefore we explicitly introduce what we include and how we have done it. Space determines we only give examples rather than provide every statement.

The following font conventions are used in this section: the class name is in **bold**, and the class defined by us is in **bold** and underline; the property we defined is in *italic* and underlined; description from "Asserted Conditions" panel in "OWL Classes" tab on the Protégé-OWL interface is in *italic*. We connect the sub class with super class by symbol "<" and super class with sub class by symbol ">". The general procedures, mentioned earlier, are given here (e.g., C, E etc).

1. ***In female adolescent cystic fibrosis patient: infertile with scanty cervical mucus, bronchiectasis***

- C) To analyze and dissect knowledge fact into basic concepts

  In female adolescent cystic fibrosis patient group, the clinical manifestations include infertile, scanty cervical mucus and bronchiectasis. Basic concepts include:

gender: female; different age groups: adolescent; diseases: cystic fibrosis and infertility; human being group: patient; quantitative concept: scanty; body substances: cervical mucus; respiratory tract diseases: bronchiectasis.

- E) To select proper terms and upper class terms used in UMLS according to their definitions

  We choose the corresponding terms from UMLS according to the concepts acquired from step (C) and their super classes. Examples include in figure 2.

Conceptual_entity	Phenomenon_or_process
Group	Pathologic_function
Population_group	Disease_or_syndrome
Female	Diseases
Age_group	Cystic_fibrosis
Adolescent_age_group	Infertility

Figure 2 - Example terms & hierarchy from UMLS

Patient - there is no proper terms in UMLS for patients in different diseases categories.

- F) To organize, create and arrange the terms with original identities and alternative vocabularies

In this case the original identity is a concept unique identifier (CUI) from UMLS, which can be used to track the concept over time and also can be used like a "primary key" in further application or communication. Alternative vocabularies are synonyms. Both of CUI and synonyms are recorded in the "Annotations" panel in the "OWL Classes" tab.

The exact terms found in UMLS and their upper class terms can be organized hierarchically from bottom up. We create the new term **"Patient_CF**" as a subclass of **"Human_being"** to represent patient diagnosed with cystic fibrosis. Because the cystic fibrosis patients have different properties, such as different age groups, different mutations and different therapies, and they are the subjects for representation as knowledge facts, we split the concept **"Patient_CF"** further. An example hierarchy for this concept is:

Patient_CF>Patient_CF_with_age_group>Adolescent_CF>

Adolescent_female_CF

We define the **"Patient_CF"** as intersection of *(1) Human_being; (2) has_diagnosis some Cystic_fibrosis*.

"has_diagnosis" is a property used to connect human being with diseases or syndrome.

"Patient_CF_with_age_group": this class is only an abstract and interim concept used for keeping its subclasses tidy.

"Adolescent_female_CF" is an intersection of *(1) Adolescent_CF; (2) occur_in_gender_group some Female*.

"occur_in_gender_group" is a sub property of *"occur"* to describe human being in male or female group.

- G) To represent the knowledge fact by combination of the terms and properties with logic symbols

The final representation of this fact is under the **"Adolescent_female_CF"** concept as a necessary condition: *has_manifestation some (Infertility and Scanty_cervix_mucus)*. There are also some inherited descriptions from upper classes. The final representation of this knowledge fact showed in figure 3.

"has_manifestation" is a property to describe human beings manifestation specifically related with diseases or syndromes.

We define **"Scanty_cervix_mucus"** as in intersection of *(1) Cervix_mucus; (2) has_quantitative_property some Scanty.*

Figure 3 - Representation of example 1

2. *Introduce the hierarchy of the bio concepts in the model*

We explain the construction idea top down. There are two major parts: **"Nucleobase_mutation**, **Translation_change"**, both are subclasses of **"Phenomenon_or_process"** for mutation related concepts; **"Amino_acid_peptide_or _protein"**, **"Nucleic_acid_nucleoside_or_nucleotide"**, both are sub classes of **"Chemical"**. We will introduce them separately.

- Hierarchy of **"Nucleobase_mutation"** and **"Translation_change"**(Figure 4, 5)

Nucleobase_mutation	Translation_change
Nucleobase_deletion	Amino_acid_deletion
Nucleobase_insertion	Amino_acid_insertion
Nucleobase_transition	Amino_acid_substitution
Nucleobase_transversion	

*Figure 4 - Hierarchy of **Nucleobase_mutation** and*

Translation_change

Nucleobase_transversion	Nucleobase_deletion
A_transversion_C	Del_A
A_transversion_T	Del_C
C_transversion_A	Del_G
C_transversion_G	Del_T
G_transversion_C	Del394_TT
G_transversion_T	Del_U
T_transversion_A	
T_transversion_G	

*Figure 5- Hierarchy of **Nucleobase_transversion** and **Nucleobase_deletion***

"Nucleobase_insertion" is similar with **"Nucleobase_deletion"**, and Ins_A, C, G, T, U are its subclasses.

There are four possible nucleobase transitions, **"A_transition_G"** , **"C_transition_T"** , **G_transition_A**, and **T_transition_C"**.

Hierarchy of **"Translation_change"** is shown in figure 6.

Hierarchy of **"Amino_acid_peptide_or_protein"** and **"Nucleic_acid_nucleoside_or_nucleotide"**

The subclasses of **"Amino_acid_peptide_or_protein"** are: **"Amino_acids"** and **"Human_CFTR_protein"**. Under **Amino_acids** we list all the three letters abbreviation names. Hierarchy of **"Nucleic_acid_nucleoside_or_nucleotide"** is shown in figure 7.

We choose **"Nucleobase"** as subclass of **"Nucleic_acid_nucleoside_or_nucleotide"** and super class of all the nitrogenous bases abbreviation name. We also classified the nucleobases into **"Purines_and_derivatives"** or **"Pyrimidines_and_derivatives"** correspondingly.

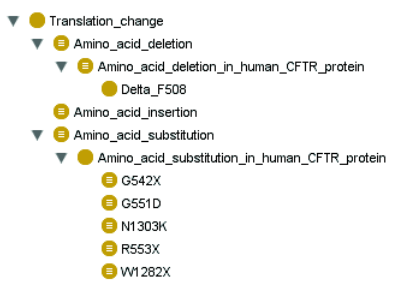


Figure 6 - Hierarchy of Translation_change

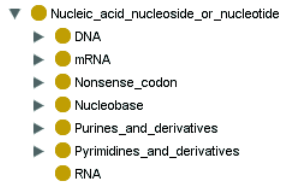


Figure 7 - Hierarchy of Nucleic acid

Discussion

The complete OntoKBCF is a subset of cystic fibrosis clinical and biological knowledge facts. These are appropriate to our model border so as to support our applications aim. We have provided some comments about the construction, its limitations, and the upper ontology.

Explanation of the construction approach

The construction idea starts from the top-down analysis of the knowledge facts and for the dissection from general to specific; construction work starts from basic concepts, from bottom to top and from specific to general-to modify step by step until the basic concept is turned into complex one with more meaning or restrictions. For example we may need to represent patient with the G542X mutation, and need several steps to achieve it. Thus: we define **"G542"** as an amino acid location in the human CFTR protein; we define **"G542X"** as an amino acid substitution in human CFTR protein; finally we define **"Patient_CF_with_G542X"** as a cystic fibrosis patient group with amino acid change G542X.

Nucleotides are units of nucleic acid[18] (p47). According to its chemical structure, nucleotides can be broken into nucleosides and phosphate. The nucleosides can be broken down further into nitrogenous bases and ribose (RNA) or deoxyribose (DNA). It is the nitrogenous base (A, T, C, G, or U) decides the nucleotide type. The one letter abbreviation can be used for either the bases alone or for the nucleotides containing them[18] (p185). So when mutation occurs in DNA or RNA chain, it is mutation in nitrogenous base, also named nucleobase. We choose Nucleobase in the classes naming for description of bases and their mutations.

There are 22 types of amino acids, so there are many more possibilities for amino acid changes; we only list those involved in our border or scope. There are 5 types of nucleobases, so there is less possibility for nucleobase mutations; we list the entire possible point mutation types.

The amino acid three letters names are listed and the single letter names and full names are recorded in the Annotation panel as reference. The nomenclature follows the IUPAC recommendation. For description of amino acid changes we use single letter abbreviation name (such as **"G551D"**) and three letters abbreviation name (**"Gly_551_Asp"**) is recorded in the Annotation panel as well. The single abbreviation letter name also is used for amino acid change location, such as: **"G542"**, which is shorter and simpler.

Single letter abbreviation of amino acid can be confused with nucleobase: A, C, G, T have different meanings. In the OntoKBCF amino acids change with three letter abbreviation names and full names in the Annotation panel. Those nucleobase mutation location name with minus or plus, or start from Ins or Del, all are labels for nucleobase mutation location (such as **"G621_plus_1"** or **"Ins3905"**); it is not easy to distinct the location between nucleobase mutation and amino acid change if they dont carry the labels, although we did not encounter an example in construction.

There is a dilemma: according to IUPAC recommendation X can refer to any amino acid; according to recommendation of protein sequence variants[19] X is used to designate a translation termination codon. In the OntoKBCF we use X to represent stop codon. Because we list all the 22 amino acids, we can use the specific name if required.

Protégé-OWL can not accept the class name start from number, and also can not accept ">" or "+" symbols in the class name, which is the necessary symbols for nomenclature of mutation[19], so we adjust the way it represented. We use **"Del_A"** and **"Ins_A"** to represent nucleobase deletion or insertion. Specific name like **"Ins3905_T"** means insertion in nucleotides 3905 (position) with thymine. For nucleobase substitutions name, we use minus or plus to replace the symbol and start from the replaced nucleobase and end with the new nucleobase (avoid to use ">"): such as **"G1717_minus_1_A"**, which means at nucleotides 1717-1(means the end of the intron, in this case it is intron 10, 1717 is the first nucleotide for exon 11) guanine change to adenine. To name amino acid substitution is in the similar way, such as **"Gly_551_Asp"**, amino acid 551- glycine changes to aspartate .

Although we didnt follow strictly the nomenclature recommendation[19] about mutation, we believe it is acceptable considering our aim for the model, which attempts to explain explicitly the meaning of mutation, instead of a reference dictionary to map all the possible mutations. We hold to the principle of the "recommendation" and adjust it according to the naming rules of Protégé-OWL.

In OntoKBCF and its further application within an EHR system, we only consider the human being as the subject of care, not other species. So we narrow down the classes into human being type, such as: under **"Gene_or_genome"** , we list **"Human_CFTR_gene"**; under **"Exon"** we list **"Human_CFTR_gene_exon"**. At this stage all the basic concepts or combined concepts are treated as class. We only consider individual if it belongs to a specific person, but this can be revised later if required.

Scope: border, granularity and axes

It is important to decide the border and granularity of the OnKBCF as knowledge itself is endless; here we only construct the OntoKBCF within the defined scope to complete our goal. Indeed, construction of the OntoKBCF is only part of our research and we need use it in the next stage. We prefer to extend the properties and constraints in future according to the application scenario rather than complete all properties, some of which will never be used. Without establishing clear granularity, it would also be an endless process to split the knowledge into subunits. The main axes are helpful in deciding the position of terms and will provide clarity and keep the knowledge unit organization consistent.

Upper ontology

There are some higher level formal ontologies which can be followed: Medical Ontology, SnapBFO (snapshot ontologies, indexed by times) and SpanBFO (videoscopic ontology) [20].

However at this stage it is difficult to say if the upper ontology view from SnapBFO and SpanBFO can make sense in this application. The OntoKBCF tests with the EHR system may provide more practical evidence for upper ontology mapping.

Representation limitation

An attempt was made to represent all the selected content fully in the OntoKBCF, but there are some precise details that cannot be represented. For example, "nasal polyps, especially if recurrent", is a highly suspected evidence for cystic fibrosis for older children. But it is difficult to represent especially within OWL. And it is also difficult to represent the difference between "very common" and "sometimes", both of which are quite common in bio and clinical knowledge description.

Conclusion

The OntoKBCF includes knowledge facts from nucleobase mutation, amino acid change to clinical phenotype. Although the horizontal layers of knowledge in the model are far from complete, the vertical relationships and structures partially succeed in bridging the knowledge from different levels. The link to a record system as modelled by openEHR will be the next stage of our work.

Acknowledgements

Sincere thanks to Matthew Horridge and Hai Wang from Manchester University for their kind help in Protégé-OWL.

References

[1] W3C. Semantic web. [cited 2006 Oct. 10]; Available from: http://www.w3.org/2001/sw/

[2] Davies J, Fensel D, Harmelen Fv. Introduction. In: Davies J, Fensel D, Harmelen Fv, eds. *Towards the semantic web: Ontology -driven knowledge management*. West Sussex UK: John Wiley & Sons, 2003;4.

[3] NLM NIH. Unified Medical Language System. [cited 2006 July 6]; Available from: http://www.nlm.nih.gov/research/umls/

[4] Gene Ontology Consortium. Gene Ontology Home. [cited 2006 July 6]; Available from: http://www.geneontology.org/

[5] W3C. OWL Web Ontology Language Overview-W3C Recommendation 10 February 2004.DL McGuinness,F van Harmelen, editors [cited 2006 Oct. 10]; Available from: http://www.w3.org/TR/owl-features/

[6] W3C. RDF/XML syntax specification (revised). [cited 2006 Nov. 5th]; Available from: http://www.w3.org/TR/rdf-syntax-grammar/
[7] Bechhofer S, van Harmelen F, Hendler J, Horrocks I, McGuinness D, Patel-Schneider P, et al. OWL Web Ontology Language Reference-W3C Recommendation 10 February 2004. [cited 2006 Oct. 10]; Available from: http://www.w3.org/TR/owl-ref/
[8] Rector A. Foundations of the Semantic Web: Ontology Engineering. 2005 [cited 2006 Aug. 1st]; Available from: www.cs.man.ac.uk/~rector/modules/CS646/Ontology-building-2005-Lect-1.ppt
[9] Smith R. What clinical information do doctors need? BMJ. 1996;313:1062-1068.
[10] Seol YH, Kaufman DR, Mendonca EA, Cimino JJ, Johnson SB. Scenario-based assessment of physicians' information needs. In: Fieschi M, editor. MEDINFO 2004; 2004: Amsterdam: IOS Press; 2004.p. 306-310.
[11] Lappa E. Undertaking an information-needs analysis of the emergency-care physician to inform the role of the clinical librarian: a Greek perspective. Health Information and Libraries Journal. 2005;22:124-132.
[12] The Cocharane Collaboration. Cochrane review topics: cystic fibrosis. [cited 2006 July 7]; Available from: http://www.cochrane.org/reviews/en/topics/55.html
[13] Harris A, Super M. Cystic fibrosis: the facts. 3rd ed. Oxford: Oxford University Press, 1995.
[14] Alton EWFW. Use of nonviral vectors for cystic fibrosis gene therapy. Proc Am Thorac Soc. 2004;1:296-301.
[15] Flotte TR, Laube BL. Gene therapy in cystic fibrosis. CHEST. 2001;120:124S-131S.
[16] Orenstein DM, Rosenstein BJ, Stern RC. Cystic fibrosis: medical care. Philadelphia: Lippincott Williams & Wilkins, 2000.
[17] Zielenski J, Tsui LC. Cystic fibrosis: genotypic and phenotypic variations. Anu Rev Genetics. 1995;29:777-807.
[18] Campbell NA, Reece JB, Taylor MR, Simon EJ. Biology-concepts & connections. 5th ed. San Francisco, CA, USA: Benjamin Cummings, 2006.
[19] den Dunnen JT, Antonarakis SE. Recommendation for the description of protein sequence variants. Hum Mutat. 2000;15:7-12.
[20] Smith B. Home page of Barry Smith. [cited 2006 Mar. 8]; Available from: http://ontology.buffalo. edu/smith/

Address for correspondence

Prof. Stephen Kay
Room PO33,
SHIRE, Brian Blatchford Building
Frederick Road Campus,
University of Salford,
M6 6PU, UK
Email: S.Kay@salford.ac.uk

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Integrating Biological Pathways in Disease Ontologies

Julie Chabalier[a], Jean Mosser[b], Anita Burgun[a]

[a] *EA 3888 - IFR 140, Faculté de Médecine, Université de Rennes I, 35033 Rennes, France*
[b] *UMR CNRS 6061- IFR 140, Faculté de Médecine, Université de Rennes I, 35033 Rennes, France*

Abstract

Anatomy, clinical features, etiology, and morphology are the major organizing principles in existing disease ontologies. Assuming that biological pathways (including protein physical interactions, metabolic reactions, regulatory networks) will be in the near future key components in classifications of diseases, we have analyzed how information about pathways can be integrated into disease ontologies. We designed a disease ontology in OWL. SNOMED CT was used to provide the initial disease descriptions. In a second step, we integrated information from the KEGG PATHWAY and the GO annotation databases into the disease ontology. In the last step, we analyzed the classification of diseases. For example glioma of brain shares 30 pathways with other cancers, and 19 pathways with Alzheimer's disease. As our knowledge about biological pathways is constantly evolving, this approach can be used for integrating automatically this knowledge in existing ontologies. Thanks to the automatic classification associated with formal ontologies, this approach helps identify physio-pathological classes and taxonomic relations in diseases ontologies. It can therefore be used to create new partitions, focusing on pathways, in biomedical ontologies.

Keywords:

ontology, biomedical domain, knowledge representation, knowledge base, KEGG

Introduction

In most disease ontologies, the major organizing principles relate to clinical features, etiology, location and morphology. Etiology of diseases, when present, is mainly represented by extrinsic agents such as micro-organisms, while the genetic etiology and the metabolic pathways behind diseases remain absent. For example, in SNOMED Clinical Terms® (SNOMED CT®)[1], the characterization of diseases is based on roles such as *definitional manifestation,* which relates disorders to clinical findings, *causative agent*, which relates disorders to organisms, *finding site*, which relates disorders to anatomical entities, and *associated morphology*, which relates disorders to morphologic abnormalities.

Nowadays, associations between classes of genes and diseases as well as associations between pathways and diseases are key components in the characterization of diseases. Different phenotypes may share common pathways and different biological processes may explain the different grades of a given disease. Therefore, we assume that knowledge about pathways will become a major organizing principle in the next versions of disease ontologies.

The Kyoto Encyclopedia of Genes and Genomes (KEGG) PATHWAY database is a collection of pathways maps representing our knowledge on the molecular interaction and reaction networks for metabolism and cellular processes [1]. As the Gene Ontology (GO) [2] does not provide direct association with pathways, Mao et al have proposed to use the KEGG Orthology (KO) as a controlled vocabulary for automated gene annotation and pathway identification [3]. Recently, information about the pathways involved in human diseases has been added to KEGG. The Nov 26th, 2006 release of KEGG contains pathway information for six neurodegenerative disorders (including Alzheimer's disease), three infectious diseases, three metabolic disorders, and four malignant diseases (including glioma and chronic myeloid leukemia).

The objective of this study is to investigate how information about *pathways* can be integrated into disease ontologies. In this paper, we will use the term 'pathway' for metabolic pathways, regulatory pathways and biological processes. We have evaluated the degree to which, the knowledge provided by KEGG and GO annotation can be used to populate the *hasPathway* role in disease ontologies. We selected SNOMED CT as the reference for disease definitions because it is the most comprehensive biomedical terminology recently developed in native description logics formalism, which enables automated classification. To illustrate our results, we selected glioma, because the KEGG PATHWAY database contains information about the pathways involved in glioma. Glioma is the most common primary brain tumor. To investigate how information about pathways can serve disease classification purposes, we compared glioma to other neurological diseases, including Alzheimer's disease, and other cancers, including chronic myeloid leukemia.

1 http://www.snomed.org/

Selected for best paper award.

Materials and methods

The methods of this study can be summarized as follows. First, we identify the diseases present in the KEGG PATHWAY database and create a disease ontology in OWL[2] (Web Ontology Language). Disease descriptions are based on SNOMED CT. A *hasPathway* role is added to the ontology. We then use the KEGG PATHWAY database to (i) extract the relations between a given disease and the biological pathways which are related to it, and (ii) populate the *hasPathway* role. Using the *hasPathway* property, we identify the associations among the diseases in the ontology and evaluate them. An overview of the method is given in figure 1.

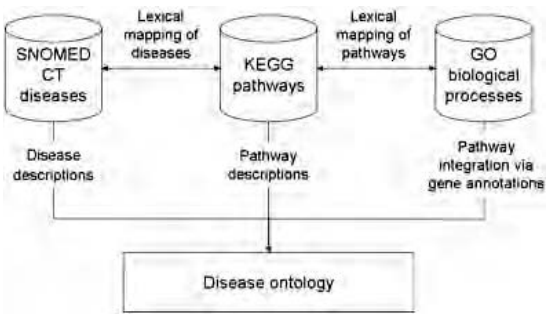


Figure 1 - Conception of the disease ontology

Designing the ontology

We described the ontology in OWL, a description logics language designed specifically to provide a common way to process the content of web information. The SNOMED concepts and roles needed to describe the diseases present in the KEGG PATHWAY database as well as their ascendants were turned into, respectively, OWL classes and properties. We added a *hasPathway* property and its inverse, *isPathwayOf,* in the disease ontology.

The KEGG PATHWAY database was used as the reference database for biochemical pathways. It contains most of the known metabolic pathways and some regulatory pathways. KEGG pathways are structured according to a four-level hierarchy with five major root classes: *Metabolism, Genetic Information Processing, Environmental Information Processing, Cellular Processes,* and *Human Diseases*. The third level corresponds directly to the pathway maps, and the fourth level consists of the leaf nodes, which represent the genes involved in the pathways. The first three levels of this hierarchy were integrated in the disease ontology.

In order to obtain a disease ontology consistent with the KEGG hierarchy and GO, the KEGG pathways were aligned with the concepts of SNOMED CT and GO terms using lexical mapping. As the KEGG pathways often correspond to composite terms, e.g. *Fructose and mannose metabolism*, we segmented and reconstructed these terms according to the coordination conjunctions in order to improve the mapping. For example, *Fructose and mannose metabolism* is present as such neither in GO nor in SNOMED CT. The segmentation-reconstruction operation results in two terms, *Fructose metabolism* and *Glucose metabolism*, which are both present in GO.

We used the MetaMap Transfer program (version 2.4.B) [4] which is part of the UMLS Knowledge Sources [5]. The MetaMap program maps a list of terms to UMLS concepts with the possibility to restrict the output to selected sources (*restrict_to_sources* option). As SNOMED CT and GO are part of the UMLS source vocabularies, we used this option to restrict the mappings to respectively SNOMED CT and GO. Once the mapping was done, we extracted all the ascendants of these concepts.

Populating the *hasPathway* role in the ontology

Diseases and pathways were linked through the *hasPathway* properties. We used the KEGG PATHWAY database and the Gene Ontology Annotation files (GOA) [6] to retrieve these relations.

First, for each disease present in KEGG, we extracted the relations between the genes and the disease pathways from the fourth level of the KEGG hierarchy. We retrieved the biological pathways where these genes are involved by mining the KEGG hierarchy and the GOA files. We have considered that a pathway is related to a disease if a gene is involved in both the disease and the pathway according to KEGG and GOA. The question is then to quantify the strength of this association. For a given disease D, we assume that the more genes are associated with D and involved in a same pathway, the more this pathway is likely to be related to D. The relationships and the gene-dependent distance between diseases and pathways were described in the DOT language which is a plain text graph description language. We used Graphviz[3], a graphical layout package, for visualization of the resulting graphs (especially the NEATO module). Each graph is centered on a disease node, which is linked to all the pathways that are related to it. The length of each edge is inverse-proportional to the number of genes that are shared by the disease and the pathway. For a given disease, the pathways that are graphically close to it are considered related to this disease very strongly.

Reasoning with the disease ontology

Once the disease ontology was modeled and populated, we used Pellet, an open-source OWL-DL reasoner to infer new knowledge on the diseases [7]. We focused on the classification of diseases according to their shared pathways.

Results

Of the 16 disease entities modeled in the KEGG hierarchy (third-level subclasses of the *Human Disease* class), 15 disease terms were successfully mapped to SNOMED CT concepts through MetaMap. The remaining disease, *Epithelial cell signaling in Helicobacter pylori infection*, was partially mapped to the concept *Helicobacter pylori infection (Helicobacter Infections)*.

2 http://www.w3.org/TR/owl-features/

3 http://www.graphviz.org

Of the 213 KEGG pathways (third-level subclasses not derived from *Human Disease* class), 112 have an exact correspondence in GO. This relative low number is due to the complexity of the KEGG classes which are often represented by composite terms. The segmentation-reconstruction operation resulted in 17 supplementary classes strictly mapped to GO terms. Finally, 61% of the classes were successfully mapped to GO.

Figure 2 shows the pathways associated with glioma. 25 pathways involve at least two genes that are associated with glioma. The four pathways in the first circle around the glioma node, *i.e.* those that are closely associated with this disease, correspond to cell proliferation and signal transduction mechanisms. Those mechanisms are well-known for their implication in cancers. On the second circle, two pathways, *Cell motility and insulin receptor pathway*, illustrate the invasive aspect of glioma which is one of the most deadly tumors, with median survival of about one year [8].

Through the reasoner, we proposed an automatic classification of the diseases modeled in the ontology according to the pathways that they shared. First, we compared two neurological disorders, namely glioma and Alzheimer's disease. 19 pathways involved in glioma were also associated to Alzheimer's disease (see figure 3). Then we compared two cancers, namely glioma and chronic myeloid leukemia; 30 pathways were shared by these two cancers, including the four pathways previously introduced as closely related to glioma (see figure 4).

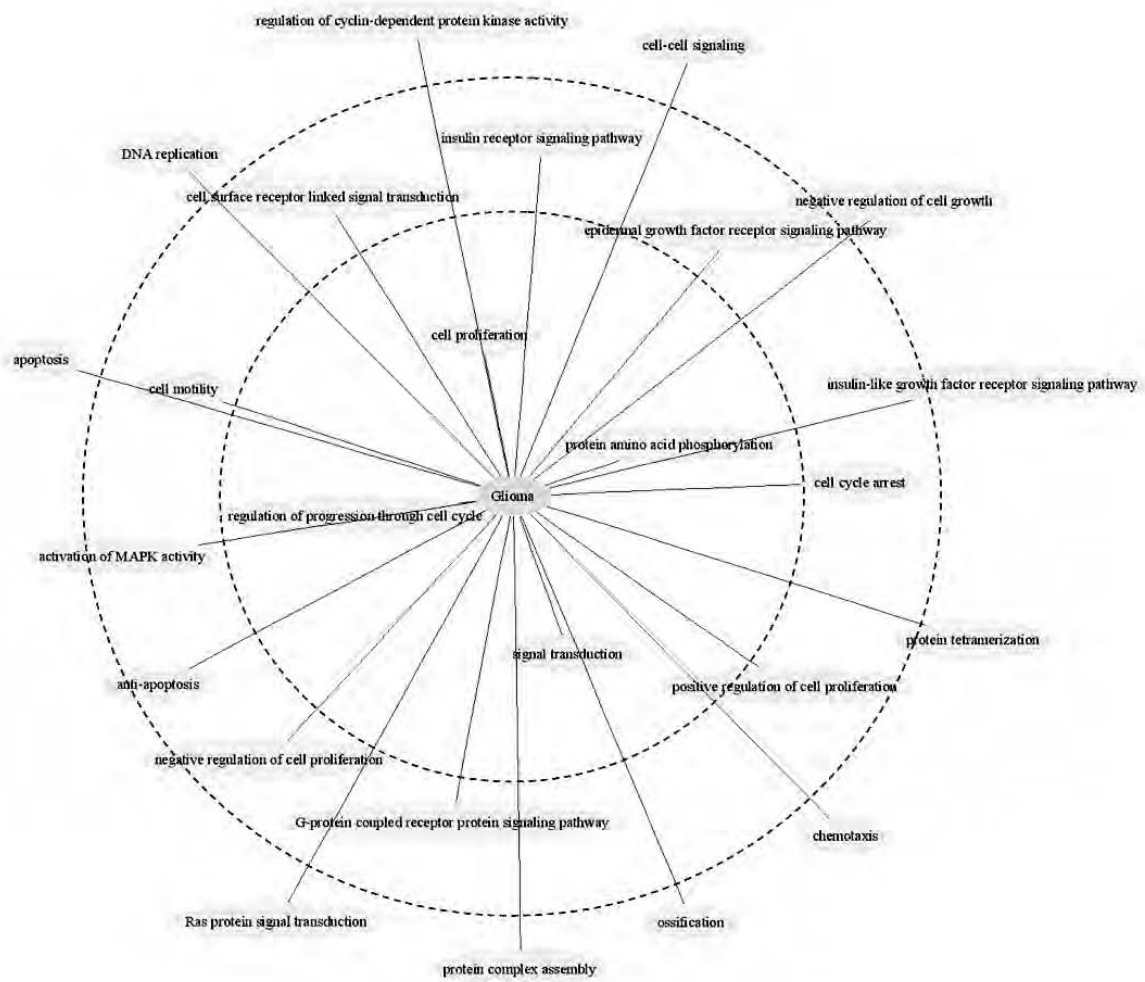


Figure 2 - Pathways related to glioma

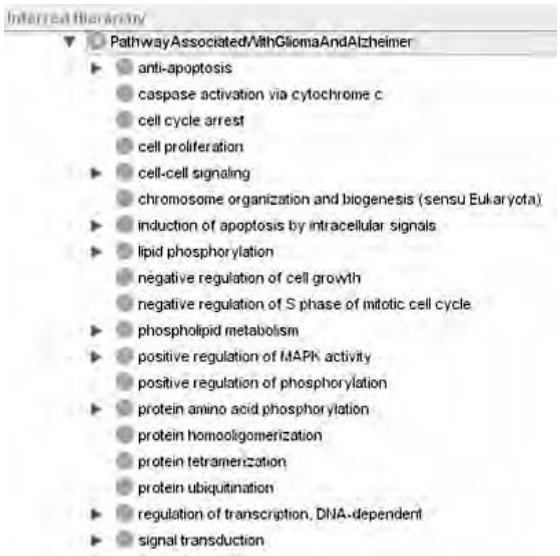


Figure 3 - Pathways shared by two neurological disorders

Discussion

Integration of genotype and phenotype information is crucial for biomedical research *e.g.* [9]. This paper presents a study about the integration of information on biological pathways in disease ontologies. To address this issue, we modeled a disease ontology based on SNOMED CT. Each disease class was enriched with the pathway information contained in GO and KEGG, which enables automated classification based on pathways.

Generalization

This study has focused on KEGG, GO and classification of diseases with respect to pathways. In this paper, we present results that are limited to a few diseases. This limitation is due to the limited coverage of the KEGG database in terms of disease pathways -- the Nov 26th, 2006 release of KEGG contains pathway information for 16 diseases. As more information about disease pathways becomes available, this approach can be applied to larger sets of diseases. This approach can be used to classify diseases with respect to the pathways involved. Given an ontology of biological pathways, a database providing relations between genes and diseases, and a database providing functional annotation of genes, these resources can help identify new relations between the diseases. Knowing that a disease D is related to a pathway P, it is possible to infer that D may be classified with other diseases sharing the same pathway. Moreover, since the whole process of integrating pathways in the ontology is automated, the classification of diseases evolves as the KEGG knowledge base is updated.

Disease classification

In our approach, the relation between a disease and a pathway is based on the number of genes involved in both. These relative distances are visualized as a set of undirected graphs which are centered on the disease. However, in order to illustrate how information about pathways can be integrated in disease ontologies, we decided to model the relations in the ontology independently of these distances. Therefore, the resulting model is made of two complementary representations. First, in the ontology, all the relations between pathways and diseases that can be extracted from KEGG are represented. Second, by the means of graph-based representations the strength of the associations between diseases and pathways is taken into account and pathways are organized according to the distances between the pathways and the diseases.

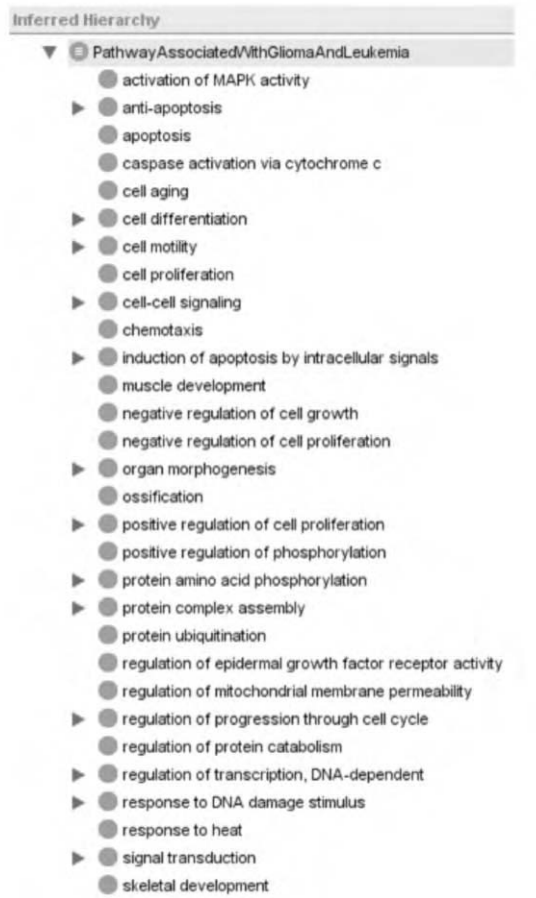


Figure 4 - Pathways shared by two cancer disorders

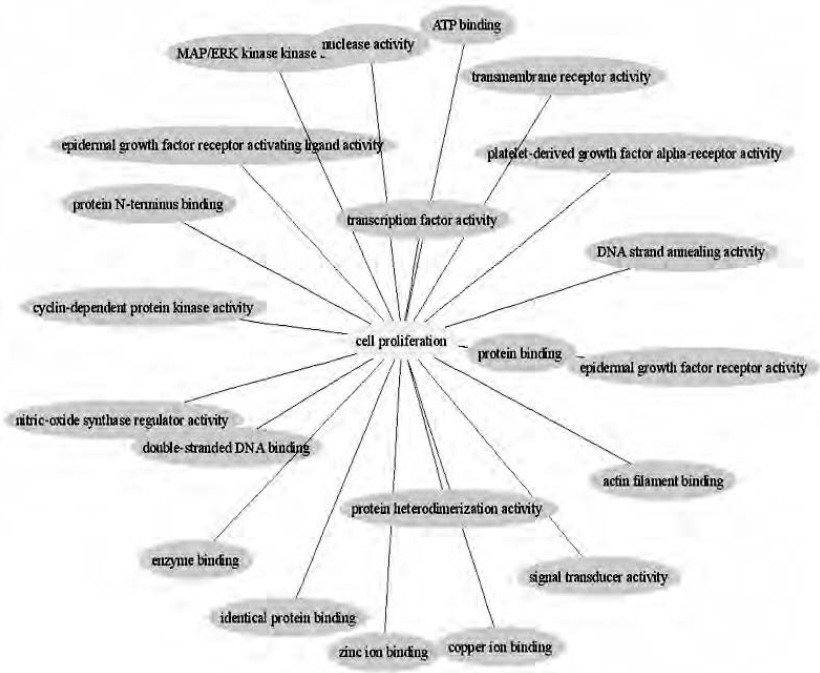


Figure 5 - Molecular functions related to the cell proliferation pathway

Molecular function integration

This study focused on the integration of the biological pathways in disease ontologies. Currently, we work to integrate the molecular functions related to the pathways in these ontologies. For instance, we have studied the pathway-function relationships based on the number of shared genes (see figure 5). However, to extend our approach to molecular functions, we have to first address some issues in reasoning in large, complex, nested ontologies with the current classifiers.

Conclusion

Biological processes and pathways related to gene expression are underlying mechanisms for diseases (phenotypes). Beyond the traditional classificatory principles employed in the construction of disease ontologies, we have represented the relation between pathways and diseases and we have used KEGG to populate this ontology. Thanks to the automatic classification associated with formal ontologies, this approach helps identify relations between diseases based on their underlying physio-pathological mechanisms

Acknowledgments

This work was supported by a grant from the Region Bretagne (PRIR).

References

[1] Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M: The KEGG resource for deciphering the genome. Nucleic Acids Res 2004, 32:D277-280.

[2] Consortium GO: The Gene Ontology (GO) project in 2006. Nucleic Acids Res 2006, 34:D322-326.

[3] Mao X, Cai T, Olyarchuk JG, Wei L. Automated genome annotation and pathway identification using the KEGG Orthology (KO) as a controlled vocabulary. Bioinformatics. 2005 Oct 1;21(19):3787-93. Epub 2005 Apr 7.

[4] Aronson AR. Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program. Proc AMIA Symp. 2001:17-21.

[5] Bodenreider O. The Unified Medical Language System (UMLS): integrating biomedical terminology, Nucleic Acids Res 2004,32, Database issue:D267-70.

[6] Camon E et al. The Gene Ontology Annotation (GOA) Database: sharing knowledge in Uniprot with Gene Ontology. Nucleic Acids Res 2004, 32:D262-266.

[7] Parsia B., Sirin E. Pellet: An OWL DL Reasoner. In 3rd International Semantic Web Conference (ISWC2004), 2004.

[8] Quigley MR, Post C, Ehrlich G. Some speculation on the origin of glioblastoma. Neurosurg Rev. 2007 Jan;30(1): 16-21. Epub 2006 Nov 23.

[9] Goh CS et al. Integration of curated databases to identify genotype-phenotype associations. BMC Genomics. 2006 Oct 12;7(1):257.

Address for correspondence

Julie Chabalier, Laboratoire d'Informatique Médicale, Faculté de Médecine, 2 av. Pr. Léon Bernard, 35043 Rennes, Cedex, France.
julie.chabalier@univ-rennes1.fr

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Reconciliation of Ontology and Terminology to Cope with Linguistics

Robert H. Baud[a], Werner Ceusters[b], Patrick Ruch[a], Anne-Marie Rassinoux[a], Christian Lovis[a], Antoine Geissbühler[a]

[a] *University Hospitals of Geneva, Service d'Informatique Médicale, Switzerland*
[b] *Center of Excellence in Bioinformatics & Life Sciences, SUNY at Buffalo, USA*

Abstract

Objectives: *To discuss the relationships between ontologies, terminologies and language in the context of Natural Language Processing (NLP) applications in order to show the negative consequences of confusing them.*

Methods: *The viewpoints of the terminologist and (computational) linguist are developed separately, and then compared, leading to the presentation of reconciliation among these points of view, with consideration of the role of the ontologist.*

Results: *In order to encourage appropriate usage of terminologies, guidelines are presented advocating the simultaneous publication of pragmatic vocabularies supported by terminological material based on adequate ontological analysis.*

Conclusions: *Ontologies, terminologies and natural languages each have their own purpose. Ontologies support machine understanding, natural languages support human communication, and terminologies should form the bridge between them. Therefore, future terminology standards should be based on sound ontology and do justice to the diversities in natural languages. Moreover, they should support local vocabularies, in order to be easily adaptable to local needs and practices.*

Keywords:

terminology, ontology, natural language processing

Introduction

Recently, Cimino expressed 12 new desiderata for controlled vocabularies in the twenty-first century [1]. The second desideratum says that for each object of a domain, there should be one and only one representative term, which is unique and non ambiguous. Such a term is often call the *preferred term* and the collection of all such terms is a *controlled terminology* (formerly named a *controlled vocabulary*). A revision paper adds 6 new desiderata [2], in response to the recent trend in favor of more formal ontological foundations for terminologies in which terms refer to entities in reality, rather than to concepts in the mind of domain experts [3]. Although controlled terminologies based on preferred terms may contribute to semantic interoperability amongst applications, they are also the source of a modern myth according to which the world could do without the freedom of expressivity that is possible in natural language [4]. Clearly, natural languages are not controllable. Human beings use natural language not only for professional reasons, but also for cultural and social reasons. The patient record, for example, whether paper- or computer based, is usually dictated or written directly in the care provider's natural language so that direct local expressiveness of the language and not conformity to academic desiderata or controlled terminologies is given priority.

This is not to say that Cimino is wrong, but that his paper is written from the perspective of a terminologist: the primary goal is to make data annotated by means of terminologies computer understandable and to enable interoperability between different systems. The desiderata he proposes thus encode these constraints. Despite his expertise in terminology-based lexicon development such as the MED [5], the clothes that Cimino wears in his papers are not those of the linguist. Nevertheless, he also asserts that *'synonymy is a type of redundancy which is desirable'* (desideratum 12). This statement, so we argue, supports the conclusion of this paper.

If a terminology is intended to be disseminated widely and to facilitate communication between human agents and computer systems, then it should take the modus operandi of both these players into account. Thus in order to accommodate the needs imposed by computer systems, terminologies should be based on adequate ontologies. To service human agents, terminologies should be linked to lexicons containing multiple synonyms, eponyms, acronyms, and local jargon. Perhaps, they should even come equipped with facilities to deal with common spelling errors.

This paper deals primarily with the design issues of terminologies related to human agents, and more specifically with the following paradox: on one side, the terminologists recommend unique preferred terms for referring to the entities in a domain and on the other side the linguists and the majority of users are in favor for a full diversification of natural languages, including all kinds of synonyms. To overcome this paradox this paper will first consider the two points of view (that of the terminologist and that of the linguist/end-user) and their supporting arguments. Acknowledging that each expert is guided by specific con-

straints and is accordingly acting coherently, the constraints are made explicit. In a second step, conflicting constraints are examined and an enlarged perspective able to accommodate both points of view is developed. Finally, guidelines based on this enlarged perspective are formulated.

State of the art

Numerous terminologies have been found to suffer from major inconsistencies, examples being SNOMED CT [6], the Terminologia Anatomica (TA) [7] and the NCI Thesaurus [8]. Rector, who raised the question "Clinical terminology: Why is it so hard?" [9], argues that *'clinical terminology concerns the meaning, expression and use of concepts in statements in the medical record'* and assumes the reason for the inconsistencies in terminologies to be the result of a failure in *'separating language and concept representation'* (#4 of 10 difficulties). He describes it as the *'confusion of concepts and words used to express those concepts.'*

However, others argue that the inconsistencies are introduced by the lack of a sound ontological basis for these terminologies, misled as their authors are by the concept orientation advocated by Rector [3] thereby confusing ontology with epistemology [10]. Therefore, in order to avoid such inconsistencies and to coordinate the efforts of several groups by means of commitment to an agreed upon set of principles for best ontology practices, the Open Biomedical Ontology (OBO) Foundry - a new paradigm for biomedical ontology development - has been created [11]. Currently, a set of 10 principles is published and all OBO members are committed to follow them. This initiative is clearly a collaborative experiment in the quest for discovery of best practices in ontology and terminology development. It rests on the principle that high quality representations should be built out of representational units that refer exclusively to three sorts of entities that exist in reality - universals (e.g. person, disease), defined classes (e.g. employee of a hospital, patient under treatment) and particulars (this tumor, the World Health Organization) – and that are connected by means of relationships that mirror ontological relationships [12]. In this context, a terminology is defined as a representational artifact consisting of representational units which are the general terms of some natural language used to refer to entities in some domain. As a result of this effort, quality ontologies in the biomedical domain start to exist. The Gene Ontology (GO) [13] is in constant development, following the evolution of genetic sciences and the provisions of the OBO Foundry [11]. On the side of clinical practice, the Foundational Model of Anatomy (FMA) [14], supporting the TA [15], is an example of a high quality ontology following the OBO principles.

This is in contrast with mainstream work on terminology during the last 20 years which was dominated by a linguistic and normative perspective, primarily driven by the English speaking community. This has had (at least) two inconveniences. First, language itself is not able to make predictions about what exists: although the term "king" might refer to the universal *king*, and "France" to the particular *France*, language rules allow us to create terms such as "the king of France", even though no such entity currently exists. Second, emphasis on English language terminologies hampered creating translations and this for many reasons: 1) human translation is resource dependant and time consuming; 2) the terminology is in principle subject to major revision; 3) questions arise about the role and quality of terms; 4) sets of useful synonyms are not often available and difficult to collect.

The terminology line

The most widely prevailing view on terminology holds that *'terms are the linguistic representation of the concepts in a particular subject field and are characterized by special reference'* as opposed to words that *'function in general reference over a variety of codes'* [16]. According to Lauriston, a term is *'the intersection between a conceptual realm (a defined semantic content) and a linguistic realm'* [17].

In order to ensure that terms are used with the correct meaning, the terminologist may provide definitions which allow the terms to be organized in a hierarchy. However, because this task is difficult, long and expensive, he may not be able to provide explicit definitions for all entities. Therefore, he uses the entity's name as subsidiary definition. But at the same time he is constrained to limit the names to short terms (say up to 5 words) for pragmatic reasons. Using terms alone as implicit definitions is a design error, leading to severe problems[1].

Most terminologies have also adopted usage of what is called *preferred terms*, although SNOMED CT, GO and FMA have added non significant numeric identifiers. Preferred terms are subject to desideratum 2 (one and only one meaning) and terminologists try to respect this constraint.

However, in our view, the primary goal of the terminologist is not to relate the terms in a domain of discourse to 'concepts', but to organize them in such a way that it is clear which terms refer to what entities in reality, and which do not. Furthermore, for the purpose of communication he has to identify the relevant entities that are not yet properly named. This means that the terminologist must play partly the role of the ontologist, or rely on services offered by ontologists. However, he must be aware that the ontology underlying any terminology is universal and language independent. There is a significant risk of mixing the ontologist's role with the terminologist's role, which is mainly language dependent. As an example, it is not because medicine identified the two distinct diseases formerly named "diabetes type 1" and "diabetes type 2", that there exists a universal referred to by the term "diabetes".

The terminologist should also be aware that in contrast to definitions in ontologies that describe the necessary and sufficient conditions for an entity to be what it is, termino-

1 In TA the code a05.6.02.010 is for *hidden part of duodenum*. Not every physician knows about this object and the term is non existent in most atlas of anatomy!

logical definitions should focus on the conditions under which a term may appropriately be used.

This in turn puts a different perspective on the notion of preferred terms the effectiveness of which has been questioned:

- preferred terms have been selected according to the needs of the experts, not the casual users,
- they may not be adequate in specific contexts,
- they contradict local habits and usages,
- they are not easily accessed by non expert users,
- they vary from one language to another,
- human beings fancy to be different and are reluctant to standardize their language.

In his paper [4], Ceusters gives a voice to the users. He demonstrates that less than 50% of them are using the defined preferred terms in practice, this trend being augmented by usage of speech recognition tools. He says that *'clinicians do not face major problems in understanding terms derived from clinical narratives generated by peers'*. He also suggests that *'preferred terms are merely an academic artifact rather than a reality'*. Moreover, Ceusters shows that the mean number of variants for any term is superior to 5 in practice, in concordance with another study showing even higher figures [18]. In order to do justice to the users, it should be recognized that they *'want (and get) back the freedom of expression with all delicate yet important nuances that are required for individual patient care'* [4].

In theory, the need for preferred terms is artificial: with only numeric identifiers and explicit definitions of entities, the design and publication of a source terminology is feasible. From a pragmatic point of view, the usage of unique text identifiers (the *knowledge names* in Galen) is possibly a good choice, but their usage should be limited to the experts and discouraged to the end-users!

The language line

The primary goal of the computational linguist is to build applications that can analyze and understand medical texts. A first requirement for an application of this type is that it is able to recognize which terms or phrases in a text, including any linguistic variants or forms, refer to domain entities [19]. There is a practically unlimited number of variants for any given term, as shown in [18], thus the construction of an explicit list is not feasible. Therefore, the linguist tries to design intelligent algorithms that can take a short list as input, and deduce other variants based on it.

Variants come in many flavors and for various reasons, as witnessed by the TA [20]: 1) different sources for naming an entity (for a02.1.05.053 *pterigoid canal*, there is also *recurrent canal*); 2) use of Latin terms (for a02.1.06.022, there are two terms: the Latin *tegmen tympani* or simply *roof of tympanum*); 3) usage of eponyms (*Eustachian tube* for a15.3.02.073 in place of *pharyngotympanic tube*); 4) representation by different part of speech (for a15.2.07.024 *eyelids*, there is also an adjective *palpebral* and a prefix root *blepharo*); 5) use of old or layman terms (for a02.4.01.001 *scapula* there is also *shoulder blade*); 6) difficulties with orthography and/or usage of keyboard (for a11.3.00.001 *tyroid gland* in place of *thyroid gland*); 7) usage of local expressions; 8) differences between professionals, typically surgeons and anatomists; 9) order of word segments (for a03.1.02.006 both *frontosphenoid suture* and *sphenofrontal suture* are correct!); 10) morphological changes and usage of plural (terminology should avoid plural terms, but not the medical texts: a02.2.05.001 *sacral vertebrae* should preferably be *sacral vertebra*); etc.

In free text, rather than terminologies, for instance in the context of data registration in electronic health records, there are even more degrees of freedom such as the use of local idioms, non-academic sentences or orthographic errors. If NLP applications aim for total understanding of the content, these lexical aspects must be taken into account in a way that fortunately has become almost feasible today.

A second requirement for NLP applications is to identify to what domain entity a specific term refers. This often requires the resolution of ambiguities. One type of ambiguity is that the same term may refer to two or more distinct domain entities. Another type is brought about by the use of more generic terms when previously a more specific term has been used. Furthermore, algorithms that generate lexical variants to obtain better recall typically sacrifice on precision, thereby introducing more ambiguity. This calls for taking context into account. If the author of the text is willing to be non ambiguous, he could apply ambiguity free paradigms such as Referent Tracking [21], but this requires considerable future developments and fine tuning of computer applications. This is the cost for an improved situation.

Conjunction of the needs

In [3], Smith introduced a formal definition of a terminology based on philosophical realism. Taking into account the follow-up work reported on in [12], this definition may be rewritten as:

A **terminology** T is a triple <N, L, v>

where:

- N is a set of triples <p, S_p, d>, called *nodes*, with p a unique *label*, S_p a set of *synonyms* and d a *definition* of a node,
- L is a set of ordered pairs <r, L_r>. called *links*, consisting of a relation designation r (*is_a, part_of,* etc), together with a set L_r of ordered pairs <s, s'> of those terms for which srs' represents a consensus assertion of biomedical science about corresponding universals and defined classes at the time the given terminology is prepared,
- v is a *version* number, which encodes the time.

But in order to harmonize the terminology line and the language line, an extension of the definition of synonym is necessary[2]:

This approach involves a shift from the preferred term paradigm to the *bag of terms* paradigm, a bag of terms being by definition an unordered set of terms, each being qualified by different properties governing its usage. No term is above the others and the terminology does not favor any language. Nevertheless, the parameter k below may include a tag indicating whether a string is to be considered appropriate for use in text generation applications. This will avoid that such an application would introduce terms with typographic errors or layman terms in a professional document. But from a pragmatic point of view, any implementation is free to select any term from the set according to local preferences, for the sake of designing user friendly interfaces.

S_p is a set of synonyms, where each **synonym** is a five-tuple <s, k, p, l, t>

where:

- s is a string of characters in some regimented language l,
- k specifies the sort of string (perferred term, acronym, eponym, local idiom, old term, common orthographic error, etc), governing its usage,
- p is the *part of speech* argument of the term t (noun, adjective, verb, etc, augmented by information on gender, number, and so forth)
- l is the language,
- t is the time (or time period) when the string is (or was) considered appropriate to refer to the corresponding entity.

This is in line with, Rector's *'principle of separability between clinical linguistics and clinical pragmatics'* [9]: building appropriate bags of terms is a matter of clinical pragmatics, preferred terms a matter of clinical linguistics.

The bag of term principle is not new and has been advocated by UMLS (Unified Medical Language System), where each CUI (Concept Unique Identifier) may be considered as acting as a recipient or identifier for several variant terms. But UMLS is not a terminology in itself; it is rather a collection of terminologies. The paradigm shift presented here remains a valid recommendation for any individual terminology. Wordnet [22] presents a similar approach with the synsets.

2 The above definition of terminology is borrowed from Smith in [3]; the definition of synonyms' set below is original, in complement.

Guidelines

Based on the above arguments, we propose the following guidelines for future terminologies (emphasizing the language point of view):

- let the label p for a node N in terminology T be a meaningless unique identifier,
- let definition d for node N be such that it reflects the necessary and sufficient conditions for a particular to be an instance of the universal, or a member of the defined class, referred to by N,
- prepare an open bag of terms S_p able to contain any number of strings, preserving the principle of an unordered set (this is the main complement to the initial Smith proposal in [3]),
- qualify any string in S_p for its usage, in particular specify any preferred terms depending on the contexts (be aware that terms do not strictly need to be unique),
- fill in the bag of terms for any language, corresponding to the locations where the terminology is to be disseminated,
- add for any string its parts of speech,
- develop the links of the terminology (this includes the taxonomy) in total independence of the bag of terms, in a language independent way, and on the basis of sound ontological principles,
- consider the maintenance aspect of the terminology, including free addition of new terms and their attributes at any time,
- consider quality assessment tests for the terminology, including validation of the completeness of the set of synonyms.

Discussion

Although ontology is by definition language independent, terminology is not. We argue that terminologies should be made available for each language by using an underlying ontology as a reference, rather than by relying on a structured organization of preferred terms in a specific language.

We believe however that the ontology, terminology and language points of view can be reconciliated through a simple move: the renunciation of preferred terms and the replacing of them with a bag of representative terms, together with qualifiers defining their usage. In addition, the availability of explicit definitions is recommended. This should lead to a win-win situation: the terminology line is totally preserved and the language line is adequately presented. The multilingual aspect underlying any universal terminology is now explicitly mentioned. But of course, there is a cost: the guidelines proposed here will certainly be labor intensive if adhered to. However, we believe that the approach will help save efforts in the long run, which will be necessary anyway before widespread usage of the best terminologies will be possible. Further-

more, the open bag of terms may remain partially empty for a long time, without direct inconvenience for most of the users. The main point is that the bag remains open and is progressively filled in. Natural language generation of compound terms might be helpful here: if the ontology is expressed in some form of logic, then the ontology itself can be used to generate parts of the terminology automatically, and this in many languages. This approach should progressively become the rule for terminologies, because numerous entities are composed from more atomic entities referred to by means of single words.

In order to show how this works, consider the SNOMED CT code SN 285344007, whose preferred term in English is *viral gastritis*. A graph representation of this entity could be the following (square brackets are for objects, round brackets are for relations):

[InflammationProcess]

- (hasLocation) - [Stomach]

- (hasAgent)-[Virus]

The generator program may find in its English lexicon the following words: for [Stomach]: { gastr-, stomach, stomachal}, for [InflammationProcess]: {-ite, inflammation, inflammatory}, for [Virus]: {virus, viral}. In any other language the lexicon may be more or less extensive and the generator will have to cope with this. The generator may use any combination of words, one of them coming from each subset. However, some rules specific to the language will give more weight to some combinations and others will be excluded. Each word will be selected according to its own specificity, computed from its frequency in a large corpus of representative medical texts. At the end of the process, the generator is left with 2 combinations (the first being considered as the best if the strategy is to prefer the short terms): *viral gastritis, inflammation of stomach caused by a virus*. Other terms are theoretically possible, they are correct but they are discarded because they are unusual: *viral stomachal inflammation, gastritis by virus*.

A successful generation experiment has already been conducted for surgical procedures [23, 24]. The same approach is considered by WHO for preparing ICD-11.

Conclusion

Terminologies should not be developed by reference to a system of preferred terms, rather they should be developed in such a way that their individual nodes and relations amongst these nodes are modeled on an underlying formal ontology, where the linguistic content of these nodes will be filled in based on a system of terms and synonyms (from many different languages) that is associated with each node based on the intended ontological interpretation of that node. The idea is that this will give terminologies the rigor of scientific theories, while also making them understandable in natural language.

Acknowlegments

Thanks to James Cimino, Barry Smith and Olivier Bodenreider for their kind comments to this paper. Thanks to Andrew Spear for proof reading of the manuscript.

References

[1] Cimino JJ. In Defense of the desiderata. J Biomed Inform 2006;39(3): 299-306.

[2] Smith B. From concepts to Clinical Reality: An Essay on the Benchmarking of Biomedical Terminologies. J Biomed Inform 2006;39(3):288-298.

[3] Moerkerke C, Ceusters W. The myth of preferred terms in medical sublanguage and its impact on natural language understanding applications: an empirical study. In De Moor G, De Clercq E (eds.) Proc 18th MIC Conf, 2000: pp55-62.

[4] Cimino JJ. From Data to Knowledge through Concept-oriented Terminologies: Experience with the Medical Entities Dictionary. J Am Med Inform Assoc 2000;7(3): pp288-97.

[5] Ceusters W, Smith B, Kumar A, Dhaen C. Ontology-based error detection in SNOMED-CT. Medinfo. 2004;11 (Pt 1):482-6.

[6] Rosse, C. Terminologia Anatomica; Considered from the Perspective of Next-Generation Knowledge Sources. *Clinical Anatomy* 14:pp 120-133.

[7] Ceusters W, Smith B. A Terminological and Ontological Analysis of the NCI Thesaurus. Methods Inf Med 2005; 44: pp498-507.

[8] Rector AL. Clinical Terminology : Why Is it so Hard ? Methods Inf Med 1999; 38: pp147-57.

[9] Bodenreider O, Smith B, Burgun A. The Ontology - Epistemology Divide: A case study in Medical Terminology. Procs 3rd conf on Formal Ontology in Information Systems (FOIS 2004): IOS Press; 2004. p185-195.

[10] The OBO Foundry at: http://obofoundry.org/

[11] Smith B, Kusnierczyk W, Schober D, Ceusters W. Towards a Reference Terminology for Ontology Research and Development in the Biomedical Domain. Proceedings of KR-MED 2006, Biomedical Ontology in Action, November 8, 2006, Baltimore MD, USA

[12] See: http://www.geneontology.org/ or the NLM browser GenNav: http://mor.nlm.nih.gov/perl/gennav.pl

[13] Rosse C et als. Foundational Model of Anatomy. See: http://sig.biostr.washington.edu/projects/fm/AboutFM.html

[14] Federative Committee on Anatomical Terminology. Terminologia Anatomica: International Anatomical Terminology. Thieme Ed. 1998.

[15] Sager J.C. A Practical Course in Terminology Processing. John Benjamins Publishing Company, 1990.

[16] Lauriston, A. Automatic recognition of complex terms: Problems and the TERMINO solution. In Terminology 1994, 1:;147-170.

[17] In TA the code a05.6.02.010 is for *hidden part of duodenum*. Not every physicians knows about this object and the term is non existent in most atlas of anatomy!

[18] Baud RH, Ruch P, Gaudinat A, Fabry P, Lovis C, Geissbuhler A. Coping with the variability of medical terms. Medinfo. 2004;322-6.

[19] Bodenreider O. Lexical, terminological and ontological resources for biological text mining. Chap 3 in Ananiadou S et al: Text mining for biology and biomedicine; Artech House; 2006. pp43-66.

[20] Federative Committee on Anatomical Terminology. Terminologia Anatomica: international anatomical terminology. Stuttgart; New York: Thieme; 1998
[21] Ceusters W, Smith B. Strategies for Referent Tracking in Electronic Health Records. J Biomed Inform. 2006 Jun;39(3):362-78.
[22] The above definition of terminology is borrowed from Smith in [3]; the definition of synonyms' set below is original, in complement.
[23] Fellbaum C. An Electronic Lexical Database. MIT Press, 1998.
[24] Trombert-Paviot B, Rodrigues JM, Rogers JE, Baud RH et al. GALEN : a third generation terminology tool to support a multipurpose national coding system for surgical procedures. Int J Med Inform 58-59 (2000); pp71-85.
[25] Rodriges JM, Rector A, Zanstra P, Baud R & al. An Ontology driven collaborative development for biomedical terminologies: from the French CCAM to the Australian ICHI coding system. MIE 2006 Procs. A Hasman et al. (Eds), IOS Press, 2006, pp 863-8.

Address for correspondence

Robert H Baud, PhD
Service d'Informatique Médicale
Hôpitaux Universitaires de Genève
rue Micheli-du-Crest, 22
CH-1211 Genève Suisse
E-mail: Robert.Baud@sim.hcuge.ch

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

SNOMED CT's Problem List: Ontologists' and Logicians' Therapy Suggestions

Stefan Schulz[a,b], Boontawee Suntisrivaraporn[c], Franz Baader[c]

[a] *Freiburg University Hospital, Medical Informatics Department, Freiburg, Germany*
[b] *Pontificial Catholic University of Paraná (PUCPR), Master Program of Health Technology, Curitiba, Brazil*
[c] *Dresden University of Technology, Faculty of Computer Science, Dresden, Germany*

Abstract

After a critical review of the present architecture of SNOMED CT, addressing both logical and ontological issues, we present a roadmap towards an overall improvement of this terminology. In particular, we recommend the following actions: Upper level categories should be rearranged according to a standard upper level ontology. Meta-class like concepts should be identified and removed from the taxonomy. SNOMED concepts denoting (non instantiable) individual entities (e.g. geographical regions) should be kept separate from those concepts that denote (instantiable) types. SNOMED binary relations should be reduced to a set of canonical ones, following existing recommendations. Taxonomies should be cleansed and split into disjoint partitions. The number of full definitions should be increased. Finally, we propose a new approach to modeling part-whole hierarchies, as well as the integration of qualifier relations into the description logic framework.

Keywords:

SNOMED CT, ontologies, description logic

Introduction

The globalization of SNOMED CT offers a unique opportunity to bring together the following tendencies:

- The urgent need for a global standardized terminology for medicine and life sciences, suitable to cope with an immense flood of clinical and scientific information;
- An impressive legacy of systematized terminology;
- Ongoing efforts toward an ontological foundation of the basic kinds of entities in the biomedical domain;
- The increasing availability of logic-based reasoning artifacts suited for large terminological knowledge bases.

After a long-lasting embryonic and fetal period, the Standardized Nomenclature of Medicine, SNOMED, has seen the light of day as SNOMED CT (Clinical Terms). Since then, it has grown rapidly, in spite of some congenital and infancy problems which have raised the attention not only of domain experts, medical terminologists and software engineers but also of computer scientists and ontologists. Concerned about the conditions under which SNOMED CT is now reaching its adolescence, we recommend an in-depth health check.

Clinical advice will be required from specialists of Ontology and Logic. A careful follow-up of their counseling will be crucial for making SNOMED CT fit for the next decades.

According to the present status of its growth chart [1], SNOMED CT counts 300,000 concepts, about 770,000 English language descriptions, and 900,000 defining relationships. Each SNOMED concept is assigned to one of 19 top-level categories and each relation to one of 49 attribute types.

Methodology

The ontologists' investigation

Amongst the multiple definitions of what an ontology is [2], we narrow the scope to *formal* ontologies. We consider a formal ontology a formal theory that accurately describes a domain in light of the kinds of entities it contains. Ontologists evaluate a terminology system by its agreement with ontological principles, such as the use of well-defined, unambiguous, and non-idiosyncratic types and relations, in accordance with existing standards.

The logicians' investigation

When dealing with ontologies or terminological systems, logicians are primarily interested in formal description languages in which the whole content of the terminology can be expressed. Such languages need to have sufficient expressivity for describing the domain entities. It must be suited for supporting truth-maintenance in the terminology design and maintenance process. It should be sufficiently intuitive for being used in a proper way by the curators of the terminology.

In the style of a clinical problem list, we will present first a thorough analysis of SNOMED CT's current status, under both ontological and logical perspectives. After that we will present a plan for adequate therapeutic measures which will be as little as invasive as possible, cost-effective, and with little impact on the patient's working capacity.

Selected for best paper award.

SNOMED CT's problem list

Dystrophic upper level

Ontologically oriented terminology systems should be grounded upon an upper level that introduces fundamental distinctions such as between endurants and perdurants. Endurants are those entities that exist in their entirety at any point in time, perdurants, however, are never completely present at one single moment. Examples for endurants are objects and spaces, such as organs, cells, spaces, molecules, body fluids etc. Perdurants are traumatic events, courses of diseases, surgical interventions. Functions, dysfunctions, states, and processes, are frequently classified as non-material endurants. SNOMED CT's 19 top level categories preserve the legacy of the former SNOMED axes which do not easily agree with any formal upper level ontology.

Concept borderline disorder

The Biomedical Informatics community has used the term "concept" to an extent that "Concept Systems" became a synonym of any ontology or knowledge engineering artifact. In the last years there has been a strive toward more terminological clarity which has challenged this tradition, arguing that the term "concept" is too ambiguous and obscures the representation of real-world entities [3,4] by ontologies.

We use the term "concept" as a synonym for the nodes in SNOMED CT – as SNOMED CT does, but we normally refine it by talking about "classes", "meta-classes" and "individuals", which is – in our opinion – the most neutral parlance. Problematic under ontology scrutiny is that the wording of numerous SNOMED CT concepts suggests a meta-class interpretation, i.e. classes which classify classes, not individual entities of the world. Examples are *"SNOMED CT concept"*, *"Navigational concept", and "Additional values"*. This causes strange conclusions such as that *London* is an *Additional value*; my particular *Adverse reaction to premedication* is a *Navigational concept*; and my particular *Heartburn* is a *SNOMED CT concept*. Unfortunately – for lack of documentation – it remains open, whether these examples should be interpreted as normal classes (which are identified by rather sloppy and misleading labels), or indeed as meta-classes.

Infestation by individuals

There is no principle objection to why an ontology should not represent individuals, well distinguished from classes. There is, however, a broad consensus that the subsumption relation between two kinds, types or classes (*subtype-of* or *is-a*) is completely different from the instantiation relation between an individual and its type or class (*instance-of*). There are numerous SNOMED CT concepts that stand for individual entities (such as *Europe*, *Greater London*, *Binge eating scale*), most of them being instances of the class *Geographic Location*. However, an *instance-of* relation in missing in SNOMED CT.

Relation idiosyncrasy

Relations should be consistent and unambiguous in order to assist ontology developers and users in avoiding errors [5]. SNOMED CT does not formally define its relations, and by their names they can rarely be mapped to other ontologies. Some relations such as *Finding Site/Procedure Site* or *Specimen Substance* obviously specialize standard relations (e.g. *has-location, has-part*) which, on their part, are missing in SNOMED CT. Finally, there are a number of utterly fuzzy relations such as *Subject Relationship Context*. The problem here is that the more relations exist, the less one can expect agreement among users. Last but not least, there is the ontologically obscure *Relationship group* which was found to correspond to *has-part* between perdurants in most cases but which can also have other hidden meanings [6].

Taxonomic dystrophy

Subclass hierarchies (taxonomies) should obey certain principles such as described in [7]. Accordingly, it must be criticized that numerous non-terminal SNOMED CT classes have one subclass only and that the number of classes with multiple parents is higher than necessary. Another kind of taxonomic dystrophy is the so-called *"is-a overloading",* i.e., the use of taxonomic subsumption in order to express roles rather than generic properties, as already analyzed by [8], using the OntoClean methodology [9]. Sadly, this and other problems found nearly three years ago have not been fixed since then. An example for this is that *Bacterium* is subsumed by *Infectious Agent:* Not every individual bacterium is an infective agent, this is rather a role that can be played by bacteria under certain circumstances. Finally, as pointed out by [10], epistemological aspects should be kept apart from a clinical terminology. In SNOMED CT, there are numerous cases for "epistemology intrusion" such as *Newly diagnosed diabetes*. The point here is that the diabetes as such is no way of a different type by the fact that is has recently been diagnosed.

SEP implants

The so-called SEP triplets [11] are modeling artifacts which expand taxonomies by reified relations. For instance, $Femur_P$ is defined as the class of everything that is part of a *Femur*. $Femur_S$ is introduced as a common taxonomic parent of $Femur_P$ and *Femur*. Together, *Femur*, $Femur_P$ and $Femur_S$ form an SEP triplet. Such structures allow one to express part-whole relationships without explicitly using partitive relations. The main reason why SNOMED CT uses such structures is to enable the propagation of attributes along the aggregation hierarchy (part-whole relationships) in a parsimonious way. E.g., *Femur fracture* defined as a *Fracture* located at some $Femur_S$. Since $Femur_S$ subsumes *Neck of femur*, *Fracture of the neck of femur* is classified as a *Femur fracture*. SNOMED CT is replete of such reified classes, yet in an unsystematic and incomplete way. They are undefined and the terms assigned to them are often misleading. More precisely, we can consider taxonomic A_S-B_S links as kind of prostheses for missing *part-of* relations in the anatomy branch. However, they serve the needs of attribute propagation which is seen as an important asset in medical terminological reasoning.

Partition agenesis

Partitioning means that sibling classes are declared mutually disjoint, as it can be assumed with the 19 SNOMED CT top-level categories. On lower levels, sibling overlap is a general phenomenon that gives rise to complex taxonomic graphs, making the maintenance of the ontology difficult.

Description asthenia

As advocated by [12, 13] for biomedical ontologies, taxonomies should be founded upon the Aristotelian principle of *genus* (the common properties of members in the subsuming class) and *differentiae* (the properties that distinguish each instance of the subsumed class from the genus). According to [7], half of the classes are primitive ones, i.e. they have no criterion which distinguishes them from their super-classes. Besides cases in which Aristotelian definitions are difficult or impossible (e.g. in anatomy), other ones are obviously defined as primitive in order to obviate undesired inferences, such as classifying *amputation of the foot* as a kind of *amputation of the lower limb.* In other cases, there is no obvious reason for primitive definitions: For instance, *severe asthma* could be fully defined as *asthma* (genus), characterized by the value *severe* of the attribute *severity* (differentia).

The qualifier syndrome

SNOMED CT qualifiers, such as *laterality, severity, onset* and *course* are relations used for constraining post-coordination for a further refinement of a class [14]. For example, *asthma* allows 12 different values for the qualifier *course* and six for the qualifier *severity.* Only a small subset of all SNOMED CT relations are used as qualifiers, and it seems that these relations are never used for different purposes. On the other hand, those relations which are used in definitions never feature as qualifiers. So we have the strange situation in which the qualifier *severity* is allowed for the class *asthma*, but is not used for defining its subclass *severe asthma*. For the latter one, *severity* is allowed with its whole range of values, so that the formation of a post-coordinated class *severe asthma* with *severity = mild* is possible. There are innumerous examples that show that the value ranges of the qualifiers are not well adapted to the characteristics of the class they belong to.

SNOMED CT's treatment plan

The assessment of SNOMED CT's health status by both specialties has revealed the following trade-off: Ontologists strive for a comprehensive account of reality, they want to use the whole inventory of logics for describing it with the precision and expressiveness they deem adequate. In contrast, logicians point at the computational properties of full logics, which are prohibitive for any large scale implementation, let alone the issues of modeling and maintenance expenses. A good compromise is given by description logics (DLs), a family of decidable fragments of the first-order logic, which have a clean and intuitive syntax (without the need for free variables), cf. [15]. The description logics implementations available now as well as in future only cover part of this expressiveness. This results in an only partial ability of the system to perform the inferences which should ideally be expected. It is therefore not desirable (and not even feasible from a practical point of view) to use the whole logical machinery for describing entities in a large terminology system. As a result of these underspecifications, one has to be aware of unintended models, but this is unavoidable. However, description logics are relatively well studied, and computationally cheap dialects can be tailored to the actual needs of the ontology under scrutiny. We here sketch the specification of a logic which seems to be well suited to support most modeling and reasoning requirements of SNOMED CT.

It is the *computationally tractable* Description Logic $\mathcal{EL}^{++}$ [16], which is the underlying DL $\mathcal{EL}$ (conjunction and existential quantification) of SNOMED CT enhanced by general class inclusions (GCIs), complex role inclusions (CRIs), the empty class $\bot$, nominals (individuals as classes), restricted concrete domains, and ABox (extensional knowledge about individuals). To give an example, the $\mathcal{EL}$ class expression $Inflammation \sqcap \exists has\text{-}location.Appendix$ denotes the set of all individuals belonging to the class *Inflammation* and relating via the relation *has-location* to some individual of the class *Appendix.* The example constitutes both necessary and sufficient conditions for the class *Appendicitis* and thus can be used as its definition. Not only can GCIs be used to add supplement constraints that are beyond capacity of definitions, together with $\bot$, class disjointness can also be expressed ($BodyPart \sqcap ClinicalFinding \sqsubseteq \bot$). Briefly speaking, CRIs make feasible several ontologically important constraints on roles, including role hierarchy ($proper\text{-}part\text{-}of \sqsubseteq part\text{-}of$), role reflexivity ($\varepsilon \sqsubseteq part\text{-}of$), role transitivity ($part\text{-}of \circ part\text{-}of \sqsubseteq part\text{-}of$), and right identity on role ($has\text{-}location \circ part\text{-}of \sqsubseteq has\text{-}location$). Concrete domains help connect classes in the medical domain to values in concrete domains such as numbers and strings. A simple example is $Minor \equiv Person \sqcap <_{18year}(age)$, i.e., a minor is defined to be a person with less than 18 years of age. Considering the scale of SNOMED CT, it is inevitable to pay attention to the complexity of the logic employed. It has been shown both theoretically [16] and empirically [17] that the $\mathcal{EL}$ DL family is computationally cheap and adequate in terms of ontological expressive power.

Rectification of the upper level

There are several proposals of upper level ontologies, such as BFO, DOLCE, GFO, and SUMO, which roughly coincide in their upper level categories. The ontologists' recommendation is to refer as much as possible to a commonly accepted upper ontology. From the logicians' point of view, the choice is of minor importance. A major preoccupation is, however, that the upper level be expressible in terms of the logics supported. Note that with regard to upper-level organization there are still several controversial points, first of all the ontological account for disease (delimited from *courses of disease*, but also from *sign* and *symptom*). This is currently subject to ontological inquiry, and SNOMED CT could be a good testbed for this.

Isolation of meta-classes

The mentioned borderline problems derive from the fact that most SNOMED CT concepts coincide with classes, describing real-world entities, whereas a minor part corresponds to meta-classes, defining and describing the proper SNOMED CT terminology. These two aspects must be kept apart. However, the proposed logic does not properly support a meta-class level for reasoning. This is not really a problem because we see the necessity of meta-classes more as a kind of housekeeping feature for which annotation functions such as in Protegé and OWL can be used and in which additional *rdf* attributes can be introduced.

Increasing tolerance of individuals

Geographic locations, scales, etc., which are individual entities, should be related as individuals to the corresponding classes and could be related via a certain relation to other individuals. As mentioned, $\mathcal{EL}^{++}$ supports nominals and ABoxes. Both are closely related and are helpful when information about individuals is to be included. For instance, the Siamese cat could be specified to have been originated in Thailand, with Thailand a nominal, not a class ($SiameseCat \subseteq \exists country\text{-}of\text{-}origin.\{Thailand\}$). An ABox comprises assertions about individuals by means of *instance-of* (concept assertions) or *related-by* (role assertions) relations, e.g., $London \in GeographicLocation$ and $(London, England) \in has\text{-}location$, respectively. To keep it simpler without the need of introducing individuals, we could limit ourselves to the reference to geographical entities in terms of location classes. In this case, reifications of the type $A_L \sqsubseteq \exists has\text{-}location.A$ with $A \sqsubseteq GeographicLocation$ may be discussed for the sake of parsimony. Then, for instance, the fact that London is located in England could be indirectly expressed by $London_L \sqsubseteq England_L$.

Reconstruction of relations

SNOMED CT relations should be reduced to a minimum of canonical ones, starting with the OBO relations [5]. Relationship groups should be substituted by the corresponding relation, most likely *has-part*. One should also give a clear account of the algebraic features of each relation in terms of reflexivity, symmetry, and transitivity. Furthermore, relations should be further defined in terms of domain ($\exists r.\top \sqsubseteq D$) and range ($\top \sqsubseteq \forall r.R$)[1] restrictions.

Cleansing of taxonomies

All classes should have at least one sibling, otherwise they should be merged with their super-class. Multiple inheritance should be reduced to a minimum. Wherever an *is-a* link is inferable from defined classes it should be omitted for the sake of brevity and clarity. For instance, *Acute type B viral hepatitis* is fully defined as *Type B viral hepatitis* which is *acute.* By the definitions of *Type B viral hepatitis, Hepatitis,* and *Acute hepatitis* a DL classifier can compute the subsumption between *Acute type B viral hepatitis* and *Acute hepatitis,* so that this *is-a* link does not have to be included.

SEP explanation and substitution

The extra nodes should be fully defined. Although irrelevant under ontological scrutiny, they may be preserved for reasons of backward compatibility. A full definition of S and P nodes, however, requires distinguishing between *proper-part-of* which is transitive and irreflexive, and the broader relation *part-of* which is transitive and reflexive.

So we can fully define $A_P \equiv \exists proper\text{-}part\text{-}of.A,$ together with $A_S \equiv \exists part\text{-}of.A$. However, our language does not allow to enforce irreflexivity of a relation, so that the following GCI might be added where required: $\exists proper\text{-}part\text{-}of.A \sqsubseteq A \sqsubseteq \bot$. With the right identity rule $has\text{-}location \circ part\text{-}of \sqsubseteq has\text{-}location$ we then get the right inference in the femur example. False inferences, such as the classification of an *amputation of the foot* as an *amputation of the lower limb*, can then be prevented by introducing a subrelation of *has-location, viz. has-exact-location* for which the right identity rule dos not apply.

Taxonomy partitioning operation

Although we do not advocate pure single hierarchies as an ontological engineering dogma, we nevertheless recommend the use of one classifying principle for all subclasses of every class as far as possible. This will yield clear partitions at each level, which helps maintain and better understand the terminology. Partitioning a taxonomy generally requires negation statements of the form $A \sqsubseteq \neg B$. Such a restricted negation statement is in fact equivalent to a disjointness axiom of the form $A \sqcap B \sqsubseteq \bot$ which is available in our DL dialect.

Revitalization of full definitions

We suggest the revision of primitive SNOMED CT classes, especially the elimination of misspecifications that are, obviously, the reasons that SNOMED CT classes which could be fully defined, are still kept as primitive ones. Where possible, full definitions should be introduced. The introduction of full definitions generally brings to light hidden misspecifications as soon as the ontology is classified. The use of a terminological classifier is therefore of utmost heuristic importance in the process of building and maintaining SNOMED CT. This requires, however, that SNOMED CT moves to the DL format as the primary format in which all editing is performed.

Qualifier transplant

The realm of qualifiers which is kept somewhat apart from the rest of SNOMED CT sheds light on an intricate problem which complicates the move from a database of frame-like view towards description logics. Whereas the former systems assume a closed world, description logics assume an open world. As a practical consequence, this means that once there are relation types and classes, any relation may be asserted between any individuals unless this is explicitly precluded. SNOMED CT's approach of providing qualifiers with well-defined value restrictions for controlling the building of post-coordinated classes in

1 A controlled use of the universal quantifier $\forall$ in these cases has no negative impact on the computational properties.

description logics extends the capabilities of the logics we use. Strictly spoken, the job of description logics is to define classes by logic descriptions but not to provide constraints for the definition of new classes or the handling of individuals. There are in principle two different ways we can deal with this problem. Firstly, we can handle the constraints as provided by the qualifiers outside description logics. This would mean that we abandon the goal of proposing that the whole content of SNOMED CT be expressed in description logics. The second way is to resort to a more expressive DL dialect, at the price of performance of the implementations. As a possible way out of this dilemma we suggest the following. On the one hand, we maintain the $\mathcal{EL}$ specification for SNOMED CT class definitions, but on the other hand we add an additional layer using DL value constraints. This second layer would then be invisible for the DL reasoner, but it can be used as a resource for those applications in which this information is needed, e.g., to constrain data entry by adaptive pick lists etc., which had been the main rationale for the SNOMED CT qualifiers. Similar to what for meta-classes, this kind of housekeeping information can be realized with the help of annotation functions.

Conclusion

We have subjected the current version of SNOMED CT to an in-depth diagnostic examination under the aspects of ontology and logic. SNOMED CT's clinical picture exhibits mostly chronic problems most of which can be treated in a conservative but yet determined fashion. Some of the problems require a more invasive intervention. We recommend the elaboration of a treatment plan, the definition of priorities, and the allocation of resources. Altogether, the cost of this treatment will be considerable, and it requires specialists from the field of ontology and description logics; nonetheless, it is a good investment for assuring the SNOMED CT's long lasting fitness and its increasing ability to stand the upcoming challenges of medical documentation and standardization.

Acknowledgments

This work was supported by the EU Network of Excellence Semantic Interoperability and Data Mining in Biomedicine (NoE 507505). Additionally, the first author was supported by a research fellowship (550830/05-7) from the Brazilian Research Council (CNPq/Brazil).

References

[1] SNOMED CT® Technical Reference Guide – Jan. 2006, College of American Pathologists, Northfield, IL.

[2] Kunierczyk, W. Nontological Engineering. FOIS 2006 - Proc of the 4th Intl Conf on Formal Ontology in Information Systems; 2006. p. 39-50.

[3] Klein GO, Smith B. Concept Systems and Ontologies. [updated 2006 Oct 6; cited 2006 Dec 4]. Available from http://ontology.buffalo.edu/concepts/ConceptsandOntologies.pdf

[4] Smith, B. Beyond concepts, or: Ontology as reality representation FOIS 2004 - Proc of the 3rd Intl Conf on Formal Ontology in Information Systems; 2004. p. 73-84.

[5] Smith B, Ceusters W, Klagges B, Khler J, Kumar A, Lomax, J et al. Relations in biomedical ontologies, Genome Biology. 2005;6(5).

[6] Schulz S, Hanser S, Hahn U, Rogers, J. The semantics of procedures and diseases in SNOMED CT, Methods Inf Med. 2006; 45(4): 354-358.

[7] Bodenreider O, Smith B, Kumar A, Burgun A. Investigating subsumption in DL-based terminologies: A case study in SNOMED CT. KR-MED 2004 - Proc of the 1st Int Workshop on Formal Biomedical Knowledge Representation; 2004. p. 12-20.

[8] Spackman KA, Reynoso G.Examining SNOMED from the perspective of formal ontological principles. KR-MED 2004 - Proc of the 1st Int Workshop on Formal Biomedical Knowledge Representation; 2004. p. 72-80.

[9] Guarino N, Welty CA. An overview of OntoClean. In: Staab S, Studer R. editors. Handbook on Ontologies, Berlin: Springer; 2004. p. 151-171.

[10] Ingenerf J, Linder R. Ontological Principles Applied to Biomedical Vocabularies. FCTC 2006 - Workshop on Foundations of Clinical Terminologies and Classifications; 2006 Apr 8: Timioara, Romania.

[11] Schulz S, Hahn U. Part-Whole Representation and Reasoning in Biomedical Ontologies, Artificial Intelligence in Medicine. 2005: 34(3), p. 179-200.

[12] Michael J, Mejino JL, Rosse C. The role of definitions in biomedical concept representation. Proc AMIA Symp; 2001. p. 463-467.

[13] Smith B, Rosse C. The role of foundational relations in the alignment of biomedical ontologies. Proc MEDINFO; 2004. p. 444-448.

[14] Dolin RH, Spackman KA, Markwell D. Selective retrieval of pre- and post-coordinated SNOMED concepts. Proc. AMIA Symp; 2002. p. 210--214.

[15] Baader F, Calvanese D, McGuinness DL, Nardi D, Patel-Schneider PF, editors. The Description Logic Handbook. Theory, Implementation, and Applications, Cambridge, U.K.: Cambridge University Press; 2003.

[16] Baader F, Brandt D, Lutz C. Pushing the EL Envelope. IJCAI-05 - Proc of the 19th Intl Joint Conf on Artificial Intelligence; 2005. p. 364-369.

[17] Baader F, Lutz C, Suntisrivaraporn B. CEL–A Polynomial-time Reasoner for Life Science Ontologies. IJCAR06 - Proc of the 3rd Intl Joint Conf on Automated Reasoning; 2006. p. 287-291.

Address for correspondence

Stefan Schulz, Department of Medical Informatics, Freiburg University Hospital, Stefan-Meier-Str. 26, 79104 Freiburg, Germany, stschulz@uni-freiburg.de

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Automatic Checking of the Correctness of Clinical Guidelines in GLARE

Paolo Terenziani[b], Luca Anselma[a], Alessio Bottrighi[b], Laura Giordano[b], Stefania Montani[b]

[a] *DI, Università di Torino, Corso Svizzera 185, 10149 Torino, Italy, E-mail: anselma@di.unito.it*
[b] *DI, Univ. del Piemonte Orientale "Amedeo Avogadro", Spalto Marengo 33, 15100 Alessandria, Italy*

Abstract

Representing clinical guidelines is a very complex knowledge-representation task, requiring a lot of expertise and efforts. Nevertheless, guideline representations often contain several kinds of errors. Therefore, checking the well-formedness and correctness of a guideline representation is an important task, which can be drastically improved with the adoption of computer programs. In this paper, we discuss the advanced facilities provided by the GLARE system to assist physicians to produce correct representations of clinical guidelines.

Keywords:

artificial intelligence, clinical guidelines, syntactic and semantic correctness

Introduction

Clinical guidelines are a means for specifying the "best" clinical procedures and for standardizing them. Despite the fact that they are usually produced as a result of long-term cooperative efforts of large teams of experts, clinical guidelines representations can nevertheless contain different forms of errors and/or inconsistencies. This fact seems to us a natural consequence of the intrinsic difficulty of organizing and representing huge amount of both explicit and implicit knowledge. The dimension of guidelines usually makes it infeasible to conduct an extensive human check for correctness. On the other hand, advanced Artificial Intelligence techniques can be used to automate large parts of such a check. In the rest of the paper we describe the advanced facilities provided by GLARE (Guideline Acquisition, Representation and Execution), a domain-independent prototypical system for acquiring, representing and executing clinical guidelines, to check the correctness of guidelines being represented.

Background

In recent years, the medical community has started to recognize that computer-based systems dealing with clinical guidelines provide relevant advantages, since, for example, they can be used to support physicians in the diagnosis and treatment of diseases, or for education, critical review and evaluation aims [5]. Thus, many different approaches and projects have been developed in recent years to create domain-independent computer-assisted tools for managing clinical guidelines (see e.g., Asbru [13], EON [9], GEM [14], GLIF [10, 11], GUIDE [12], PROforma [3], and also [5, 7, 8, 16]). Besides the above-mentioned advantages, we believe that computer-based tool might also play a crucial role in assisting experts in the extremely difficult task of producing correct guideline representations. In the following, we first sketch the main features of our system, and then we show how it has been extended in order to provide several forms of automatic checks, to enhance the production of terminologically, syntactically and semantically correct representations of guidelines.

The GLARE system

GLARE (Guideline Acquisition, Representation and Execution) is a domain-independent tool for acquiring, representing and executing clinical guidelines [4, 15]. It has been built, starting from 1997, in a long-term cooperation between Dipartimento di Informatica of Università del Piemonte Orientale, Alessandria, Italy, and Azienda Ospedaliera S. Giovanni Battista, Torino, Italy, one of the largest hospitals in Italy. In the rest of this section, we sketch some of the more interesting general features of the GLARE's approach, while in the following section we focus on GLARE's advanced approach to check the correctness of guideline representations.

Representation formalism

In order to guarantee usability of the program to physicians not expert in Computer Science, in GLARE we aimed at defining a *limited set* of clear representation primitives, covering most of the relevant aspects of a guideline [15]. We distinguish between *atomic* and *composite actions* (*plans*), where *atomic actions* represent simple steps in a guideline, and *plans* represent actions which can be defined in terms of their components via the *has-part* relation. The *has-part* relation supports top-down refinement: a guideline itself can be seen as a composite action. *Control relations* establish which actions can be executed next, and in what order. We distinguish between four different control relations: *sequence, controlled, alternative* and *repetition*.

Four different types of *atomic actions* have been defined as well: *work actions* (actions to be performed at a certain step of the guideline), *query actions* (requests for information), *decisions* (selections among alternatives) and

conclusions (explicit output of a decision process). Actions are described in terms of their attributes.

Acquisition and execution tools

As in most approaches in the literature, GLARE distinguishes between the *acquisition* phase (when a guideline is introduced into the system –e.g., by a committee of expert physicians) and the *execution* phase (when a guideline is applied to a specific patient). Therefore, the system is composed by two main modules, the *acquisition tool* and the *execution tool*. The tools strictly interact with a set of databases, including the terminological database (during acquisition) and the patient database (during execution).

The acquisition tool provides a graphical interface to acquire atomic actions, *has-part* relations and control relations between the components of plans. The guideline is depicted as a graph, where each action is represented by a node (different forms and colours are used to distinguish among different types of actions), while control relations are represented by arcs. By clicking on the nodes in the graph, the user can trigger other windows to acquire the internal descriptions (attributes) of the nodes. The interface also shows the hierarchical structure of the guideline in the form of a tree, where plans can be seen as parents of their components (see figure 1).

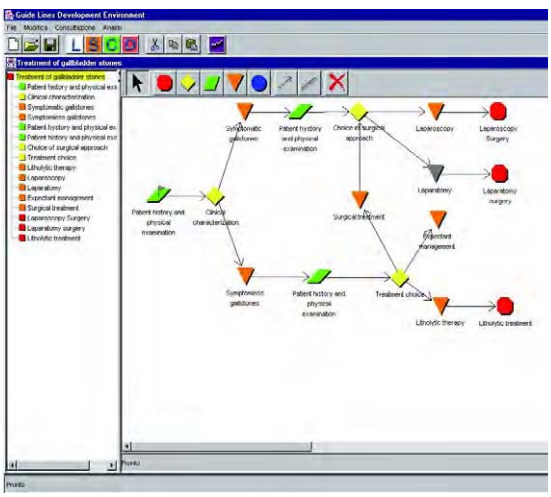


Figure 1 - A window of GLARE's acquisition tool graphical interface (concerning part of the gallbladder stones treatment guideline): on the left, the hierarchical structure of the guideline is displayed; on the right, the representation of control relations is shown in form of a graph

The execution tool is typically used "on-line": a user physician applies a guideline with reference to a specific patient. This method is used for integrating guidelines into clinical practice. Moreover, GLARE is available for "off-line" execution, i.e. for education, critical review and evaluation purposes. The execution tool also provides a decision support facility, which allows physicians navigate through the guideline to see and compare alternative paths (stemming from decision actions).

Testing

We have already tested our representation formalism and acquisition tool prototype. Several groups of expert physicians, following a few-hour training session, used GLARE to acquire algorithms concerning different clinical domains (e.g., bladder cancer, reflux esophagitis, and heart failure), with the help of a knowledge engineer. In all the tests, our representation formalism and acquisition tool proved expressive enough to cover the clinical algorithms, and the acquisition of a clinical guideline was reasonably fast (e.g., the acquisition of the guideline on heart failure, starting from a non-structured textual representation, required only 3 days).

Methods

Our acquisition module provides an "intelligent" interface to expert-physicians, in the sense that it helps them in the task of acquiring a consistent guideline. In order to achieve this goal, our acquisition module supports different types of consistency checking, which automatically operate whenever the expert-physician modifies (typically, with the addition of new nodes —actions—, arcs —control relations— or descriptions of the attributes of a node) a guideline. Other checks can be applied afterwards, when an entire guideline has been acquired, to verify its semantic consistency.

Checking terminological correctness

This first type of consistency checking is automatically triggered whenever the expert physician introduces a new term or value within the description of an action in a guideline. The acquisition module strictly interacts with the clinical vocabulary Database in order to provide expert-physicians with a standard terminology, and with a standard range of values for clinical findings. We currently support two modalities for the execution of the acquisition tool: the "safe" and the "advanced" modalities. In the "safe" mode, the expert physician is only allowed to introduce in the guideline terms/values that have already been defined within the clinical vocabulary Database. In order to make this task easier, the acquisition tool provides a means of browsing the clinical vocabulary Database, on the basis of the hierarchical organization of the data it contains. For instance, in the decision "GERD differential diagnosis" one criterion regards duration of heartburn. This datum can be found by browsing the database as follows: patient history (section) → subjective symptoms (class) → specific complaints (category) → heartburn (datum) → duration (attribute). The possible values of this attribute, as stored in the Clinical Database, are "null", "less than 3 months", "more than 3 months".

Thus, in the "safe" mode, the acquisition tool enforces all data and values to be consistent with the dictionary provided by the Clinical Database.

In the "advanced" mode is we also allow expert physicians to introduce new terms/values, that are not already contained within the Clinical Database. In such a case, a warning is shown to the expert physician. If s/he decides to

go on, s/he is directly responsible for the correctness of the new term/value.

Checking syntactic consistency

The second type of consistency checking is automatically triggered whenever the expert physician introduces a node or arc within a guideline. The acquisition module checks whether the new element being introduced is consistent with several "logical design criteria" of guidelines we want to enforce with our tool. For example, the acquisition module:

i) checks that each alternative is preceded by a decision; whenever alternative arcs exit a node, the acquisition module checks that such a node is a decisional one. This check allows for the fact that whenever alternative ways of achieving goals are considered in the guideline, the guideline also contains an explicit way of discriminating between them. This property is important especially at execution time, since, in such a way, the execution tool can provide specific support to the user physicians whenever they have to choose from alternatives;

ii) checks that decision actions are preceded by query actions specifying all the data involved in the decision criteria. Thus, at execution time, a decision action is executable only after all necessary data for that decision are available.

Checking semantic consistency: temporal constraints

In most therapies, actions have to be performed according to a set of temporal constraints concerning their relative order, their duration, and the delays between them. Additionally, in many cases, actions must be repeated at regular (i.e., periodic) times. Furthermore, it is also necessary to carefully take into account the (implicit) temporal constraints derived from the hierarchical decomposition of actions into their components and from the control-flow of actions in the guideline. Checking the consistency of such a set of implicit and explicit constraints is a very hard task, which cannot be performed manually by any expert. On the other hand, within the Artificial Intelligence community, several approaches to perform automatically the propagation of temporal constraints and to check their consistency have been developed [17].

Despite the large amount of valuable works, there still seems to be a gap between the range of phenomena covered by current AI temporal reasoning approaches and the needs arising from clinical guidelines management. In particular, in clinical guidelines,

1. qualitative and quantitative constraints, as well as repeated/periodic events need to be considered at the same time; all types of constraints may be *imprecise and/or partially defined*;
2. a structured representation of complex events (in terms of part-of relations) must be supported, to deal with structured descriptions of the domain knowledge;
3. the distinction between *classes* of actions (e.g. an action in a general guideline) and *instances* of such actions (e.g., the specific execution of an action in a guideline) has to be supported;

Obviously, the interplay between issues (1)-(3) needs to be dealt with, too. For example, the interaction between composite and periodic events might be complex to represent and manage. In fact, in the case of a composite periodic event, the temporal pattern regards the components, which may, recursively, be composite and/or periodic events. For instance, consider Ex.1.

(Ex. 1) *The therapy for multiple mieloma is made by six cycles of 5-day treatment, each one followed by a delay of 23 days (for a total time of 24 weeks). Within each cycle of 5 days, 2 inner cycles can be distinguished: the melphalan treatment, to be provided twice a day, for each of the 5 days, and the prednisone treatment, to be provided once a day, for each of the 5 days. These two treatments must be performed in parallel.*

In Ex. 1, e.g., the instances of the melphalan treatment must be repeated "twice a day, for 5 days". However, since the melphalan treatment is part of the general therapy for multiple mieloma, such a repetition pattern must be repeated for six cycles, each one followed by a delay of 23 days.

Unfortunately, no current approach in the AI and in the guideline literature proposes a comprehensive approach in which all the above phenomena can be represented, and correct, complete and tractable temporal reasoning can be performed. In GLARE, we define an approach addressing all the above-mentioned issues [1].

A complete automatic treatment of temporal constraints involves, besides the design of an expressive representation formalism, also the development of suitable temporal reasoning algorithms operating on them, to be applied both at acquisition and at execution time. However, subtle issues such as the trade-off between the expressiveness of the representation formalism and the tractability of correct and complete temporal reasoning algorithms have to be faced in order to deal with temporal constraints in a principled and well-founded way; few works in the area of computerized guidelines have deeply analyzed this topic so far.

As a starting point, we have chosen to rely as much as possible on STP (Simple Temporal Problem) [2], a well known and consolidated Artificial Intelligence approach coping with different types of temporal constraints. However, we had to extend it, in order to cope with the beforementioned additional temporal issues. Specifically, We have chosen to model the constraints regarding repeated actions into separate STPs, one for each repeated action. Thus, in our approach, the overall set of constraints between actions in the guideline is represented by a tree of STPs (STP-tree henceforth). The root of the tree (node N1 in the example in Fig. 2) is the STP which homogeneously represents the constraints (including the ones derived from the control-flow of actions in the guideline) between all the actions in the guideline (e.g., in N1, the fact that the duration of the chemotherapy is 168 days), except repeated actions. Each node in the tree is an STP, and has as many

children as the number of repeated actions it contains. Each edge in the tree connects a pair of endpoints in an STP (the starting and ending point of a repeated action) to the STP containing the constraints between its subactions, and is labeled with the list of properties describing the temporal constraints on the repetitions. For example, in Fig. 2, we show the STP-tree representing the temporal constraints involved by the example Ex. 1.

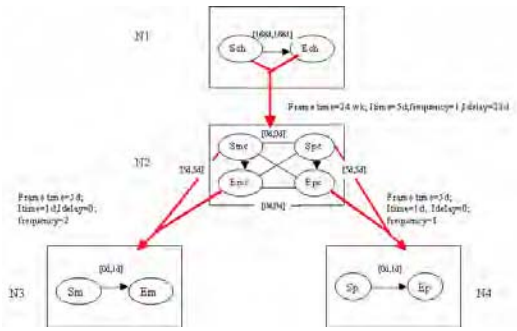


Figure 2 - STP-tree for the multiple mieloma chemotherapy guideline. Thiny lines and arcs between nodes in a STP represent bound on differences constraints. Arcs from a pair of nodes to a child STP represent repetitions. Arcs between any two nodes X and Y in a STP of the STP-tree are labeled by a pair [n,m] representing the minimum and maximum distance between X and Y

In order to check the consistency of the STP-tree, it is not sufficient to check the consistency of each node separately. In such a case, in fact, we would neglect the repetition/ periodicity information. Temporal consistency checking, thus, proceeds in a top-down fashion, starting from the root of the STP-tree. Basically, the root contains a "standard" STP, so that the Floyd-Warshall's algorithm can be applied to check its consistency (as shown in [2]). Thereafter, we proceed top down towards the leaves of the tree. For each node in the tree, we first check that the constraints on the arcs, considered alone, are consistent. If so, we then merge such constraints with the constraints in the son node, and propagate the resulting constraints to verify the joint consistence. We also formally proved the following.

Property 1. Our algorithm to check the consistency of STP-trees is *correct, complete,* and operate in *polynomial time* with respect to the number of actions in the guideline.

Property 1 grants that all and only the correct inferences are drawn by our temporal reasoning algorithm, and that such inferences are performed in tractable time.

Checking semantic "logical" consistency

Besides the property of being temporally consistent, several other semantic properties (e.g., logical consistency, safeness) should be checked on clinical guidelines. Specifically, we have identified four different classes of properties relevant in the clinical guideline context, and we have proposed a general and task-independent approach to check all of them.

i) **Properties concerning a guideline "per se"**. One can check if the guideline contains a path of actions satisfying a given set of conditions (e.g., a path including actions X, Y and Z, or a path in which no action of type X is executed, or a path nor requiring a given laboratory test, or a path requiring only a given set of resources, and so on);

ii) **Properties of a guideline in a given context.** Specific contexts of execution may impose several limitations on the executable actions of guidelines, related, e.g., to the lack of certain resources (e.g., laboratory instruments). The consequences of such limitations may be automatically investigated. For instance, one can check whether there is or not a therapy for a patient affected by a given disease, in the case a specific set of resources is available (not available).

iii) **Properties of a guideline when applied to a specific patient.** For instance, the feasibility of a given action or path of actions on the specific patient can be checked.

iv) **Integrated proofs.** Any combination of the above types of checks can be performed. For instance, one may ask whether, given a patient with a specific disease and set of symptoms, and given an hospital with a specific set of resources, there is a path in the guideline which applies to the patient and satisfies a given set of properties.

In our approach, we provide a general way of proving all the above properties by loosely coupling GLARE with the model checker SPIN [6]. Roughly speaking, we have devised a tool to map clinical guidelines acquired by the GLARE system into the Promela language, which is the language used by the SPIN model-checker. Promela allows a high level model of a distributed system to be defined by modelling each agent in an extended pseudo-C code, including synchronization primitives and message exchange primitives. Specifically, GLARE's guidelines are translated into a set of agents (e.g., the agents representing the guideline actions, the agent representing the physician executing the guideline, and so on). Once we have the translation of GLARE's guidelines in Promela code, we can use SPIN as a general-purpose engine to prove any property that can be expressed in the temporal logic LTL. In fact, SPIN translates each process (each agent) into a finite automaton, and the global behaviour of the system is obtained by computing an asynchronous interleaving product of automata. The resulting automaton represents the global state space of the system and can built on-the-fly during the verification process. The property which has to be verified on the system is passed to the verifier through an interface, which maps it into a temporal formula, as required by SPIN. SPIN converts the negation of the temporal formula into a Büchi automaton and computes its synchronous product with the system global state space. If the language of the resulting Büchi automaton is empty then the property is true on all the possible execution of the system, otherwise the verifier provides a

counterexample for the property (an execution path on which it is false).

In such a way, we provide a general-purpose approach to automatically check all the types of properties discussed above.

Results

GLARE provides a set of facilities to help physician during the acquisition of clinical guidelines, and to verify a-posteriori the correctness of guidelines being represented. To the best of our knowledge, no other guideline system in the literature provides such a large set of facilities. GLARE's facilities have proven to be quite effective in our testing activities. In particular, in the cases in which a textually written guideline had to be entered into the GLARE systems, the adoption of GLARE facilities allowed us to detect different kinds of errors in the original guidelines. In certain cases, these errors were simply due to omissions (e.g., lack of a decision action to discriminate between alternative paths of actions). In other cases, they were due to the human impossibility to propagate constraints (and, in particular, temporal constraints) along long paths of actions. In all cases the physician experts agreed that the errors detected by the GLARE systems were "genuine" errors in the original guidelines, and agreed to correct them.

Acknowledgments

We acknowledge Prof Gianpaolo Molino and Dr. Mauro Torchio of Azienda Ospedaliera S. Giovanni Battista, Turin, Italy for their cooperation in defining and testing the GLARE system. The research reported in this paper has been partially supported by a grant from Koine Sistemi, Torino, and by PRIN'06.

References

[1] Anselma L Terenziani P, Montani S, and Bottrighi A. Towards a Comprehensive Treatment of Repetitions, Periodicity and Temporal Constraints in Clinical Guidelines. Artificial Intelligence in Medicine 2006; 38(2)*:* 171-195.

[2] Dechter R, Meiri I, and Pearl J. Temporal Constraint Networks, Artificial Intelligence 1991; 49: 61-95.

[3] Fox J, Johns N, Rahmanzadeh A, and Thomson R. Disseminating medical knowledge: the PROforma approach, AI in Medicine 1998; 14: 157-181.

[4] Giordano L, Terenziani P, Bottrighi A, Montani S, and Donzella L. Model Checking for Clinical Guidelinesç an Agent-Based Approach, Proc. AMIA 2006, Washington D.C.; November 2006. p.289-293.

[5] Gordon C and Christensen JP, eds., Health Telematics for Clinical Guidelines and Protocols (IOS Press, Amsterdam); 1995.

[6] Holzmann GJ. The SPIN Model Checker. Primer and Reference Manual. Addison-Wesley; 2003.

[7] JAMIA, Focus on Clinical Guidelines and Patient Preferences, JAMIA 1998; 15(3).

[8] Special Issue on Workflow Management and Clinical Guidelines, D.B. Fridsma (Guest ed.), JAMIA 2001; 22(1):1-80.

[9] Musen MA, Tu SW, Das AK, and Shahar Y. EON: A component-based approach to automation of protocol-directed therapy. Journal of the American Medical Information Association 1996; 3(6): 367-388.

[10] Ohno-Machado L, Gennari JH, Murphy S, Jain NL, Tu SW, Oliver DE, et al. The GuideLine Interchange Format: A Model for Representing Guidelines, JAMIA 1998; 5(4):357-372.

[11] Peleg M, Boxawala AA, et al.. GLIF3: The evolution of a Guideline Representation Format, in: Proc. AMIA Annual Symposium; 2000.

[12] Quaglini S, Stefanelli M, Cavallini A, Miceli G,. Fassino C, and Mossa C. Guideline-based careflow systems, Artificial Intelligence in Medicine 2000; 20(1):5-22.

[13] Shahar Y, Mirksch S, Johnson P. The Asgaard Project: a Task-Specific Framework for the Application and Critiquing of Time-Oriented Clinical Guidelines, Artificial Intelligence in Medicine, 1998; 14:29-51.

[14] Shiffman RN, Karras BT, Agrawal A, Chen R,. Menco L, and Nath S. GEM: a proposal for a more comprehensive guideline document model using XML, JAMIA 2000; 7(5):488-498.

[15] Terenziani P, Molino G, and Torchio M. A Modular Approach for Representing and Executing Clinical Guidelines. Artificial Intelligence in Medicine 2001;23: 249-276.

[16] Tu SW and Musen MA. A Flexible Approach to Guideline Modeling, in: Proc. AMIA'99; 1999. p. 420-4.

[17] Vila L. A Survey on Temporal Reasoning in Artificial Intelligence, AI Communications 1994; 7(1):4-28.

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Knowledge Zone: A Public Repository of Peer-Reviewed Biomedical Ontologies

Kaustubh Supekar, Daniel Rubin, Natasha Noy, Mark Musen

Stanford Medical Informatics, Stanford University School of Medicine

Abstract

Reuse of ontologies is important for achieving better interoperability among health systems and relieving knowledge engineers from the burden of developing ontologies from scratch. Most of the work that aims to facilitate ontology reuse has focused on building ontology libraries that are simple repositories of ontologies or has led to keyword-based search tools that search among ontologies. To our knowledge, there are no operational methodologies that allow users to evaluate ontologies and to compare them in order to choose the most appropriate ontology for their task. In this paper, we present, Knowledge Zone – a Web-based portal that allows users to submit their ontologies, to associate metadata with their ontologies, to search for existing ontologies, to find ontology rankings based on user reviews, to post their own reviews, and to rate reviews

Keywords:

biomedical ontologies, ontology search, ontology ranking, knowledge management

Introduction

The medical community has seen widespread application and use of ontologies. With the ever-increasing amount of competing knowledge available for computation, workers in the medical domain who wish to reuse this computational knowledge find it difficult to assess and keep track of all available ontologies. Consequently, more often than not, institutions expend their valuable time and resources to develop their own ontologies, thus creating an assortment of ontologies of varying degree of quality that are not interoperable. In the healthcare system, this translates to the creation of solutions that are not interoperable, thereby hampering the information flow between these solutions, which is critical to the operation and long-term sustainability of the system.

There have been some efforts to develop Web-based libraries of ontologies: Table 1 provides a listing of some of the popular systems. These efforts have served varying purposes, from developing repositories specific to a knowledge representation (DAML Ontology Repository, Protégé-OWL Library) to supporting collaborative development of ontologies (Ontolingua). Most of these resources are mere listings of ontology resources and use different logical and structural organization to present the information relevant to the ontology. A survey [1] of these systems has rightly identified their key inadequacies, and also has suggested important requirements for structuring an ontology library system to enhance ontology management, adaptation, and standardization. Furthermore, with the limited search facility available in these resources, finding the desired ontology requires manual perusal of these Web resources, a cumbersome and a time-consuming process.

Table - Ontology libraries

DAML Ontology Library	http://www.daml.org/ontologies
WebOnto	http://kmi.open.ac.uk/project/Webonto
Ontolingua	http://www.ksl.stanford.edu/software/ontolingua
Protégé-OWL Library	http://protege.stanford.edu/plugins/owl/owl-library
OntoWeb	http://www.ontoWeb.org
OBO	http://obo.sourceforge.net

Researchers at the University of Maryland have developed Swoogle [2] - a Web based tool that provides a keyword based query facility to search and access ontologies. Through its basic and advanced query interfaces, users can search across a large collection of knowledge resources (1,898,651 Semantic Web documents, ~10,000 ontologies)[1] that exists in the cyberspace. This tool is noteworthy in its ability to search a large collection of ontologies, and to present the user with a list of relevant ontologies that match his query.

We argue that simply listing relevant ontologies is not sufficient for users who need to select the most appropriate ontology for their task. Users need a facility to help them evaluate ontologies and to compare them. Users need to ascertain how well an ontology or a part of an ontology covers the domain of interest, what is the maturity of ontology content, and how the content is related to stan-

1 Statistics obtained from http://swoogle.umbc.edu/ as of November 28, 2006

dard ontologies such as CYC, GO, and UMLS. Most of this information, if available, is provided by the creators of the ontology or by the institution that hosts the ontology, and is usually made available to the public through their Website. For evaluating an ontology, it is also valuable for a user to find out what were the experiences of other users with that ontology. Unfortunately, most of this important metadata is usually not included as part of the ontology content. This problem can be attributed to the limitations of the underlying knowledge representation language to specify such support information and the lack of tool support to associate metadata with an ontology.

To facilitate ontology reuse, we have developed Knowledge Zone, a resource where users can submit their ontologies, associate metadata with their ontologies and search for existing ontologies that satisfy their requirements. Unlike existing libraries, our system allows users to create and view peer reviews, ratings, experience reports and to find out rankings of an ontology. In Knowledge Zone, we have implemented an Open Rating System model [3] that uses reviews and ratings to compute a Web of Trust, which is then used to compute ranking of an ontology and to filter reviews.

Methods

Ontology submission: associating metadata with an ontology

The Knowledge Zone portal provides a Web interface (figure 1) that allows users to associate metadata with their ontologies. We have developed a comprehensive ontology of features (metadata) that characterize an ontology. Some of the categories included in the metadata ontology are:

- Domain of the ontology (using controlled terminology, when possible); informal description of the content; intended use of the ontology;
- Version number; contact and author information; supporting institutions; availability and licenses; citations and references;
- Verification tools used and development methodology;
- Naming policy; reliance on other ontologies
- Peer reviews; experience reports; usage data; ratings along different axes, such as correctness; coverage; degree of formality.

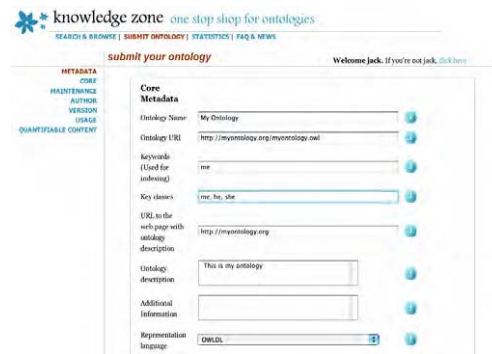


*Figure 1 - **Knowledge Zone - Ontology submission interface.** Users can enter textual information such as name and URL of the ontology, and specific information about the ontology, such as its representation language.*

These metadata features are principally categorized along two axis (1) source metadata, which include metadata provided by the ontology authors, and (2) third-party metadata, which are provided by the ontology users.

The Knowledge Zone portal employs a dynamically generated UI mechanism that leverages the semantics encoded in the Metadata ontology to provide a contextual interface based on the representation format of the ontology being submitted, level of expertise of the user, type of user (author vs. third-party user), and importance of a particular metadata element. Accordingly, the ontology submission interface is dynamically constructed to generate a Web form that will solicit the user to enter all the metadata information that has been categorized as "source metadata". This process of ontology-driven dynamic UI generation, which is used to generate every page in the Knowledge Zone portal, is depicted in figure 2. Through the dynamically generated ontology submission interface, users can enter, among other things, textual information such as the name of the author and a description of the ontology, as well as controlled information such as intended use, conceptual representation {e.g. Directed Acyclic Graph, Logic based, Frames} and the representation language {e.g. XML, Text, RDF, OWL} can be entered through the drop down boxes, which are populated from the metadata ontology.

The ontology development community in the medical domain comprises of a wide range of researchers, from those who have little or no formal training in logic, to expert knowledge engineers. To accommodate these diverse users, it is imperative to develop a solution that would not overwhelm the naïve user, while at the same time enabling capture of detailed information from the expert user. The dynamic UI generation process employed by Knowledge Zone, and the semantics encoded in our comprehensive metadata ontology allows us provide this flexible behavior in the user interface. For example, based on the axioms in the metadata ontology, an expert user who wishes to annotate description-logic ontology is presented with a different set of metadata information than would be a novice user.

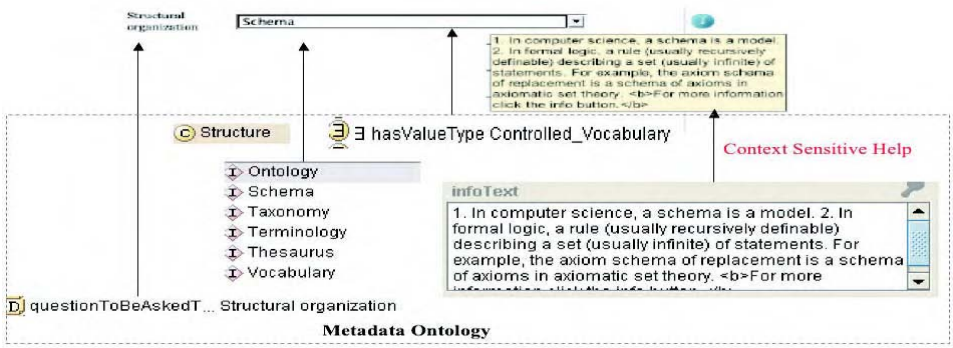


*Figure 2 - **Dynamic UI Generation in Knowledge Zone**: Process of generating a drop-down form element to capture information about a "Controlled_Vocabulary" metadata element - structural organization of ontology is shown. The label of the form element, tool tip text, and the drop-down box elements, is retrieved from the "questionToBeAskedToAuthor", "infoText", and the "Structure" properties associated with the "Structural_Organization" class in the metadata ontology.*

The metadata ontology employed in Knowledge Zone is comprehensive, and has undergone several revisions through a rigorous ontology development process. However, the inclusion and exclusion of metadata features from our ontology could, potentially, evolve; similarly, users might have different opinions about the features attributed to the novice users, controlled terms for a metadata feature, and so on. To accommodate these diverse user opinions and to ensure wide usability of Knowledge Zone, we allow the user to extend the metadata ontology. Currently, these extensions can be made on one or more of the following sections of our ontology: Ontology features (metadata), Controlled Vocabulary, and Tool Tip.

If users extend the metadata ontology, Knowledge Zone can be configured to use their extended ontology (by simply specifying its location). The metadata information that we collect is stored as instances of the metadata ontology.

Ontology search

Most of the current ontology repositories provide a simple keyword search facility. Even though keyword searches are preferred by many users, and work partially for searching Web pages, they are not desirable for searching ontologies, which have an inherent structure. Also, keyword-based searches suffer from poor precision. For example, a search for "anatomy" in Swoogle yields 59 "hits". It is then up to the user to scour through these XML files to look for the ontology that suits his requirements. Also, keyword-based searches are likely to miss out on ontologies that do not have those keywords. For example, the GALEN ontology, one of the popular anatomy ontologies, which, even though crawled by Swoogle, does not appear in the search results when users enter "anatomy" into Swoogle.

Knowledge Zone provides not only a simple keyword-based search interface, but also a structured query interface. For example, through the structured query interface, users can query for anatomy ontologies, which have been encoded in OWL, have been used for data integration, have a GNU license, and have heart, lung, and kidney among its classes. The search engine executes this query by looking up the values of the associated metadata terms across ontologies stored in our repository, and returns all the ontologies that satisfy the search criteria.

Ontology evaluation

Peer-review of ontologies

As more ontologies become available, even when users make structured and specific queries, they will obtain multiple "hits". These results will include ontologies that are developed by different institutions, and that are of varying quality. Currently, there are no operational methodologies to compare and evaluate quality of ontologies to select the best ontology for the purpose. We argue that deciding whether some ontology is "best" for a particular purpose is a purely subjective matter.

In Knowledge Zone, users can enter their subjective evaluations and reviews of an ontology. Our idea is akin to that employed by Epinions and Amazon to collect user reviews and ratings on products so that users can subjectively evaluate which "book" or "MP3 player" is best suited for their needs. This peer-review approach is well suited to ontologies, as, like "books" or "mp3 players", ontologies lack universally agreed upon criteria to denote "goodness", and thus users have to rely on other user evaluations and opinions to select a good one.

To evaluate ontologies, it is not only useful to have user evaluations of the ontology itself, but also it is particularly useful if users can comment on particular aspects of an ontology. For example, it is valuable to know what other users think about the "correctness" of a particular ontology. In Knowledge Zone, currently, users can provide numeric ratings and a free text review along the following dimensions: Syntactic Correctness of the ontology, Maturity of the ontology content, Expressibility, Semantic Consistency, Degree of Formality, Availability of documentation, and Usability of the Ontology.

Open rating systems and computing ontology rank

A user review of ontologies is a time consuming process and to have its maximum utility it is desirable to keep the review process "open". However, opening the review process to everyone causes the problem of trust. Furthermore, users have varying degree of expertise: biologists are more likely to comment better on the "domain coverage" of the ontology compared to a logician, who is likely to review better on the "semantic correctness" of the ontology. Consequently, we are bound to get some reviews that are of poor quality. To have a "usable" system, it is imperative to be able to filter them.

In Knowledge Zone, we have used the Open Rating System model [3] to accomplish this and to be able to rank ontologies. Users of our system can not only rate the ontology content, but also can rate the reviews as well as the reviewers. We use these reviews and "trust" statements made by the users to compute a "Web of Trust" (WOT), which is then used to compute rank of ontologies. The Web of Trust model [3] has been successfully used and implemented in systems such as Epinions and Amazon to rank products and to filter reviews.

The Web of Trust model as adapted and implemented in Knowledge Zone has six major components and can be summarized as follows:

- Set of Ontologies O: {O1, O2…On) that are being rated.
- Set of Users U: {U1, U2…Un} that either participate in reviewing the Ontology or providing ratings on other users.
- Set of possible ratings on an Ontology D: {0,1,2,3,4,5}
- Set of possible ratings by a user on another user T: {useful (positive), not-useful (negative)}
- Function that stores the ontology ratings provided by the user R: $O \times U \rightarrow D$
- Function that stores the ratings of users provided by other users: W: $U \times U \rightarrow T$

As mentioned earlier, a structured query by a user might return more than one ontology. We define "ranking" as a problem of being able to rank these ontologies in the descending order of their importance for a particular user. This is equivalent to finding top N ontologies, given all the six major components of our Web of Trust model. Similarly, reviews can be filtered by finding the top N reviews that are relevant to the user.

Given the WOT model; there are many approaches [4] to compute the top N objects. In our case, due to the relatively small size of the user set as compared to that available for systems such as Amazon and Epinions, we envision that functions R and W would be relatively sparse, consequently, most of the users will not have a WOT. To tackle this issue, in Knowledge Zone, we have implemented a TrustRank [5], an approach that is well suited to scenarios where most of the people do not have a WOT. However, TrustRank is limited by its applicability to compute rankings based on reviews and ratings on ontology as a whole. To accommodate for reviews that are collected on specific aspects of an ontology (for example: maturity, correctness) we augmented the TrustRank with a topic-specific trust model. This augmented approach is used to compute ontology rankings and to rank reviews associated with an ontology. Space limitations prohibit us from providing details of this approach; interested reader is encouraged to read the supplementary paper [6].

Results and implementation

Knowledge Zone is a java-based Web portal that is hosted on an Apache Tomcat server, and is publicly available. The portal uses AJAX, DHTML, and XSLT for dynamic UI generation, and MySQL database to store the user information. User reviews and ontology annotations are stored as instances of the Metadata Ontology in OWL format; Protégé-OWL API is used to access, query, and retrieve and store these instances.

From its inception, Knowledge Zone has consistently attracted a sizeable number of hits (Total hits: 12,456) from the user community, with a large portion of it being unique hits. At the time of this writing, Knowledge Zone is host to a total of twenty eight ontologies, with some notable submissions like BioPAX, Foundational Model of Anatomy, OBO Relationship Ontology and GALEN, and a total of eleven user reviews. Most of the current ontologies in our repository cover different domains. Finding the ontology rank in that case is trivial because a keyword-search or a structured query will at the most return one or two ontologies as the search result. For e.g. a structured query made in Knowledge Zone that represents the user query "Find all the anatomy ontologies" returns two ontologies – the Foundational model of Anatomy and GALEN. The number of ontologies and the number of reviews in Knowledge Zone are a limiting factor in doing a real-world evaluation of our Web of trust model to rank ontologies. However, the Web of trust model, implemented in Knowledge Zone has been successfully tested on non-ontological data, which in terms of our model has similar properties compared to the ontological data.

In future, we envision Knowledge Zone to be host to a sizeable number of ontologies and their reviews. To facilitate this process, we investigated factors that may hamper the growth of Knowledge Zone as a system. From informal interviews with the curators of biomedical ontologies, we postulate that the following may be the rate-limiting factors,

Associating metadata with ontologies is a time consuming process. To tackle this issue we are working on methods to automatically compute the quantifiable metadata. Another approach is to hire curators to enter information about ontologies on behalf of the ontology authors, and get it verified from them. We have used this approach to seed Knowledge Zone with ontologies from the Protégé-OWL library. For these ontologies we obtained metadata information through manual perusal of relevant publications and Web sites. Even though we were able to capture most of the metadata information, such an approach is not scalable, and we would miss out on critical

mass of ontologies, which we are not aware-of or are not hosted on one of the existing ontology repositories. We found that there are approximately 500 ontologies in the popular ontology libraries (see Table 1); on the other hand, the number of ontologies crawled by Swoogle is 10K.

Social issues. There are a number of social issues involved, predominantly in the biomedical domain, which hampers information sharing among researchers. These issues are widely studied and solutions involve making the user aware of the benefits of sharing information, ontologies in our case.

Peer-pressure. Currently, the biomedical ontology development community is small and comprises of closely knitted group of researchers. Our approach results in exposing their work (ontologies) for critiquing from colleagues, an approach to which they are receptive to in a private setting such as through email conversations rather than in a public forum. Having anonymous ontology submissions and reviews could possibly alleviate this concern. However, having large number of anonymous users would need substantial modifications to the current Web of Trust model.

In order to overcome these issues, work is in progress to automatically compute the metadata information associated with an ontology, and in future we plan to disseminate information in the ontology community about incentives of sharing ontologies.

Discussion and conclusion

We have developed Knowledge Zone – a Web-based portal that allows users to submit their ontologies, associate metadata with their ontologies, to search for existing ontologies, to find out their rankings based on user reviews, to post their own reviews, and to rate reviews. Our hypothesis is that having a substantial number of ontologies and a large number of reviews provided by the user community would support the user in selecting the suitable ontology for his purpose, which, consequently, would facilitate ontology reuse.

The infrastructure developed as part of Knowledge Zone: metadata ontology, ontology search and indexing mechanisms, and peer-review approach for ontology evaluation, is currently being ported for similar purposes in the Bioportal application [7], which is maintained and developed by the National Center for Biomedical Ontology. With the migration of Knowledge Zone functionality and implementation to Bioportal, planned activity by our Center to disseminate information in the ontology community about "incentives" of sharing ontologies, and current efforts to seed the Bioportal with all the ontologies from the Open Biomedical Ontology (OBO) repository, we believe that there would an increase in the number of ontology submissions, and consequently number of reviews collected in, thereby increasing the utility of our system to facilitate ontology reuse and consequently enabling the building of sustainable, interoperable health systems.

Acknowledgments

This work was supported by the National Center for Biomedical Ontology, under roadmap-initiative grant U54 HG004028 from the National Institutes of Health.

References

[1] Ding Y, Fensel D. Ontology library systems: The key to successful ontology reuse. SWWS. (2001)
[2] Ding L, Finin T, Joshi A. Swoogle: A search and metadata engine for the semantic Web. CIKM (2004).
[3] Guha R. Open rating systems. http://citeseer.ist.psu.edu/694373.html. (2003)
[4] Noy N, Guha R, Musen M.User ratings of ontologies: Who will rate the raters? AAAI Spring Symp (2005)
[5] Haveliwala T. Topic-sensitive pagerank. WWW (2002)
[6] Lewen H, Supekar K, Noy N, Musen M. Topic Specific and Open Rating Systems: An Approach for Ontology Evaluation. EON (2006).
[7] Bioportal. http://www.bioontology.org/resources-bioportal.html

Address for correspondence

ksupekar@stanford.edu

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

PICO Linguist and BabelMeSH: Development and Partial Evaluation of Evidence-based Multilanguage Search Tools for MEDLINE/PubMed

Paul Fontelo[1], Fang Liu[1], Sergio Leon[1], Abrahamane Anne[2], Michael Ackerman[1]

[1]High Performance Computing and Communications, Lister Hill National Center for Biomedical Communications
[2]National Library of Medicine Associate Fellow
National Library of Medicine, Bethesda, MD 20894

Abstract

PICO Linguist and BabelMeSH are multilanguage search tools intended for users whose native language is not English. A database of medical terms was created using concept identification equivalents of English terms to other languages. The primary sources of vocabularies were UMLS, MeSH, WHO EMRO and UMLF. The search interface changes according to the language selected which allows search terms to be entered in the native language. The user can limit the search output according to the language of publication but citations retrieved are in English only. Links may be provided to journals if published online. Evaluation of the French and Spanish versions using journal key words and a list of common diseases showed 77.5% and 86.5% accuracy respectively. User feedback was positive. PICO Linguist and BabelMeSH could be useful and convenient tools in finding current evidence sources in the medical literature especially for non-English medical terms that may be difficult to express in English.

Keywords:

evidence-based medicine, multilanguage search, MeSH, MEDLINE/PubMed, PICO, UMLS

Introduction

*ask*MEDLINE was developed as a free-text search tool for MEDLINE/PubMed.[1] We noticed that about 8% of queries early in its development were non-English terms. This is not unexpected because when searching the medical literature, it is often easier to use the language in which one is most accustomed to.[2] This observation motivated us to design search tools for MEDLINE/PubMed. PICO Linguist and BabelMeSH are the results of these efforts. Using their own native language could help find certain concepts that are difficult to translate into English. MeSH translations could be useful to healthcare providers and researchers who are not too comfortable with English in locating journal articles in MEDLINE/PubMed.

The PICO (Patient, Intervention, Comparison, Outcome) format is a method of searching medical literature that promotes the use of a focused, structured question. This search strategy leads to more precise searches[3] and facilitates the practice of evidence-based medicine using the well-built clinical question.[4] There are reports[5,6] in the literature of cross-language or multi-language projects for medical information retrieval but not many active Web sites. We propose PICO Linguist and BabelMeSH, concept-based search tools, for healthcare personnel and researchers who are not too familiar with English.

Materials and methods

BabelMeSH is designed as a transparent multilanguage and cross-language interface. Users can submit medical terms in their native language (currently, Arabic, Chinese, English, French, German, Italian, Japanese, Portuguese, Russian and Spanish) then a parser translates the query into English using a multi-language MySQL database. BabelMeSH then sends the query to PubMed through E-Utilities and returns English citations to the user.

PICO Linguist includes all the features that BabelMeSH have but instead of a single input box, PICO Linguist provides users with the structured PICO format consisting of Patient/Problem (P), Intervention (I), Comparison (C) and Outcome (O) input forms. The interface changes according to the input language selected.

Multi-language database

The major source for most translation records in the databases is the UMLS Metathesaurus, which contains MeSH translations in French, German, Italian, Japanese, Portuguese, Russian and Spanish. Permissions were obtained from the contributing organizations. In addition, Dr. Stephan Darmoni and Dr. Patrick Ruch provided French MeSH translations and the unified medical lexicon for French (UMLF). Dr. Najeeb Al-Shorbaji, Regional Office for the Eastern Mediterranean, World Health Organization, kindly provided Arabic translations of MeSH and the Unified Medical Dictionary. Chinese terms were collected from multiple open source web sites.

The concept unique identifier (CUI) and its concept source were used to find translations from UMLS. Briefly, all the concepts in one non-English language and corresponding CUIs in UMLS were identified. If the English concept in UMLS from MeSH links to the same CUI, the English concept and the foreign language concept were paired.

Otherwise, the English concept from another vocabulary was selected.

A hierarchal scheme was devised in order to create separate translation tables for each foreign language using MySQL. The priority of source vocabulary from high to low in our system is: MeSH, UMLS-Meta, SNOMED and any other vocabulary. Table 1 illustrates the structure of each table, which includes CUI, English term, non-English term, and accented non-English term (if available).

Table 1 – Translation table structure (French)

CUI	English	French	Accented French
C0702166	Acne	Acne	Acné
C0019686	HIV Antigens	Antigenes VIH	Antig_nes VIH

For translation pairs that are not from UMLS where the foreign term and English term do not map to the same CUI, an internally generated concept ID is created by the system. An example is shown in Table 2 (P-45-52.)

Character-based language processing

Character-based languages require special handling for string processing in Web applications and in the database. First, we unified character settings by encoding in UTF-8 (Unicode). The default character set in MySQL configuration file is UTF-8. Next, we installed the Multi-byte String Extension package for PHP because each letter or character in these languages may be represented by more than one byte in storage. Multi-byte String Extension can perform string processing for all of these languages. The character setting for internal encoding, HTML input and output in this extension is also in UTF-8.

Parsing algorithm

The parser identifies terms found in the translation database and deletes all others not in the database. The corresponding English translations are sent to PubMed through E-utilities. For terms entered in the search form with accents, the parser searches the accented terms column in the database (Table 1.) If no matches are found, it will transform the accented input to the unaccented form, then search the database again and return the unaccented term. If no exact matches for some terms are found, BabelMeSH will suggest records that contain part of the non-English query. The user may modify the suggestions before sending to PubMed.

An algorithm was developed for complex searches such as combined MeSH terms or concepts. In order to compel the parser to find the optimal translation, the algorithm browses the input by two pointers, recursively (Figure 1.) First, translations of PubMed "stop words" are ignored. A multi-word input is then split into an array where each array element stores an input word. Two pointers (Pointer 1 and Pointer 2) are inserted at the beginning and the end of the array. Comparing them with the database identifies individual elements between the two pointers in the array. If no match is found, Pointer 2 moves towards the front of the array one element at a time, until a match is found, or Pointer 1 meets Pointer 2. Once a translation match for the multi-word term or phrase is found, these elements are removed from the array and Pointer 1 will move towards the end to the next unmatched element in the array. If two pointers meet before the translation processing finishes, Pointer 2 will reposition to the end of the array, and Pointer 1 will again move to the end, one element at a time. The translation search will repeat until Pointer 1 and 2 meets at the end of the array. Table 2 shows the records in the database related to the search illustrated in Figure 1.

Table 2 – Records involved for translating the example query

CUI	English	Spanish
C0004916	bed	cama
C0231290	after	despues
C0442022	lumbar	lumbar
C0033119	puncture wound	puncion
C0553794	spinal tap	puncion lumbar
C0004910	bed rest	reposo en cama
P-42-52	rest	reposo

User interface

An auto-complete feature (Figure 2, in Chinese) using XMLHttp object request JavaScript, provides suggestions as the query is entered. Suggestions can be selected from the drop down list.

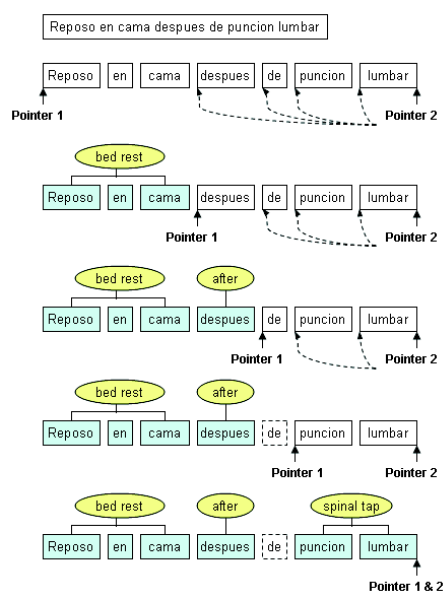


Figure 1 – How the parser works

Figure 2 – BabelMeSH (Chinese) with autosuggestion

In PICO Linguist, the embedded JavaScript code makes the Web interface change dynamically with the input language choices, shown in Figure 3. The JavaScript code changes the text orientation to right to left when Arabic is selected (Figure 4.)

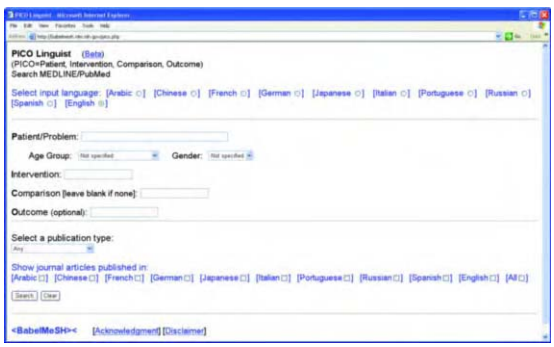


Figure 3 – PICO Linguist English search page

Figure 4 – PICO Linguist interface in Arabic

Citation result filter

The user can select to view the titles of citations by language of publication. An output filter can display the results in one or more publication languages. The default is for articles published in all languages in PubMed journals. If a filter of publication language other than English is selected, the title of citation results will be shown in both English and the original language. However, this feature will only work for Latin languages. PubMed does not record original titles in character-based language, such as Arabic, Chinese, Japanese, and Russian.

BabelMeSH and PICO Linguist transmit the translated query to PubMed through E-Utilities. Retrieved citations are shown in English only because citations of foreign language journals indexed in MEDLINE are in English only. Links are provided to the foreign journal if published online. Journals may require subscription for full-text access. Search terms in the result's page are highlighted for the user's easy identification. Figure 5 demonstrates the result page of a query in Arabic.

Evaluation

The full list of *BMJ Clinical Evidence* systematic reviews[7] was translated into Spanish by one of the authors (SL). These terms were searched in BabelMeSH then compared to the Spanish translations. The accuracy of BabelMeSH's English translation was compared to the original list. User opinion was obtained through a Web form. The evaluation of the French version consisted of three parts: (1) comparison of the author's own translation of keywords to English in journal articles published in French and the translations of the same terms to English by BabelMeSH; (2) actual user search terms from the our Web server's log files; (3) online user feedback questionnaire using a 5-point Likert scale to evaluate user opinion on the usefulness of BabelMeSH.

Results

The search strategy in Figure 4 retrieved results shown below in Figure 5.

Figure 5 – Retrieval of a PICO Linguist search in Arabic

Table 3 shows the evaluation results of the French and Spanish versions of BabelMeSH and PICO Linguist. The two test sets used for evaluating the French version journal keywords and user query terms from server logs showed an accuracy of 75% and 79.9% respectively (mean=77.5%). Partial matches were obtained in 13.2% and 7.3% respectively. Full term matches and partial term matches equal to 88.2% and 87.2% of the test sets respectively. There were no matches in approximately 12% for both sets.

Table 3 – Evaluation of the French and Spanish versions

	French		Spanish
	Journal keywords	**User query from server log**	
Total number of terms	**174**	**179**	**221**
Exact or suggestions match	130	143	191
% Total	**75%**	**79.9%**	**86.4%**
Partial match	23	13	20
% Total	**13.2%**	**7.3%**	**9.0%**
Incorrect translation	21	23	10
% Total	**12%**	**12.8**	**4.5%**

One hundred ninety one of the 221 terms from the disease list had exact matches for the Spanish version or 86.4% accuracy. For 20 terms (9%) partial matching was found but some translations missed one or two key words that could negatively affect a search. No translations were found for 10 (4.5%) search terms. We analyzed the complexity of search term or phrase as a measure of the parsing algorithm. The results (Table 4) show that accuracy is inversely related to the number of terms used.

Of the six responses received from the French feedback form, the average ratings (5=agree, 1=disagree) for the following statements were: 1) that BabelMeSH was useful 4.3/5; 2) the overall quality of citations retrieved was excellent 4.3/5; and 3) that they would continue to use BabelMeSH 4.6/5. All stated that they had previously searched MEDLINE in English and all except one declared that they would recommend it to others.

Table 4 – Effect of search term count on translation accuracy

Number of terms	Total	Translated	% Total
1	104	85	81.7
2	150	69	46
3	67	11	16.4
4	24	2	8.3
5	6	1	1.7
6	2	0	0

Comments from the Spanish users were mostly positive. Figure 6 shows the results of the two questions asked: (1) is searching MEDLINE/PubMed in your native language useful? and (2) how would you rate the results obtained? The mean responses were 4.3 and 3.9 respectively in a 5-point Likert scale.

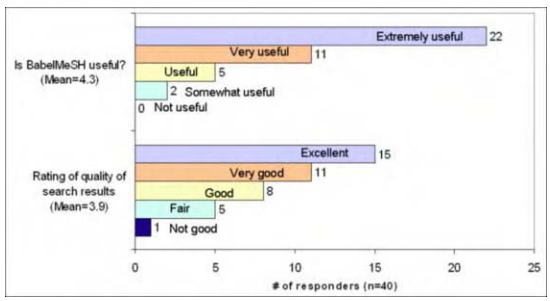


Figure 6 - User feedback on Spanish version

Discussion

The aim of this project, as in most medical translations is to bring out the same cognitive "equivalent effect"[8] that an individual might be searching for in MEDLINE/PubMed. This is challenging in one language, even more when dealing with many languages as we have attempted here. The quality of sources is also a major determinant of success, one that we have no control over.

The evaluation of PICO Linguist and BabelMeSH demonstrates the usual challenges of medical translation due to the highly technical and scientific nature of medical language. The parsing algorithm is an attempt to overcome Newmark's "transparent collocation" phenomenon[8] as well as grammar. This is quite a challenge as shown in Table 4 where translation success drastically plummets as the complexity of the search term increases.

During development, we also found that the simple change in use of singular or plural nouns adversely affected the result of a query. This was common among all language versions. This is related to the MeSH translation in other languages, some translations only include descriptors but not all the valid entries. For example, the MeSH heading, "heart diseases", has 8 valid records, such as "heart disease" and "cardiac disease". For this medical concept, our database has differences among languages: six records in French, five in German, two in Japanese and one record each in Italian and Russian. If "heart disease", instead of "heart diseases", in Italian or Russian is entered, there will be no translation result. We have done some machine-normalization work on plural and non-plural words in those foreign languages, but it could also bring errors. MeSH contains more than 22,000 descriptors, but it also has greater than 130,000 additional valid entries that can help find the appropriate MeSH heading.

Table 3 shows the results of the evaluation of the French and Spanish versions. Using two different test sets, the results in French are comparable. Contrasted with the Spanish version the accuracy of translation was higher in Spanish, although this is really not a one-on-one comparison since they use entirely different databases. This is also

likely related to the size of the database, the Spanish database contains 788,835 entries while the French only 190,330.

We were unable to compare translation accuracy with other medical translation studies. Cross-language reports such as the one done by Volk,[9] evaluated their algorithms using precision and recall from MEDLINE searches. In this project we opted to use keywords and disease list to maximize the number of terms tested. It would be difficult to use precision and recall studies. We chose user satisfaction questionnaires to evaluate real-world searches.

Usability approval was encouraging. Thirty-eight responders in Spanish said that BabelMeSH was 'useful' to 'extremely useful' while only two found it 'somewhat useful' and none responded that it was not useful. The high rating on the quality of search results directly supports the cognitive equivalent effect of translations. This would have been possible only with accurate translations. Almost all of the 40 responders submitted comments that were even more enthusiastic like, "I've been waiting for it" and "There's nothing like mother language." There were also comments that expressed a preference to be able to read the citations and abstracts in their native language. The French evaluators were equally enthusiastic about these resources.

The highly positive feedback obtained from French and Spanish users showed us that there is need for multilanguage resources and the medical community may be eager to adopt utilities like PICO Linguist and BabelMeSH in their daily activities. We suspect the same type of response from users in other languages not tested in this phase of the project. Medical translation in not a simple matter, nevertheless, it is very important to have tools available for immediate use at the point-of-care when the healthcare professional is looking for current evidence in the field.

Future challenges and developments

This resource is only as good as its database. It is a product of international collaboration and we invite collaborative research to make it even better. We also invite evaluation studies especially languages that are not covered by this report. The databases will be updated yearly as new editions of UMLS are released. Non-English translations of the abstract would be desirable. We will explore possible solution to this goal.

Conclusion

PICO Linguist (http://babelmesh.nlm.nih.gov) and BabelMeSH (http://babelmesh.nlm.nih.gov/pico.php) provide templates for multilanguage search tools including character-based languages. They can be alternative search resources for healthcare personnel searching for current evidence in the medical literature for whom English is not their primary language. We invite collaborative research especially in improving its multilanguage database and evaluating its potential.

Acknowledgments

We gratefully acknowledge colleagues[10] who provided advice, content and translation assistance for this project especially Dr. Najeeb Al-Shorbaji, EMRO, WHO.

References

[1] http://askmedline.nlm.nih.gov
[2] Nelson SJ, Schopen M, Savage AG, et al. The MeSH translation maintenance system: structure, interface design, and implementation. Medinfo. 2004;11(Pt 1):67-9.
[3] Booth A, O'Rourke A, Ford N. Structuring the preseach reference interview: a useful technique for handling clinical questions Bull Med Libr Assoc. 2000 July; 88(3): 239246.
[4] Richardson WS, Wilson MC, Nishikawa J, Hayward RS. The well-built clinical question: a key to evidence based decisions. ACP J Club 1995;123(3):A12–3.
[5] Boyer C, Baujard V, Scherrer JR. HONselect: multilingual assistant search engine operated by a concept-based interface system to decentralized heterogeneous sources. Medinfo. 2001;10(Pt 1):309-13.
[6] Hersh WR, Donohoe LC. SAPPHIRE International: a tool for cross-language information retrieval. Proc AMIA Symp 1998:673-7
[7] A full list of *BMJ Clinical Evidence* systematic reviews. http://www.clinicalevidence.org/ceweb/topics/topics-static.jsp
[8] Newmark P. A layman's view of medical translation. Br Med J. 1979 Dec 1;2(6202):1405-7.
[9] Martin Volk, Bärbel Ripplinger, Spela Vintar, Paul Buitelaar, Diana Raileanu and Bogdan Sacaleanu: Semantic Annotation for Concept-Based Cross-Language Medical Information Retrieval. In: International Journal of Medical Informatics. 67(1-3). 2002. 97-112.
[10] http://babelmesh.nlm.nih.gov/ac.php

Address for correspondence

Paul Fontelo
High Performance Computing and Communications
National Library of Medicine
Bethesda, MD 20894
fontelo@nlm.nih.gov

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Lessons Learned from Cross-Validating Alignments between Large Anatomical Ontologies

Songmao Zhang[1] and Olivier Bodenreider[2]

[1]Institute of Mathematics, Academy of Mathematics and Systems Science
Chinese Academy of Sciences, Beijing, P. R. China
[2]U.S. National Library of Medicine, NIH, Bethesda, Maryland, USA

Abstract

Objectives: To compare the alignments of two large anatomical ontologies (the Foundational Model of Anatomy and GALEN) produced by three ontology alignment systems (AOAS, FALCON and PRIOR) in the framework of the Ontology Alignment Evaluation Initiative during its 2006 campaign. Materials: Number of mappings identified by AOAS: 3,132, FALCON: 2,518 and PRIOR: 2,589. Methods: Three approaches to analyzing and comparing the results were utilized: computing the overlap among result files, manual review of some 2,000 mappings and structural validation. Conclusions: The generic systems FALCON and PRIOR identify many false positives and false negatives. With a stricter and domain-specific lexical similarity model, AOAS has a better precision, but is more sensitive to missing synonyms and misspellings.

Keywords:

anatomy, ontology alignment, collaborative evaluation, cross-validation

Introduction

A given domain is often represented by multiple ontologies, providing overlapping, yet different coverage of the domain knowledge. Anatomy, for example, is represented in ontologies such as the Foundational Model of Anatomy (FMA) and GALEN. There is a need for creating mappings among such ontologies in order to facilitate the integration of data annotated with these ontologies and reasoning across ontologies. Ontology alignment is the identification of correspondences among entities (i.e., concepts and relationships) across ontologies with overlapping content. Ontology alignment is an active field of research and many approaches to aligning ontologies have been developed in the past decade [1].

Like other research communities, such as information retrieval (TREC[1]) and information extraction (BioCreAtIvE[2]), ontology alignment researchers have set up a competitive evaluation: the Ontology Alignment Evaluation Initiative (OAEI[3]), with the goal of comparing systems and algorithms and gaining insights from the best matching strategies [2].

Evaluating ontology alignment is generally challenging. Except for ontologies of limited size (and significance), no reference alignment ("gold standard") is usually available for most ontologies, particularly in specialized domains such as anatomy. In the absence of a gold standard, the traditional framework of recall and precision cannot be used as the basis for the evaluation. Instead, the organizers used cross-validation as a surrogate. The assumption here is that mappings identified by several teams have a better chance of being valid.

At the 2006 edition of OAEI, five teams – including ours – presented the results of their alignment of the FMA and GALEN. The objective of this study is to review some of the results of the 2006 OAEI campaign for anatomy and to analyze the strengths and weaknesses of the various approaches.

Background

Anatomical ontologies

The two anatomical ontologies under investigation in the 2006 OAEI campaign are the Foundational Model of Anatomy (FMA) and GALEN.

The **Foundational Model of Anatomy**[4] (FMA) is an evolving ontology that has been under development at the University of Washington since 1994 [3]. Its objective is to conceptualize the physical objects and spaces that constitute the human body. The underlying data model for the FMA is a frame-based structure implemented with Protégé[5]. Over 70,000 concepts cover the entire range of macroscopic, microscopic and subcellular canonical anatomy. In addition to preferred terms (one per concept), some 50,000 synonyms are provided (up to 6 per concept). For example, there is a concept named *Uterine tube*, which has two synonyms: *Oviduct* and *Fallopian tube*. Because single inheritance is one of the modeling principles used in the FMA, every concept (except for the root) stands in a

1 http://trec.nist.gov/
2 http://biocreative.sourceforge.net/
3 http://oaei.ontologymatching.org/
4 http://fma.biostr.washington.edu/
5 http://protege.stanford.edu/

unique *is-a* relation to other concepts. Additionally, seven kinds of partitive relationships are used to connect anatomical concepts (e.g., *part of, constitutional part of*, *regional part of*, and their inverses *part, constitutional part, regional part*). Beside hierarchical relationships, there are 81 kinds of associative relationships between concepts in the FMA. While most of them have inverses (e.g., *branch of* and *branch*), a few do not (e.g., *input from*).

The Generalized Architecture for Languages, Encyclopedias and Nomenclatures in medicine[6] (**GALEN**) has been developed as a European Union AIM project led by the University of Manchester since 1991 [4]. The GALEN common reference model is a clinical terminology based on description logics. GALEN contains some 25,000 concepts and intends to represent the biomedical domain, of which canonical anatomy is only one part. Only one name is provided for each non-anonymous concept (e.g., *Lobe of thyroid gland*). There are over 3,000 anonymous concepts (e.g., *SolidStructure which <isPairedOrUnpaired leftRightPaired>*). GALEN supports multiple inheritance and every concept in GALEN (except for the root) stands in at least one *IS-A* relation – and often several – to other concepts. Relationships in GALEN are generally finer-grained than in the FMA. There are 41 kinds of *PART-OF* relationships (e.g., *isStructuralComponentOf, IsDivisionOf)*, and 536 associative relationships (e.g., *isBranchOf*, *isServedBy*). All relationships have inverses (e.g., *hasStructuralComponent*, *HasDivision*, *hasBranch*, *serves*).

OWL Full representation. As mentioned earlier, the FMA and GALEN were created using different knowledge representation formalisms: frames for the FMA and description logics for GALEN. In order to facilitate the alignment, the organizers converted the FMA and the anatomy subset of GALEN into OWL Full, the most expressive version of the Web Ontology Language. The resulting representation includes the class hierarchy and relations between classes for both ontologies. Additionally, concept names (including synonyms) and textual definitions for classes are represented for the FMA. The datasets provided by the organizers contain 72,560 concepts for the FMA (with 44,597 synonyms), and 9,566 concepts for GALEN (anatomy subset), of which 1,035 are anonymous.

Alignment systems

The three alignment systems analyzed in this study are the Anatomical Ontology Alignment System (AOAS), the Propagation and InfOrmation Retrieval based ontology mapping system (PRIOR) and Falcon-AO (FALCON). Two other systems participating in the 2006 OAEI campaign are not included in this review for the following reasons. Almost all mappings identified by COMA++ were specific to this system and could therefore not contribute to cross-validation. The result files contributed by IsLab were not available when this study was performed. As most alignment systems, the three systems under investigation rely on a combination of lexical and structural methods, based on the assumption that equivalent concepts across ontologies have similar names and similar relations to other concepts. A brief description of the three systems analyzed follows.

AOAS is a domain-specific ontology matching system for anatomical entities. Its lexical component compares concept names using a model of lexical resemblance developed for biomedical terms and exploits additional synonyms from an external resource: the Unified Medical Language System® (UMLS®). The presence of shared hierarchical paths among concepts across ontologies is then used as positive evidence for the mappings identified lexically. AOAS also identifies incompatible concepts, which receive negative structural evidence [5, 6].

PRIOR is a domain-independent, generic ontology matching system, based on an information retrieval approach. The features used to establish the profile of a concept include all lexical information available (concept name, label, comments, property restriction, etc.). Profile propagation is used to integrate structural information. To the profile of a concept is added, with different weights, the profiles of its ancestors, descendants and siblings. A search engine is then used to compare profiles in a vector space model [7].

FALCON is a domain-independent, generic ontology matching system. It combines three alignment methods, evaluating concept similarity based on strings (lexical similarity of concept names), "documents" (concept names and definitions treated as bags of words and compared in a vector space model) and graph structures (structural similarity based on a bipartite graph, exploiting all relations represented in the ontology for a given concept). Other features of FALCON include the partitioning of large ontologies into smaller blocks and the strategy used for combining the three mapping approaches [8].

Noticeably, both PRIOR and FALCON allow partial matches between concept names (e.g., *Adductor magnus of thigh* matches *Adductor magnus*), while only minor term variations are allowed between matches by AOAS. Unlike AOAS or FALCON, PRIOR can exploit the anonymous concepts in GALEN. And while AOAS only identifies mappings between concepts, PRIOR and FALCON also find mappings between relationships.

Table 1 - Number of mappings from the three systems

		AOAS	FALCON	PRIOR
Measure of confidence	1.0	3,029	2,115	2,583
	[.95-1.0]	0	397	0
	0.5	81	0	0
Incompatible		22	0	0
Relationships		0	6	6
Total		3,132	2,518	2,589

6 http://www.opengalen.org/

Materials

The result files for the OAEI 2006 campaign for anatomy were downloaded from the participants' web sites. The reporting format required from the organizers imposes four fields: entity1, entity2, measure (of confidence) and relation. Most mappings identified by the 3 systems are between equivalent concepts (relation: =). Incompatible mappings despite lexical similarity (negative evidence) are also reported by AOAS (relation: !=). A measure of confidence (0-1, continuous) is attached to each mapping and thresholds determined heuristically are used to select valid mappings. All mappings reported by PRIOR have a measure of 1.0, while FALCON also reported mappings with measures between .95 and 1.0. The mappings identified by AOAS are accompanied by a two-valued measure of confidence (1 when supported by positive evidence, 0.5 otherwise). The number of mappings identified by the three systems is summarized in Table 1.

Methods and results

Overlap. We first computed the intersection among the lists of mappings obtained by the three systems, thus partitioning the set of all mappings into subsets with respect to their origin, i.e., according to the system or systems that identified them (e.g., mappings identified by AOAS and FALCON, but not by PRIOR). The number of mappings with respect to their origin is summarized in Figure 1. 1,429 matches were identified by all of the three alignments, accounting for about 46%, 57% and 55% of concept matches in AOAS, FALCON and PRIOR, respectively (Figure 2). The proportion of mappings specific to one system varies largely, from 14% for FALCON to 39% for PRIOR, with 27% for AOAS.

Manual validation. Then, one of us (OB) manually reviewed for accuracy all mappings not identified by AOAS. There are several reasons for explaining our bias towards this system. Unlike the other two systems, AOAS was developed specifically for aligning anatomical ontologies. In previous work, we evaluated it against a gold standard established manually and against other systems. Recall was about .9 and most mappings identified specifically by AOAS were deemed valid [6]. The mappings were classified into the following categories: certain, possible (requires additional domain knowledge) and wrong. The objective of this cursory evaluation is primarily to quantify the false positive for FALCON and PRIOR and the false negatives for AOAS. As shown in Table 2, 1,183 of the 1,383 (86%) mappings not identified by AOAS were deemed invalid. More knowledge is required to establish the validity of half of the remaining 14%.

Structural evidence in AOAS. Another element of validation is provided by the presence of positive structural evidence, i.e., the existence of shared hierarchical paths to other tentative matches (anchors) across systems. We analyze the presence of structural evidence in the various subsets for which no manual review was performed. Overall, 97% of the lexical matches identified by AOAS were supported by positive evidence. Detailed results are reported in Table 3. Except for a larger proportion of conflicts (negative evidence) – 3.1% – in the mappings identified by AOAS and PRIOR, no major differences can be observed in the various subsets.

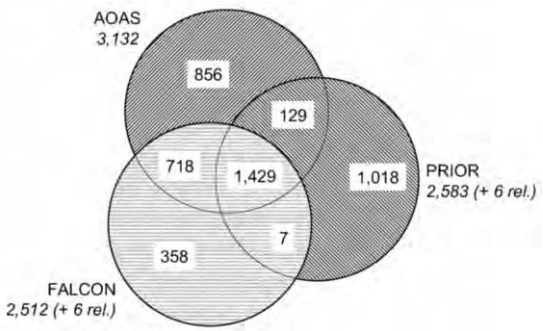


Figure 1 - Number of mappings with respect to their origin

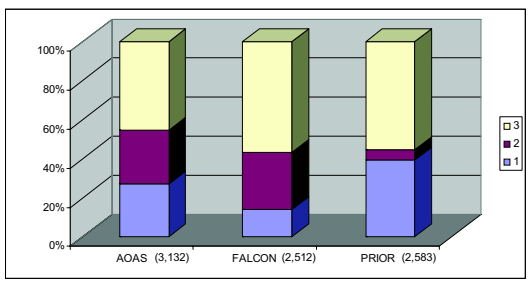


Figure 2 - Proportion of mappings identified by 1, 2 and 3 systems

Table 2 - Manual review of the mappings not identified by AOAS (excluding relations)

	Certain	Possible	Wrong	Total.
FALCON only	48	13	297	358
PRIOR only	53	80	885	1,018
FALCON + PRIOR	5	1	1	7
Total	106	94	1,183	1,383

Table 3 - Structural evidence in AOAS

	Positive	None	Negative	Total.
All 3 systems	1,382	44	3	1,429
AOAS only	819	30	7	856
AOAS + FALCON	705	5	8	718
AOAS + PRIOR	123	2	4	129
Total	3,029	81	22	3,132

Table 4 - Strengths and weaknesses of the three alignment systems

	AOAS	FALCON	PRIOR
Strengths	• Use of synonyms • Domain knowledge • Strict model of lexical resemblance • Few false positives	• Relaxed model of lexical mapping • Handles misspelling and minor term variation	• Mapping to anonymous concepts in GALEN • Information retrieval approach (more tolerant than edit distance)
Weaknesses	• Dependence on synonyms from FMA and UMLS • Very sensitive to misspelling and term variation and segmentation issues • Some false negatives	• Allows approximate matching • Many false positives • Some egregious mappings	• Allows approximate matching • Mostly false positives • No term matching • Many false negatives

Discussion

Limitations of the evaluation

The evaluation presented in this paper is partial (only the mappings not identified by AOAS were reviewed manually), cursory (some mappings were simply characterized as "possible", awaiting further review by a domain expert) and non-independent (since the person performing the review was also involved with AOAS). However, for the purpose of cross-validating three alignment systems and in conjunction with other techniques, we believe that it is appropriate in the context of this study.

Strengths and weaknesses of each system

The strengths and weaknesses of each system are summarized in Table 4 and discussed in details below.

Lexical matching constitutes an important step in ontology alignment. Systems such as PRIOR focusing on bag-of-word matching rather than term matching miss many mappings identified by the other two systems on the basis of exact matches of concept names. Examples include the match of *Arm*, the match of *Eyeball*, and the match of *Neck of mandible*.

Compared to AOAS, FALCON uses a relaxed model of lexical similarity, based on edit distance. AOAS missed some mappings due to improper segmentation of the original GALEN strings. For example, the string "SupraHyoidMuscle" was segmented at points where case changes, leading to the term *supra hyoid muscle*. However, the proper spelling for this term is *suprahyoid muscle* and the normalization algorithm used by AOAS could not match the two terms. In contrast, the relaxed approach to string matching employed by FALCON identified the two strings as a match. The analysis of the mappings identified by FALCON and not AOAS revealed about 10 segmentation issues and 15 misspellings in GALEN (e.g., "Mensicus" for "Meniscus").

Conversely, the relaxed model of lexical resemblance can lead to "egregious" mappings and therefore be extremely detrimental to the alignment. For example, FALCON identified a mapping between *Axillary artery* (in the armpit) and *Maxillary artery* (near the mandible).

Both FALCON and PRIOR allow approximate matching to happen, typically resulting in mappings where one term is more specific than the other, because one term contains a modifier while the other does not. These modifiers are often indicative of laterality (left/right) or level (in the vertebrae). For example, FALCON identified a match between *Zygomatic process of maxilla* and *Zygomatic process of left maxilla* and PRIOR a match between *Spinous process of thoracic vertebra* and *Spinous process of tenth thoracic vertebra*. While related, terms from these pairs should not be identified as equivalent. Of note, in many cases of inaccurate mapping, one term is a proper substring of the other (e.g., *Acetabulum* and *Right acetabulum*). Finally, other examples of mismatches involve not a modifier, but the head of the noun phrase or prepositional phrase, as in the mapping between *Posterior tibial nerve* and *Posterior tibial vein* (PRIOR) and between *Posterior cutaneous nerve of arm* and *Posterior cutaneous nerve of forearm* (FALCON).

AOAS is the only system to fully take advantage of synonymy for the alignment. Some synonyms are provided by the FMA, but others come from the UMLS Metathesaurus. In fact, we verified that most of the 856 mappings identified by AOAS are indeed valid and involve such synonyms. This is the case, for example, of the mapping between *Aortic orifice* and *Ostium of aorta*, and between *Shoulder joint* and *Glenohumeral joint.* In some cases, however, reliance on synonyms is an issue when the synonyms fail to be represented in the ontologies or external terminologies. For example, the mapping between *Heel of foot* and *Heel* and between *Cartilage of larynx* and *Set of cartilages of larynx* was missed by AOAS, but identified by PRIOR.

Not relying on lexical similarity, but using an information retrieval paradigm instead makes it possible for PRIOR to identify 56 matches to anonymous concepts in GALEN. Some of them are valid, including the mapping between *Artery which <serves Brain>* and *Set of arteries of brain*, and between *Bone which <IsDivisionOf (Skull <hasTopology actuallyHollowTopology>)>* and *Skull bone*. Others are not, for example, the mapping between *Kidney which <hasLeftRightSelector leftSelection>* (i.e., left kidney) and *Kidney*, or between *Potential Cavity which <locativelyContain Pus>* and *Pus*.

The more conservative and linguistically-motivated approach to lexical similarity adopted by AOAS [9] prevents a large number of false positives. As mentioned earlier, however, it is also more sensitive to misspellings and segmentation issues, as well as missing synonyms. Overall, we believe that the benefit of preventing many false positives largely outweighs the few false negatives.

We understand why generic and domain-independent systems such as FALCON and PRIOR have adopted relaxed lexical models. The resources available for biomedicine (UMLS synonyms, domain-specific model of lexical resemblance) are not available for most domains. However, pairs of long terms encountered in anatomy often differing by one qualifier (e.g., for laterality) have an artificially high similarity value when compared with edit distance or in a vector space model. Calibrating the models for a particular domain is an issue that remains to be addressed. Some mappings with a measure of confidence equal to 1 are less than perfect (e.g., between *Lamina* and *Suprachoroid lamina* in PRIOR), while near perfect matches fail to have perfect scores (e.g., between *Surface Of Calcaneum* and *Surface of calcaneus* in FALCON, with measure=0.962).

Structural validation is specific to AOAS and is designed to operate in combination with a model of lexical resemblance. In fact, structural validation is essentially used to confirm that the terms matching lexically bear some common semantics. Because our model of lexical resemblance is strict, there was no need to calibrate the structural resemblance too strictly. The minimum requirement for positive evidence is that one compatible path to another mapping be shared across ontologies. These requirements are not adapted to the validation of a more relaxed model of lexical similarity. Two terms differing solely by laterality are likely to share paths to many other mappings across systems. For this reason, it would not be sufficient to use our model of structural similarity to validate the mappings identified only by FALCON or PRIOR, which is why we performed a manual review instead.

Consequences for the OAEI Campaign for Anatomy

Evaluating the alignment of large scale, real world ontologies is an interesting, but very challenging endeavor. We showed that the absence of a reference alignment cannot be adequately compensated by the use of cross-validation. A cursory review also leaves many open questions. On the other hand, establishing a gold standard alignment would require the collaboration of domain experts and adequate funding. This study also illustrated that the conversion of the FMA and GALEN into OWL Full and, particularly, different uses of instances, classes and metaclasses by the two models tend to confuse users and impair alignment systems.

Acknowledgments

This research was supported in part by the Intramural Research Program of the National Institutes of Health (NIH), National Library of Medicine (NLM), and by the Natural Science Foundation of China (No.60496324), the National Key Research and Development Program of China (Grant No. 2002CB312004), the Knowledge Innovation Program of the Chinese Academy of Sciences, MADIS of the Chinese Academy of Sciences, and Key Laboratory of Multimedia and Intelligent Software at Beijing University of Technology.

References

[1] Noy NF. Tools for mapping and merging ontologies. In: Staab S, Studer R, editors. Handbook on Ontologies: Springer-Verlag; 2004. p. 365-384

[2] Euzenat J, Mochol M, Shvaiko P, Stuckenschmidt H, Šváb O, Svátek V, et al. First results of the Ontology Alignment Evaluation Initiative 2006. Proceedings of the Ontology Alignment Evaluation Initiative 2006 Campaign (OAEI 2006) 2006:73-90

[3] Rosse C, Mejino JL, Jr. A reference ontology for biomedical informatics: the Foundational Model of Anatomy. J Biomed Inform 2003;36(6):478-500

[4] Rector AL, Bechhofer S, Goble CA, Horrocks I, Nowlan WA, Solomon WD. The GRAIL concept modelling language for medical terminology. Artif Intell Med 1997;9(2):139-71

[5] Zhang S, Bodenreider O. NLM anatomical ontology alignment system: Results of the 2006 ontology alignment contest. Proceedings of the Ontology Alignment Evaluation Initiative 2006 Campaign (OAEI 2006) 2006:145-156

[6] Zhang S, Bodenreider O. Experience in aligning anatomical ontologies. International Journal on Semantic Web and Information Systems 2007:(in press)

[7] Mao M, Peng Y. PRIOR system: Results for OAEI 2006. Proceedings of the Ontology Alignment Evaluation Initiative 2006 Campaign (OAEI 2006) 2006:165-172

[8] Hu W, Cheng G, Zheng D, Zhong X, Qu Y. The results of Falcon-AO in the OAEI 2006 Campaign. Proceedings of the Ontology Alignment Evaluation Initiative 2006 Campaign (OAEI 2006) 2006:118-125

[9] McCray AT, Srinivasan S, Browne AC. Lexical methods for managing variation in biomedical terminologies. Proc Annu Symp Comput Appl Med Care 1994:235-9

Address for correspondence

Olivier Bodenreider, National Library of Medicine
8600 Rockville Pike, MS 3841, Bethesda, MD 20894, USA.
Email: olivier@nlm.nih.gov. Phone: (301) 435-3246.

Chapter 5.

Decision Support and Workflow

This page intentionally left blank

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

How Updating Textual Clinical Practice Guidelines Impacts Clinical Decision Support Systems: a Case Study with Bladder Cancer Management

Jacques Bouaud[a], Brigitte Séroussi[b], Ambre Brizon[a], Thibault Culty[c], France Mentré[d], Vincent Ravery[c]

[a]*AP-HP, DSI, STIM, Paris, France; INSERM, UMR_S 872, eq. 20, Paris, France*
[b]*Université Paris 6, UFR de Médecine, Paris, France; AP-HP, Hôpital Tenon, Département de Santé Publique, Paris, France*
[c]*Université Paris 7, UFR de Médecine, Paris, France; AP-HP, Hôpital Bichat, Service d'Urologie, Paris, France*
[d]*INSERM, U738, Paris, France; Université Paris 7, UFR de Médecine, Paris, France; AP-HP, Hôpital Bichat, Paris, France*

Abstract

Guideline-based clinical decision support systems (CDSSs) can be effective in increasing physician compliance with recommendations. However, the ever growing pace at which medical knowledge is produced requires that clinical practice guidelines (CPGs) be updated regularly. It is therefore mandatory that CDSSs be revised accordingly. The French Association for Urology publishes CPGs on bladder cancer management every 2 years. We studied the impact of the 2004 revision of these guidelines, with respect to the 2002 version with a CDSS, UroDoc. We proposed a typology of knowledge base modifications resulting from the update of CPGs making the difference between practice, clinical conditions and recommendations refinement as opposed to new practice and new recommendations. The number of formalized recommendations increased from 577 in 2002 to 1,081 in 2004. We evaluated the two versions of UroDoc on a randomized sample of patient records. A single new practice that modifies a decision taken in 49% of all recorded decisions leads to a fall from 67% to 46% of the compliance rate of decisions.

Keywords:

clinical decision support system, practice guideline modelling, knowledge base revision, bladder cancer

Introduction

Clinical practice guidelines (CPGs) are developed and disseminated to promote interventions of proven benefits. CPGs are produced by health professional societies or national health agencies relying on evidence-based medicine principles. They are made of structured, explicit, motivated, textual statements for managing patients according to the state of the art. However, in the "knowledge crisis" era, the amount of new medical knowledge endlessly increases [1] and the meaning of "state of the art" is a relative and evolving notion. As a consequence, CPGs are involved in a life-cycle and must be regularly updated. This process is handled by CPG providers when new research results modify current best practices. It takes at least one year to publish an updated version of CPGs.

Despite their broad dissemination, there is still considerable variation in the effectiveness of guidelines to change physicians' behavior. Many studies have shown that clinical decision support systems (CDSSs) can be effective in increasing physician compliance with recommendations [2, 3]. Defined as any software in which characteristics of individual patients are matched to a computerized knowledge base (KB) for the purpose of generating patient-specific assessments or recommendations that are presented to clinicians for consideration, CDSSs rely on formalized models of textual guidelines. Hence, as these knowledge sources are revised, it is mandatory that the delivery instrument be revised accordingly in a continuous knowledge management process. However, encoding original guideline texts into a computerized format is still an issue many authors are addressing [4, 5, 6, 7]. The management of multiple, successive, versions of a guideline is a problem CDSS developers are aware of for a long time.

In this paper, we have studied the impact of the publication of a new version of a guideline on a CDSS. We used the bladder cancer guidelines developed by the French Association for Urology ("Association Française d'Urologie" or AFU) in its revisions of years 2002 and 2004. For each textual versions, we built the corresponding KB of a CDSS, named UroDoc, and obtained 2 versions of the system: UroDoc-2002 and UroDoc-2004. After a study of both versions of textual CPGs, we first compared the two corresponding KBs. Then, we assessed how actual practices were impacted by the publication of the 2004 CPGs. We carried out a retrospective study on a randomized sample of medical records from patients hospitalized in 2004 at the Department of Urology of the Bichat-Claude Bernard Hospital, Paris, France.

Selected for best paper award.

Material and method

Textual bladder cancer guidelines

With more than 10,000 cases yearly, bladder cancer is the forth most common cancer among French men and the eighth in women. Bladder cancer represents a range of disease from relatively benign surface tumors to highly malignant life threatening carcinomas which require radical treatment. If approximately 70% of patients with newly diagnosed bladder cancer will present with superficial bladder tumors, most of them will eventually have recurrent cancer, and 10 to 20% will progress to muscle-invasive or metastatic disease. Treatment of superficial tumors is mainly based on transurethral resection (TUR). Adjuvant bladder instillation of mitomycin C (MMC) or BCG may be indicated. A strict and prolonged follow-up is mandatory to detect recurrence or progression. The standard treatment for patients with invasion remains cystectomy. However, surgical cure rates are only in the range of 10 to 30% in locally advanced bladder cancers. At present, standard therapy is still missing for numerous patient clinical profiles, hence making bladder cancer management a very active area of research.

Founded in 1896, the French Association for Urology represents and supports most of the French urologists. An expert group, the AFU Oncology Committee, publishes CPGs updated every two years to disseminate new management standards and improve healthcare quality and survival rates. Bladder cancer CPG revisions have been regularly published in 1998, 2002 and 2004. We studied bladder cancer CPGs evolution between 2002 and 2004.

Documentary-based decision support paradigm

Numerous studies have shown that the sole dissemination of textual CPGs is inefficient to modify physician behavior. On the contrary, computer-based reminders, intervening at decision time, to automatically provide patient-specific recommendations from coded patient data, seem to be the most efficient means to influence healthcare professionals in the adoption of CPGs [2, 3].

However, the success of reminder-based CDSSs is not warranted [3]. Relevance and accuracy of the advices automatically provided are often weak: Judge *et al.* [8] reported that only a little half of alerts were directly relevant to medication order when using a CDSS, Van der Sijs *et al.* [9] explained that this may lead clinicians to override 49 to 96% of the alerts. In addition, the lack of flexibility of automated approaches is often criticized: if reducing the complexity of a given patient to a set of *data* is a necessary step to feed the electronic medical record, the medical reasoning process involved in medical decision making requires a flexible and contextual interpretation of involved medical *notions*.

That is why other approaches, where guideline knowledge is structured in a way a user could retrieve patient-specific recommendations more easily than within texts, have been proposed to provide physicians with guidance [10]. The OncoDoc system, applied to breast cancer therapeutic management, has been developed according to these principles.

We developed a comparable system, UroDoc, to implement the bladder cancer CPGs of year 2002 (UroDoc-2002) and 2004 (UroDoc-2004). In each case, the KB is formally structured as a decision tree. The system need not be automatically executed from an electronic medical record. The physician can browse the KB hence controlling how each patient criterion is instantiated; she has the opportunity to assess the patient clinical states that do not need to be encoded and can be available on different supports. Thus, she identifies the path of the decision tree that corresponds to her patient and obtains in a flexible way the relevant patient-specific recommended therapy.

Modelling textual CPGs into structured decision trees

The two textual CPGs were studied in chronological order. For each CPG, the modeling work has been handled independently, *i.e.* we didn't try to build the version $d + 1$ of the CPG from its differences with version d. Following the approach to guideline implementation proposed by Shiffman *et al* [4], we first proceeded to the atomization step to extract single concepts from the recommendation's natural language text (*microscopic hematuria, grade, intravesical therapy*, etc.). Then, we carried on the deabstraction step to adjust the level of generality at which decision variables or actions were described in the CPGs to permit the operationalization in a clinical setting ("*early recurrence*" has been translated in "*recurrence at the same stage as the initial tumor or at a more advanced stage within 3 months*"). Then the disambiguation step tried to establish a single semantic interpretation for the recommendation statement (how "*biopsies may be performed on suspicious area*" should be interpreted?). The most important step was the verification of completeness to assure that both implicit decision criteria and implicit modalities of identified criteria have been explicited as decision variables and actions, and that there was a treatment (either a recommendation or a professional agreement depending on the availability of evidence) proposed for all logically possible combinations of condition states meaning that the whole set of clinical situations described by the complete expansion of the decision tree is addressed. Two different decision trees have thus been built from the two CPGs.

Comparison of the two versions of CPGs

Theoretical study

We first compared the original two textual documents from a macroscopic point of view, in terms of title, authors, general outlook, logical structure, length of comparable sections. Then we compared the two KBs built from the texts. To evaluate the evolution of CPGs between 2002 and 2004, we proposed a typology of the changes that may be found between consecutive versions of documents. CPGs can be considered as a set of recommendations $\{R_i\}_i$. Each recommendation R_i is characterized by a pair (S_i, T_i) with $S_i \rightarrow T_i$, denoting that a treatment plan T_i is recommended in the clinical situation S_i. As usual, a clinical situation is defined by a set of instantiated decision variables, $S_i =$

$\{c_{i,j}\}_j$, and a treatment plan is defined by an ordered sequence of therapeutic actions, $T_i = \{t_{i,j}\}_j$.

The set of recommendations has to be dated, thus we have to compare recommendations $\{R^d_i\}_i$ in their *d*-version to their *d+1*-version $\{R^{d+1}{}_i\}_i$. For simplicity reasons, we have dropped indices in the rest of the section. We consider that two clinical situations S^d and S^{d+1} are comparable, noted $S^d \cong S^{d+1}$, when both situations correspond to the same clinical profile in terms of therapeutic history, staging and grading of the tumor. We consider that two treatment plans T^d and T^{d+1} are comparable, noted $T^d \cong T^{d+1}$, when both treatment plans correspond to the same sequence of care in terms of therapeutic modalities (surgery, immunotherapy, chemotherapy, radiation therapy) and, in case of a surgery, when it applies to the same anatomical part. The set of all possible modifications that can occur in the update of a given CPG is illustrated by figure 1.

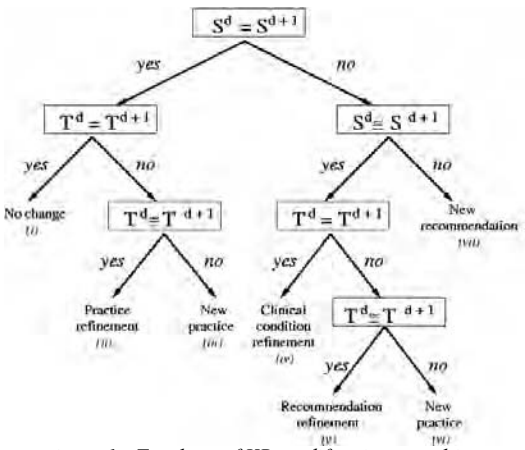


Figure 1 - Typology of KB modifications resulting from CPGs updating

Four main situations can be observed in the update process:

1. No change: an identical clinical situation lead to an identical treatment plan *(i)*.
2. Refinement of an existing recommendation. The new recommendation shares some common parts with the former one. Refinement may concern: the description of the treatment plan but not the situation *(ii)*, only the description of the clinical situation *(iv)*, or both descriptions *(v)*.
3. New practice. A totally new therapy (non comparable) appeared in an already identified clinical situation *(iii)*, or in its corresponding refined expression *(vi)*,
4. New recommendation: extending the CPG coverage, a new clinical situation is identified with its corresponding therapy leading to a new recommendation *(vii)*.

When comparing the two KBs, decision variables, single therapeutic actions, and treatment plans were analyzed literally to determine those which were common, *i.e.,* used in both KBs, those which were 2002-specific, *i.e.,* obsolete in 2004, and those which were 2004-specific, *i.e.,* new in 2004. In this first step, we used the identity relationship between objects (=). The same process was applied at a more abstract level where decision variables and single therapeutic actions were substituted by their upper concept in a kind of ontological reasoning. The relationship used for comparison was the comparability (≅). For instance, "chemo-III" and "chemo-IV" correspond to two therapeutic actions, both abstracted as "adjuvant-chemotherapy", and therefore considered comparable.

Practical analysis

We studied the impact of the publication of the new version of the bladder cancer management guideline in June 2004 upon the clinical practice of the Department of Urology of the Bichat-Claude Bernard hospital in Paris, France. We collected a random sample of patients diagnosed with bladder cancer during the year 2004. For each therapeutic decision following the diagnostic of bladder cancer, we used both systems UroDoc-2002 and UroDoc-2004 to get the recommendations of 2002-guidelines and of 2004-guideline. These two therapeutic recommendations have been compared to the treatment stored in the medical record and actually received by the patient.

Results

Comparison of textual CPGs

Both 2002 and 2004 CPGs have the same title anc coverage. They have the same number of authors (15), 80% of them participated to both CPG developments. Both guidelines are published as textual double-column documents. 2002 CPGs are shorter (13 pages) than 2004 CPGs (40 pages). General outlook is comparable (Table 1), starting with a description of the staging system (tumor, node, metastasis, TNM), followed by 3 sections describing diagnostic indications, therapeutic guidelines and follow-up protocols. However, some sections especially those describing the management of patients with positive nodes and metastatic disease are more elaborated in the 2004 version. Besides, a new sub-section specifying pathologic staging modalities has been added in the 2004 version.

Table 1 - CPG structure and volume in columns (half-pages)

	Structure	*Number of columns*	
		2002	*2004*
0.	**Introduction**	**1**	**1**
I.	**TNM staging system**	**1**	**1**
II.	**Diagnostic indications**	**1.5**	**5**
1.	*Clinical presentation*	*0.5*	*1*
2.	*Imaging tests*	*0.5*	*1.5*
3.	*Urine cytology*	*0.25*	*0.5*
4.	*Cystoscopy*	*0.25*	*1*
5.	*Pathologic staging*	*n/a*	*1*
III.	**Therapeutic guidelines**	**8**	**21.5**
1.	*Therapies*	*5*	*6*
2.	*Management principles*	*3*	*15.5*
2.1.	*Superficial tumors*	*1.5*	*2.5*
2.2.	*Invasive tumors (N0M0)*	*1*	*5*
2.3.	*Positive nodes (N+M0)*	*0.25*	*6*
2.4.	*Metastatic disease (M+)*	*0.25*	*2*

Comparison of structured KBs

Table 2 reports the formal comparison of the two KBs. It should be noticed that at the literal level, the 2004 version, although sharing common elements with the 2002 version, introduced many new ones and left apart some. Although reducing the importance of the effect, the abstraction step confirms the trend. When considering decision trees, the 2002 version identifies 366 different clinical situations whereas the 2004 version identifies 584 clinical situations. The average path length is 10 in 2002 but 11 in 2004. There are 577 recommendations in 2002 and 1,081 in 2004.

Table 2 - Comparisons of decision variables (DVs), therapeutic actions (TAs), recommendations (Recos), and of their abstracted correspondents in 2002 and 2004 KBs

	2002 specific	*Common*	*2004 specific*
DVs	8	32	23
Abstracted DVs	6	28	13
TAs	9	27	22
Abstracted TAs	4	17	5
Recos	530	47	1,034
Abstracted Recos	339	238/358	723

When comparing the two sets of recommendations, $\{R^{2002}{}_i\}_i$ and $\{R^{2004}{}_i\}_i$, we adopted two different points of view. From the 2002 point of view, each recommendation has been tested against the 2004-recommendation set to automatically determine its status: whether it remained unchanged, it was refined, it has evolved towards new practices, or it became obsolete (*i.e.,* it has no more equivalent in 2004). The same algorithm was applied for each 2004 recommendations as compared to 2002 recommendations. Table 3 reports these results and figures 2 and 3 illustrate the percentages of the different subsets.

Table 3 - Distribution of recommendation evolution status between 2002 and 2004 versions of the KB

	2002 KB	*2004 KB*
Obsolete recommendations	339	—
No change	47	47
Refinement	74	91
New practices	117	220
New recommendations	—	723
Total	577	1,081

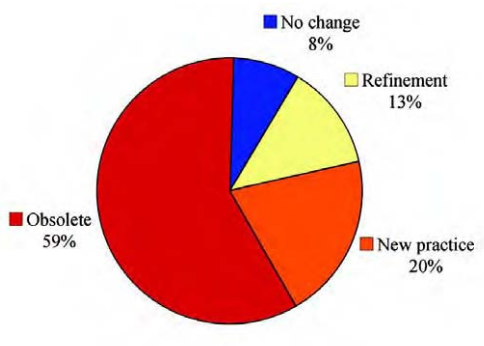


Figure 2 - Status of 2002 recommendations (n=577)

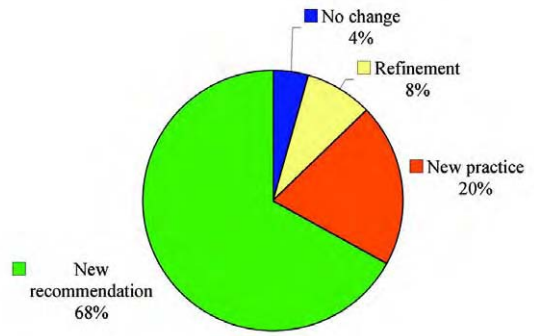


Figure 3 - Status of 2004 recommendations (n=1,081)

The overlapping of the 2004 KB on the 2002 KB is quite limited. The overlapped part (i.e., no change, refinement, new practice) is made of 238 recommendations (41%) in 2002 which projection is made of 358 recommendations (33%) in 2004. From this analysis, 339 recommendations of 2002 (59%) became obsolete in 2004 and 723 recommendations of 2004 (68%) may be considered as new.

Impact of CPG update on clinical practices

We analyzed medical records of a randomized sample of 45 patients leading to a total of 106 therapeutic decisions. Only 97 decisions, corresponding to 38 patients, were retained for the survey: 9 decisions were ruled out either because the patients were long lost (6), or decisions also involved the management of prostate cancer (3). Each therapeutic decision has been retrospectively matched to the 2002 CPGs by using UroDoc-2002 and to the 2004 CPGs by using UroDoc-2004. With respect to the 2002 CPGs, we found that 65 decisions, *i.e.,* 67% of the cases, were compliant with the CPGs. When the same therapeutic decisions were evaluated using UroDoc-2004, we found that 45 decisions, *i.e.,* 46% of the cases, complied with the 2004 CPGs. Among the non-compliant decisions, only a few cases (5 as compared to 2002 CPGs and 4 as compared to 2004 CPGs) can be explained by specific patient conditions (age) or "acceptable" physician personal choice. Table 4 synthesizes the results of the analysis.

Table 4 - Practical compliance with 2002 and 2004 CPGs

(n=97)	*# non compliant*	*# explainable*	*Compliance*
2002	32	5	67%
2004	52	4	46%

Discussion and conclusion

The 2004 evolution of bladder cancer CPGs essentially concerned the extension of CPGs coverage and thus increased the number of clinical situations for which recommendations are provided. This is particularly true for the management of patients with positive nodes and metastatic disease. This is also the case for the control of BCG side-effects which is not mentioned in 2002 while very much detailed in 2004. A particular new practice has to be

noticed: TUR which was recommended as such in 2002 has to be followed by an early post surgical endovesical instillation (EPSEI) of MMC in 2004. When analyzing the compliance of therapeutic decisions with respect to Uro-Doc-2002 and UroDoc-2004, we got the results summarized in table 5.

Table 5 - Evolution of decision compliance with 2002 and 2004 CPGs

	2002 compl.	*2002 non compl.*	*Total*
2004 compl.	41	4	45
2004 non compl.	24	28	52
Total	65	32	97

The 24 decisions that are 2002-compliant, but not 2004-compliant, are made of 21 TUR with no EPSEI, 1 decision of bladder instillation of BCG with no endoscopic evaluation, and 2 incomplete surgeries as compared to 2004 recommendations. When studying the clinical practices of the Department of Urology of the Bichat-Claude Bernard hospital, it appears that 49% of the therapeutic decisions for bladder cancer correspond to TUR decisions (figure 5). Since 2004 recommendations modified the TUR decision, this illustrates the fact that, although they were published, new 2004 CPGs were not applied in routine care. The side-effect of this delay in CPG implementation is a fall from 67 to 46% of the compliance rate with CPGs.

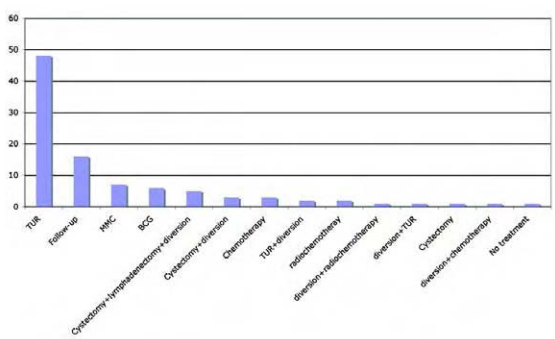


Figure 5 - Distribution of actual therapeutic decisions (n=97)

Besides, the formal comparison of KBs showed that the number of recommendations increased by a factor of 2. It is remarkable that more than half of the 2002 recommendations (339/577) have no formal equivalent in the 2004 KB since this is not consistent with CPG content. The 2004 version is supposed to cover every aspects of the 2002 version and the 339 "unmatched" recommendations should have their equivalents at the semantic level in the 723 "identified as new" 2004 recommendations. The reason is that every 2002 recommendation which contains at least one of the 8 2002-specific variables never matches any 2004 recommendation. This emphasizes the fact that in 2002 and 2004, some similar notions expressed differently have been modeled differently. As a conclusion, KB revisions following CPG updating could be improved if the new CPG version respects the terminology and point of view used in the former version. This could be achieved by explicitly using, when possible, standard terminologies to describe guideline-covered specific clinical situations in CPGs.

References

[1] Wyatt J. Use and sources of medical knowledge. Lancet 1991;338(8779):1368–73.

[2] Garg AX, Adhikari NKJ, McDonald H, Rosas-Arellano MP, Devereaux PJ, Beyenne J, et al. Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review. JAMA 2005;293(10):1223–38.

[3] Shiffman RN, Liaw Y, Brandt CA, and Corb GJ. Computer-based guideline implementation systems: a systematic review of functionality and effectiveness. JAMIA 1999;6(2):104–14.

[4] Shiffman RN, Michel G, Essaihi A, and Thornquist E. Bridging the guideline implementation gap: a systematic, document-centered approach to guideline implementation. J Am Med Inform Assoc 2004;11(5):418–26.

[5] Svatek V and Ruzicka M. Step-by-step formalisation of medical guideline content. Int J Med Inf 2003; 70(2- 3):329–35.

[6] Votruba P, Miksch S, Seyfang A, and Kosara R. Tracing the formalization steps of textual guidelines. In: Kaiser K, Miksch S, and Tu S, eds, Computer-based Support for Clinical Guidelines and Protocols. Proc Symposium on Computerized Guidelines and Protocols (CGP 2004), (vol 101) of Studies in Health Technology and Informatics, Amsterdam. IOS Press, 2004:172–6.

[7] Shahar Y, Young O, Shalom E, Galperin M, Mayaffit A, Moskovitch R, et al. A framework for a distributed, hybrid, multiple-ontology clinical-guideline library, and automated guideline-support tools. Journal of Biomedical Informatics 2004;37(5):325–44.

[8] Judge J, Field TS, DeFlorio M, Laprino J, Auger J, Rochon P, et al. Prescribers' response to alerts during medication ordering in the long term care setting. J Am Med Inform Assoc 2006;13(4):385–90.

[9] van der Sijs H, Aarts J, Vulto A, and Berg M. Overriding of drug safety alerts in computerized physician order entry. J Am Med Inform Assoc 2006;13(2):138–47.

[10] Séroussi B, Bouaud J, and Antoine EC. OncoDoc, a successful experiment of computer-supported guideline development and implementation in the treatment of breast cancer. Artif Intell Med 2001;22(1):43–64.

Address for correspondence

Jacques Bouaud
STIM/DSI/AP-HP
3, avenue Victoria, F-75004 Paris, France
jacques.bouaud@sap.aphp.fr

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Improving Compliance to Guidelines through Workflow Technology: Implementation and Results in a Stroke Unit

Silvia Panzarasa[a], Silvana Quaglini[b], Giuseppe Micieli[c], Simona Marcheselli[d], Mauro Pessina[e], Corrado Pernice[e], Anna Cavallini[d], Mario Stefanelli[b]

[a] *Consorzio di Bioingegneria e Informatica Medica, Pavia, Italy*
[b] *Dip. Informatica e Sistemistica, Università di Pavia, Italy*
[c] *Neurology and Stroke Unit, IRCCS Istituto Clinico Humanitas Rozzano (MI), Italy*
[d] *IRCCS Fondazione Mondino, Stroke Unit, Pavia, Italy*
[e] *TSD Project, Milan, Italy*

Abstract

This work describes the results of the implementation of a workflow management system integrated into the electronic clinical chart of a Stroke Unit. The workflow logic is based on the rules provided by the SPREAD guidelines for stroke management. In this way, the already existing clinical chart has been transformed into an evidence-based, real-time decision support system, meanwhile maintaining the same look the users were familiar with. Since the final aim of the work was to improve evidence-based behavior and detect possible organizational bottlenecks, non-compliance to the clinical practice guidelines, before and after the system introduction, have been analyzed, as well as the accuracy of the clinical chart compilation, some care process variables, and system usability. Results show that the system enhances the clinical practice without boring users. Moreover, non-compliance analysis gives rise to ideas for further improvement.

Keywords:

decision support systems, practice guideline, information systems, workflow management, stroke

Introduction

Insufficient or erroneous information at point-of-care, both for diagnosis and treatment, is a frequent and significant cause of medical error, as well as communication problems among healthcare operators and the different healthcare organizations, where patients' data are spread. Prevention of possible errors during a patient's management is a critical issue. Recent studies have highlighted the dramatic dimension of the problem, from the well-known report of the US Institute of Medicine [1], assessing between 44,000 and 98,000 deaths per year as due to medical errors in American hospitals, to an increasing number of country-based studies reporting similar figures.

Evidence-based clinical practice guidelines (GLs) have been widely promoted as a way of improving health outcomes. In the stroke management area, in particular, the study described in [2] demonstrates an association between stroke outcome and compliance with GLs. In current healthcare systems, however, scientific knowledge about best care is not applied systematically or expeditiously to clinical practice. GLs usually capture both literature-based and practice-based evidence into a textual format, which can be easily diffused but uneasily used in daily work. Thus there is a great effort to disseminate them in computer-interpretable representations, more suitable for individual clinical decision support [3]. Contemporary, there is increasing need of smooth integration of GLs into the existing hospital information systems.

The current workflow technology [4] seems to offer a convenient solution to build a cooperative system in which the activities of a care providers' team can be coordinated within a process properly designed on the basis of available best medical knowledge. This paper presents an approach to the design, implementation and evaluation of an evidence-based workflow management system (WfMS), from here on called Careflow management system (CfMS). On the basis of a general methodology, we describe a CfMS implementation in the area of Stroke management, particularly focusing on user interaction, presentation of GL suggestions and management of non-compliance to GL. We show how this technology, especially in fields where acute patients treatment is critical, helps in timely collecting the correct and most useful data and avoiding errors. The test-bed for the proposed methodological and technological solutions is the Stroke Unit (SU) in Pavia, where we implemented the system based on the SPREAD guidelines (Stroke Prevention and Educational Awareness Diffusion) [5].

The evaluation methodology

Outcomes of the CfMS have been tested following the steps below, described in detail in the subsequent sections:

1. analysis and update of the Computerized Clinical Chart (CCC) that physicians were using since many years in the SU, with the aim of ensuring that all the necessary data for testing the care process could be stored;

2. utilization of the updated CCC, without any decision support utility: during this phase, compliance to GLs was exclusively based on physicians' personal knowledge and on hospital organization;
3. development of the decision support tool, i.e. the CfMS, and its integration with the CCC;
4. utilization of the integrated system for 8 months (from its installation until November 2006). During this phase, compliance could have been affected by system reminders; and
5. analysis of collected data.

Updating the stroke unit data model

The Stroke Unit was already equipped with a CCC implemented some years ago, with the commercial tool Wincare® (by TSD-Projects). System-user interaction is based on so-called *Events*: an event is a single data entry form of the clinical chart, and it may be activated by users with competent role (controlled access). Before starting the project we analyzed all the events and their associated graphical user interface, with the aim of checking the presence of the data required for both properly executing the decision support system and evaluating GL compliance. As a result, the data model has been updated both by increasing the number of data input fields and by changing the nature of some existing data, mainly shifting from free text to encoded data. In fact, data-entry was mainly limited to free-text forms, because the initial purpose of the CCC was only to limit paper-based communication and to improve patients' data retrieval. The advantage was that information was ready to be printed *as-it-was* for summaries, discharge letters, etc.. Moreover, such CCC gave users the flexibility to represent patient's condition in the desired order and granularity. On the other hand, integration with a decision support system was not possible because of almost complete lack of structured data. In other words, free-text information stored in the CCC didn't allow a proper interpretation of the rules embedded in the GL: natural language processing is a hot research field, but results are not mature enough to guarantee full recognition of crucial data in critical patients' management [6].

Eventually, a deep examination of SPREAD GLs allowed determining the minimum data set required for implementing all the recommendations. These data have been encoded and, when possible, standard terminologies such as ICD9-CM for diseases, have been used. To maintain flexibility, in addition to encoded items, the interface still allows entering free text notes, useful to input additional details but not essential to decision support purposes.

Careflow implementation and integration

After the analysis and the consequent update of the data model, we implemented the CfMS, based on all the GL recommendations from SPREAD chapters 9 and 10, related to diagnosis and treatment of the stroke acute phase, with some site-specifications decided by the neurologists involved. Technology used to implement the whole system and to integrate it with Wincare® is based on DBMS Oracle™, Oracle Workflow™, Pl/SQL and Visual Basics (see [7] for details). Briefly, the CfMS manages the execution of the care process for a particular patient, reading his/her clinical data, verifying the GL rules and creating specific recommendations. These latter, translated in to-do lists, messages and alerts, are communicated to the right professional role via the end-user application.

Since the users were satisfied with their CCC, our choice was not to create an additional and specific interface, but to integrate all the needed functionalities within the existing CCC, making it more "dynamic" and "smart" thanks to the CfMS execution. Our philosophy was that users must perceive the new system just as an update, with some new functionalities. To obtain this goal, a middleware layer for the data sharing has been developed in order to keep the two systems (Wincare® and CfMS) independent, while granting communication. In particular, whenever there is a new data entry, the CCC transfers these data into a buffer database, usable by the decision support system and designed on the basis of the minimum data set required by the GL. Similarly, when the CfMS generates a suggestion, or a new to-do list, they are put into special tables of the same database, and Wincare® can read and show them through its interface.

Relevant functionality

A picture of the updated system interface is shown in Figure 1.

Facilitating data input

First of all, data relevant for GL rules interpretation appear as yellow input fields, in order to encourage a complete editing of the essential fields of the CCC.

Smart graphical user interface

With respect to the former system, the list of the CCC events has been transformed into a "to-do" list, i.e. events are now listed in a patient-specific, dynamic way, taking into account the patient medical condition and the time spent since his/ her admission. Different icons and colors are used to mean that an event is completed (icon with a green happy smiley), it is currently being done (orange thoughtful smiley), it has not been executed due to an exception (red sad smiley), it has been filled on the user initiative (Wincare® icon), or it still has to be done (no icon). Degree of the scientific evidence of recommendation is shown, as in the SPREAD book, by letters A-D. According to the physician needs, special attention has been put in avoiding over-information: particularly in acute situations, only events necessary to face the urgency are listed. For example, if a patient is potentially eligible for r-tPA treatment (a thrombolitic drug very effective but that requires a very careful administration), the list of events only refers to actions to be done in order to detect possible contraindications. Again, if CT scan shows an intracranial hemorrhages, that is one of the contraindications, "r-tPA Treatment" is switched off (from yellow to red) and less urgent events appear in the to-do list (see difference between lists (a) and (b) of Figure 1).

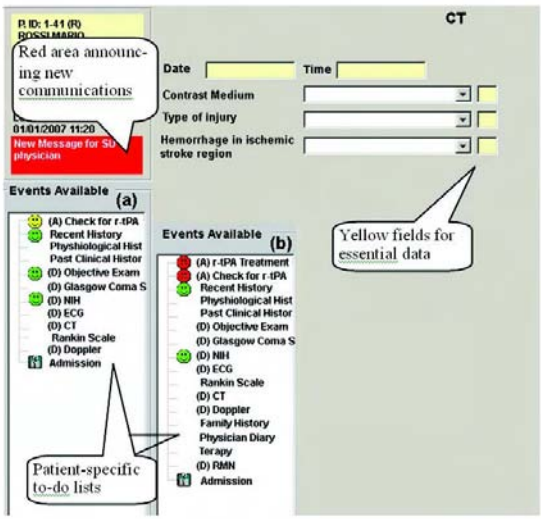


Figure 1 - A smart form of the Clinical Chart integrated with the CfMS. (a) and (b) are two different to-do lists for the same patient shown in two different points in time

Showing CfMS reminders

How to show GL reminders is a big issue in the implementation of a decision support system into a real healthcare setting. In general, pop-up windows are not well-accepted by physicians, because they are considered too intrusive. Thus, we preferred to rely upon the above described colored visualization of the to-do list: clicking the task icon, the data input form immediately appears. Moreover, a red area informs the users about new communications (red rectangle in Figure 1). This modality is minimally intrusive because users may open the communication box asynchronously, on their own will. The communication list, in chronological order, shows the priority level of each recommendation, its type (whether it is a diagnostic process, a treatment, a variable monitoring, or a message) and its positive or negative sign (if it is something to do or to rule out). Different users receive different communications according to their role (physician, nurse, physiotherapist, etc). Users may read communications for a specific patient or for all the patients currently admitted in the SU. If an answer is required, a checkbox appears near the text, allowing the receiver to say whether he/she agrees or not with the SPREAD recommendation. If a physician accepts a suggestion about a drug prescription, a record with the drug name, dosage and timing is stored and at the same time the CfMS takes care to inform the role "nurse", responsible for drug administration, sending a message with the useful details. Thus CfMS is also able to manage automatically a large amount of communication acts among personnel involved in patient care, making between-users interoperability more efficient by supporting medical and organizational knowledge sharing.

Ancillary tools

Natural language reporting

With the new system release, interpretation of GL rules is easier and safer but information is no more immediately available for nice printed free-text reports. To meet this physicians' need we developed a tool for the generation of such reports. Briefly, code descriptions are combined with associated notes (if any), according to grammatical rules, to generate the correct words and suitable introduction and conjunction sentences. The resulting reports are structured in distinct paragraphs for physiological history, recent and past clinical history, family-related history; information about past history is grouped by year and ranked in chronological order; once generated, the report may be edited by physicians before storing or printing.

Detecting and showing non-compliance

In a real-world setting, it is extremely difficult to detect non-compliance in real time. There are too many variables that could affect timely data storing and, consequently, timely automatic detection of non-compliance. This is particularly true in acute units. So, demanding that physicians justify their non-compliance has some drawbacks: first, detected non-compliance could be not real, i.e. an action has been performed but related data have not been entered in the system yet; second, users perceive the system not only a gentle reminder, but as a controller; eventually, if he's in a hurry, physician may have no time to write down justification. As a consequence, provided justification may be not reliable. For these reasons, in agreement with our clinical partners, we decided to manage non-compliance in a less invasive way. The point in time where we decided to accomplish this task is the patient discharge. This choice is based on the consideration that when physicians prepare patient's discharge, they are in team and they are not in a hurry. In addition, compiling the discharge letter requires summarizing the patient care process, and reasoning about it. Thus, presenting the non-compliance list at that moment has two advantages: overcoming the previous issues and facilitating bethinking about the case. In practice, at the patient discharge, a set of queries is activated on the patient record, and a report is produced. It shows the non-compliance list detected automatically, starting from the CCC, and gives the physicians the possibility to provide motivations that can be useful for further revision of the SPREAD GL (that is regularly updated every 2 years).

Results

The CfMS, after an intensive testing in our university laboratory, has been implemented in the SU (semi-intensive six beds unit) on April 2006. Three physician's and two nurse's workstations have been installed. Since the CfMS did not introduce dramatic changes in the end-user application, it has been accepted by healthcare personnel and entered the daily medical routine. We administered the Likert Scale SUS (Software Usability Scale [8]) to the five users of the Stroke Unit, obtaining an average value of 82.5/100, ranging from 80 to 85.

In order to evaluate the impact of the system in the daily activity, a set of queries has been developed on CfMS data, for checking both data completeness and GL compliance. Reports are produced by analyzing all medical data available for ischemic stroke patients admitted within a user-defined period. Reports contain statistics about:

- encoded data input: evaluation of the completeness of the minimum data set required by the GL;
- GL compliance: evaluation of the adherence to all the recommendations of the acute stroke phase;
- additional process variables, useful to capture the system utilization rate and modality, system intrusiveness and additional information about the clinical routine.

We compared the care processes of ischemic stroke patients admitted from April to November 2006 (113 patients managed with the help of the CfMS) with those of the 141 patients admitted in the same period of 2005, when the CCC was the same, but without decision support functionalities. Homogeneous periods avoid seasonal biases.

Table 1 - Encoded data input completeness

Event	April-November 2005	April-November 2006
Personal data	97%	98%
Physiological history	96%	97%
Recent history	50%	65%
Past history	94%	97%
Neuroradiology diagnosis	1%	39%
Neurosonologic exam	0%	0%
Cardiologic diagnosis	2%	35%
Diagnostic hypothesis	0%	35%
Objective examination	46%	56%
NIH stroke scale	76%	85%
Hematochemical exam	97%	97%
Rankin scale	35%	35%

Encoded data input

Completeness of encoded data input is increasing, since the CfMS introduction, for most of the events useful for GL interpretation (Table 1). The apparently astonishing result for neurosonologic examination highlights a double-input problem: this test is executed outside the SU and its result is automatically stored in the CCC as free text. This observation led to an update of the interface, which now shows, at the user's request, the free text, and prompts for encoding. Of course the best solution should be an update of the radiology department information system. In fact, the possibility of the CfMS of *identifying* and *quantifying* organizational problems is very important because it represents a sound starting point for further refinements of the hospital information system.

Guideline compliance

Of course completeness of encoded data affects the possibility of automatically evaluating compliance. Thus, due to poor completeness in the past, only few recommendations could be compared. In fact in many cases we have not enough encoded data to verify the recommendation and for other recommendations we could recognize too few eligible patients. Fortunately, sufficient data exist to compare compliance for three important recommendations: the SPREAD 10.5 (ASA in acute phase, grade A), 10.18 (Deep Venous Thrombosis prevention, grade B) and 10.7 (anticoagulant treatment with heparin, acute phase, grade D). Table 2 shows the number of eligible patients and the percentage of compliance for both periods, 2005 and 2006. For recommendation 10.7 the percentage of compliance remains the same and in both cases there is a total adherence to the SPREAD GL. There is a significant increase in compliance for recommendations 10.5 and 10.18.

Table 2 - GL compliance

SPREAD Recommendation	April-November 2005		April-November 2006	
	Eligible patients	% compliance	Eligible patients	% compliance
R 10.18	21	57%	21	86%
R 10.5	93	67%	48	71%
10.7	29	100%	8	100%

Additional process variables

Adoption of a WfMS has the main objective of improving efficiency and effectiveness of the organization processes; however, as a by-product, it also greatly facilitates obtaining interesting statistics on the system performance and on process variables. Since the system installation, 5643 records have been exchanged between the CCC and the WfMS components (average 36/day). These records concern to-do lists, messages, and clinical patients' data. Table 3 shows that the system is not too intrusive, since the amount of messages arriving to the communication box is very reasonable. The different percentage of therapy vs alert messages in the two patients groups (severe vs mild) is explained by the fact that severe patients have a longer hospital stay, therapy recommendations are concentrated in the first days, and then it is necessary to monitor side effects and complications. As another example, Table 4 shows, on the basis of encoded data entered in the CCC, the evaluation of the timing for the first two tasks that must be executed, as indicated by the SPREAD GLs, as soon as the patient is admitted: patient's history and diagnostic assessment (based on Rankin Scale, NIH Stroke Scale, Neuroradiologic and Neurosonologic examination). Although computed times to accomplish tasks are obviously biased toward high values (data input could be delayed), these data highlight and quantify possible problems in performing some diagnostic tests. It's worth noting that workflow process data allow to detect bottlenecks due to lack/misuse of resources (both technological and

Table 3 - Message exchange figures

	Average hospital stay	n. of messages /patient	n. of messages /day/patient	Therapy suggestion messages	Alert messages
Severe patients (MRS[§]>2)	10.55	11	1.062	14%	86%
Mild patients (MRS<=2)	6.61	6	0.953	30%	70%

§ Modified Rankin Scale

Table 4 - Evaluation of the clinical routine timing

	Elapsed time between admission and complete history		Elapsed time between admission and diagnostic assessment	
	<6 hours	<24 hours	<6 hours	<24 hours
Severe patients	67%	85%	15%	50%
Mild patients	71%	94%	33%	56%

human), thus allowing to discover the causes of the above cited problems.

Conclusions and future developments

Despite the big potential of Information and Communication Technology to improve healthcare organizations' outcomes (in terms of efficiency and/or effectiveness), there are still several problems in achieving successful real-world implementations. It is estimated that 60-70% of all software projects fail [9]. In this paper we have described the implementation of a decision support system aimed at improving compliance to GLs for stroke management. Results of the first eight months of the system use in the Stroke Unit are very encouraging. We think that key factors to the system success were manifold. First of all, users' commitment was high: the need of improving clinical practice was manifested by the neurological medical community, and all the system functionalities have been designed together with physicians. Second, the principle of minimum intrusiveness has been followed, and we took care in introducing minimal changes to the interface that users were familiar with. Last, we strongly collaborated with third partners providing the CCC, in order to achieve full system integration, and promptly intervene on users' feedback about system bugs and flaws. About future developments, we will focus on advanced technologies for data input. As a matter of fact, a well-known bottleneck for the optimal use of decision support systems is the difficulty to insert data in real time. We mentioned that some statistics could be biased by the delayed interaction. Ubiquitous computing technologies, provided for example by PDA, vocal interaction, and wireless connections, could greatly increase the CfMS performance and results reliability.

Acknowledgements

This work has been partially funded by projects FIRB (Italian Ministry of University and Research) and STAGE (Italian Ministry of Health).

References

[1] Kohn LT, Corrigan JM, and Donaldson MS, eds., To Err Is Human: Building a Safer Health System (Washington: National Academy Press, 1999).

[2] Micieli G, Cavallini A, Quaglini S; for the GLADIS Study Group. Guideline compliance improves stroke outcome: a preliminary study in 4 districts in the Italian region of Lombardia. Stroke. 2002,33(5):1341-7.

[3] Peleg M, Tu S, Bury J, Ciccarese P, Fox J, Greenes RA, Hall R, Johnson PD, Jones N, Kumar A, Miksch S, Quaglini S, Seyfang A, Shortliffe EH, Stefanelli M. Comparing computer-interpretable guideline models: a case-study approach. J Am Med Inform Assoc. 2003 Jan-Feb;10(1):52-68.

[4] 2006 Workflow Handbook Layna Fischer Ed., Future Strategies Inc.Lighthouse Point, FL, USA, 2006.

[5] SPREAD 2005 - Stroke Prevention and Educational Awareness Diffusion (2005) Edition of march 15th, www.spread.it.

[6] Baud RH, Weber P, Lovis Ch, Coding in context. PCSE 2001 Brugge.

[7] Panzarasa S, Quaglini S, Micieli G, Pernice C, Stefanelli M, Pessina M. Workflow technology to enrich a computerized clinical chart with decision support facilities. Proc. AMIA 2006, November 10-15, Washington, 2006.

[8] Brooke J. (1996) SUS: a "quick and dirty" usability scale. In P W Jordan, B Thomas, B A Weerdmeester & A L McClelland (eds.) Usability Evaluation in Industry. London: Taylor and Francis.

[9] Ammenwerth E, Iller C, Mahler C. IT-adoption and the interaction of task, technology and individuals: a fit framework and a case study. BMC Medical Informatics and Decision Making 2006, 6:3.

Address for correspondence

Silvia Panzarasa
Consorzio di Bioingegneria e Informatica Medica
UPIT,
Via Ferrata,
1 - 27100 Pavia (Italy).

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Towards a Decision Support System for Optimising Clinical Pathways of Elderly Patients in an Emergency Department

Marc Cuggia [a], Delphine Rossille [a], Aude Arnault [b], Jacques Bouget[c], Pierre Le Beux [a]

[a] *EA3888, Faculté de Médecine, Université de Rennes1, France,* [b] *ENSAI, Bruz, France*
[c]*Service de l'accueil des Urgences, CHU Pontchaillou, France*

Abstract

Data stored in Healthcare Information Systems correspond potentially to a vast source of information for supporting decisions in management or public health issues. The presented study illustrates clinical data valuation, through the analysis of clinical pathways of elderly patients at the Emergency Department (ED) of Rennes hospital. Method: Relevant data were extracted from the Emergency Department database. Several analysis (e.g., cusum method) and representation tools (e.g., Graphviz) were used to study the patients' pathways, the dynamics of flows and the patients' characteristics. Results: 4951 admissions were analysed and visualized. The representations provided a synthetic, global and comprehensive view of the department activities, to the satisfaction of the clinicians. Limitations of the ICD-10 coding of the diagnoses at the ED were pointed out as well as syntax and semantic interoperability issues. A solution is proposed for automating and scaling the Decision Support System.

Keywords:

management information systems; emergency service, hospital; critical pathways; decision making, computer-assisted

Introduction

The first departments to introduce and use computers in French public hospitals were administration and finance departments. The main purpose of collecting these medico-economic data was to help monitor the hospitals. Specific analytical systems were developed, some using for instance data mining methods, and this allowed data to be used as evidence in annual reports to determine budget allocation. One of our hospitals' main challenges is now to computerize clinical departments, hence healthcare processes. Thanks to the development of healthcare tools (i.e., for medical prescription, medical records, nursing records), it is possible to collect clinical data in the Hospital Information System's (HIS) databases on a day-to-day basis. These data are exhaustive, up-to-date and in a storage format facilitating their exploitation. They potentially represent a vast source of information and knowledge of prime interest for Public Health studies within hospitals. From the HIS databases, analysis and decision support tools could be developed to address Public Health issues: quality of care, hospital organisation and epidemiological research studies.

The objective of this paper is to illustrate the exploitability of the clinical and medico-economic data produced by the HIS for decision support in hospital management by studying the activity of a specific department. Several problems were identified.

The present study focuses on the care of elderly patients (aged 75 or over) in the emergency department (ED) of Rennes hospital. It is well known that the current conditions of hospital care of elderly patients are not satisfactory: waiting time at the ED is usually too long and patients are not always transferred to the hospital departments relevant to their conditions. The situation becomes worse in times of crisis, such as during the long spell of very hot weather in the summer of 2003 when the ED could not handle the patients flow and an unexpectedly high mortality rate was reached. In this context, the Emergency Department is in demand of a better understanding of its activity related to elderly patients, especially in terms of patient pathways.

This paper presents the results of the study from the decision support point of view. The aim is to provide feedback to the Emergency Department in the form of comprehensive reports of quality and effectiveness of care. In addition, the study should contribute to a better understanding of the delivery of care to elderly patients within the hospital in a public-health perspective.

The paper will focus on three areas of interest for the ED: the characterisation of patients, dynamics of flows and patients' pathways within the Emergency Department (that is the patients' flow within the ED). The three points will be discussed from a public health perspective, stressing the lack of formal semantics in the medical concept representations in use at the hospital.

Background

Several papers describe differentways to exploit data produced by ED's information systems for decision support in medical, management and public-health issues.

While knowledge-based systems aim at supporting medical decisions in the ED at the time of care (e.g.,. QMR, Iliad), other focus on management issues, such as the coordination, resource allocation and documentation aspects of

the Emergency Department operations [1]. Some specialize in the triage process, helping to prioritize incoming patients. For instance, the MET-AT system [2] provides triage plans for acute paediatric abdominal pain. Others propose methods (based on the cumulative sum chart) to manage hospital bed occupancy crisis [3].

Other papers relate experiences of exploiting data for public-health purposes. In Sydney [4], Australia, during the 2003 Rugby World Cup in a context of bioterrorism threats and emergent diseases, national public health authorities tested a near real-time syndromic surveillance system. In this system, EDs in the Sydney metropolitan area automatically transmitted surveillance data from their existing information systems to a central database in near real-time. This information included patient demographic details, presenting problem and nursing assessment entered as free-text at triage time, physician-assigned provisional diagnosis codes, and status at departure from the ED. Automated processes were used to analyse both diagnosis and free text-based syndrome data and to produce web-based statistical summaries for daily review. An adjusted cumulative sum (cusum) and Bayesian methods were used to assess the statistical significance of trends. In France, at a national level, the Heat Health Watch Warning System [5] was developed in 2003 to anticipate heat waves that may result in a large excess of mortality. The system was developed on the basis of a mortality analysis, healthcare activities in EDs and meteorological data. Several meteorological indicators were tested in relation to levels of excess mortality. The system requires close cooperation between the French Weather Bureau (Meteo France), the National Institute of Health Surveillance (InVS) and the Ministry of Health. The system is supported by a panel of preventive actions, to prevent the sanitary impact of heat waves.

Such systems illustrate the current interest in exploiting healthcare data for public health and management purposes. They rely on information whose format and sources are heterogeneous. The collection of clinical data from EDs is meant to improve the delivery of care for patients in primary care.

Materials and methods

The choice of the emergency department was motivated by its central dispatching role in the hospital and the daily contribution of medical records in its clinical database. Elderly patients only were selected as they represent an ever increasing part of the population in developed countries. Most of them have poly-pathologies.

The emergency department is structured in three medical wards: medicine, surgery and small traumatology. A proximity unit (named Duhamel ward) is specific to the ED. Patients staying in the Duhamel ward (theoretically for a maximum of 24 hours) either need to stay in observation or wait for an available bed in an appropriate hospital department.

Materials: Data were extracted from the ED clinical database (RESURGENCE) and the medico-economic database (DRG) produced by the HIS. The clinical medical records include general data (e.g., sex, age), data on arrival, departure and transfer (e.g., arrival mode, arrival data), data on the follow-up at ED, on the diagnoses and medical status. Medico-economic data include ICD-10-coded diagnoses, medical procedures (coded with a French classification named CDAM), date of hospitalisation and the various hospitalisation departments where the patient went. Hence only patients transferred in a hospitalization department appear in the medico-economic database. All the extracted data are limited vocabularies. Only data of elderly patients were selected. Over a period of 6 months (from October 2005 to may 2006), 32883 patients came to the emergency department, 4951 were aged 75 or over.

Method: After extracting all the data from the HIS databases (SQL requests), relevant data to the study's objectives were identified and their quality control checked through statistical analysis (SAS). The analysis methods and representational views were chosen to reach the objective of providing useful, easy-to-read and synthetic views of the ED activities.

Descriptive statistics provided characterisation of the population, more specifically on the length of stay at ED and the diagnoses. To study the dynamics of flows, the cusum method was used on the admission rates. A cusum chart [3] is the cumulative sum of consecutive differences between an individual measurement and a selected target. It reveals trends otherwise not obviously detectable in the corresponding time-series data. This analysis technique has been applied to manage hospital bed [3] or to monitor trends in hospital-acquired infection. For visualizing patient pathways within ED, the data were pre processed according to a new arrangement of the operational structures. Patterns of patient pathways were displayed using the visualization tool Graphviz (http://www.research.att.com/sw/tools/graphviz/). The chosen visualization tool was aimed at providing a global representation of the most frequent (and typical) pathways for the population, giving insights on the use of the different ED wards as well as on the waiting times.

Results

The demographic study shows that the admitted elderly population is representative of the elderly in France and corresponds to 17.9% of the total number of patients admitted at the ED, slightly higher than national percentages (15 to 17%).

Characterisation of the population

The **ED length of stay** (expressed in hours) depends strongly on the week day of arrival and the patient's medical status at arrival. The medical status is expressed in a scale from 1 (stable prognosis) to 5 (vital prognosis in danger, immediate intensive care required). The highest mean length of stay is found on Sunday (Figure 1), a week-end day during which the medical manpower is currently reduced. Not surprisingly, the length of stay is correlated with the patient's medical status ($p>0.005$), to the exception of cases of extremely bad medical status corresponding to patients with potentially lethal conditions (Figure 2).

From the distribution of the **diagnoses** as a function of the ICD-10 main chapters (Figure 3), the three most repre-

sented principal diagnoses at the ED are Chapter XVIII "Symptoms,

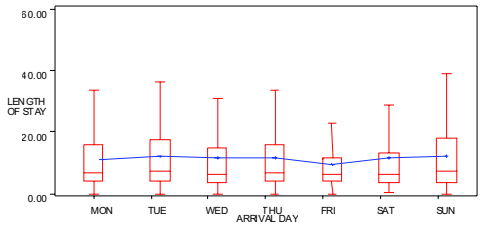


Figure 1 - Tukey box for the ED length of Stay vs. the week of the day at arrival

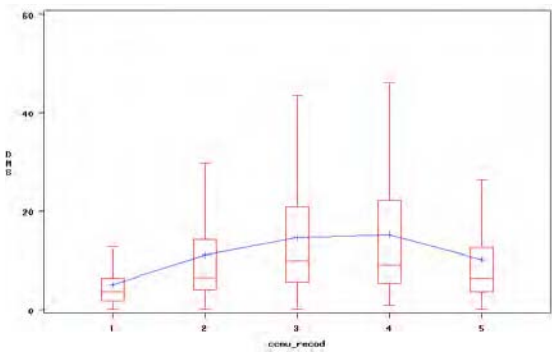


Figure 2 - Tukey box for the ED length of Stay vs. the patient's medical status at arrival

signs and abnormal clinical and laboratory findings, not elsewhere classified" (22.7%), Chapter XIX "Injury, poisoning and certain other consequences of external causes" (20.9%) and Chapter IX "Diseases of the circulatory system" (13.3%).

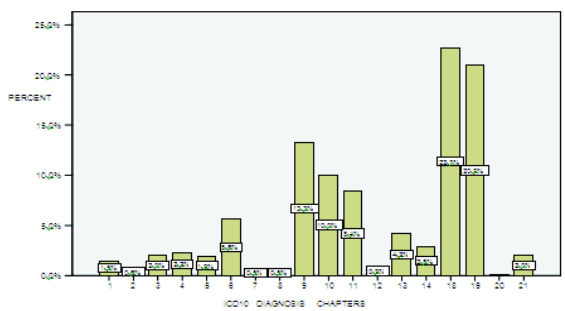


Figure 3 - Distribution of the principal diagnoses at ED

Table 1 details diagnoses for the three principal diagnoses at ED. It shows for instance that 48.5% of the diagnoses related to Chapter XVIII correspond to "General symptoms and signs".

Dynamics of flows

Figure 4 represents the number of admissions at ED with respect to the date, for the considered period of time. The time-series data appear to be very complex to interpret. Using the cusum method provides a comprehensive and global view of the variations (Figure 5). Indeed, the slope changes correspond to admission rate higher than the mean (that is the selected target). Hence three periods show an increased admission rate: in December, February and April.

Chapter 18		Chapter 19		Chapter 9	
48.5 %	General symptoms and signs	26.4 %	Injuries to the hip and thigh	41.8 %	Cerebrovascular diseases
14.5 %	Symptoms and signs involving the circulatory and respiratory systems	20.1 %	Injuries to the head	29.9 %	Other forms of heart disease
11.3 %	Symptoms and signs involving nervous and musculoskeletal systems	10.4 %	Injuries to the shoulder and forearm	11.8 %	Diseases of arteries, arterioles and capillaries

Table 1 - Detailed diagnoses for the three principal ICD-10 diagnoses at ED

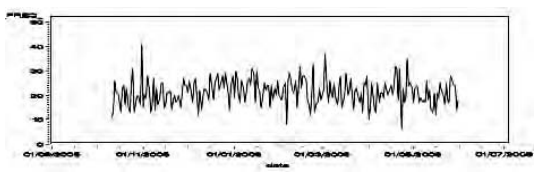


Figure 4 - Number of admissions per date

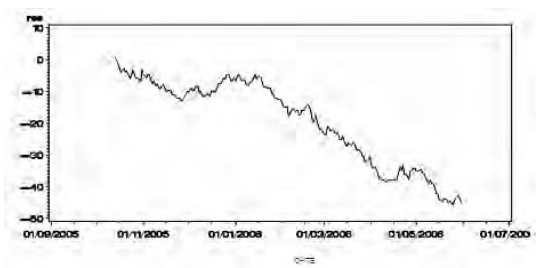


Figure 5 - Cusum chart of the number of admissions per date

Patient pathways

Figure 6 provides a visual representation of the patient pathways within the Emergency Department. Only pathways of type <admission – step 1 - step 2> are considered. They represent alone 82.7% of the total patient pathways.. A "step" consists of one of the four ED services or an exit mode. The pathways not represented correspond to complex pathways of more than three steps, or to pathways taken by very few patients. At each step is given the median waiting time. On each branch of the graph the percentage of the patients following this path between the two nodes is displayed, 100% referring to the father node. The most followed pathway corresponds to <admission – medicine service – exit>(49%, represented in red). Then come the pathway <admission – surgery service – exit> (30%, in grey), far behind <admission – traumatology

service – exit> (2.42%, in yellow) and <admission – exit> (0.67%, in blue).

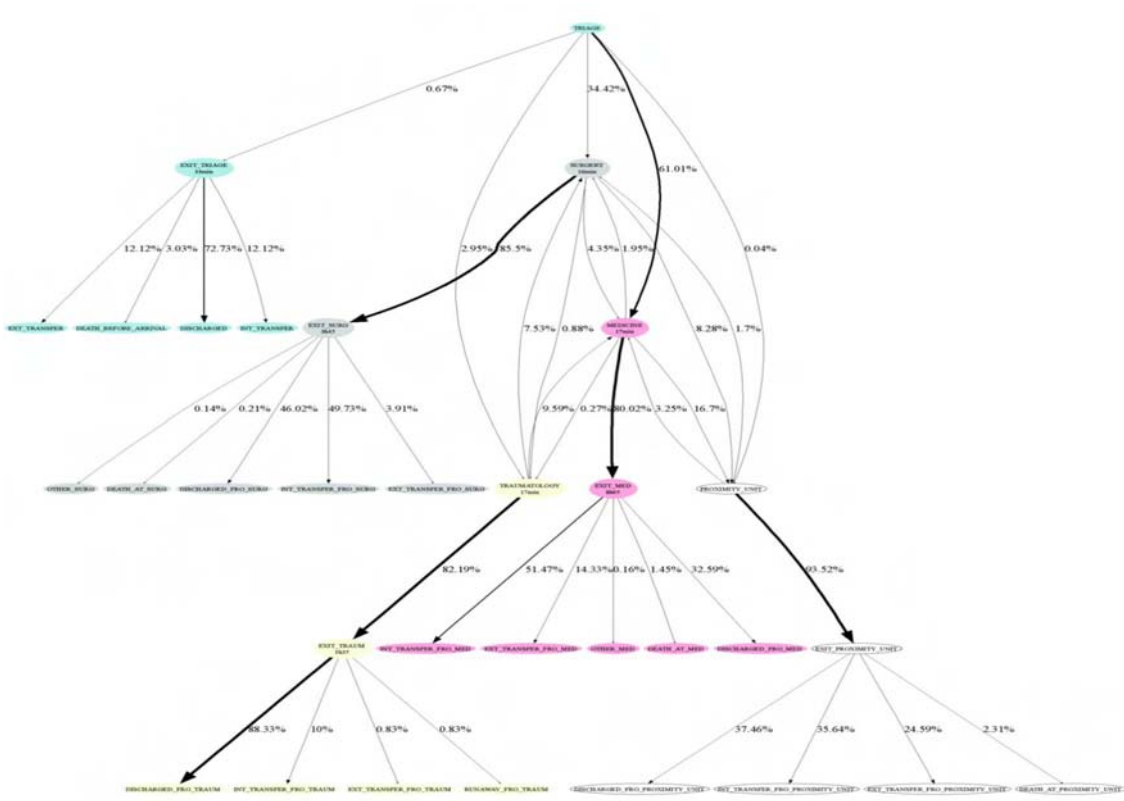


Figure 6 - Patients' pathways at the Emergency Department

Discussion

The study showed that the vast source of medical data stored for healthcare purposes in the HIS databases can provide valuable insights to the clinicians in a comprehensive and straightforward manner. Clinicians at the Emergency Department were very satisfied with this study of elderly patients and with the chosen representation tools. No actual satisfactory survey was done, but the team responsible for the optimisation of the patients' flows within the hospital has decided to use this work as it provides a quantitative measure of the ED activities, especially in terms of patients' pathways.

Thanks to these representation tools used, the patients' characteristics, the dynamics of flows and the patients' pathways provide a multifaceted and complete view of the Emergency Department easily readable by the clinicians for decision support purposes, as for improving the department structural organisation and service management.

Dynamics of flows : The temporal analysis (cusum method) provides a yearly view of the dynamics of flows. It is complemented with the periodic representation per week days (Tukey boxes). This has led the Emergency Department to look at increasing medical manpower during week-ends, in order to face the higher patients flow.

Patients' pathways : Graphviz allowed a relatively clear representation of the patient pathways. However many pathways had a frequency of almost one. It was decided to represent graphically only the most important pathways in terms of frequency. Hence a subjective threshold was applied to patients' pathways. The Galois lattice calculates the threshold based on statistics [6]. However the lattice structure and its graphical representation are much more difficult to interpret than the Graphviz representation that was proposed in this paper.

Medical characterisation of the population : Limitations of the exploitation of healthcare data appear to lie in the use of ICD-10 coding for diagnoses. As shown in Table 1, the most frequently-used code for diagnoses at the ED corresponds to "Symptoms and general signs" of Chapter XVIII of ICD10 "Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified". Indeed determining precise diagnosis at the ED is often difficult as for instance several complementary exams may be required during hospitalisation. Priority is first of all to deal with acute situation, to start symptomatic treatment and to transfer the patient to the more appropriate hospitalisation department when needed. It is not surprising that ED clinicians often code the symptoms (s.a. abdominal pain, migraines) instead of precise diagnoses. However the ICD-10 is primarily intended to classify diseases while the

symptoms are described only in one chapter out of 22, therefore not being linked in any way (into the ICD-10 structure) to the underlying disease.

To solve this problem of coherency of ICD-10, a recent work suggests SNOMED CT to be the coding reference. Indeed, a comparative study within the UMLS shows that SNOMED CT covers better the needs for coding medical records in the Emergency Department [7]. Another paper [8] shows that better semantic coherency is possible with SNOMED CT when comparing diagnoses codes at the different steps of the patient pathway in the hospital.

Perspective

To be scalable to the hospital level, regional or even national level, such study requires the setting up of an automatic Decision Support System. With the presented preliminary work, several issues were brought to light on the system's architecture and the syntax and semantic interoperability within the HIS.

First a data warehouse should be built in order to gather all the heterogeneous data extracted from the HIS. This approach facilitates the transformation and connexion of healthcare data for decision support purposes. Calculating decisional indicators can be very time and resource consuming and having already all the required data in the proper format in the data warehouse will guaranty the complete availability of healthcare resources for medical activities. The process of loading data from heterogeneous sources into the data warehouse assumes syntax interoperability issues to be solved. This can be achieved by using international exchange standards, such as the HL7 [9] and should be easier if the HIS is built on an SOA architecture (using web services)[10].

However semantic interoperability issues remain to be considered. For instance, taking the case of a Decision Support System at the hospital level, the adequacy between the diagnoses coded at the ED (often "symptoms") and the diagnoses coded at the hospitalisation departments is not straightforward as shown in the study. This could be solved in two steps. First the "symptom" would be related to an appropriate "disease" as hospitalisation departments are specialised to cure specific pathologies (s.a. cardiology or pneumology departments). The SNOMED CT, with its semantic relationships between symptoms and diseases, could be used [8]. The second step consists of checking the adequacy between the hospitalisation department and the patient's disease. This task is not straightforward and several ambiguities exist. For instance, the internal medicine department is in charge of a large number of different pathologies. Overlapping of competency between two separate departments may also occur. For instance, a patient with an acute oedema of the lung may be directed either to the cardiology or the pneumology department. One solution is to describe automatically the competency fields of departments based on their economic activities, for instance by using the main Diagnosis Related Groups (DRGs) of each department.

Conclusion

The objective of the paper is to show, through the study of the delivery of care to elderly patients at the Emergency Department, how data produced by healthcare processes are valuable for decision support systems. This approach is related to the data mining approach whose objective is to extract relevant knowledge for the hospital management.

The results indicate that providing a synthetic and global view of the department's activities, thanks to appropriate representation tools comprehensive to clinicians, is particularly relevant for the department's managers and clinicians.

Finally, it is explained that for scaling the application to the hospital level, issues of syntax and semantic interoperability have to be considered. A solution is proposed based on SNOMED CT for its semantic medical representation and DRGs to allow the automatic description of the competency fields of clinical departments.

References

[1] R. Ceglowski, L. Churilov, J. Wassertheil, "Facilitating decision support in Hospital Emergency Departments: A process-oriented perspective", in *Proceedings of the 13th European Conference on Information Systems*, Regensburg, Germany, 2005

[2] K.J. Farion, W. Michalowski, S. Rubin, S.Wilk, R. Correll and I. Gaboury, "Prospective evaluation of the MET-AP system providing triage plans for acute pediatric abdominal pain", *Int J Med Inform*., 2007 Feb 21.

[3] C. M. Burns, C. J. Bennett, C. T. Myers, and M. Ward, "The use of cusum analysis in the early detection and management of hospital bed occupancy crises," *Med J Aust*, vol. 183, pp. 291-4, 2005.

[4] D. J. Muscatello, T. Churches, J. Kaldor, W. Zheng, C. Chiu, P. Correll, and L. Jorm, "An automated, broad-based, near real-time public health surveillance system using presentations to hospital Emergency Departments in New South Wales, Australia," *BMC Public Health*, vol. 5, pp. 141, 2005.

[5] M. Pascal, K. Laaidi, M. Ledrans, E. Baffert, C. Caserio-Schonemann, A. Le Tertre, J. Manach, S. Medina, J. Rudant, and P. Empereur-Bissonnet, "France's heat health watch warning system," *Int J Biometeorol*, vol. 50, pp. 144-53, 2006.

[6] N. Jay, A. Napoli, and F. Kohler, "Cancer patient flows discovery in DRG databases," *Stud Health Technol Inform*, vol. 124, pp. 725-30, 2006.

[7] J. C. McClay and J. Campbell, "Improved coding of the primary reason for visit to the emergency department using SNOMED," *Proc AMIA Symp*, pp. 499-503, 2002.

[8] P. L. Elkin, A. P. Ruggieri, S. H. Brown, J. Buntrock, B. A. Bauer, D. Wahner-Roedler, S. C. Litin, J. Beinborn, K. R. Bailey, and L. Bergstrom, "A randomized controlled trial of the accuracy of clinical record retrieval using SNOMED-RT as compared with ICD9-CM," *Proc AMIA Symp*, pp. 159-63, 2001.

[9] K. M. Coonan, "Medical informatics standards applicable to emergency department information systems: making sense of the jumble," *Acad Emerg Med*, vol. 11, pp. 1198-205, 2004.

[10] J. Mykkanen, A. Riekkinen, M. Sormunen, H. Karhunen, and P. Laitinen, "Designing web services in health information systems: From process to application level," *Int J Med Inform*, 2006.

Address for correspondence

Dr Marc CUGGIA
Département d'information médicale - CHU Pontchaillou
2, rue H. Le Guilloux - 35033 Rennes Cedex

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Ontology-based Modeling of Clinical Practice Guidelines: A Clinical Decision Support System for Breast Cancer Follow-up Interventions at Primary Care Settings

Samina R. Abidi, Syed SR. Abidi, Sajjad Hussain, Mike Shepherd

NICHE Research Group, Faculty of Computer Science, Dalhousie University, Halifax, Canada

Abstract

Breast cancer follow-up care can be provided by family physicians after specialists complete the primary treatment. Cancer Care Nova Scotia has developed a breast cancer follow-up Clinical Practice Guideline (CPG) targeting family physicians. In this paper we present a project to computerize and deploy the said CPG in a Breast Cancer Follow-up Decision Support System (BCF-DSS) for use by family physicians in a primary care setting. We present a semantic web approach to model the CPG knowledge and employ a logic-based proof engine to execute the CPG in order to infer patient-specific recommendations. We present the three stages of the development of BCF-DSS—i.e., (a) Computerization of the paper-based CPG for Breast Cancer follow-up; (b) Development of three ontologies—i.e., the Breast Cancer Ontology, the CPG ontology based on the Guideline Element Model (GEM) and a Patient Ontology; and (c) Execution of the Breast Cancer follow-up CPG through a logic-based CPG execution engine.

Keywords:

clinical practice guidelines, breast cancer,
decision support system, medical ontology, semantic web

Introduction

Clinical Practice Guidelines (CPG) entail medical knowledge intended for clinical decision-making and standardization of clinical practice [1, 2]. Despite the potential benefits of CPG, reviews show that CPG are underutilized in clinical practice [3, 4], largely due to problems associated with their dissemination to physicians [2, 5]. CPG computerization involves the modeling and conversion of a paper-based CPG into an electronic and executable format that can both be accessed by physicians and be embedded within clinical decision-support systems at the point of care. CPG guided decision support systems are particularly useful in clinical settings where non-specialist health practitioners, such as family physicians or nurses, are required to deal with complex or unusual cases. In such situations, CPG based decision support systems can guide the healthcare practitioner's actions and suggest appropriate recommendations. One such situation is the discharge of Breast Cancer (BC) follow-up care by family physicians. Note that in Nova Scotia follow-up care sis currently being provided by cancer care specialists at tertiary care centers.

Recent advancements in BC treatment have significantly improved the rate of BC survivors in Nova Scotia. The follow-up care for patients in remission entails periodic visits for history, physical exams and mammogram surveillance [6]. Although specialized cancer clinics provide long-term follow-up care, there is a case for formally involving family physicians in breast cancer follow-up care. In fact, trials conducted in Canada and Britain show that family physician offer a viable alternative to specialized care clinics for offering follow-up care to women who are in remission from breast cancer [6]. However, for most family physicians BC follow-up care is a new and added responsibility, therefore they need clear clinical guidelines to effectively perform the follow-up activities, make correct decisions and provide the right recommendations. The Canadian Steering Committee on CPG for the Care and Treatment of Breast Cancer has developed and recently updated the guideline on follow-up care after treatment for BC [7] with special emphasis on the needs of primary care physicians. The challenge was to disseminate the CPG to the family physician and to integrate it within his/her clinical workflow so that the CPG is seamlessly executed whenever a patient undergoes BC follow-up in a primary care setting.

In this project, we collaborated with Cancer Care Nova Scotia to address the abovementioned challenges by promoting the (knowledge) translation of the CPG for the Care and Treatment of BC to family physicians, to support them in the delivery of BC follow-up care and patient education at their clinics. This will reduce the workload of specialist cancer centers within Nova Scotia. Our approach was to develop an interactive Decision Support System (DSS) that enables family physicians to (a) access and utilize the said CPG at the point of care to provide standardized follow-up care; and (b) offer customized patient educational information targeting disease management, lifestyle behaviours and psychosocial support.

In this paper, we present an ontology-based Breast Cancer Follow-up Decision Support System (BCF-DSS) based on the CPG for the Care and Treatment of Breast Cancer. We

take a semantic web approach to model the CPG knowledge and to reason over the ontology to provide 'trusted' CPG-driven recommendations. We have developed three ontologies: (a) *CPG ontology* that models the structure of the CPG based on the Guideline Element Model (GEM); (b) *Breast Cancer Ontology* that represents the medical knowledge encapsulated within the CPG and general BC related concepts; and (c*) Patient Ontology* that models the patient's parameters. The ontologies are developed using Protégé and are in OWL format. We have developed a logic-based reasoning engine that reasons over the knowledge from these three ontologies. Our BCF-DSS allows family physicians to collect patient data and assists them to make CPG mediated decisions, recommendations and referrals for BC survivors. We present the three stages of the development of BCF-DSS—i.e. (a) Computerization of the paper-based CPG for the Care and Treatment of Breast Cancer; (b) Development of the ontologies, in particular the Breast Cancer Ontology; and (c) Execution of the BC follow-up CPG through a logic-based CPG execution engine.

Computerization of BC follow-up CPG

Computerization of the CPG involved (a) selection of a CPG modeling formalism; and (b) capturing and representing the CPG knowledge based on the modeling formalism. We selected the Guideline Representation Model (GEM) to model the BC follow-up CPG. GEM is based on XML that renders it operable in a semantic web environment by allowing semantically salient indexing and searching of CPG knowledge. We used the GEM Cutter tool to annotate the BC follow-up CPG with GEM tags (or elements). The conversion task involved determining the function of a specific CPG text and annotating it using the relevant GEM tag. It may be noted that GEM constitutes 100 tags covering a wide variety of concepts. For our purposes, the most salient concepts were the 'Knowledge Components' that store and categorize the knowledge present in a CPG. The knowledge components have subcomponents called recommendations which are categorized as either *imperative*—i.e., directed towards entire target population, or *conditional*—i.e., act on the decision variables and results in appropriate actions.

From a knowledge modeling perspective the main challenge was to resolve the (medical and semantic) ambiguities inherent within the BC follow-up CPG. To resolve the ambiguities we (a) consulted with BC oncologists, in particular the author of the BC follow-up CPG; (b) reviewed available literature; and (c) applied our personal clinical experience. Examples of ambiguities included phrases such as, 'vaginal bleeding is present in the absence of obvious cause', 'physiological causes of fatigue' and 'other risk factors of osteoporosis'. The phrase 'vaginal bleeding in the absence of obvious cause' was resolved to the term 'Menstruation', and other ambiguous phrases were similarly resolved by mapping them to explicit concepts. Finally, through the use of GEM we managed to create an executable representation for the BC follow-up CPG.

Development of breast cancer ontology

The BC ontology models the knowledge encapsulated within the BC follow-up CPG. We used Protégé ontology editing environment to build our BC ontology in OWL (Web Ontology Language) using Protégé OWL.

The BC ontology is largely derived from the contents of the knowledge components—i.e. the 'Imperative' or Conditional recommendations—in the GEM representation of the BC follow-up CPG. More specifically, the conditional recommendation element, which comprises sub-elements such as 'decision.variable', 'action' and 'logic' elements, was used to develop the BC ontology. Given conditional recommendations, the challenge was to identify the decision variables, the actions to be taken and the Boolean logical operations in the recommendations, so that the resultant ontology was compatible with our logical reasoning engine. In this regard, two design constraints were addressed: (a) Our CPG execution engine does not processes statements containing 'OR' and 'NOT' logical operators. Therefore a rule such as:

> "**IF** *age* >65 **OR** *family history of* osteoporosis **OR** *menstrual status of* premature menopause due to treatment,
>
> **THEN** screen with bone mineral density and treat accordingly with bisphosphonates",

was required to be decomposed into three smaller rules, each with a single decision variable so that the OR operator was eliminated; and (b) The BC ontology classes that have multiple domains or ranges could not be executed safely. Therefore, we ensured that all properties have a single domain and range.

Specifying BC ontology classes

Considering the above constraints we defined eight main classes, namely; Patient_Type, Physician_Type, Illness, Menstrual_Status, Recommendation, Symptom, Diagnostic_Tests, Treatment, Age, Risk_Factor, Weight_Status and Patient_Wish Next, we specified the disjoint classes, where classes are disjoint when an individual cannot be an instance of more than one of these classes. In the BC ontology the only classes which are not disjoint are the Recommendation and Diagnostic Tests, and Recommendation and Treatment since they share some instances, for example "Screening with bone mineral analysis" is an instance of two classes i.e., Diagnostic_Test and Recommendation; and "Bisphosphonates" belong to class Treatment and Recommendation.

Specifying properties for the BC ontology classes

Properties for patient class

The *'Patient_Type'* is the most important class because most conditional recommendations are targeted towards the patient. To specify different patient types we defined a range of properties. The patient properties represent the patient profile—i.e., an instance of the class *Patient_Type*. We defined object properties to establish link between the classes so that recommendations can be associated with a patient profile. The class *Patient_Type* has the most object

properties with their domain being *Patient_Type* but their range includes instances from other classes in the BC ontology, for example, the properties 'has_*history_of'* and *'has_illness'* have individuals of the class Illness as their range.

In our CPG execution engine most of the properties are treated as *decision variables* that serve as the premise of a logical rule. The conclusion of the rule is an *action variable* that corresponds to a recommendation, treatment, or statement directed towards a patient. To account for the patient-centric action variables, we specified two *Patient_Type* properties—i.e. *is_Recommended* and *possible_cause_can_be* with an unspecified range as their range can be any individual from any class. The action variable *is_Recommended* refers to the any recommendation, diagnostic test or treatment suggested to a patient. The action variable *possible_cause_can_be* provides the physician information regarding the cause of certain sign or symptom, for example the CPG statement "emotional distress, may be the underlying cause of subjective complaints of impaired cognitive functioning."

Properties for other classes

The properties for the other classes were derived from conditional statements in the recommendation element of the GEM representation of the BC follow-up CPG. These conditional statements specify variables that non-patient specific data, for example consider the CPG statement "If the purpose is to detect distant metastasis, then routine lab and radiographic exam should not be carried out". Such as statement was modeled by the class 'Diagnostic_Test' through two properties; *has_purpose_to_detect* whose range is Illness and *test_apply_to* whose no specified range. In total, we specified 40 properties for all other classes excluding the patient class.

Properties for statements having the not logical operator

Modeling of some statements in the BC follow-up CPG required the use of the 'NOT' operator. For instance consider the statement, "When such bleeding (vaginal bleeding) is present in the absence of obvious cause, endometrial biopsy should be carried out". Here the phrase 'in the absence of obvious cause', really means *not* obvious cause. We handled such situations by specifying a new property, for instance *is_not_caused_by,* for the class Symptom. Note that the rationale for creating such as property is because our execution engine is unable to handle the 'NOT' logical operator.

Specifying property characteristics

Certain properties such as *has_age*, *has_weight_status* and *has_menstural_status* are functional properties since a patient can have only one age, weight status (i.e., can either be over-weight, under-weight or have correct weight) and menstrual status (i.e., can either be premenopausal, postmenopausal or premature menopause due to treatment). Most of the properties are not functional and allow multiple values. We also specified some inverse object properties such as 'is_*recommended_for_illness*' which is the property of class 'Treatment' is the inverse property of '*is_treated_by*' which is the property of class 'Illness'. This means that 'Bisphosphonates' *is_recommended_for_illness* 'Osteoporosis' and 'Osteoporosis' *is_treated_by* Bisphosphonate. Note that Osteoporosis is the individual of the class Illness and Bisphosphonate is the individual of the class Treatment.

Specifying individuals (instances)

In the next step of BC ontology development we specified the individuals (or instances) from the conditional recommendations of the BC follow-up CPG. For example, individuals for class Symptom include Anxiety, Back_Pain, Cognitive_Impairment, Fatigue, Impaired_sexual_function, Menopausal_Symptoms, Vaginal_Bleeding, and Vaginal_Dryness. The class Patient_Type has the most individuals since each recommendation is valid for a patient with a particular set of clinical characteristics, thus each patient type refer to a particular patient profile in accordance to the said CPG.

Specifying relationships between the classes

The relationships among different classes were modeled using the class properties, and can be best understood by the following example. In order to model the recommendation statement "When such bleeding (vaginal bleeding) is present in the absence of obvious cause endometrial biopsy should be carried out". Here, "obvious cause of bleeding" means Menstruation. We established a relationship between the classes *Patient Type, Symptom, Menstural_Status* and *Diagnostic_Test* as follows: Patient_Type_12 is an individual of class Patient Type who has Vaginal_Bleeding which is the value for its object type property *has symptom*. Vaginal_Bleeding is also an individual of class Symptom, which has an object type property called *is_not_caused_by*, whose value is Mensturation_or_Obvious_cause, which in turn is an individual of another class called Mentrual_Status. The class Menstrual_Status has an object property called *ms_apply_to_diagnostic_test* whose values in this case will be Endometrial_Biopsy which is an instance to the class Diagnostic_Test (See Figure 1). By adding the property *is-not-caused_by* we ensured that the recommendation is logically valid i.e. endometrial biopsy is not recommended whenever the patient has vaginal bleeding. In this way we are able to establish an inter-class relationship that can be used to infer that if a patient has vaginal bleeding and bleeding is not caused by menstruation or any other obvious cause, then endometrial biopsy is the recommended test.

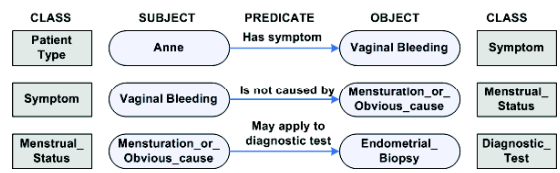


Figure 1 - RDF Triples depicting the relationships between classes Patient_Type, Symptom, Menstrual_Status, and Diagnostic_Test to model a recommendation

CPG Execution Engine

In keeping with our Semantic Web approach we developed a CPG Execution Engine (CPG-EE) that constitutes (a) multiple ontologies to model the domain and CPG knowledge; (b) a logic-based proof engine that leverages the ontologies and CPG specific rules to infer CPG mediated recommendations; and (c) a justification trace to describe the rationale for the inferred recommendations; this is to establish 'trust' in the proposed recommendations (Figure 2). The CPG-EE provides the functionality to define CPG-specific decision logic rules based on the decision variables in the CPG and to execute the rules based on patient clinical data to provide CPG based recommendations. The CPG-EE comprises two main modules: (i) Rule Authoring Module and (ii) Rule Execution Module.

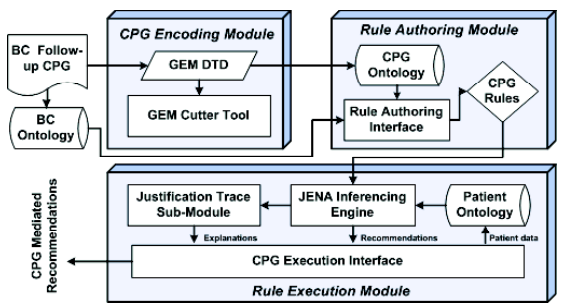


Figure 2 - System diagram of the CPG-EE

Rule Authoring Module

The Rule Authoring Module provides users an interface to specify decision logic rules, using a native CPG rule syntax, based on the decision logic inherent within the CPG. The rule authoring process is guided by the knowledge, relationships and constraints represented within the CPG ontology and the domain ontology (in our case the BC ontology). In this way, rule authoring is constrained by pre-defined knowledge and hence ensures the sanity of the decision rules.

The CPG ontology is designed to semantically model the structure of a CPG in order to annotate the decision variables and logic structures inherent within the CPG. Our CPG ontology is based on the GEM DTD which provides a characterization of different elements of a CPG. In particular we utilized the knowledge component of the GEM DTD and mapped it to a *Recommendation* class that entails the procedural, conditional or imperative knowledge of the CPG. The decision and action variables that constitute the premises and conclusions of a CPG rule, respectively, are explicitly stated in the CPG ontology, and these variables are utilized in authoring CPG rules. In the CPG ontology, the decision variables are represented as a sub-class. For execution purposes we added a new property variable.name to the decision.variable, such that its value is derived from all properties in the Domain Ontology.

Rule Authoring is performed by defining decision rules in the logic tag of CPG ontology as follows: *Step 1:* Select decision variables from the Decision Variable List, which represents the body (premises) of the rule; *Step 2:* Select the action variable from the Decision Variable List, which represents the head (conclusion) of the rule; *Step 3:* For each decision and action variable in the rule, an equality/ inequality relation is defined with either a variable, a value, a binary algebraic formula, another decision variable or list of decision variables. We give a rule authoring example, where we assign the variable names i.e. properties to *decision variables* (coded as dv and each with a unique #) as well as action variables (coded as *av* and each with a unique #). In case of a rule

> **IF dv1** i.e., *Patient_is_on_medication* = Tamoxifen (property of class Patient_Type) **AND dv2** i.e. *Rx_apply_to_recommendation* = ? (property of class Rx_Recommended)
>
> **THEN a1** i.e., *Patient_is_recommended.* (property of class Patient_Type) = **dv5**

The derivation for this rule is as follows. The Patient_Type_1 which is an instance of the class Patient_Type is *on medication*, Tamoxifen. Patient_Type_1 is the resource for this rule. The treatment i.e. Tamoxifen is an instance of class Treatment, which has a property *apply to recommendation*, whose value is 'query about vaginal bleeding'. Since we have specified in the rule that the value for *a1* (*Patient_is_recommended*) is same as the value for *dv2* (*Rx_apply_to_recommendation*), which according to the BC ontology is 'query about vaginal bleeding', the recommendation for this patient type is to query about vaginal bleeding.

Rule Execution Module

The Rule Execution Module executes the CPG rules based on a patient instance to infer CPG based recommendations. Rule execution is performed by a logic-based inference engine—i.e., JENA. The processing of this module is as follows: (i) The CPG rules are transformed from their native syntax to JENA rule syntax; (ii) The patient data is acquired through the CPG execution interface (see Figure 4) to form an instance of a patient, based on the Patient Ontology, that incorporates patient properties such as *age, gender, medical history*, etc. The values of the patient properties serve as input to the execution engine; (iii) The JENA inference engine uses the CPG rules and the patient instance to build an inference model using backward chain reasoning. The outcome is a set of inferred recommendations based on the patient data; (iv) A justification trace of the inferred recommendations is generated to explain the reasons for the proposed recommendations.

BCF-DSS in Action

We present an example to demonstrate BCF-DSS in action. The clinical case is: *"A BC patient who is overweight, complains of fatigue and is experiencing vaginal bleeding in the absence of obvious cause. She has a family history of osteoporosis and is on Aromatase inhibitors. She also wishes to get pregnant and wants to know whether this is a viable option".*

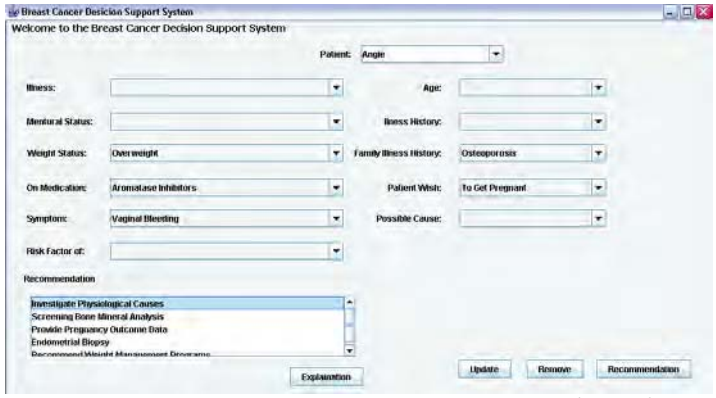


Figure 3 - The CPG execution interface for BCS-DSS, used to collect the patient data and to give recommendations

The family physician records the patient's properties using the BCC-DSS user interface (shown in Figure 3). The physician presses the *Recommendation* button and is provided five recommendations (shown in the bottom left box). The physician can seek an explanation for any recommendation by highlighting it and pressing the *Explanation* button. Figure 4 shows the explanation interface that includes the CPG description (upper left box) for the recommendation, the reasons for the proposed recommendation (upper right box) and the related references (lower middle box); all explanation material is derived from the annotated BC CPG. Finally, the justification trace (see Figure 5) for the inferred recommendations is as follows:

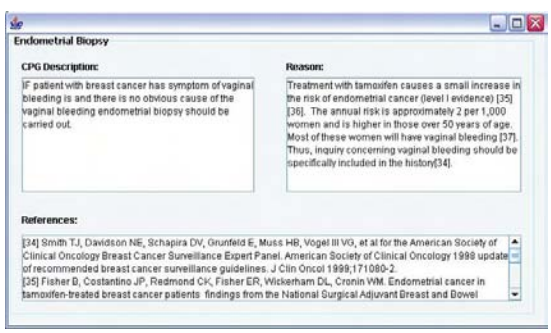


Figure 4 - The explanation interface of the BCF-DSS

Angie → has weight status = Overweight
Overweight → associated with recomm = Recommend Weight Management Programs
Angie → is Recommended = *Recommend Weight Management Programs*

Angie → has symptom = Vaginal Bleeding
Vaginal Bleeding → is not cause by = Mensturation or Obvious Cause
Mensturation or Obvious Cause → ms apply to diagnostic test = Endometrial Biopsy
Angie → is Recommended = *Endometrial Biopsy*

Angie → has family history of Illness = Osteoporosis
Osteoporosis → illness apply to recomm = Treatment Should Include Bisphosphonates
Angie → is Recommended = *Treatment Should Include Bisphosphonates*

Angie → has wish = To Get Pregnant
To Get Pregnant → associated with recommendation = Provide Pregnancy Outcome Data
Angie → is Recommended = *Provide Pregnancy Outcome Data*

Angie → is on medication = Aromatase Inhibitors
Aromatase Inhibitors → Rx apply to Diagnostic Test = Screening Bone Mineral Analysis
Angie → is Recommended = *Screening Bone Mineral Analysis*

Angie → has symptom = Fatigue
Fatigue → apply to recommendation = Investigate Physiological Causes
Angie → is Recommended = *Investigate Physiological Causes*

Figure 5 - Justification trace for the recommendations

Concluding remarks

We have developed a CPG based interactive clinical decision support system for the BC follow-up to be used in the primary care setting. Our approach is innovative since we have linked the CPG ontology to the breast cancer domain ontology from which rules were derived. This approach can also be applied to CPG in other medical specialties. The objective of this project is to promote knowledge translation to primary care settings in Nova Scotia so that family physicians can take on the responsibility for the BC follow-up care, thereby reducing the strain on specialist cancer centers within Nova Scotia. The project also aims to create an interactive environment for family physicians to facilitate customized patient management and educational information for an individual patient.

Acknowledgement

This project is funded by Nova Scotia Health Research Foundation (Canada), through an innovation research grant. Special thanks to Dr. Grunfeld for the BC-CPG and her support.

References

[1] Pryor DB, Fortin DF. Managing the delivery of health care: care-plans/managed care/practice guidelines. *Int J Biomed Comput.* 1995;39:105–9.

[2] Zielstorff RD. Online Practice Guidelines, Issues, Obstacles and Future Practice, *JAMIA*. 1998;5:227–236.

[3] Lomas J, Anderson GM, Domnick-Pierre K. Do practice guidelines guide practice? The effect of a consensus statement on the practice of physicians. *N Engl J Med.* 1989; 321:1306–1311.

[4] Greco PJ, Eisenberg JM. Changing physicians' practices. *N Engl J Med.* 1993;329:1271–1273.

[5] Sanders GD, Nease RF, Jr., Owens DK. Design and Pilot Evaluation of a System to Develop Computer-based Site-specific Practice Guidelines from Decision Models; *Med Decis Making* 2000;20:145–159.

[6] Grunfeld E, Levine M, Julian J, Prtichard K, Coyle D, Mirsky D, Verma S, Dent S, Sawka C, Ginsburg D. A randomized controlled trial of routine follow-up for early stage breast cancer: a comparison of primary care versus specialist care. *J Clin Oncol* 2004; 22:43

[7] Grunfeld E, Dhesy-Thind S, Levine M. Guidelines for the Care and Treatment of Breast Cancer: Follow-up after treatment for breast cancer (2005 update), *Can. Med. Assoc. J.*, May 2005; 172: 1319 - 1320

Address for correspondence

Syed Sibte Raza Abidi Email: sraza@cs.dal.ca

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Learning Causal and Predictive Clinical Practice Guidelines from Data

Subramani Mani[a], Constantin Aliferis[a], Shanthi Krishnaswami[b,c] and Theodore Kotchen[c]

subramani.mani@vanderbilt.edu constantin.aliferis@vanderbilt.edu shanthi.krishnas@vanderbilt.edu tkotchen@mcw.edu

[a] *Department of Biomedical Informatics, Vanderbilt University, Nashville TN 37232*
[b] *Department of Rheumatology, Vanderbilt University, Nashville TN*
[c] *Department of Medicine, Medical College of Wisconsin, Milwaukee WI*

Abstract

Clinical practice guidelines (CPG) propose preventive, diagnostic and treatment strategies based on the best available evidence. CPG enable practice of evidence-based medicine and bring about standardization of healthcare delivery in a given hospital, region, country or the whole world. This study explores generation of guidelines from data using machine learning, causal discovery methods and the domain of high blood pressure as an example.

Keywords:

clinical practice guidelines, machine learning, prediction, compliance, causal discovery, high blood pressure

Introduction and background

Clinical practice guidelines (CPG) have been developed to streamline practice of medicine in a complex healthcare environment. Guidelines are the foundation of evidence-based medicine (EBM) and provider compliance with evidence-based CPG facilitates the practice of evidence-based medicine.

Typically guidelines are developed by a committee of domain experts specifically constituted for the purpose. After a rigorous review and analysis of published evidence over a period of time, the committee publishes its set of guidelines. However, this process is laborious, time consuming and expensive. See for the example the joint national committee report on high blood pressure [1].

Based on published evaluations of CPG, Grimshaw and Russell showed that practice guidelines streamlined the process of care and contributed to better outcomes in general [2]. There has been considerable interest in guideline representation for making them computable. Many different frameworks for guideline implementation such as PRODIGY [3], Arden Syntax [4], GLIF [5], PROforma [6], Asbru [7] and EON [8] have been proposed.

The fields of machine learning (ML), data mining and knowledge discovery including methods for learning cause and effect relationships have matured over the years with applications to clinical and biological data. There have only been limited attempts at using machine learning and knowledge discovery methods for generating practice guidelines from data [9-12]. The goal of Abston et al. [9] was to discover the variation between the American College of Cardiology guidelines for management of acute myocardial infarction and documented practices in a tertiary care facility in Salt Lake city. The study did not address generation of new guidelines from data. Mani et al. [10] presented a two-stage machine learning model as a data mining method to develop clinical practice guidelines and showed its value in dementia staging. It modeled the methodology used by clinicians by deriving intermediate concepts in the first phase and using the intermediate concepts for dementia staging. However, it is not clear if the method is generalizable across different domains. Morik et al. [11] used a combination of prior knowledge from experts and learning from data for clinical protocol generation and validation. Sboner et al. [12] used machine learning techniques to model the decision making of dermatologists in melanoma diagnosis. Svatek et al. [13] describe a data mining approach based on association rule learning for checking guideline adherence. However, the clinical validity of the approach could not be ascertained due to the small sample size of the study.

Machine learning techniques have typically been explored for disease screening [14], differential diagnosis [15], and other outcome measures [16]. A method based on inductive logic programming for learning qualitative physiological models from clinical data has also been described [17]. However, to the best of our knowledge causal discovery methods have not been explored for guideline generation from data.

This paper explores automated generation and compliance checking of guidelines. Guidelines can be broadly classified as predictive guidelines or prevention/intervention guidelines based on their goals. Note that predictive guidelines are sufficient for diagnosis (diagnose disease D1 from symptoms S1 and S2). For prevention and intervention we need a cause and effect interpretation for the guidelines.

We now introduce a formal notion of causality. We define the *causal influence* of a variable A on a variable B using the *manipulation criterion* [18, 19]. The manipulation criterion states that if we had a way of setting just the values of A and then measuring B, the causal influence of A on B will be reflected as a change in the conditional distribution of B. That is, there exist values a_1 and a_2 of A such that $P(B \mid \text{set } A = a_1)$ $P(\boldsymbol{B} \mid \text{set } A = a_2)$. A causal influence of

variable *A* on variable *B* is represented as an arc from *A* to *B* i.e. *A* *B*. We say that variable *A* causally influences variable *B* if and only if *A* and *B* satisfy the manipulation criterion.

A causal influence of a variable *A* on a variable *B* is said to be *unconfounded* if and only if there is *no* measured or unmeasured variable *C* that is a common cause of variables *A* and *B*.

Materials and methods

Algorithms

For learning predictive guidelines we selected two machine learning algorithms with the following properties.

1. Perform classification (prediction) tasks well.
2. The generated models are comprehensible to humans.
3. The models can be easily implemented as computerized guidelines.

The C4.5 algorithm that uses the decision tree representation formalism [20] and RIPPER [21] that has the format of an *If ... Then rule* were selected. Decision trees and rules generate clear descriptions of how the ML method arrives at a particular classification.

For checking the cause and effect interpretation of the guideline we used the FCI algorithm [18]. The FCI algorithm takes as input a dataset D and outputs a graphical model consisting of edges between them that have a cause and effect interpretation. The FCI algorithm can handle hidden (unmeasured) variables and sample selection bias that are likely to be present in real-world datasets. There are other causal discovery algorithms (for example, PC [18]) that output a causal Bayesian network (CBN) model [22, 23], incorporating all the variables represented in a dataset. However, PC assumes that all the variables in a domain are observed and there are no unobserved variables. There are also causal discovery algorithms that take a local approach and output causal relationships of the form "variable *A* causally influences variable *B*". LCD [24] and BLCD [25] are two such algorithms. The local causal discovery algorithms are particularly suitable for large datasets.

In this study we apply C4.5, RIPPER and FCI to the high blood pressure (HBP) dataset that is described below.

Dataset

The prevlanence of high blood pressure in the US is approaching 30% and the rate of prevalence is also showing an increasing trend [26]. The dataset used in this work is part of an ongoing NIH funded study with Dr. Kotchen, T as the principal investigator and its goal is to ascertain genetic determinants and other causal factors of high blood pressure. The HBP dataset is a population based dataset consisting of data collected from consenting African Americans between the ages of 18 and 55 years in Milwaukee and neighboring areas. Anthropometric measurements included height, weight, waist, hip, arm circumferences and skinfold thickness measured at different sites. Consenting subjects who satisfied inclusion and exclusion criteria were admitted for a 2 day inpatient protocol to obtain additional hemodynamic and renal measurements under standardized controlled conditions. Exclusion criteria included secondary hypertension, diabetes, creatinine >2.2 mg/dl, body mass index (BMI) >35, recent stroke or myocardial infarction, malignancy and substance abuse including alcohol. Currently 369 people are enrolled (202 hypertensives and 167 normotensives). 47.3% of the normotensives and 53% of hypertensives were females. The average blood pressure of normotensive subjects was 114/74 vs. 147/96 in hypertensive subjects. 31% of the subjects were on antihypertensive medication. We selected 23 variables after excluding patient identifiers and redundant variables (variables derived from other variables present in the dataset). See Table 1 for the list of variables selected. The variables were categorized based on either the established risk levels of each variable for cardiovascular diseases or the study-specific cutpoints using the 90th / 10th percentile levels of values in normotensive subjects. The outcome (class) variable was coded H for hypertensive and N for normotensive based on the following guideline. If the outpatient blood pressure (OP-BP) was greater than or equal to 140/90 or the subject was on BP medication (BP-MED), the outcome variable (HBP) was coded H, otherwise it was coded N. All the independent variables were categorized as 0 and 1 (one representing risk for high blood pressure). Thus the guideline used for creating the outcome variable was our target hypothesis for learning which is given below.

If OP-BP = 1 or BP-MED = 1, HBP = H; **else** HBP = N.

We created two datasets DS1 and DS2 as follows. DS1 had 21 variables after excluding OP-BP and BP-MED that were used for generating the outcome HBP. The purpose of creating DS1 was to ascertain the causal factors for the outcome variable as a baseline before generating guidelines. DS2 included all the 23 variables. A third dataset DS3 was created from DS2 by toggling the value of the outcome variable for a randomly selected 10% of the subjects i.e. if the value was H it was changed to N and vice versa. This was to artificially create a set of 37 patients for whom the guideline was violated. DS3 was used for verification of guideline compliance. A fourth dataset DS4 was also created with data on this set of 37 patients for whom the class label was manipulated. C4.5 and Ripper were run using datasets DS2, DS3 and DS4. FCI was run on datasets DS1 and DS2.

The FCI program is available from the Tetrad project site at Carnegie Mellon (www.phil.cmu.edu/tetrad/tetrad4.html). We used the Java implementation of C4.5 and RIPPER available in the Weka machine learning software package [27]. We performed a ten fold cross-validation for C4.5 and RIPPER and report the results based on the test cases that were not used in model building. Note that for causal discovery cross-validation is not relevant because we are not performing classification or regression.

Table 1: Variables used in the study

Plasma Aldo/ plasma Renin ratio
Age
Cardiac output baseline
Creatinine clearance
Gender
High density lipoprotein
Heart rate at baseline
Insulin Resistance
Potassium excretion
Calculated Low density lipoprotein
Sodium excretion
Baseline renal Blood flow
Baseline Renal vascular resistance
Systemic Vascular resistance Index baseline
Stroke Volume baseline
Serum Triglycerides
Urine 24hrs Microalbumin
Waist circumference risk
Outpatient Hypertension yes/no
Glucose risk
Screening Height
Outpatient high BP /Normal BP
On antihypertensive medication

Results

Figures 1 and 2 present the results of the application of C4.5 and Ripper to the HBP DS2 dataset.

```
OP-BP = 0
|   BP-MED = 0: N
|   BP-MED = 1: H
OP-BP = 1: H

Number of Leaves  : 3
```

Figure 1 - A C4.5 tree from DS2

The C4.5 tree classified all the 369 instances correctly with an accuracy of 100%. The precision and recall were 1 for the H and N classes.

RIPPER

DS2 Ripper rules:

```
(OP-BP = 0) and (BP-MED = 0) => HBP=N
 => HBP = H
```

Figure 2 - A RIPPER rule set from DS2

The RIPPER classified all the 369 instances correctly with an accuracy of 100%.

Results of C4.5 and Ripper on DS3

Both C4.5 and Ripper misclassified the 37 instances for which the class assignments had been changed. All the other instances were classified correctly. Both the algorithms generated the same models shown in Figures 1 and 2.

Results of C4.5 and Ripper on DS4

The results of application of C4.5 and Ripper to the DS4 dataset are shown in Figure 3 and Figure 4 respectively. Recall that the DS4 dataset has just the 37 instances with labels manipulated.

```
OP-BP = 0
|   BP-MED = 0: H
|   BP-MED = 1: N
OP-BP = 1: N

Number of Leaves  : 3
```

Figure 3 - A C4.5 tree from DS4

The DS4 C4.5 tree classified 35 out of the 37 instances correctly with an accuracy of 95%. The precision for class N was 1 and recall for class H was 1.

RIPPER

DS4 Ripper rules:

```
(OP-BP = 0) and (BP-MED = 0) => HBP=H
 => HBP = N
```

Figure 4 - A RIPPER rule set from DS4

The RIPPER rule set classified 35 out of the 37 instances correctly with an accuracy of 95%.

FCI results

When applied to the HBP DS1 dataset, FCI output four possible causal factors for high blood pressure. Table 2 enumerates them. Note that when the relationship is categorized as "O>", there could be a common cause or a feedback loop.

Discussion

The machine learning algorithms C4.5 and Ripper recovered the study guideline that was used for assigning labels to the outcome variable when the guideline was followed in all the cases (DS2). The same guideline was also generated from the dataset when the guideline was not followed in 10% of the subjects (DS3). The subjects for whom the guideline was not followed were also identified as misclassified instances by C4.5 and Ripper. This shows that ML methods can be used for generating simple guidelines and guideline compliance checking.

Table 2 - The output of FCI from DS1

Aldo/Renin ratio O> HBP
Age O> HBP
Renal vascular resistance <> HBP
Waist risk <> HBP

A O> B Means there is definitely an arrowhead at B.
There may or may not be an arrowhead at A.
A <> B denotes that there is a common cause for A and B.

Table 3 - The output of FCI from DS2

OP-BP HBP
BP-MED HBP
HDL Cholesterol Insulin resistance
Insulin resistance Glucose risk

The results shown in Figure 3 and Figure 4 using the dataset DS4 with the 37 misclassified cases from DS3 shows the alternate guideline model generated for those cases. This is the alternate guideline that was followed for these instances. This shows that when a specified guideline is not followed, ML can be used to identify any alternate guideline that might have been used instead. Note though that the instances would be misclassified only if the application of the alternate guideline changed the outcome variable.

We also used a causal discovery algorithm (FCI) to ascertain the cause and effect basis of the guideline. Using DS1 FCI output four causal influences shown in Table 2. When DS2 was used OP-BP and BP-MED were output by FCI as unconfounded causal factors for HBP. Note that OP-BP and BP-MED are causal factors for HBP based on the manipulation criterion for causality. However, they are also causal based on the definition of the BP study guideline. Two other unconfounded causal relationships from the domain were also output by FCI (see Table 3).

Decision tree models and If … Then rules are expressive and easily interpretable by humans. Moreover, the tree and rule formats are also suitable for computerized guidelines and hence useful for incorporation as decision support tools in electronic medical record systems.

Our study addresses the question of generating new practice guidelines in a data driven way and explores the role of causal discovery along with traditional machine learning approaches for guideline generation from data.

Two relevant issues that come up in guideline application are generalizability and customization. A guideline developed in one institution or organization may not be exactly applicable in another practice setting. Likewise, a guideline developed by a committee of national or international experts might need to be customized to a local setting. Fridsma et al. have developed a knowledge-based approach to customization based on the separation of site specific and site independent factors that can be identified from the knowledge of the organization and understanding of its workflow [28]. We believe that data driven machine learning approaches could be a useful tool in the overall effort to make guidelines generalizable and customizable.

Limitations

The dataset that we used in this study was from a population-based study of high blood pressure. There were only a small number of variables in the dataset. Only two machine learning and one causal discovery algorithm were used in this study. The guidelines were also very simple.

Conclusions and future work

In this paper we presented a machine learning approach to generate guidelines from data, check for guideline compliance and if non-compliant for a set of patients, generate the alternate guideline used. We also provided a method for ascertaining whether the guideline has a causal semantics using a causal discovery algorithm.

In future we plan to apply machine learning and causal discovery algorithms to different medical datasets involving more complex guidelines for further evaluation of our approach.

References

[1] Chobanian, A, Bakris, G, Black, H, Cushman, W, Green, L, Izzo, J, Jones, D, Materson, B, Oparil, S, Wright, J, Roccella, E and NHBPEPCC. Seventh report of the joint national committee on prevention, detection, evaluation, and treatment of high blood pressure. Hypertension l2003;42: 1206-1252.

[2] Grimshaw, JM and Russell, IT. Effect of clinical guidelines on medical practice: a systematic review of rigorous evaluations. Lancet l1993;342: 1317-22.

[3] Purves, I, Sugden, B, Booth, N and Sowerby, M. The PRODIGY project--the iterative development of the release one model. In: AMIA fall symposium; 1999. p. 359-63.

[4] Clayton, P, Pryor, T, Wigertz, O and Hripcsak, G. Issues and structures for sharing knowledge among decision-making systems: The 1989 Arden Homestead Retreat. In: K. LC, editor. Thirteenth Annual Symposium on Computer Applications in Medical Care: IEEE Computer Society Press; 1989. p. 116-21.

[5] Ohno-Machado, L, Gennari, J, Murphy, S, Jain, N, Tu, S, Oliver, D, Pattison-Gordon, E, Greenes, R, Shortliffe, E and Barnett, G. The guideline interchange format: a model for representing guidelines. JAMIA l1998;5: 357-72.

[6] Fox, J, Johns, N and Rahmanzadeh, A. Disseminating medical knowledge: The PROforma approach. Artiff Intell Med l1998;14: 157-81.

[7] Miksch, S, Shahar, Y and Johnson, P. Asbru: a task-specific, intention-based, and time-oriented language for representing skeletal plans. In: Seventh Workshop on Knowledge Engineering Methods and Languages (KEML-97): Milton Keynes, UK; 1997.

[8] Musen, M, Tu, S, Das, A and Shahar, Y. EON: a component-based approach to automation protocol-directed therapy. JAMIA l1996;3: 367-88.
[9] Abston, K, Pryor, T, Haug, P and Anderson, J. Inducing practice guidelines from a hospital database. In: JAMIA Supplement; 1997. p. 168--172.
[10] Mani, S, Shankle, WR, Dick, MB and Pazzani, MJ. Two-Stage Machine Learning Model for Guideline Development. In: Artificial Intelligence in Medicine; 1998.
[11] Morik, K, Imboff, M, Brockhausen, P, Joachims, T and Gather, U. Knowledge discovery and knowledge validation in intensive care. Artiff Intell Med l2000;19: 225-249.
[12] Sboner, A and Aliferis, CF. Modeling clinical judgment and implicit guideline compliance in the diagnosis of melanomas using machine learning. In: AMIA fall symposium; 2005. p. 664-668.
[13] Svatek, V, Riha, A, Peleska, J and Rauch, J. Analysis of guideline compliance---a data mining approach. In: K. Kaiser, S. Miksch and S. W. Tu, editors. Symposium on Computerized Guidelines and Protocols: IOS Press; 2004. p. 157-161.
[14] Shankle, W, Mani, S, Pazzani, M and Smyth, P. Detecting Very Early Stages of Dementia from Normal Aging with Machine Learning Methods. In: E. Keravnou, C. Garbay, R. Baud and J. Wyatt, editors. Lecture Notes in Artificial Intelligence: Artificial Intelligence in Medicine, AIME97: Springer; 1997. p. 73--85.
[15] Ohmann, C, Yang, Q, Moustakis, V, Lang, K and Elk, van PJ. Machine Learning Techniques Applied to the Diagnosis of Acute Abdominal Pain. In: P. Barahona and M. Stefanelli, editors. Lecture Notes in Artificial Intelligence: Artificial Intelligence in Medicine, AIME95: Springer; 1995. p. 276--281.
[16] Cooper, G, Aliferis, C and Ambrosino, R. An evaluation of machine learning methods for predicting pneumonia mortality. Artificial Intelligence in Medicine l1997;9: 107--138.
[17] Hau, DT and Coiera, EW. Learning qualitative models of dynamic systems. Machine Learning l1993;26: 177-211.
[18] Spirtes, P, Glymour, C and Scheines, R. Causation, Prediction, and Search. In. Cambridge, MA: MIT Press; 2000.
[19] Cooper, GF. An overview of the representation and discovery of causal relationships using Bayesian networks. In: C. Glymour and G. F. Cooper, editors. Computation, Causation, and Discovery. Cambridge, MA: MIT Press; 1999. p. 3--62.
[20] Quinlan, J. C4.5: Programs for Machine Learning. In. Los Altos, California: Morgan Kaufmann; 1993.
[21] Cohen, WW. Fast effective rule induction. In: ICML; 1995. p. 115-123.
[22] Pearl, J. Probabilistic Reasoning in Intelligent Systems. In. San Francisco, California: Morgan Kaufmann; 1991.
[23] Neapolitan, RE. Probabilistic Reasoning in Expert Systems: Theory and Algorithms. In. New York: John Wiley and Sons; 1990.
[24] Cooper, GF. A Simple Constraint-Based Algorithm for Efficiently Mining Observational Databases for Causal Relationships. Data Mining and Knowledge Discovery l1997;1: 203--224.
[25] Mani, S. A Bayesian Local Causal Discovery Framework (doctoral dissertation). In: Intelligent Systems Program: University of Pittsburgh; 2005.
[26] Hajjar, I and Kotchen, TA. Trends in prevalence, awareness, treatment, and control of hypertension in the United States, 1988 - 2000. JAMA l2003;290: 199-206.
[27] Witten, IH and Frank, E. Data Mining: Practical machine learning tools and techniques. In. 2nd ed: Morgan Kaufmann, San Francisco; 2005.
[28] Fridsma, DB, Gennari, J and Musen, M. Making generic guidelines site-specific. In: AMIA fall symposium; 1996.

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Supporting Therapy Selection in Computerized Clinical Guidelines by Means of Decision Theory

Stefania Montani, Paolo Terenziani, Alessio Bottrighi

DI, Universiy of Piemonte Orientale,Alessandria, Italy

Abstract

Supporting therapy selection is a fundamental task for a system for the computerized management of clinical guidelines (GL). The goal is particularly critical when no alternative is really better than the others, from a strictly clinical viewpoint. In these cases, decision theory appears to be a very suitable means to provide advice. In this paper, we describe how algorithms for calculating utility, and for evaluating the optimal policy, can be exploited to fit the GL management context.

Keywords:

Decision Theory, Clinical Guidelines

Introduction

Clinical guidelines (GL) can be defined as a means for specifying the "best" clinical procedures and for standardizing them. In recent years, the medical community has started to recognize that a computer-based treatment of GL provides relevant advantages, such as automatic connection to the patient databases and, more interestingly, decision making facilities; thus, many different approaches and projects have been developed to this hand (see e.g. [1,2]). As a matter of fact, decision making is a central issue in clinical practice. In particular, supporting therapy selection is a critical objective to be achieved. Consider that, when implementing a GL, a physician can be faced with a choice among different therapeutic alternatives, and identifying the most suitable one is often not straightforward. Unlike clinical protocols [3], that specify what is the only admissible procedure in a given situation, GL are used in domains in which different choices are actually possible (see the example in the results section). Alternatives can be pruned relying both on *site-related* contextual information (e.g. due to the unavailability of certain resources in a given hospital), and on *patient-related* contextual information (due to the peculiarities of the single patient on which the GL is being applied). The works in [4, 5] show how decision models can be resorted to in order to help physicians in defining and contextualizing GL. However, also when the GL has been properly contextualized, it is frequent to find more then one option left, and sometimes no one of these remaining alternatives is really "better" than the others, from a strictly clinical viewpoint. In clinical practice, various selection parameters (such as the costs and the effectiveness of the different procedures) can be available to physicians when executing a GL. The computer-based GL systems described in the literature offer sophisticated formalizations of these decision criteria. Nevertheless, the available information is often only qualitative in nature, and "local" to the decision at hand: it does not take into account the consequences of the choice, in terms of actions to be implemented, and future decisions to be taken along the path stemming from the selected alternative. On the other hand, the possibility of obtaining a complete scenario of the decision consequences, considering the probability of the different therapy outcomes, the utilities associated to the different health states, and the money, time and resources spent, would be clearly an added value for physicians and hospital administrators. Decision theory seems a natural candidate as a methodology for covering this task. To this hand, a systematic analysis of the main GL representation primitives, and of how they could be related to decision theory concepts has been recently proposed [6]. Since, at a sufficiently abstract level, the GL representation primitives treated in that work are shared by all the systems in the literature [7], that contribution can be seen as the first step towards the implementation of a tool within any of the such approaches. In this paper, we start from such knowledge representation results, to describe how decision theory algorithms (to calculate utility and to obtain the optimal policy) can be exploited, when the goal is the one of supporting therapy selection in a GL management system. We also analyse complex situation which may arise due to the presence of certain types of control flow relations among GL actions, namely iterations and parallel executions. A practical application of this work is represented by the tool which is being implemented in GLARE, a domain-independent system for GL acquisition and execution [8, 9]. The algorithmic choices, and the technical issues discussed in this paper will therefore refer to this specific example. In particular, an earlier version of GLARE already embedded a facility able to calculate costs, time and resources required to complete paths in a GL (details can be found in [9], and are briefly sketched in the next section); the decision theory support can be seen as an extension of that work. The paper is structured as follows: in the next section we clarify the goal of our decision theory tool and we summarize the previous results in the direction of supporting therapy selection, i.e. the concept

mapping work and the main features of the GLARE's cost-collection facility. In the results section we describe technical issues about the implementation (referring to the specific example of GLARE). Finally the last section addresses some concluding remarks.

Materials and methods

Designing the main features

We envision the possibility of adopting a decision theory tool for supporting therapy selection in two fashions. First, it can operate in the on-line modality, when applying the GL actions one at a time to the patient at hand, by automatically retrieving the patient's data from the Hospital Information System (HIS). In this case, at the time at which a therapeutic decision has to be taken, the facility is able to provide local pros and cons of the various alternatives. Secondly - and more interestingly - the tool can be used off-line, if the physician wants to make a simulation of the consequences of a therapeutic alternative, by evaluating the patient's evolution along the different paths stemming from the decision at hand, typically until the end of the GL is reached. The possibility of collecting this global information is crucial to allow her making a well informed choice. This modality would be useful also for education purposes. From the algorithmic viewpoint, this working mode generalizes the first one; therefore, in the rest of the paper, we will concentrate on the off-line modality. Off-line simulation typically involves a series of temporally consequent decisions. The clinical GL can therefore be seen as a dynamic decision problem. In particular, as we will briefly motivate below, the GL can be mapped onto a (completely observable) Markov Decision Process (MDP), in which the sequence of therapeutic decisions generates the sequence of (patient) states. The typical goal of a decision theory tool is to find the optimal policy, i.e. the sequence of decisions able to maximize the expected utility. In the context of GL, it is possible to adapt and simplify this task by limiting the decisions to be considered to therapy selections among clinically equivalent alternatives (see the example in the results section); we will call them non-trivial decisions henceforth. Moreover, we propose to realize an implementation in which also costs, resources and time spent to complete any path in the GL can be obtained, and can be coupled with the calculation of the expected utility along the path itself. Note that fixing the path means fixing the policy that has to be applied, i.e. knowing which alternative will be chosen at any decision. Finally, the user should be always allowed to select the part of the GL s/he wants to focus on (as a default, the overall GL will be taken into account). The possibility of selecting only a portion of the GL seems to us particularly relevant, since we aim at supporting only non-trivial decisions, while the GL will typically include paths where therapeutic choices do not require the adoption of the decision theory facility; moreover, concentrating only on a subpart of the GL will obviously reduce the computation time. All the technical details about a concrete implementation of a decision theory tool within the system GLARE are described in the results section.

Previous work

Concept mapping

In [6], a knowledge representation contribution, aimed at mapping the GL primitives to decision theory concepts, was provided. In particular, at a sufficiently abstract level, GL representation formalisms share the following assumptions (for the terminology used here, we refer in particular to [8, 10]). First, a GL can be represented as a graph, where nodes are the *actions* to be executed, and arcs are the *control relations* linking them. It is possible to distinguish between *atomic* and *composite* actions (plans), which can be defined in terms of their atomic components via the *has-part* relation. Three different types of *atomic actions* can then be identified: (1) *work actions*, i.e. actions that describe a procedure which must be executed at a given point of the guideline; (2) *query actions*, i.e. requests of information from the outside world; (3) *decision actions*, used to model the selection among different alternatives. Decision actions can be further subdivided in *diagnostic decisions*, used to make explicit the identification of the disease the patient is suffering from, and *therapeutic decisions*, used to represent the choice of a path in the GL, containing the implementation of a particular therapeutic process (henceforth, we will concentrate on (non-trivial) therapeutic decisions, that we want to support). *Control* relations establish which actions can be executed next, and in what order. For example, actions could be executed in *sequence*, or in *parallel*. Moreover, the *alternative* relation describes how alternative paths can stem from a decision action, and the *repetition* relation states that an action has to be repeated several times (maybe a number of times not known a priori, until a certain *exit condition* becomes true). In a well-formed GL, a decision action is preceded by a query action, that is adopted to collect all the patient's parameters necessary (and sufficient) for taking the decision itself. Each decision is therefore based on an (explicit or implicit) data collection completed at decision time, and does not depend on the previous history of the patient. We can thus say that the GL describes a discrete-time first-order Markov model, since each time a query action is implemented, the patient's situation is completely re-assessed, and an (explicit or implicit) query action is always found before a decision action. This observation justifies the mapping of GL primitives to the field of decision theory, and in particular allows us to represent a GL as a MDP. Despite the fact that Markov processes have been often applied to medical domains (e.g. for disease management in [11]), the difficulties in relying on them to simulate clinical processes are well known. In particular, the resulting model can be very demanding not only with regard to the amount of data needed to specify the probability distribution underlying the stochastic process, but also computationally [12]. Nevertheless, these limitations appear to be less critical in the domain of clinical guidelines, where rather strict design policies are typically applied. Therefore, when dealing with GL, some simplifications hold. In particular, as already observed, a first-

order Markov model is sufficient to capture the GL dynamics. Moreover, the process modelled by the GL is completely observable, since in a GL a decision can be taken only if all the required parameters have been collected: if some needed data are missing, the query action will wait for them and the decision will be delayed. It is then straightforward to define the *state* as the set of patient's parameters that are normally measured for taking decisions and for assessing therapy outcomes. Query actions are the means for observing the state. *State transitions* are produced by all the work actions between two consecutive non-trivial therapeutic decisions. Finally, the *utility* of a state can be evaluated in terms of life expectancy, corrected by Quality Adjusted Life Years (QALYs) [13].

The cost-collection facility

GLARE already incorporates a decision support facility, able to assist physicians in choosing among therapeutic alternatives [9]. Relying on this tool, it is possible to compare different paths in the GL, by simulating what could happen if a certain choice was made. In particular, users are helped in calculating the "cost" of the paths themselves, in order to select the cheapest choice. Costs are not interpreted just as monetary expenses, but also as resources and time required to complete GL actions. Note that, when running the tool, if a composite action is found, it is expanded in its components, and the reasoning facility is recursively applied to each of them, by analysing all the decision actions that appear at the various decomposition levels. At the end of this process, the tool displays the values of the collected parameters (costs, resources, times) gathered along each path. The final decision is then left to the physician.

Results

Within GLARE, the facility described in the previous section is being extended, by allowing: (1) the identification of the optimal policy, and (2) the calculation of the expected utility along a path. In order to implement these functionalities, we had to take into account the following issues.

Focusing

As already observed, the possibility of selecting only a sub-part of a given GL is a fundamental issue to be addressed, since it allows one to skip the paths on which decision theory support is not required. In our tool, path selection has been conceived as the first step of the interaction with the user. Technically speaking, the mechanism works as follows: through a user-friendly graphical interface, the physician is asked to indicate the starting node (normally the decision at hand) of the paths to be compared and (optionally) the ending nodes (otherwise all possible paths exiting the starting node will be taken into consideration, until the end of the GL). For every decision action within each path, s/he is allowed to restrict to a subset of alternatives. Moreover, the selection process is recursively applied to composite actions. All the paths pruned by this procedure will be ignored by the subsequent steps of the reasoning process (i.e. mapping to the Markov model and extraction of the optimal policy, or calculation of the expected utilities).

Parallel actions

In case two or more composite actions, each one containing non-trivial decisions, have to be executed in parallel along a selected path, the mapping towards the corresponding Markov model (needed to provide functionality 1 above) is not straightforward. As a matter of fact, in this situation the order of execution of the various actions is not univocally provided. The policy we have chosen to adopt to this hand is the one of calculating just one possible order of execution, compatible with the available temporal constraints [14], and to rely on it.

Generating the Markov model

Once path selection and parallel actions management have been addressed, the next step towards the calculation of the optimal policy (see functionality 1 above) is the mapping of the GL to the corresponding Markov model. The algorithm, which relies on the concept mapping results described in the previous section, automatically produces the conversion by operating as follows: for every selected path, in correspondence to a non-trivial decision (or to an exit point of the GL) it generates a new state (unless the same state was already identified). It then collects all the work actions between two consecutive states, and builds a macro-action; the macro-action determines the transition between the two states with a given probability. Probability values are extracted from the medical literature - when possible; otherwise, in the current implementation, they can be obtained from interviews with expert physicians. Note that, for those medical fields in which the medical literature does not provide these numbers, it is reasonable to expect this information to be available in the near future. As a matter of fact, the increasing exploitation of HIS and of computerized GL management tools will allow for the collection of large amounts of clinical practice data, on which it will be easy to draw statistics, at least at the local level. Consider also that relying on local data is not necessarily a limitation: remember that a guideline always needs to be contextualized to the features of the hospital in which it has to be implemented, before its exploitation begins [15].

Repeated actions

On the Markov model, classical algorithms for evaluating the optimal policy can be relied on [16, 17]. When dealing with a finite time horizon, the dynamic programming algorithm can be easily applied [16]. Nevertheless, in the context of clinical GL, it is not infrequent that the number of states, though finite, is not known a priori. This situation may be induced by the presence of iterations to be repeated several times, until a certain exit condition becomes true. The number of repetitions could therefore vary among different executions of the same GL. To handle this problem, the choice made in GLARE is the one of relying on algorithms for calculating the optimal policy on an infinite time horizon (as a matter of fact, "infinite" can be used in

the meaning of "unknown a priori"), and in particular on value iteration [17].

Simplifications

Algorithms can be simplified in case the required output is not the optimal policy, but the expected utility of all the different paths selected by the user (see functionality 2 above). Note that following a path corresponds to apply a specific policy, i.e. to make the hypothesis of knowing what is the decision to be taken at any decision node. In case of a finite time horizon, since the policy is known, we can calculate the utility by applying the dynamic programming algorithm, avoiding to maximize the expected value of the cumulative utility function with respect to the different actions. The corresponding costs can be summed up resorting to the functionality described in the previous section. On the other hand, when iterations with an exit condition have to be tackled, we can ask the user physician the minimum and maximum number of times that the action has to be repeated, given the specific patient's characteristics. Then, we can generate the paths corresponding to these two extreme situations, thus reducing to a finite time horizon, on which it is possible to calculate utility and costs as described above. In the case of utility, however, we can also keep working on an infinite time horizon, not having to rely on the physician's estimate of the number of iterations, and resort to value iteration. Again, since the policy is fixed, the algorithm can be simplified by avoiding maximization wrt the different actions. This second strategy is clearly not applicable to costs, which are additive. It is up to the user physician to select the preferred output in these cases. As an example, we present an application of the GLARE decision theory tool to a GL for asthma treatment. Figure 1 shows part of the GL. Patients affected by mild persistent asthma (see upper branch in the figure) may be treated by four different therapies (as indicated by the four edges exiting the T1 node): inhaled beta-2-agonists (A), oral beta-2-agonists (OA), inhaled anticholinergics (IA) or theophylline (TH). Each therapy implementation consists of a daily dose. Basically, the four drugs for mild asthma are clinically equivalent; therefore, indications about implications of each alternative could be useful in deciding. If the therapy does not work, the patient could worsen to moderate asthma. In this case, another therapeutic decision has to be taken (action T2), in order to implement a more effective treatment. It is possible to select between inhaled steroids (S) and inhaled steroids plus bronchodilators (SB). Again, these drugs have to be provided daily. Periodically (e.g. weekly), the patient's state is re-assessed, and the therapeutic decision has to be repeated in a loop, until this guideline becomes not applicable for the patient at hand (because asthma is now severe, and a different guideline has to be referred to, or because asthma improves: this case is not explicitly represented in figure 1). In the GLARE formalism, the treatment of mild asthma has to be represented as an iterated plan with an exit condition (see the previous section), that corresponds to the onset of moderate asthma. Analogous considerations hold for the treatment of moderate asthma. The utility of the mild persistent asthma state is 86, while the utility of moderate asthma is 82 and the one of severe asthma is 78 [18]. Table 1 lists the probabilities of transition, that were provided by medical experts (see acknowledgements). Since the GL includes two cycles, which are repeated a number of times not known a priori, in this example the optimal policy has been calculated using the value iteration algorithm, which has identified A as the optimal therapy in case of mild asthma, and SB as the optimal therapy in case of moderate asthma. We have applied (simplified) value iteration also for calculating the utility of all the possible paths in the GL; the corresponding values are reported in table 2.

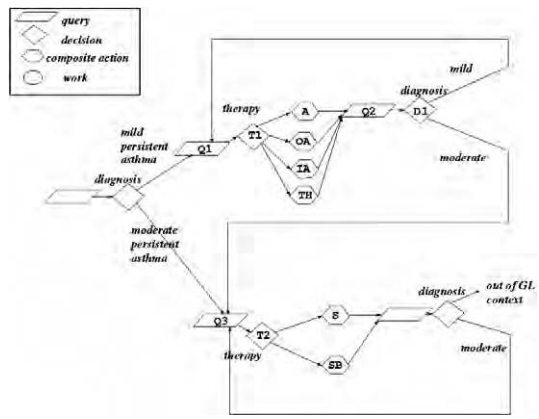


Figure 1 - Part of the asthma treatment guideline. Node types legend is provided as well.

Table 1 - Probabilities of transition between states.

Therapy	From	To	Prob.
A	Mild	Mild	0.9
A	Mild	Moderate	0.1
OA	Mild	Mild	0.85
OA	Mild	Moderate	0.15
IA	Mild	Mild	0.6
IA	Mild	Moderate	0.4
TH	Mild	Mild	0.5
TH	Mild	Moderate	0.5
S	Moderate	Moderate	0.85
S	Moderate	Severe (out of GL)	0.15
SB	Moderate	Moderate	0.95

Therapy	From	To	Prob.
SB	Moderate	Severe (out of GL)	0.05

Conclusions

In the context of clinical GL, it is not infrequent to identify actions of therapeutic selection which would benefit from a decision theory support. Embedding a decision theory facility would therefore be an added value for a computerized system for GL management. In this paper, we have discussed the main features that should characterize such a tool. In particular, we have identified what simplifications can be made with respect to classical decision theory approaches, meant to be applied to a generic domain. Moreover, we have underlined what integrations and specific choices are required, on the other hand, to deal with non trivial GL control flow constructs, namely iterations and parallel executions. These considerations appear to be general enough to be exploited when designing a decision theory tool within any of the systems described in the literature. As an example, we have presented here the features of the system GLARE's facility. We believe that a tool developed along these lines would be able to provide a valuable support to physicians, thus reinforcing the claim that the adoption of AI techniques can provide relevant advantages in the (semi)-automatic treatment of clinical GL, and favouring the actual exploitation of computer science facilities within the medical community.

Table 2 - Utilities of the asthma GL states having fixed the various policies (i.e. paths). Numbers have been normalized with respect to the maximum value.

Policy	Utility Mild asthma	Utility Moderate
A-S	0.57590	0.24231
A-SB	1	0.66641
OA-S	0.46470	0.24231
OA-SB	0.88879	0.66640
IA-S	0.32570	0.24230
IA-SB	0.74979	0.66639
TH-S	0.30902	0.24230
TH-SB	0.73311	0.66639
A-S	0.57590	0.24231

Acknowledgments

Information about utilities and state transition probabilities in the example were obtained from Prof. G. Molino and Prof. M. Torchio of Azienda Ospedaliera S. Giovanni Battista in Turin, with whose cooperation the implementation work is being carried on. This research has been partially supported by a grant from Regione Piemonte.

References

[1] Fridsma D.B. (Guest Ed.) Special issue on workflow management and clinical guidelines. Journal of the American Medical Informatics Association, 1(22), 1-80,2001.

[2] Gordon C. and Christensen J.P. editors. Health Telematics for Clinical Guidelines and Protocols. IOS Press, Amsterdam, 1995.

[3] Purves I., Computerised guidelines in primary health care: Reflections and implications, Telematics for Clinical Guidelines and Protocols, C.Gordon and J.P. Christensen, 57-74, IOS Press, Amsterdam, 1995

[4] Sanders G.D., Nease R.F., Owens D.K., Design and pilot evaluation of a system to develop computer-based site-specific practice guidelines from decision models, Medical Decision Making 20, 145-149, 2000

[5] Sanders G.D., Nease R.F., Owens D.K., Publishing web-based guidelines using interactive decision models, Journal of Evaluation in Clinical Practice, 20, 145-159, 2001

[6] Montani S., Terenziani P., Exploiting decision theory concepts within clinical guideline systems: towards a general approach, International Journal of Intelligent Systems 21(2006) 585-599

[7] Peleg M. and al., Comparing models of decision and action for guideline-based decision support: a case-study approach, JAMIA, ed. American Medical Informatics Association, 10, 52-68, 2003

[8] Terenziani P., Molino G., Torchio M., A modular approach for representing and executing clinical guidelines, Artificial Intelligence in Medicine, 23(3), 249-276, 2001

[9] Terenziani P., Montani S., Bottrighi A., Torchio M., Molino G., Supporting physicians in taking decisions in clinical guidelines: the GLARE "what-if" facility, Journal of the American Medical Informatics Association (JAMIA) Symposium supplement, 772-776, 2002

[10] Fox J., Johns N., Rahmanzadeh A., Thomson R., Disseminating medical knowledge: the PROforma approach, Artificial Intelligence in Medicine, 14(1-2), 157-182, 1998

[11] Lucas P, de Bruijn N., Schurink K, Hoepelman A., A probabilistic and decision-theoretic approach to the management of infectious disease at the ICU, Artificial Intelligence in Medicine 19 (2000) 251–279

[12] Boutilier C., Dean T., Hans S., Decision-theoretic planning: structural assumptions and computational leverage. J Artif Intell Rev 1999;11:1–94.

[13] Gold M.R., Siegel J.E., Russell L.B., Weinstein M. C., Cost-Effectiveness in Health and Medicine, Oxford University Press, New York, 1996

[14] Vila L., A survey on temporal reasoning in artificial intelligence, AI Communications, 1(7), 4-28, 1994

[15] Terenziani P., Montani S., Bottrighi A., Torchio M., Molino G., Correndo G., A context-adaptable approach to clinical guidelines, Proc. Medinfo (World Congress for Health Informatics) 2004, San Francisco, September 2004, M. Fieschi et al. eds., IOS Press, Amsterdam, 169-173

[16] Bellman R., Dynamic programming, Princeton University Press, 1957

[17] Tijms H.C., Stochastic modelling and analysis: a computational approach, Wiley and Sons, 1986

[18] Moy M.L., Israel E., Weiss S.T., Juniper E.F., Dube L., Drazen J.M., Clinical predictors of health-related quality of life depend on asthma severity, Am. J. Respir. Crit. Care Med., 163(4), 924-929, 2001

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Development, Deployment and Usability of a Point-of-Care Decision Support System for Chronic Disease Management Using the Recently-Approved HL7 Decision Support Service Standard

David F. Lobach, MD, PhD, MS[a]; Kensaku Kawamoto, PhD[a]; Kevin J. Anstrom, PhD[b]; Michael L. Russell, MD[c]; Peter Woods[c]; and Dwight Smith[c]

[a]Division of Clinical Informatics, Department of Community and Family Medicine; [b]Department of Biostatistics and Bioinformatics; and [c]Duke Health Technology Solutions; Duke University, Durham, North Carolina, USA

Abstract

Clinical decision support is recognized as one potential remedy for the growing crisis in healthcare quality in the United States and other industrialized nations. While decision support systems have been shown to improve care quality and reduce errors, these systems are not widely available. This lack of availability arises in part because most decision support systems are not portable or scalable. The Health Level 7 international standard development organization recently adopted a draft standard known as the Decision Support Service standard to facilitate the implementation of clinical decision support systems using software services. In this paper, we report the first implementation of a clinical decision support system using this new standard. This system provides point-of-care chronic disease management for diabetes and other conditions and is deployed throughout a large regional health system. We also report process measures and usability data concerning the system. Use of the Decision Support Service standard provides a portable and scalable approach to clinical decision support that could facilitate the more extensive use of decision support systems.

Keywords:

clinical decision support, web services, service-oriented architecture, decision support service, Health Level 7

Introduction

The volume of clinical knowledge reported in the biomedical literature is rapidly increasing [1]. However, the dissemination of this medical knowledge through traditional channels (e.g., publication of evidence-based guidelines, continuing medical education) has been found to be inadequate [2]. In the United States, a recent nationwide audit assessing 439 quality indicators found that American adults receive only about half of recommended care [3], and the U.S. Institute of Medicine has estimated that up to 98,000 Americans die each year as the result of preventable medical errors [4]. Similar deficiencies in care quality have been found in other industrialized nations [5].

Selected for best paper award.

In seeking to address this crisis in care quality, one of the most promising strategies for optimizing patient care and ensuring patient safety involves the use of computer systems that support clinical decision making [6]. Such clinical decision support (CDS) systems represent one of the most effective means for improving clinician compliance with evidence-based care standards [7]. The utilization of CDS systems, however, remains limited in most healthcare facilities [8]. While multiple factors have contributed to this limited adoption of CDS systems, one important factor has been the lack of an efficient method for encapsulating, processing, and delivering executable medical knowledge for use in clinical software applications [9].

In attempting to overcome the difficulty of re-using medical knowledge encoded in a machine-executable format, knowledge engineers have generally taken two approaches [10]. As one approach, systems such as PRODIGY [11], SAGE [12], and First DataBank's Drug Information Framework™ [13] provide access to their executable knowledge base using standard application programming interfaces. As a second approach, methods including GLIF3 [14], GEM [15], and Arden Syntax [16] encode knowledge using a common formalism, so that encoded rules can be consistently interpreted by system-specific interpreters. Despite these significant efforts, a dominant framework has not emerged for sharing executable medical knowledge, due in part to the following challenges. First, some formalisms, such as the Drug Information Framework™ [13], focus on specific knowledge domains and are not extendable to other domains. Second, many formalisms are designed for use in specific types of CDS applications and are difficult to adapt for use in other types of applications. Third, many formalisms are difficult to understand due to their conceptual complexity. Finally, many existing methods require significant investments in infrastructure, such as a system-specific compiler [17].

With regard to actual systems, most CDS systems are designed for specific institutional settings. Thus, despite repeated validation of their effectiveness, the utilization of CDS systems remains the exception rather than the rule in most practices across the country. This underutilization is

largely because CDS systems with demonstrated efficacy have generally been designed as an extension to a specific electronic health record (EHR) system and cannot be easily transferred to clinical sites that use a different EHR or no EHR at all [18]. Thus, while several CDS systems have validated efficacy [7], they have not provided a general, widely implementable solution to the problem of substandard care.

Given this portability problem, several CDS systems have been developed as stand-alone systems that operate independent of existing information systems [19, 20]. However, the lack of system-to-system integration introduces new problems. Stand-alone systems require substantial and continuous data entry to maintain the current and complete patient information required for the generation of accurate care recommendations. In addition, many stand-alone systems require clinicians to proactively access the system. The reliance on such proactive usage is particularly troublesome when the system only needs to be accessed for a subset of patients. As a result, stand-alone systems oftentimes fail to become a part of routine workflow, which is critical to a CDS system's ability to improve clinical practice [7, 18].

We have developed a CDS Web service that overcomes many of these challenges known as the SEBASTIAN Decision Support Service (DSS) [21]. A few other sites have also used a service-based approach to decision support [22, 23]. The service interface of the SEBASTIAN DSS has provided the foundation of a new HL7 draft standard known as the HL7 Decision Support Service standard, which was formally adopted as a draft HL7 standard during the September 2006 ballot cycle [24].

In this paper, we describe the first production use of a decision support system based on the new HL7 DSS standard. We also report on process measures related to the use of the system, the usability of the application enabled by the DSS, and lessons learned from this experience. The results of our efforts can assist others in using the HL7 DSS in other settings.

Materials and methods

SEBASTIAN DSS

We have previously developed and described the SEBASTIAN (an acronym for System for Evidence-Based Advice through Simultaneous Transaction with an Intelligent Agent across a Network) DSS, which enables machine-executable medical knowledge to be re-used across applications and institutions [21]. The SEBASTIAN DSS is implemented as a Web service, in which software functionality is provided over the Internet and extensible markup language (XML) messages are used to communicate with client systems [25].

Point-of-care chronic disease management system enabled by SEBASTIAN DSS

Implementation setting. The SEBASTIAN DSS was used to implement a point-of-care chronic disease management system within the Duke University Health System, located in Durham, North Carolina. The health system handles over 60,000 hospitalizations and 1.2 million outpatient encounters a year.

System functionality. The chronic disease management system was added as a tab in the patient summary section of the Duke electronic record viewing system, known as the eBrowser. The system currently provides care recommendations for diabetes management. The system is currently being enhanced to support preventive health reminders and disease management for hypertension, asthma, and dyslipidemia.

System architecture. The system architecture of the chronic disease management system is shown in Figure 1. For simplicity's sake, this figure and the accompanying text description focus solely on how the system generates care recommendations for diabetes management. Care recommendations for health maintenance and for other chronic medical conditions will be generated in a manner that parallels the information flow described below.

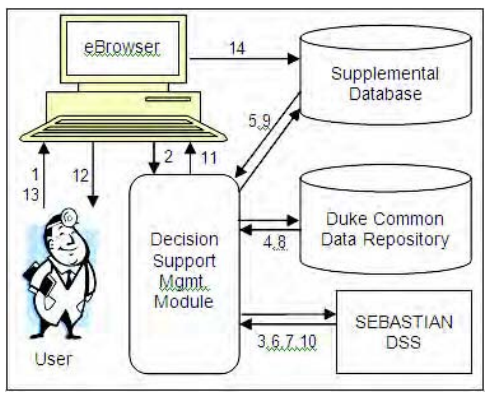


Figure 1 - Architecture of Duke Chronic Disease Management System

The information flow in the chronic disease management system is coordinated by a Decision Support Management Module. This module interacts with the SEBASTIAN DSS in order to identify the data required for evaluating a patient and to obtain machine-interpretable conclusions regarding a patient. This module also interacts with two patient data sources to retrieve the data required by SEBASTIAN.

According to this design, the clinician enters an identification number for a specific patient into the Duke eBrowser (arrow #1). The eBrowser passes this number to the Decision Support Management Module (#2). This module then contacts the SEBASTIAN DSS to find out what data are needed to determine if this particular patient has diabetes mellitus (#3). The management module then obtains the required data from the primary clinical database, known as the Duke Common Data Repository (CDR) (#4), as well as from a supplemental database that contains data pertinent to chronic disease management but not captured in the CDR (#5). The management module then sends these data to the SEBASTIAN DSS to determine if the

patient does have diabetes and receives back a result indicating whether or not the patient is diabetic (#6). If the patient is diabetic, the management module then inquires of the SEBASTIAN DSS what data are needed to run the diabetes management rules and receives back the list of required data (#7). Next, the management module obtains the required data from the CDR and the supplemental database (#8, 9). The module then sends these data to the SEBASTIAN DSS and receives back the recommendations for the chronic management of diabetes (#10). The management module formats these data and posts the results to the eBrowser (#11), where they are then viewed by the clinician (#12). If the clinician has additional information related to the diabetes recommendations, such as knowledge that an influenza immunization was given at a local pharmacy, she can then enter this data into the data entry screen on the eBrowser (#13). These data are then stored in the supplemental database (#14).

System development. The chronic disease management system was implemented through collaboration between the operational Duke Health Technology Solutions (DHTS) group and the academic Division of Clinical Informatics. The Decision Support Management Module was written in C#, and this module was interfaced with the eBrowser to fulfill clinicians' requests for disease management recommendations. The availability of the chronic disease management functionality was announced in a "What's New" pop-up window along with other eBrowser enhancements when the DSS first became available.

Evaluation

Process measures. The number of distinct system users and the number of distinct patients for whom the system was used were collected weekly. System performance was evaluated in terms of the time required for generating diabetes care recommendations for a set of representative patients.

Usability survey. We surveyed users of the SEBASTIAN system at a family medicine primary care practice within the Duke University Health System. The survey was conducted after the trial system had been in use for three months, but prior to release of the final system. The surveyed population consisted of attending family medicine physicians, family medicine residents, physician extenders, a pharmacist, and a dietitian.

The usability survey was adapted from validated survey instruments for measuring end-users' computing satisfaction [26] and their perceptions of system usefulness and ease of use [27]. Questions from these surveys were grouped into constructs related to content (precise information needed, sufficient information); accuracy; format (clear, useful); usefulness (improves job performance, increases productivity, makes job easier); and ease of use (easy to use, flexible, understandable). These questions utilized a Likert scale in which 3 is a neutral value on a 1-to-5 scale. Confidence intervals and p-values were calculated using large-sample Z tests and a null hypothesis stating that the population mean is 3. The survey respondents were also asked to provide free-text comments regarding the system.

Results

System implementation

The SEBASTIAN DSS has been implemented as a production system to support chronic disease management for the entire Duke University Health System. On August 10, 2006, the trial version of the system was made available to the 1509 attending physicians and 832 residents within the health system, as well as to physician extenders and to ancillary personnel with access to the clinical components of the eBrowser. The production version of the system was released on November 13, 2006 (Figure 2). The current system provides support for diabetes management. As discussed earlier, the system is also being expanded to provide support for health maintenance and for multiple other chronic medical conditions.

Figure 2 - Screenshot from Duke Chronic Disease Management System showing recommendations for a patient with diabetes

Process measures

Since inception, with relatively minimal promotion, the system has been accessed by 758 unique users. This user community is comprised of physicians, physician assistants, nurse practitioners, medical and allied health students, and ancillary service providers. Weekly, since inception, the system is used on average by 85 distinct users (range 72 to 104) to access chronic disease management recommendations for diabetes on an average on 176 distinct patients (range 113 to 225).

With regard to system performance, the system requires an average of 10.9 seconds to return care recommendations after a request is submitted through the eBrowser. Most of this time is spent retrieving the necessary data from the common data repository. Interactions with the SEBASTIAN DSS account for less than a second of the overall processing time.

Usability survey

Twenty (63%) of the 32 available providers responded to the usability survey. Of these 20 providers, 17 were active users of the chronic disease management system. The results

from the usability surveys are depicted in Figure 3. User responses for content, accuracy, format, and ease of use were significantly favorable ($p < 0.005$). Responses for the usefulness construct was not statistically different from the neutral response ($p = 0.76$). An analysis of users' free-text comments regarding the system revealed that this neutral perception of system usefulness arose in large part from the perception that the system took too much time to use.

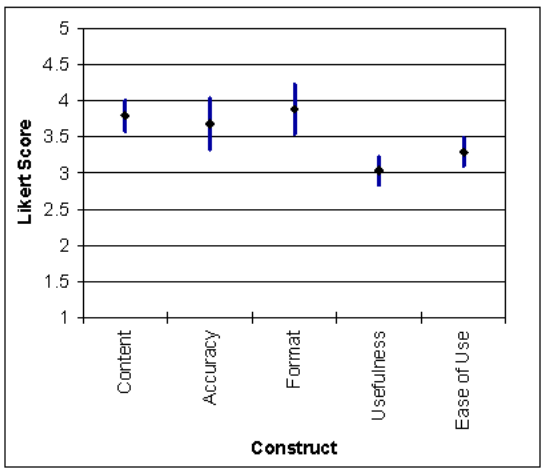


Figure 3 - Results of usability survey, presented as mean scores on usability constructs with associated 95% confidence intervals

Discussion

In this paper, we have described the first implementation of a decision support system based on the recently approved HL7 Decision Support Service draft standard. This implementation validates the operational usefulness of this new HL7 draft standard in the context of a large health system.

The SEBASTIAN DSS approach to meeting a client's CDS needs offers many important strengths and overcomes many of the limitations of other approaches to implementing CDS systems. As one strength, this approach provides a clear mechanism for various clinical information systems to leverage the SEBASTIAN CDS knowledge base. Second, the functionality and interface of the SEBASTIAN DSS is the basis of the HL7 DSS standard. As a result, SEBASTIAN is aligned with an emerging industry consensus on how to deliver CDS using a service-based approach. Third, our approach could support CDS needs at a regional or national level. Fourth, our approach is aligned with the *Roadmap for National Action on CDS* commissioned by the Office of the National Coordinator for Health IT, which calls for a service-based approach to the management and delivery of CDS content [28]. Fifth, our approach is scalable and standards-based. Finally, our approach has been vetted and validated through its production-level use in multiple settings, including in the Duke University Health System for point-of-care chronic disease management and in the North Carolina Medicaid program for population health management [21].

A limitation to our approach is that some clients may be hesitant to have a third party host a critical clinical application. However, the increasing success of vendors that provide electronic health record systems using an application service provider model points to the fact that this hesitation can be overcome with a track record of reliable service. Second, our approach has not yet been validated for several important types of CDS applications (e.g. computerized physician order entry systems). Finally, we do not yet have documented evidence that using a DSS will lead to outcomes desired by a client, such as improved performance on care quality and pay-for-performance metrics. However, there is strong evidence that CDS systems implemented in the manner described in this manuscript reliably produce significant improvements in clinical practice [7]. Also, we are currently conducting several evaluation studies to assess the impact of SEBASTIAN-enabled CDS systems on patient care.

From the limited survey results based on use of the prototype system, we conclude that the clinicians perceived the SEBASTIAN disease management system as easy-to-use, accurate, and appropriate in content and format. The main objection to the system was in the area of usefulness, because of the additional time required for obtaining the care recommendations from the system. We believe this objection was valid, as the prototype system sometimes took over twenty seconds to return the diabetes care recommendations. As noted earlier, this delay resulted from the time required to retrieve the required data from the Duke clinical data repository. In recognition of this limitation, indexing changes were made to the clinical data repository, so that the system currently takes approximately ten seconds to generate a care recommendation summary. We are in the process of introducing additional performance-enhancing strategies, including multi-threaded data retrieval and pre-caching of patient data, in order to reduce the time required for obtaining the care recommendations to under five seconds in the short-term and under one second in the long-term.

In terms of lessons learned, we have observed that the time required to extract the necessary data in order to run the decision support rules may be excessive for busy clinicians. As a consequence, we recommend the retrieval and caching of necessary data prior to the actual request to activate the DSS. This implementation has also taught us that the usefulness of a DSS could be enhanced by the availability of software libraries that facilitate a client's interaction with a DSS. Our implementation experience has also shown us the need to accommodate local configuration of the DSS. Many clinical areas lack a clear consensus for how care should be delivered, and a DSS needs to be able to allow for such "gray" areas of knowledge. We also identified the need to allow a DSS to be configured for individual patients, by allowing rules to be inactivated when clinically inappropriate or declined in deference to patient preference. Finally, we have come to recognize common patterns of CDS needs faced by client applications (e.g., need to determine if patient has condition X and is need of test Y). Recognition of these patterns may allow for more generalized approaches to DSS-client interactions.

Conclusion

This paper is the first report of a decision support application that is built upon the HL7 Decision Support Service draft standard. This application provides disease management information to clinicians for diabetes and other conditions at the point of care. The implementation of this disease management application in a large health system validates the concept of a Decision Support Service.

Acknowledgements

The authors wish to thank Ms. Sherri Gerringer for assistance with survey data collection and management. This work was funded in part by grant R01 HS015057 from the Agency for Healthcare Research and Quality and by grant R41 LM009051 from the National Library of Medicine.

Disclosure: Drs. Lobach and Kawamoto are part owners of Kedasys, LLC, a holding company that has a pending patent application for the intellectual property related to the SEBASTIAN approach to instantiating a Decision Support Service. Through prior agreement, SEBASTIAN is available for use within the Duke University Health System at no cost.

References

[1] National Library of Medicine. MEDLINE Fact Sheet. Available at: http://www.nlm.nih.gov/pubs/factsheets/medline.html. Accessed 12/4/06.

[2] Davis DA, Thomson MA, Oxman AD, and Haynes RB. Changing physician performance. A systematic review of the effect of continuing medical education strategies. AMA 1995:274:700-705.

[3] McGlynn EA, Asch SM, Adams J, Keesey J, Hicks J, DeCristofaro A, and Kerr EA. The quality of health care delivered to adults in the United States. N Engl J Med 2003:348:2635-2645.

[4] Kohn LT, Corrigan JM, and Donaldson MS, eds. To Err is Human: Building a Safer Health System. Washington, DC: National Academy Press, 1999.

[5] Vincent C, Neale G, and Woloshynowych M. Adverse events in British hospitals: preliminary retrospective record review. BMJ 2001:322:517-519.

[6] Hunt DL, Haynes RB, Hanna SE, and Smith K. Effects of computer-based clinical decision support systems on physician performance and patient outcomes: a systematic review. JAMA 1998:280:1339-1346.

[7] Kawamoto K, Houlihan CA, Balas EA, and Lobach DF. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. BMJ 2005. 330:765-768.

[8] Southern California Evidence-based Practice Center. Costs and Benefits of Health Information Technology. Evidence Report/Technology Assessment, Number 132. Agency for Healthcare Research and Quality. April 2006. http://www.ahrq.gov/downloads/pub/evidence/pdf/hitsyscosts/hitsys.pdf. Accessed 11/28/06.

[9] Kawamoto K. Integration of knowledge resources into applications to enable clinical decision support: architectural considerations. In Greenes RA, ed. Medical Decision-Making: Computer-Based Approaches to Achieving Healthcare Quality and Safety. Burlington, MA: Elsevier, Inc. 2007; pp. 503-538.

[10] Boxwala AA, Tu S, Peleg M, Zeng Q, Ogunyemi O, Greenes RA, Shortliffe EH, and Patel VL. Toward a representation format for sharable clinical guidelines. J Biomed Inform 2001:34:157-169.

[11] Johnson PD, Tu S, Booth N, Sugden B, and Purves IN. Using scenarios in chronic disease management guidelines for primary care. Proc AMIA Symp 2000:389-393.

[12] Tu SW, Musen MA, Shankar R, Campbell J, Hrabak K, McClay J, Huff SM, McClure R, Parker C, Rocha R, Abarbanel R, Beard N, Glasgow J, Mansfield G, Ram P, Ye Q, Mays E, Weida T, Chute CG, McDonald K, Molu D, Nyman MA, Scheitel S, Solbrig H, Zill DA, and Goldstein MK. Modeling guidelines for integration into clinical workflow. Medinfo 2004:2004:174-178.

[13] First DataBank. Drug Information Framework Information Page. Available at: http://www.firstdatabank.com/products/di_framework/. Accessed 12/4/06.

[14] Peleg M, Boxwala AA, Ogunyemi O, Zeng Q, Tu S, Lacson R, Bernstam E, Ash N, Mork P, Ohno-Machado L, Shortliffe EH, and Greenes RA. GLIF3: the evolution of a guideline representation format. Proc AMIA Symp 2000:645-649.

[15] Shiffman RN, Michel G, Essaihi A, and Thornquist E. Bridging the guideline implementation gap: a systematic, document-centered approach to guideline implementation. J Am Med Inform Assoc 2004:11:418-426.

[16] Pryor TA, and Hripcsak G. The Arden syntax for medical logic modules. Int J Clin Monit Comput 1993:10:215-224.

[17] Karadimas HC, Chailloleau C, Hemery F, Simonnet J, and Lepage E. Arden/J: an architecture for MLM execution on the Java platform. J Am Med Inform Assoc 2002:9:359-368.

[18] Zielstorff RD. Online practice guidelines: issues, obstacles, and future prospects. J Am Med Inform Assoc 1998:5:227-236.

[19] Hunt DL, Haynes RB, Hayward RS, Pim MA, and Horsman J. Patient-specific evidence-based care recommendations for diabetes mellitus: development and initial clinic experience with a computerized decision support system. Int J Med Inf 1998:51:127-135.

[20] Henry SB, Douglas K, Galzagorry G, Lahey A, and Holzemer WL. A template-based approach to support utilization of clinical practice guidelines within an electronic health record. J Am Med Inform Assoc 1998:5:237-244.

[21] Kawamoto K, and Lobach DF. Design, implementation, use, and preliminary evaluation of SEBASTIAN, a standards-based Web service for clinical decision support. Proc AMIA Symp 2005:380-384.

[22] Huang C, Noirot LA, Heard KM, Reichley RM, Dunagan C, and Bailey TC. Implementation of Virtual Medical Record Object Model for a Standards-Based Clinical Decision Support Rule Engine. Proc AMIA Symp 2006:958.

[23] Goldberg HS, Vashevko M, Pastilnik A, Smith K, Plaks N, and Blumenfeld BM. Evaluation of a Commercial Rule Engine as a Basis for a Clinical Decision Support Service. Proc AMIA Symp 2006:294-298.

[24] Health Level 7. HL7 Decision Support Service specification (Draft Standard for Trial Use). Available at: http://www.hl7.org/v3ballot/html/infrastructure/dss/dss.htm. Accessed 11/27/06.

[25] Cerami E. Top ten FAQs for Web services. Available at: http://webservices.xml.com/pub/a/ws/2002/02/12/webservicefaqs.html. Accessed 11/27/06.

[26] Doll WJ, and Torkzadeh G. The Measurement of End-User Computing Satisfaction. MIS Quarterly 1988:12:259-274.

[27] Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly 1989:13:319-340.

[28] Kawamoto K, and Lobach DF. Proposal for fulfilling strategic objectives of the U.S. roadmap for national action on decision support through a service-oriented architecture leveraging HL7 services. J Am Med Inform Assoc 2007:14:146-155.

Address for correspondence

David F. Lobach. email:- david.lobach@duke.edu

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

The TAR Model: Use of Therapeutic State Transitions for Quality Assurance Reporting in Chronic Disease Management

Gaikwad R[1], Warren J[1], Kenealy T[2]

[1]*Department of Computer Science, The University of Auckland, New Zealand*
[2]*Department of General Practice and Primary Health Care, The University of Auckland, New Zealand*

Abstract

Chronic disease management represents one of the challenges for health informatics and demands the appropriate application of information technology for improved patient care. This paper presents an approach to quality assurance reporting wherein the recommendations of evidence-based clinical practice guidelines are considered in the context of empirical therapeutic state-transitions (in terms of changes in individual patient prescriptions over time). We apply a Transition-based Audit Report (TAR) model to antihypertensive prescribing and related data as stored in a New Zealand General Practice Management System database. The results provide a set of quality indicators and specific patient cohorts for potential practice quality improvement with strong linkage to the selected guidelines and observed practice patterns. We see the TAR model primarily as a tool to enable internal quality improvement efforts, but also to be of relevance for focusing pay-for-performance programs.

Keywords:

hypertension, medical audit, practice guidelines, quality assurance, quality indicators

Introduction

General medical practices in New Zealand have a 99% rate of using purpose-built Practice Management Systems (PMS) software, with 89.7% using the PMS to record prescriptions and 71.8% reporting that their General Practitioners (GPs) use the PMS to record full clinical notes [1]. This high rate of PMS uptake is likely to have a positive influence on quality of care, since electronic prescribing systems have been shown to reduce the frequency of medication errors [2]; although, conversely, such systems have also been associated with specific types of increased risk in patient care [3]. The ubiquity of PMS use by GPs presents a distinct opportunity to use data mining to extract empirical information about chronic disease management practices, particularly the well-recorded prescribing practices, such as prescription of antihypertensive drugs. Cardiovascular Disease (CVD) is the number one killer in New Zealand [4], so the motivation to improve the management of related risk factors is, of course, considerable. Understanding actual practice in antihypertensive prescribing serves to inform our understanding of CVD risk management.

Therapeutic state transitions are points in time when the status of key aspects of a patient's therapy change [5]. Analysis of therapeutic state transitions in General Practice – notably, the pattern of GPs' prescribing acts for individual patients over time – shows promise as the basis for development of high-specificity interactive decision support alerts [6]. In this paper we examine the potential for therapeutic state transitions to inform the development of an audit report, making it a Transition-based Audit Report (TAR). We examine the TAR model in the context of antihypertensive prescribing using the PMS data from a medium-size metropolitan General Practice in New Zealand.

Methodology

This work utilizes the approach of therapeutic state transition based analysis as set forth by Warren et al. [5]. We examine how therapy (in this case, antihypertensive prescribing, qualified by coded observations and diagnoses) changes over time for the individual patients that make up the cohort of interest. In this paper we focus on a TAR model of report formulation to support internal practice quality improvement efforts, and to inform broader discussion of appropriate performance metrics.

Setting and data

Data was extracted from the PMS of a metropolitan New Zealand General Practice which employs about four full-time equivalent GPs and two full-time equivalent nurses, belongs to the large Primary Health Organization in the country, and is sited in an area that is socio-economically and ethnically diverse. The investigation related to all prescribing undertaken by the practice between January 1, 2005 and September 25, 2006. Analysed PMS data included age, gender, prescriptions, "classifications" (problem codes, largely as Read Codes), and observations including blood pressure (BP), serum creatinine and serum potassium measurements. Characteristics of the data extract are given in Table 1. Identity of patients was blinded to the analysts (RG and JW), but could be recombined for follow-up by the GP author (TK, one of the main GPs of the practice) for purposes of follow up in patient

care. As such, the orientation of the investigation was on internal practice quality assurance and improvement with a secondary goal of methodology development. Some data cleaning was necessary; notably, to compensate for mis-punctuation and trailing zeros in entry of BP measurements (which can be accomplished through any of multiple pathways in the user interface of the PMS, including via free-text notes). About 98% of BP measurements were usable.

Table 1 - Characteristics of data extract

General description of data extract (investigated period: Jan 1, 2005 - Sept 25, 2006)	No. of records
Patients[#]	14835
Encounter	84393
Classifications	
Total	8704
Distinct patients classified	3383
Hypertension alone or as comorbidity	481 (23%)
Hypertension without diabetes mellitus	396 (82%)
Hypertension with diabetes mellitus	85 (18%)
Prescriptions	
Total	59473
Antihypertensive prescriptions*	8466 (14%)
Measurements - Blood pressure	10418

[#] distinct patients in data extract classified irrespective of being hypertensive or not * prescriptions with antihypertensive agents

Choice of guidelines and therapeutic state variables

Within the scope of this paper, the focus was on the quality of antihypertensive therapy (AHT) prescribing, thus emphasizing CVD risk management and appropriate variation for comorbidities, notably diabetes. We chose clinical practice guidelines on the criteria that they were evidence-based, applicable to the clinical practice in the area of study and current [7, 8, 9, 10]. Generic and/ or brand names for AHT agents were identified from the data and classified by referring to [11]. There were 44 generic and/ or brand names of drugs identified as belonging to six AHT drug groups (see Table 2). The therapeutic drug groups (admittedly broad groupings) were given compact labels – A, B1, B2, B3, C and D – based on their general order in [7]. These six groups identify our therapeutic state variables for analysis; that is, changes in the presence or absence of prescription of drugs from these groups defines therapeutic state transitions.

Therapeutic state transitions

The process of prescribing by the GP produces two events; one marking the start of the prescription (therapy) and the second, an implicit event marking the expected end of the prescription if directions given by the physician have been properly adhered to. The instructions (*signatura*) given by the GP consist of dosing, frequency, repeats (refills) and a duration (computed by the PMS) per prescription. We identified 290 distinct *signatura* used by the GPs for the drugs of interest in the data extract. The PMS-computed duration was accurate except where the GP had overridden default dosing instructions, wherein we computed a corrected duration; e.g. "take 10mg od" of Plendil ER was interpreted as a directedconsumption of 4*(2.5mg Tab Plendil ER)*1 because Tab Plendil ER is available in strength of 2.5mg [11]. Comparison of the therapeutic state transitions (both for the practice overall and for individuals) with respect to the selected guidelines was used to formulate criteria for a quality audit report.

Results

A State-Transition Overview Diagram (STOD) for the investigated period was computed based on all AHT state variable changes. Figure 1 shows all transitions that occurred at least 5 times. *Init_out* denotes the state when the patient has had no antihypertensive prescriptions for a minimum of 100 days before the start of AHT during the investigated period, otherwise the patient commences with *Init_in*. It is hypothesized that Init_out patients are more likely to be commencing AHT, whereas Init_in are more likely to be already in AHT at the time of their first prescription in the data extract. A "lapse" in antihypertensive therapy implies the period which commences when all antihypertensive medications, if taken as directed from the day of prescribing, should have run out, and is indicated by the *Zero* state. States are heuristically processed to avoid over-sensitivity [5], with Zero states of less than 90 days being coalesced into the prior state, as is any other state of less than 30 days duration.

Individual Path Diagrams (IPDs) were computed to illustrate the therapeutic experiences of individual patients as identified under a variety of exploratory criteria. Criteria focused on: (a) cycles (returning to the same state one or more times) and (b) transitions that are difficult to align with guidelines. With respect to the latter case, effective combinations for AHT include ACEi/ARB agents with diuretic, beta-blocker with diuretic and beta-blocker with DCCB [8]. Paths that break effective combinations may possibly adhere to best-practice, but are considered worthy of further scrutiny. A selection of interesting IPDs from the data set is given in Figure 2. Graphviz (URL: http://www.graphviz.org/) was used to produce Figures 1 and 2.

The STOD and individual IPDs were analyzed to identify relevant statistics for an audit report, both supportive and cautionary with respect to AHT quality. States and transitions were compared to the guidelines (note both RG and TK are physicians) as a means of knowledge engineering a TAR that aligns with the selected guidelines. In the present study, this was an exploratory exercise drawing firstly from the STOD and subsequently by assessing specific IPDs to confirm appropriate focus areas for quality improvement. For example, a transition of Init_in A for patients with diabetes as well as hypertension is readily aligned as an attribute of best-practice. Presence in the B1 state (or any combination state that includes B1) for

Table 2 - Classification of Therapeutic state variables and their respective Anatomical Therapeutic Chemical (ATC) classifications

AHT state variables	AHT drug groups (drug names for matching to PMS data)	ATC codes (from [12])
A	Angiotensin converting enzyme inhibitors(ACEi) and Angiotensin receptor blockers(ARB) (Enalapril, Candesartan cilexetil, Losartan potassium, Cilazap, Lisinopril, Quinapril, Accupril, Trandolapril, Perindopril erbumine, Captopril, Inhibace)	ACE inhibitors, plain: C09A and Angiotensin II antagonists, plain C09CA
B1	Beta-blockers (Atenolol, Esmolol hydrochloride, Acebutolol hydrochloride, Nadolol, Timolol maleate*, Propranolol, Carvedilol, Labetalol hydrochloride, Celiprolol hydrochloride, Betaloc, Metoprol, Oxprenolol hydrochloride, Pindolol)	Beta blocking agents, non-selective: C07AA and Beta blocking agents, selective: C07AB
B2	Diuretics (Bendrofluazide, Hydrochlorothiazide, Amiloride, Frusemide, Triamterene, Spironolactone, Bumetanide, Indapamide hemihydrate)	Low-Ceiling Diuretics, Thiazides: C03A, Low-Ceiling Diuretics, Excl. Thiazides: C03B, High-Ceiling Diuretics: C03C, Potassium-Sparing Agents: C03D, Diuretics And Potassium-Sparing Agents In Combination: C03E
B3	Non-dihydropyridine calcium channel blockers (Non-DCCB) (Verapamil, Dilatiazem)	Calcium-channel blockers: Selective Calcium Channel Blockers With Direct Cardiac Effects: C08D
C	Dihydropyridine calcium channel blockers (DCCB) (Felodipine, Isradipine, Nifedipine, Amlodipine)	Calcium-channel blockers: Dihydropyridine derivatives: C08CA
D	alpha blockers, hydralazines, Clonidine (Prazosin hydrochloride, Terazosin hydrochloride, Doxazosin mesylate, Clonidine, Hydralazine hydrochloride, Diazoxide)	alpha-adrenoreceptor antagonists: C02CA, Hydralazine: C02DB02, Clonidine: C02AC01

* prescribed as an Antihypertensive agen

patients with asthma is contraindicated and can contribute to a cautionary statistic.

The alignment of specific therapeutic states and transitions to the selected guidelines and best-practice were discussed within the author team. At times all or a subset of patients in a state- transition-based cohort were rematched by TK and reviewed by staff of the practice. The states and transitions were found to act as 'frames' for relatively simple sets of additional qualifiers. That is, almost no state or transition is simply good or bad; however, many clearly fit a supportive or cautionary category after the identification of a qualifier in terms of comorbidity or observation values.

Quality indicators (QIs) identified for a TAR fell into three broad categories:

1. Guideline based – QIs stemming from outcome or process requirements with limited relationship to states and transitions (e.g., achieving target BP).
2. State-transition based – QIs based more-or-less entirely on states and transitions with limited reference to other PMS data; notably, breaking of effective combination therapies.
3. Hybrid – a large set of QIs where state or transition is qualified by comorbidity or observations; these subdivide into 'bad' states (e.g., presence in Zero state without support of acceptable BP observations) and problem-drug interactions (e.g., asthma with beta-blocker).

The resulting report based on the TAR model has the following sections which combine the three types of QIs above:

- I. Description of practice as per PMS data (15 QIs):
  - General (e.g., patient volume)
  - Hypertension (e.g., prevalence)
  - Antihypertensive Therapy and Monitoring
- II. Criteria to support AHT quality (15 QIs):
  - Blood pressure control in patients classified with Hypertension
  - Continuity of therapy in patients classified with Hypertension

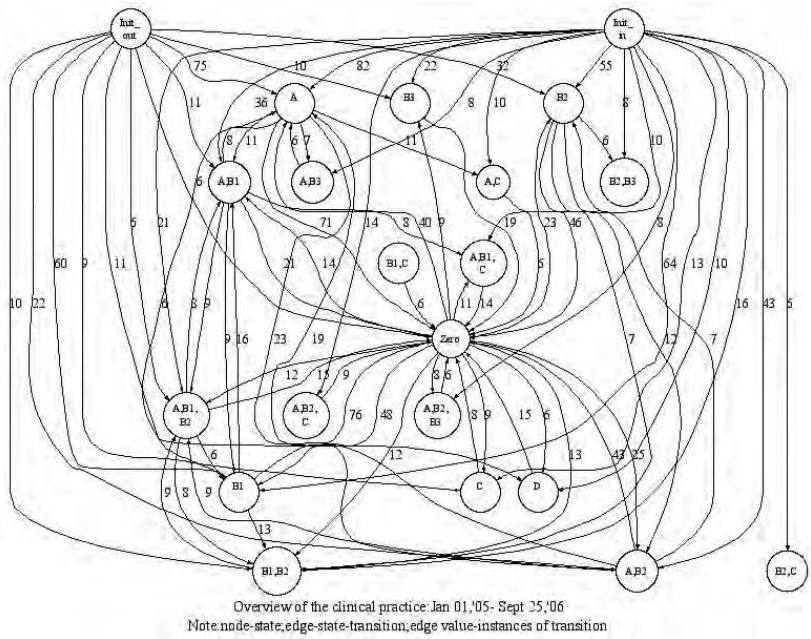


Figure 1 - State-Transition Overview Diagram (STOD) of antihypertensive therapy from 1 Jan 05 to 25 Sep 06, showing therapeutic states (circles) and transitions (arcs) with number of transitions made by patients during the investigated period.

  - Effective combination therapy in patients classified with Hypertension
  - Drug-problem indication in patients classified with Hypertension
  - Monitoring
- III. Criteria for improvement/caution of AHT (20 QIs):
  - Blood pressure management and monitoring
  - Lapses in antihypertensive therapy
  - Drug-drug interaction (e.g., concurrent beta-blocker and Diltiazem with or without atrial fibrillation and/or flutter)
  - Drug-problem interaction

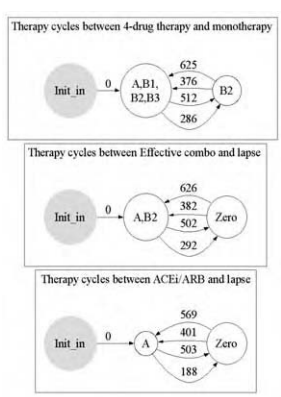


Figure 2 - Individual Path Diagrams (IPDs) showing cyclical patterns in antihypertensive therapy for select patients (circles indicate states; arcs labelled with day of transition relative to start of therapy).

Discussion

The TAR and associated analysis process supports a variety of quality assurance related activities. Broadly, these concern: (a) improvement of care for specific patients in the near term; (b) reflection on guidelines, practice and alerts; and (c) the use of QIs in relation to benchmarking and performance incentives.

Direct use of the TAR is for follow-up on the cohorts of patients responsible for critical indications. Actions may include review of the PMS record, and possible patient recall or review of therapy at next visit. The practice with which we have participated in this study is currently in the process of utilizing our report in this fashion.

A second use of the TAR is to support practice introspection beyond the immediate needs of specific patients. TAR QIs are tightly linked to concordance with guidelines. The TAR formulation process encourages clinicians to put the individual clauses of a guideline into context and to specifically qualify and interpret each clause for their practice. The extent of non-concordance to the guideline interpretation is provided in the report, and thus the potential value and impact of employing strategies such as local information programmes or online alerts can be assessed.

The tertiary use of the TAR is for benchmarking – that is, to compare the QIs to those for other practices. QIs emerging from the TAR model have several good characteristics in that they are tightly bound to clinical practice guidelines, have precise and compact definitions, and can be computed from PMS data without further clinician effort.

Relatively low performance against comparable clinics on such QIs should prompt consideration for quality improvement efforts.

The most controversial use of QIs (not particularly those from the TAR model) is as a basis for pay-for-performance funding incentives. Such incentives are already a reality in the UK [13] and substantial portions of the US [14] primary healthcare systems. Moreover, Teich et al. [15] suggest funding and incentives to reward the use of advanced decision support features in electronic prescribing, such as online alerts. Use of the TAR model may help to inform decision makers on the universe of possible QIs to achieve those with strong alignment to evidence-based best practice.

The method presented herein has a number of limitations. A number of manual steps were involved; however, we believe these could be minimized with replication, especially for key domains such as AHT. The method as applied in this study is limited by the quality of coding practice – our current results are based on the (questionable) assumption that all relevant conditions, comorbidities and observations can be found in the PMS record. Moreover, of course, the method is only applicable in an environment that has adopted General Practice computing.

Finally, it should be acknowledged that therapeutic state-transition based analysis covers some of the same ground as any other approach to guideline engineering and, conversely, that not all aspects of quality-of-care are based on therapy in the narrow sense we have applied it (e.g., there is quality of monitoring, and achieving patient ownership of care plans).

Conclusions

The TAR model report formulation process can be utilized to develop quality indicators (QIs) for audit of AHT in a clinical practice on the basis of changes in a patient's therapy over time as indicated in electronic records resulting from routine care activities. Uses can be characterized as including immediate review of therapy for specific patient cohorts, broader-view refinement and introspection of the uptake of guidelines into a practice, benchmarking to other practices, and strategic considerations on which QIs may be the best basis for incentives to promote evidence based care.

The direction of the current research is to look at areas for roll-out of the current TAR to a spectrum of practices; in New Zealand, this is best pursued through PHOs (Primary Health Organisations). We are also examining applicability to other CVD risk factors, notably dyslipidaemia agents.

References

[1] Didham R, Martin I, Wood R, and Harrison K. Information Technology systems in general practice medicine in New Zealand. N Z Med J. 2004 Jul 23;117(1198):U977.

[2] Kuperman GJ, Teich JM, Gandhi TK, and Bates DW. Patient safety and computerized medication ordering at Brigham and Women's Hospital. Jt Comm J Qual Improv. 2001 Oct;27(10):509-21.

[3] Koppel R, Metlay JP, Cohen A, Abaluck B, Localio AR, Kimmel SE, and Strom BL. Role of computerized physician order entry systems in facilitating medication errors. *JAMA*. 2005 Mar 9;293(10):1197-203.

[4] Turley M, Stefanogiannis N, Tobias M, van der Hoorn S, Lawes C, Ni Mhurchu C, and Rodgers A. Nutrition and the burden of disease: New Zealand 1997-2011. Wellington, Ministry of Health, August 2003. Public Health Intelligence Occasional Bulletin Number 17.

[5] Warren J, Gadzhanova S, Stanek J, and Iankov I. Understanding caseload and practice through analysis of therapeutic state transitions. AMIA Annu Symp Proc. 2005;:784-8.

[6] Gadzhanova S, Iankov I, Warren J, Stanek J, Misan, G, Baig Z, and Ponte L. Developing High-Specificity Anti-hypertensive Alerts by Therapeutic State Analysis of Electronic Prescribing Records. J Am Med Inform Assoc. 2006 Oct 26; [Epub ahead of print]

[7] Texas Department of State Health Services: Hypertension Algorithm for Diabetes Mellitus in Adults, Revised 1-26-06. URL: http://www.dshs .state.tx.us/diabetes/PDF/algorithms/HYPER. PDF. Retrieved July, 2006.

[8] Therapeutic Guidelines Ltd., Australia. Therapeutic Guidelines: Cardiovascular. Version 4. 2003.

[9] Chobanian AV, Bakris GL, Black HR, Cushman WC, Green LA, Izzo JL Jr, Jones DW, Materson BJ, Oparil S, Wright JT Jr, and Roccella EJ. Seventh report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure. Hypertension. 2003 Dec;42(6):1206-52.

[10] The National Heart Foundation of Australia: Hypertension Management Guide for Doctors 2004. URL: http:// www.heartfoundation.com.au/downloads/ hypertension_management_guide_2004.pdf. Retrieved July, 2006.

[11] MIMS Online. URL: http://www.library.auckland.ac.nz/ databases/learn_database/public.asp?record=Mims. Accessed Oct, 2006.

[12] WHO Collaborating Centre for Drug Statistics Methodology: ATC/DDD Index 2006. URL: http:// www.whocc.no/atcddd /indexdatabase/. Accessed June, 2006.

[13] Doran T, Fullwood C, Gravelle H, Reeves D, Kontopantelis E, Hiroeh U, Roland M. Pay-for-performance programs in family practices in the United Kingdom. N Engl J Med. 2006 Jul 27;355(4):375-84.

[14] Rosenthal MB, Landon BE, Normand SL, Frank RG, and Epstein AM. Pay for performance in commercial HMOs. N Engl J Med. 2006; 355(18):1895-902.

[15] Teich JM, Osheroff JA, Pifer EA, Sittig DF, Jenders RA; The CDS Expert Review Panel. Clinical decision support in electronic prescribing: recommendations and an action plan: Report of the joint clinical decision support workgroup. J Am Med Inform Assoc. 2005 Jul-Aug;12(4):365-76.

Address for Correspondence:

Prof. Jim Warren,
Chair in Health Informatics
Department of Computer Science – Tamaki
University of Auckland, Private Bag 92019,
Auckland, New Zealand
Tel: +64 9 373 7599/ Fax: +64 9 373 7503
Email: jim@cs.auckland.ac.nz

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Development of Case-based Medication Alerting and Recommender System: A New Approach to Prevention for Medication Error

Kengo Miyo[a], Yuki S. Nittami[a], Yoichiro Kitagawa[a], Kazuhiko Ohe[a]

[a] *Department of Planning, Information and Management, the University of Tokyo Hospital, Japan*

Abstract

The purpose of this study was to develop a new alerting and recommender system for preventing medication errors. In recent years, alerting systems have been widely implemented, but because these systems apply a same static threshold for all patients in all cases, they produce excessive alerts and subject physicians to "alert fatigue". We believe that the most commonly-written prescription for a patient's status is the safest one. From this standpoint, we developed a real-time case-based medication alerting and recommender system linked to a database of past prescriptions. When a physician issues his or her prescription, our system dynamically compares it with past ones for similar patients in the database. An analysis of the 10 most frequently-used drugs in the University of Tokyo Hospital revealed that our system reduced the number of false alerts compared to the traditional static alert method. Our system contributes to the creation of alerts that are appropriate for patients' clinical conditions and based on physicians' empirical discretion.

Keywords:

case-based alerting; decision support systems, clinical; medical order entry systems; medication errors; prescriptions, drug

Background

In recent years, computerized physician order entry (CPOE) systems have been introduced to health care institutions worldwide [1, 2]. In Japan, the use of CPOE systems for medication has become widespread [3], and their implementation rate in hospitals with 500 or more beds was 70% as of 2005 [4]. The use of CPOE system is expected to contribute to efficient health care delivery and reduce physicians' time costs [1, 3, 5].

CPOE systems also contribute to improvements in health care quality, particularly in regard to safety [6–8]. To decrease errors in medical treatments, several systems with real-time data input checks have been developed, for example, to detect inappropriate dosages or drug combinations. In these systems, as soon as a physician clicks the "issue prescription" button, the system compares input data on dose regimen and concomitant drugs to data on dosage limitations and contraindicated drugs stored in a master table file. If the prescription contains inappropriate data, the system displays an alert. Alert systems that use static threshold data stored in a master table file are called *static alert systems*.

Although static alert systems are generally useful, excessive alerts can be produced because the systems check off patient status and treatment policies. Excessive alerts cause physicians to pay less attention to the alerts [9], which then lose their effectiveness. In other words, the few important alerts are overlooked amid a lot of meaningless ones. Peterson et al. described physicians in this situation as being in a state of "alert fatigue" [7, 10]. Particularly in university hospitals, where doctors treat many patients whose cases run counter to standard treatment, the risk of overlooking alerts cannot be ignored.

Physicians need to receive appropriate alerts. In this paper, the term *appropriate alert* refers not to an alert generated based on whether the data match standards determined by drug notes, but rather to an alert generated based on whether the present treatment differs greatly from actual treatment records. Our purpose in this paper is to suggest a new approach for generating appropriate prescription alerts. We believe that most commonly written prescription is the safest one, and we considered past records stored in HIS as the gold standard. From this standpoint, we developed a real-time case-based system that alerts physicians when their prescription deviates from this gold standard.

System design

System structure

The overview of our system is shown in Figure 1. The system consists of existing HIS and a database of past prescription records, as well as an alert engine. MySQL version 5.0 was chosen as the database for our system. All prescription data from the University of Tokyo Hospital (UTH) for 2000 through 2005 were extracted from HIS and stored in the database. For the alert engine, both Perl version 5.8, PHP version 5.1, and Apache version 2.0 were used.

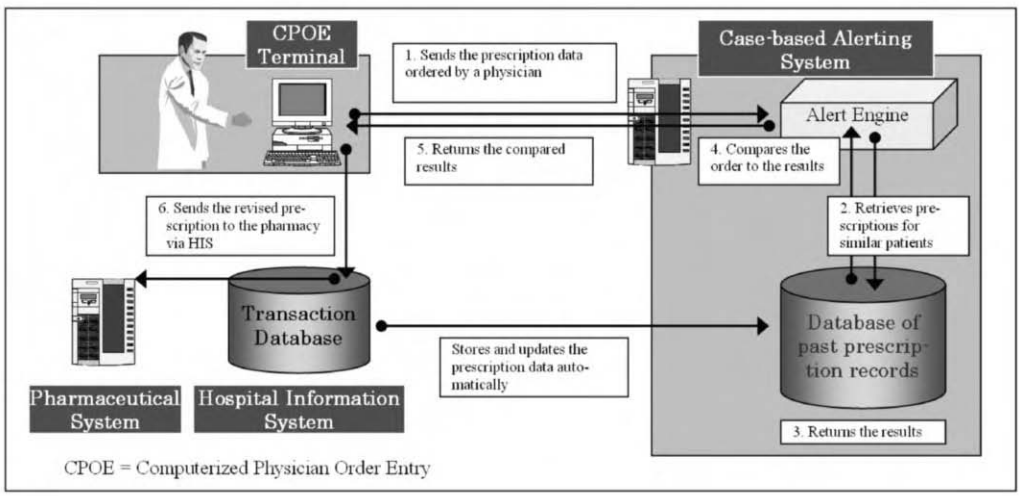


Figure 1 - The structure and workflow of the case-based alert system.

System feature

Our system has two major functions: alert function and decision support.

Alert function

The alert engine receives the prescription data entered by a physician before sending them to the pharmaceutical system. The alert engine compares these data to the statistical data for that same drug, compiled from past records in the database. If the alert engine finds that the entered data exceed the threshold described below, it sends an on-screen alert to the physician. Figure 2 presents an example of such a prescription alert. At present, the items that can trigger an alert are dosage, dose regimen, duration of drug administration, and concomitant drugs.

Statistical data on past prescriptions (such as 95th percentile and standard deviation) serve as the alert threshold. Prescription data in the database are also linked to the patient in question and to his or her diseases. Therefore, the alert thresholds classified according to clinical department (specialty), disease, sex, and age can be also used.

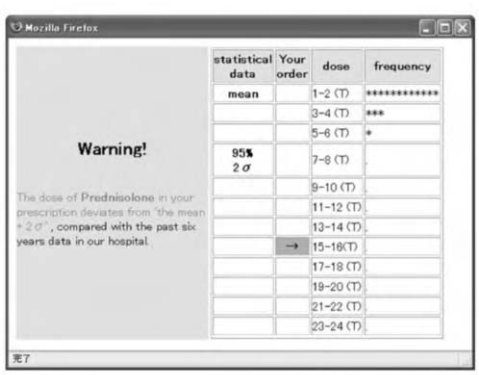


Figure 2 - A pop-up alert on the computerized physician's order entry terminal.

Decision support function

The recommendation function shown in Figure 3 is used when a physician wants to know the appropriate dosage and duration of drug administration. This function provides the most frequently used value by analyzing a distribution of values of past records. The physician can reexamine his or her prescription with these recommendations in mind. As in the case of the alert function, the recommendation function takes-the patient's attributes into consideration.

Figure 3 - The window of the recommendation function

Interim system evaluation

The final evaluation of our system should be done by physicians who actually use it. The present system, however, is still in the prototype phase. Therefore, we here clarify some interesting aspects of our method by comparing it to traditional static alert systems. We used data on dosage for the evaluation.

Evaluation method

We used the September 2006 the UTH prescription records for the system evaluation. We compared cases using our method to those using information about dosage limitation provided by drug notes. The alert function of our system used two thresholds. One was the 95th percentile threshold

of each drug, and the other was the mean + 2. We analyzed the 10 most frequently prescribed drugs at the UTH.

Evaluation results

Results are shown in Table 1. Under both the mean + 2 and the 95th percentile thresholds, we found that the number of alerts declined for nearly one-third of the drugs compared to the existing static alert method. However, because the value of the mean + 2 for prednisolone is lower than the value described in the drug note, the number of alerts increased. The reason is that the range of prescribed dosages of prednisolone is broad (i.e., the maximum prescribed dosage was up to 120 mg, whereas the values of prescribed dosages were concentrated around 5 and 10 mg). Thus, our research showed that our system, based on physicians' empirical discretion, set different thresholds than the existing static alert methods.

Discussion

System evaluation

One advantage of our method is that the number of alerts distinctly decreased from that produced by traditional static alert methods because unlike these traditional methods, our thresholds have upper values (e.g., for drugs with high variance of dosage). Loxoprofen sodium and amlodipine are good examples. In the case of prednisolone and aspirin, the mean + 2 thresholds were lower than those described in the drug notes, which caused the increase in alerts. It is critical to warn physicians when they are about to prescribe drug dosages that they do not often prescribe, even if that dosage is within the normal range described in a drug note. This is another advantage of our case-based method. It is one of our key future tasks to determine which threshold to use: mean + 2 or 95th percentile.

Another advantage of our method is that it can issue alerts differently for each clinical department. Prednisolone, whose dosage varies greatly among departments, is a good example. Applying a threshold across the board as traditional methods do results in the production of unnecessary alerts in certain departments. Our system, however, avoids this issue by issuing alerts based on department features, and on values such as disease, sex, and age, even for one drug. This ensures that alerts are appropriate for each patient's clinical conditions.

Although our system is in the prototype stage at present, we will start actual operation in 2007 and survey physicians' prescribing behavior and degree of satisfaction.

Availability of case-based approach

Master table maintenance is unnecessary

The traditional method, which keeps static descriptions of dosage limit and contraindicated drugs in master tables, requires a huge amount of work to maintain these tables. Our system, however, requires less maintenance because information on dosage limit and contraindicated drugs is dynamically created using already stored data. In addition, our method keeps up with medical advances. When dose regimens change, our method automatically updates the alert thresholds and consequently can issue alerts based on the current prescription trends.

Table 1 - Results of Interim Evaluation

Drug	Total number of orders	Threshold from drug note	# of alerts	*M* + 2*	# of alerts	95th percentile*	# of alerts
Teprenone	3424	150 mg	14	202.3 mg	0	150 mg	14
Loxoprofen Sodium	2623	180 mg	27	244.8 mg	5	240 mg	5
Brotizolam	2666	0.25 mg	460	0.485 mg	447	0.5 mg	1
Famotidine	2238	40 mg	3	51.2 mg	3	40 mg	3
Prednisolone	2110	60 mg	32	41.9 mg	77	40 mg	77
Rebamipide	2094	300 mg	12	398 mg	12	300 mg	12
Amlodipine	1915	5 mg	392	10.2 mg	2	10 mg	2
Aspirin	1950	300 mg	0	125 mg	24	100 mg	24
Sennoside AB	1820	48 mg	26	47.2 mg	154	48 mg	26
Mecobalamin	1849	1500 µg	1	1815 µg	1	1500 µg	1

* The system issues an alert when the dosage is greater than this value.

Oriented to physicians' experiences

Our system promises to clarify various dose regimens in a way that is consistent with physicians' empirical discretion. By using past records, our system can alert physicians to drug combinations not empirically prescribed, even if the drugs are not described in drug notes as being contraindicated. In addition, our system can present information about recommended concomitant drugs, similar to the way interns receive advice from supervisors.

Possibility of graded alert

Our method provides distributions of various parameters such as dosage and dose period. As a result, our system can present graded alerts, such as 1: attention, 2: warning, and 3: prohibited. Gradings of how extensively prescription data diverge from past records enable the physician to quickly decide whether he or she should heed the message.

Limitations of our system

One limitation of our system is that it does not function well with an insufficient amount of past data. For drugs on which a hospital has few or no past records, it is necessary to combine our method with other existing techniques. Small hospitals that do not have a sufficient number of past records must import prescription data from authorized large-scale hospitals. Using data from other hospitals is equivalent to asking "In your hospital, how is this drug commonly used?"

Future direction

The alert and recommendation function can also be applied to laboratory test and radiological records. We are now considering how to make the system available for other medical practices.

Conclusions

We developed a real-time case-based medication alert system that alerts physicians when a prescription deviates from similar commonly written prescriptions in a database of stored records. We set past records stored in HIS as the gold standard. An analysis of the 10 most frequently prescribed drugs at the UTH revealed that our system reduced the number of unnecessary alerts compared to the traditional static alert method. Our easy-to-maintain system creates alerts appropriate for patients' clinical conditions and alerts based on physicians' empirical discretion. In addition, it can provide more appropriate alerts than traditional methods, and as a result, will contribute to the prevention of errors in medical treatments.

Acknowledgments

This study was supported in part by a Grant-in-Aid to Young Scientists (B) 18790353 (2006) from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

References

[1] Kuperman GJ, and Gibson RF. Computer physician order entry: benefits, costs, and issues. Ann Intern Med 2003: 139(1): 31-9.

[2] Rothschild J. Computerized physician order entry in the critical care and general inpatient setting: a narrative review. J Crit Care 2004: 19(4): 271-8.

[3] Haruki Y, Ogushi Y, Okada Y, Kimura M, Kumamoto I, and Sekita Y. Status and perspective of hospital information systems in Japan. Methods Inf Med 1999: 38(3): 200-6

[4] Ministry of Health, Labour and Welfare, Japan. Survey of medical institutions in 2005. Oct 1, 2005; Available from: http://wwwdbtk.mhlw.go.jp/toukei/data/160/2005/toukeihyou/0005605/t0123581/A0064_001.html. Accessed: Dec 4, 2006. (in Japanese)

[5] Koppel R, Metlay JP, Cohen A, Abaluck B, Localio; AR, Kimmel SE, and Strom BL. Role of computerized physician order entry systems in facilitating medication errors. JAMA 2005: 293: 1197-203.

[6] Judge J, Field TS, DeFlorio M, Laprino J, Auger J, Rochon P, Bates DW, and Gurwitz JH. Prescribers' responses to alerts during medication ordering in the long term care setting. J Am Med Inform Assoc 2006: 13(4): 385-90.

[7] van der Sijs H Aarts J, Vulto A, and Berg M. Overriding of drug safety alerts in computerized physician order entry. J Am Med Inform Assoc 2006: 13(2): 138-47

[8] Kaushal R, Shojania KG, Bates DW. Effects of computerized physician order entry and clinical decision support systems on medication safety. Arch Intern Med 2003: 163(12): 1409-16.

[9] Glassman PA, Simon B, Belperio P, and Lanto A. Improving recognition of drug interactions: Benefits and barriers to using automated drug alerts. Med Care 2002: 40(12): 1161-71.

[10] Peterson JF, and Bates DW. Preventable medication errors: identifying and eliminating serious drug interactions. J Am Pharm Assoc 2001: 41(2): 159-60.

Address for correspondence

Kengo Miyo, Ph.D., R.N., P.H.N.
Hongo 7-3-1, Bunkyo-ku,
Tokyo 113-8655, JAPAN;
Email: miyo-sup@h.u-tokyo.ac.jp>

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Is the Future Evidence-Based?

Ben W.I. Sissons[a], W. Alex Gray[a], Tony Bater[b], Dave Morrey[b]

[a] *School of Computer Science, Cardiff University, UK*
[b] *Clinical Information Unit, Velindre NHS Trust, UK*

Abstract

This paper is concerned with how the future information needs of the medical community should be met. The current dominant belief within medicine is that these information needs should be met from bespoke research studies. The necessity of this approach is far from certain. Health organisations worldwide are currently investing vast resources into centralising and amalgamating every day patient data. Is there a future for these Electronic Medical Records in informing medical decisions? This paper describes the challenges to be met in using both research studies and every day patient data to inform medical decisions. It then describes an ongoing practical project to evaluate these information sources' ability to meet the information needs of cancer care decision makers. Details of background, methodology and initial promising results are presented.

Keywords:

evidence-based medicine, clinical decision support systems, computerized medical records

Introduction

For healthcare providers worldwide the creation of national Electronic Medical Records (EMR) is seen as an integral part of meeting the challenges healthcare will face over the next 20 years. These record systems are being designed primarily to make administrative patient information available at all points of care. These systems will also develop into unparalleled sources of potential information for decision makers. This paper presents challenges and opportunities for the use of EMRs to inform decisions about patient treatment. The paper also highlights the need for re-assessment of current best decision making practice in light of these issues. It then describes a project to explore the role of EMRs in our healthcare future and its initial results.

The information needs of treatment decisions

The primary goal of informing treatment decisions is to support clinician and patient in identifying the most beneficial therapy for a patient. Throughout this paper the expression decision makers is intended to mean clinicians and patients.

Clinicians have a professional obligation to make all decisions in the best interests of the patient. Sentiments to this effect have been an integral part of medicine since the days of Hippocrates, circa 400BC. Making decisions primarily to suit the best interests of the patient was part of the original Hippocratic Oath and is a central tenet of the modern version, the declaration of Geneva.

In modern medicine it is necessary that any conclusions reached or advice given by health professionals be justifiable. Over the past few decades the professional obligations of clinicians to work in the best interests of patients has developed into a legal obligation. It is theorised that mass education, mass media and a changing legal culture have led to an increasing trend amongst patients to question the decisions of health professionals and seek legal redress for clinical negligence [1-3]. Evidence seems to support this, in the United Kingdom, for example, the number and cost of legal actions against the National Health Service has risen dramatically in the past decade [4, 5]. This increase in legal action, together with a number of high profile mistakes, has created a need for clinicians and healthcare providers to be able to demonstrate that decisions made or advice given to patients are based on relevant and valid information [6]. For ethical, moral and legal reasons evidence to support treatment decisions needs to be auditable, relevant and reflect how beneficial or harmful therapies are to patients.

Evidence-Based Medicine and the research study

Evidence-Based Medicine (EBM) arose to meet the information needs of medical decision making. It advocates basing decisions primarily on randomised and prospective data obtained from research studies. It is now the dominant methodology for informing medical decisions.

The EBM movement arose due to the need to better justify medical decisions. EBM was originally put forward as a paradigm shift from a model of medicine justified by practitioner experience to one based on the "best available" scientific evidence [1].

EBM advocates that treatment decisions should be based on analysis of specially designed studies. The proponents of EBM recommend the use of a strict hierarchy of evidence for informing medical decisions. For treatment decisions this hierarchy is topped by analysis of data from Randomised Controlled Trials (RCT). Where such analy-

ses are not available or cannot be used it is recommended that analyses of specially designed studies that collect data on "cohorts" of patients be used. Only in the absence of such research evidence should the analysis of retrospective, every-day patient data be considered as a basis for medical decisions [2-5]. EBM is very clear that evidence from research studies constitutes the best basis for making medical decisions.

EBM currently dominates the way in which medicine is practiced in much of the world.

Evidential support for the hierarchy of evidence

EBM's hierarchy of evidence is based on theories of good statistical practice; it is not backed by empirical evidence.

The modern hierarchy of evidence for treatment appears to have been proposed in a paper by several of the most prominent proponents of the EBM movement, *"Clinical Recommendations Using Levels Of Evidence For Antithrombotic Agents"* [3]. Preference is given to randomised experimental data because it is believed that without randomisation any conclusions about the relative benefits of treatments to patients would be clouded by the effects of unknown variables on outcomes [4, 5]. No primary evidence is provided for this theory in any of the above papers. This lack of evidence continues through the "users' guide to medical literature" series which represents the best practice manuals of EBM. However, all other factors being equal, it is logical to pay the most attention to the evidence which makes the greatest attempt to filter out bias from its results, in this case RCTs.

The use of analyses from prospective cohort studies is recommended where it is not feasible to use analyses of RCTs. The logic behind this is that, again all other factors being equal, a well planned study that collects its own data should yield more accurate results than a study making secondary use of data collected for other purposes. If all other factors were equal it would be difficult to disagree with this, a prospective study should be at least as sound as a retrospective one.

EBM's hierarchy of evidence seems therefore to be based on reasonable arguments but there is a lack of primary evidence to support these arguments.

Challenges of evidence-based decision making

It is difficult to assess the extent to which research evidence can be used to make inferences about patients in a separate, everyday clinical environment. This represents the greatest challenges in using research evidence to justify medical decisions.

The limited ability to generalise from research studies arises from the manner in which patients enter studies. Patients are recruited in medical centres that are asked to, and agree to, take part in the study. Eligible patients are then required to give their informed consent to take part in the study; in the case of RCTs this means doctor and patient must agree to let chance decide how the patient is treated. This selection process means the patients sampled for such studies are unlikely to be representative of the population they are drawn from, let alone a separate population of patients. Making inferences to decisions about patients in daily practice under such circumstances is poor statistical practice. Consequently the representativeness of medical studies is the major challenge faced by the Evidence-Based approach to informing medical decisions.

The problems with study recruitment and representativeness can be confirmed from practice. In the UK official figures show the participation of UK cancer patients in clinical trials is low and those running trials can find it difficult to recruit a full complement of patients [1]. A recently completed analysis of over 20,000 US cancer patients that looked at how representative participants in trials are of the population they are sampled from showed there is good reason for concern [2].

The assumption that it is possible to generalise conclusions from research studies, especially RCTs to patients in everyday practice appears to have little grounding in statistical practice or be supported by evidence.

The rise of the Electronic Medical Record

Healthcare organisations worldwide are investing vast resources into developing national EMRs. For example in England this takes the form of the National Program for Information Technology (NPFiT) [3], in Wales the Informing Healthcare project [4] and in Australia HealthConnect [5].

These national EMRs provide great opportunities for improving the information available to healthcare decision makers. The eventual goal of these projects is to make all relevant information about each patient available when and where it is needed. As a side effect massive standardised datasets covering the majority of encounters between patients and healthcare providers will be created. These massive datasets potentially represent all patients treated by each healthcare provider and will almost certainly contain information that can support medical decision making.

Challenges of Electronic-Medical-Record based decisions

Analysis of EMR data has the potential to be much more representative of local circumstance and patient outcomes than analysis from research studies. This data is retrospective and documents un-randomised treatment decisions. Consequently in an Evidence-Based Medical world EMR data is only fit for informing medical decisions in the absence of analyses from research studies. This section examines the challenges that would have to be addressed if EMR Data were to be used to inform treatment decisions.

Justifiability of results

It is very difficult to say from analysis of EMR data what causes any observed differences in patient outcomes. Hence it is difficult to compare how beneficial different treatments are from analysing such data. EMRs document everyday healthcare encounters. In most of these encounters decisions are unlikely to have been taken randomly. Consequently the decision about which treatment a patient receives will have been based on what is known about that

patient, both recorded and unrecorded data. It is difficult to know if any apparent benefit from a treatment is due to the treatment itself or the factors used to select patients for the treatment. For example if one treatment is given to physically fit patients and one to less fit patients, is it patient fitness or the treatment that causes those receiving the first treatment to have better outcomes?

Data quality

Data from EMRs are unlikely to be as suitable for a given analysis as data collected specifically for that analysis. Everyday clinical data is likely to be entered and amended in a high pressure environment where treatment of patients is a higher priority than following data entry protocols. Consequently the collected data may contain frequent errors and omissions. Additionally such data are, in general, recorded primarily for supporting daily care, not analysis. This means that any analysis is restricted to working with what has been recorded, not what may be the most relevant information. Therefore without great effort and foresight, such retrospective data is unlikely to be as comprehensive and suitable for analysis as a prospective dataset.

Representativeness

Although EMR data may document all appropriate past-patient cases, these are still not entirely representative of new patients. Any inferences from the data must still be made across time. It can be argued consequently that EMR data is potentially a more representative base for real-world treatment decisions, but it is still far from perfect in this regard.

Summary

With the rising interest in National EMRs come new opportunities to better inform medical decisions. Unresolved and unmeasured problems exist in adopting either an Evidence-Based or an EMR-Based approach to informing medical treatment decisions. The need for justifiability in medical decisions must be born in mind. Currently an Evidence-Based approach can be justified by its acceptability within medicine. An EMR-Based approach must go to greater lengths to justify its conclusions. It is therefore necessary to develop an EMR-Based approach to informing medical decisions which is capable of addressing the challenges identified in the previous section. Any approach developed then needs to be evaluated against alternative methods for informing medical decisions. The remainder of this paper is concerned with a project that attempts to achieve these objectives.

Methods

A description of an ongoing project in South East Wales, UK, to assess the future role of EMR data in informing cancer care decisions is presented.

Objectives

The broad objectives of this project are:

- To analyse how data from EMRs can justifiably inform medical treatment decisions.
- To compare the practical usefulness of decision information models based on EMRs with models based on best available research evidence.

Methods

Specifically the project examines the role EMRs and best available research evidence could have played in informing past cancer care decisions in South East Wales, UK.

A series of decision information models based on analysis of EMRs and best available research evidence will be used to inform simulated decisions about cancer patients whose outcomes are already known.

Each model will be evaluated against three overall criteria:

1. **Information Provided** - What information would the model have provided to decision makers?
2. **Justifiability** - How justifiable a basis for decisions would the model have been?
3. **Accuracy** - How accurate would the outcomes information provided have been?

Evaluation

As stated above, models will be built and evaluated for information provided, justifiability and accuracy of predictions. This section describes the planned evaluation of each of these factors.

Information provided

This evaluation is designed to determine what benefits or drawbacks each information source has for decision makers. Each of the models will be examined to see what types of information they provide or fail to provide, what form this information is provided in and how specific the information is. This evaluation will be carried out on a qualitative basis with the help of clinical professionals and patient representatives.

Justifiability

This evaluation is to determine to what extent models provide an auditable and defensible basis for decision making. Models will be evaluated for the clarity of conclusions reached and the guarantees or measures of error they offer. Particularly important considerations will be how effectively each model addresses criticisms levelled at it and if the assumptions made in employing the model are explicit and testable. This evaluation will also be qualitative, weighing the pros and cons of basing decisions on each model. Much of this discussion has already been covered above.

Accuracy

This is intended to be a quantitative evaluation of the ability of each model to correctly predict the outcomes of previously unseen patients. The evaluation will assess, with the benefit of hindsight, the practical usefulness of these models to real world decision makers.

In order to evaluate each model's predictive accuracy, a decision simulation will be carried out using historical data of cancer patients in South East Wales. Each model will be

used to make predictions about the outcomes of patients diagnosed in 2003 and 2004 over the two year period following the diagnosis. The average predictive error of the models will be compared to a control model to assess which would have been the most useful for predicting patient outcomes.

Materials

Datasets

Healthcare providers of the South East Wales Cancer Network have operated a shared EMR system for over a decade. The system is appropriately named Information System for Clinical Organisations (ISCO). The purpose of this system has been to collect high quality data for analysis as well as administrative purposes.

The study will be carried out on three data sets from ISCO; Colorectal cancer patients, Breast cancer patients and Oesophageal cancer patients. These datasets have been chosen for the availability and quality of data and research evidence

Models

For each type of cancer, a set of decision information models will be built and evaluated. One model will be based on the best available research evidence as assessed by the Cochrane Collaboration [1] and the National Institute for Health and Clinical Excellence (NICE) [2], the UK's Evidence-Based Guideline body. A second model will be designed to draw as many justifiable conclusions as possible from data of patients from before 2003 contained in the ISCO EMR system. These models will be compared to two control models, also based on the pre 2003 patient data. One will be designed purely to achieve a good predictive accuracy in order to evaluate if useful information is lost by employing the justifiable past-patient data based model. The second will simply predict all outcomes as the average of the past patient dataset in order to establish if any of the other models can provide a benefit over this.

Measuring accuracy

The predictive accuracy of models will actually be measured as predictive error. Models will be tested against each other, making survival predictions for random samples of patients from the test dataset. The null hypothesis that there exists no difference in the underlying predictive error of each model will be assessed using matched pair t-tests. If there is strong evidence against the null hypothesis the alternate hypothesis that some models are better predictors than others will be accepted.

The measure of error will be Graf et al.'s Censored Brier Score [3]. This metric has been chosen as it is a measure of mean squared error and can cope with the special conditions of survival data. The objective of this evaluation is to measure the usefulness of decision information sources to those making decisions about individual patients. A measure of mean squared predictive error will be used as this takes greater account of accuracy for every individual case than non-squared error metrics. As the main outcome to be predicted is death from cancer, the metric must also be able to handle censored observations. These are cases in which the patient is still alive, out of contact or has died of other causes and so no outcome has been measured. Censoring rates can be over 50% of patients in cancer datasets. Ignoring censored cases is consequently likely to detract from the validity of the results. For a full description of the metric together with formulas see [3].

Initial results and discussion

This paper describes an ongoing investigation; the full investigation cannot be completed until 2007 as two years of outcomes data is needed for all test patients. Initial experimentation however has yielded some encouraging results.

Prototype systems based on the above models have been reviewed by clinicians to assess their ability to inform decision making. Feedback from this process has been good but a formal evaluation of finalised models has yet to be carried out.

The pros and cons of the justifiability of each data are discussed extensively above. Initial prototyping suggests an EMR-based system that addresses justifiability criticisms is feasible. Our current approach is to try and isolate effects of individual variables, especially the effects of treatment. The main challenge faced here is that resulting models tend to provide vague information, but progress is being made.

Promising results have been forthcoming from building the control model that attempts to get the best predictive accuracy from the pre 2003 EMR data (see above). The Cox Proportional Hazards model [1], a regression model specifically designed to cope with survival data, was compared to an average past-patient outcome model in 10-fold cross validation. The Cox model was fitted by systematically testing combinations of variables on the training set and choosing the simplest model that was the best fit. The cross validation was run 20 times, with each model using the same 9 folds of training data to make predictions for the test fold. For each of the three datasets a t statistic based on the difference between average Integrated Brier Scores was calculated. The resulting probability that the Cox model had the same predictive error as the average model was less than one in 4 billion chance ($p < 2.5*10^{-10}$) for each dataset. In all cases the Cox models produced a lower predictive error than the control models. The reduction in the predictive error was 3.2%, 5.6% and 3.4% respectively for the Colorectal, Oesophageal and Breast cancer patient datasets. These benefits from the fitted Cox model may seem small in terms of percentages, but these results show that there is an extremely high probability that the EMRs contain information that can be of benefit to decision makers.

Criticisms of approach

This project tests models against data from EMRs and consequently also faces the challenges described above. This does not detract from the fact however, that this is a very practical study, measuring information models against

real-world data. At the end of the day these models are designed to inform such real-world decisions.

Conclusion

For ethical, moral and legal reasons it is important to have reliable and justifiable sources of information for decisions over patient care. Currently Evidence-Based Medicine is the dominant methodology for meeting these needs and advocates basing decisions primarily on analysis of well designed, bespoke research studies. Many countries see national, and perhaps international, Electronic Medical Record systems as having a pivotal role in meeting future healthcare needs. This means that increasingly large collections of everyday clinical data are becoming available to decision makers. Should this data be used to inform medical decision or are we to write off a potentially valuable information resource as bad evidence? The study described here represents a step along the road to resolving this issue.

Although the project is still in its early stages, the initial results are very promising. There is almost certainly information contained in the South East Wales Cancer Networks Electronic Medical Records that could benefit decision makers; the major challenge is unlocking it.

Acknowledgments

We would like to thank clinicians and information staff at Velindre NHS Trust and the University Hospital of Wales for their valuable insight and support.

References

[1] Eysenbach G and Jadad A. Evidence-based patient choice and consumer health informatics in the Internet age. J Med Internet Res 2001: 3(2): p. e19.

[2] Risk A and Dzenowagis J. Review of Internet Health Information Quality Initiatives. J Med Internet Res 2001: 3(4): p. e28.

[3] Sastry S and Carroll P. Doctors, patients and the Internet: time to grasp the nettle. Clin Med 2002: 2(2): p. 131-133.

[4] NHS Litigation Authority. NHS Litigation Authority, 2006. http://www.nhsla.com/, Accessed: 28/09/2006

[5] National Audit Office. Handling clinical negligence claims in England. London: National Audit Office, 2001.

[6] Changing Attitudes and the Response to clinical claims and Complaints. Clin Risk 2001: 7(2): p. 124-125.

[7] Guyatt G. Evidence-Based Medicine - A new approach to teaching the practice of medicine. JAMA 1992: 268(17): p. 2420-2425.

[8] Levels of Evidence and Grades of Recommendation. Centre for Evidence-Based Medicine, 2005. WWW: http://www.cebm.net , Accessed: 15/11/2006

[9] Cook DJ, Guyatt GH, Laupacis A, Sackett DL, and Goldberg RJ. Clinical recommendations using levels of evidence for antithrombotic agents. Chest 1995: 108(4): p. S227-S230.

[10] Guyatt GH, Sackett DL, and Cook DJ. Users' guides to the medical literature. II. How to use an article about therapy or prevention. A. Are the results of the study valid? Evidence-Based Medicine Working Group. JAMA 1993: 270(21): p. 2598-601.

[11] Guyatt GH, Sackett DL, and Cook DJ. Users' guides to the medical literature. II. How to use an article about therapy or prevention. B. What were the results and will they help me in caring for my patients? Evidence-Based Medicine Working Group. JAMA 1994: 271(1): p. 59-63.

[12] Vale C, Stewart L, and Tierney J. Trends in UK cancer trials: results from the UK Coordinating Committee for Cancer Research National Register of Cancer Trials. Br J Cancer 2005: 92(5): p. 811-4.

[13] Elting LS, Cooksley C, Bekele BN, Frumovitz M, Avritscher EBC, Sun C, and Bodurka DC. Generalizability of cancer clinical trial results. Cancer 2006: 106(11): p. 2452-2458.

[14] NHS Connecting for Health. The National Programme for IT Implementation Guide. London: DoH, 2005.

[15] Welsh Assembly Government. Informing Healthcare: Transforming Healthcare Using Information and IT. Cardiff: Welsh Assembly Government, 2003.

[16] HealthConnect. HealthConnect: An overview. Canberra: HealthConnect, 2004.

[17] The Cochrane Collaboration: The reliable source of evidence in health care. The Cochrane Collaboration, 2006. http://www.cochrane.org/, Accessed: 20/10/2006

[18] National Institute for Health and Clinical Excellence. A Guide to NICE. London: NICE, 2005.

[19] Graf E, Schmoor C, Sauerbrei W, and Schumacher M. Assessment and comparison of prognostic classification schemes for survival data. Stat Med 1999: 18(17-18): p. 2529-45.

[20] Cox D. Regression Models and Life Tables. J R Stat Soc 1972: Series B(34): p. 187-220.

Address for correspondence

Ben Sissons
Email: B.W.I.Sissons@cs.cf.ac.uk
School of Computer Science, 5 the Parade, Cardiff, CF24 3AA, UK

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Comparing Decision Support Methodologies for Identifying Asthma Exacerbations

Judith W Dexheimer[a], Laura E Brown[a], Jeffrey Leegon[b], Dominik Aronsky[a,c]

[a] *Department of Biomedical Informatics, Vanderbilt University, Nashville, TN, USA*
[b] *Vanderbilt Center for Better Health, Vanderbilt University, Nashville, TN, USA*
[c] *Department of Emergency Medicine, Vanderbilt University, Nashville, TN, USA*

Abstract

Objective: To apply and compare common machine learning techniques with an expert-built Bayesian Network to determine eligibility for asthma guidelines in pediatric emergency department patients.

Population: All patients 2-18 years of age presenting to a pediatric emergency department during a 2-month study period.

Methods: We created an artificial neural network, a support vector machine, a Gaussian process, and a learned Bayesian network to compare each method's ability to detect patients eligible for asthma guidelines. Our outcome measures included the area under the receiver operating characteristic curves, sensitivity, specificity, predictive values, and likelihood ratios.

Results: The data were randomly split into a training set (n=3017) and test set (n=1006) for analysis. The systems performed equally well. The area under the receiver operating characteristic curve was 0.959 for the expert-built Bayesian network, 0.962 for the automatically constructed Bayesian network, 0.956 for the Gaussian Process, and 0.937 for the artificial neural network.

Discussion: All four evaluated machine learning methods achieved high accuracy. The automatically created Bayesian network performed similarly to the expert-built network. These methods could be applied to create a real-time detection system for identifying asthma patients.

Keywords:

asthma, decision support techniques, decision support systems, evaluation

Introduction

Asthma is one of the most common pediatric illnesses. It causes an estimated 14 million missed school days and 14.5 million missed workdays yearly in the United States [1]. Asthma exacerbations account for more than 1.8 million emergency department (ED) visits annually [1].

Adherence with guidelines has been shown to improve the clinical care patients receive [2, 3]. A common barrier to guideline initiation is determining eligible patients [4]. Nurses in addition to their normal workload are often charged with identifying eligible patients, a task that is frequently forgotten and leads to lower guideline adherence. Automatically detecting patients could help improve guideline adherence. Ideally, an electronic system that requires no additional data entry would be used to detect guideline eligible patients.

The goal of our study was to evaluate several machine learning techniques using a verified asthma dataset and to compare the techniques with an existing expert-built Bayesian Network (BN) used to identify asthma patients with asthma exacerbations in a pediatric ED [5]. We compared the expert-built BN with a BN automatically learned from data, a Support Vector Machine (SVM), an Artificial Neural Network (ANN), and a Gaussian Process (GP).

Background and methods

Various computerized methods have been developed for identifying asthma patients [6, 7]. These studies have focused on clinical data - such as peak flow, clinic notes or discharge summaries, and computer-based questionnaires [6]. Although some of the studies used ED data, they were not stand-alone systems or integrated into the clinical workflow or with existing information technology infrastructures.

Setting

The study data were collected at Vanderbilt Children's Hospital ED, a 29-bed pediatric facility with more than 40,000 visits annually. The ED information system infrastructure includes an electronic whiteboard, an electronic triage system, an electronic medical record, and a computerized provider order-entry system in the ED. The study was approved by the institution's IRB.

Population

All patients presenting to the pediatric ED during the 2-month study period were screened for inclusion. Patients were included if they were 2-18 years of age and seen in the pediatric ED. Patients were excluded if they did not have a coded chief-complaint, were not treated in the ED (such as left without being examined) or had no final diagnosis for their visit in the paper or electronic patient record.

Design objectives

Our objective was to apply and compare machine learning techniques to an expert-built Bayesian Network. The result of each technique could be developed into a real-time system for detecting asthma patients presenting to a pediatric ED. We adapted the design objectives of a prior BN study to fit our additional techniques [5].

Data sources

We used the database of the previous study [5] which included 4,023 patient encounters. The randomized splits for training and testing of the original BN were preserved. The data included commonly available variables from the electronic medical record, electronic triage, and electronic whiteboard such as 141 ICD-9 coded chief complaints, a past medical history of asthma, billing codes, and asthma medications including beta-agonists, steroids and others. Chief complaints and asthma medication were identified using searches of free-text. Text matching has been shown to be an accurate method to determine the past medical history of asthma [8]. The past medical history of asthma was determined by searching the billing records for prior asthma diagnosis codes (ICD-9 493.*). The patient's chief complaint, acuity level, age, respiratory rate, and oxygen saturation came from the electronic triage. All of these variables are regularly captured and stored in the hospital's information systems. All data elements were available through the computerized hospital information systems, and no additional data collection was necessary.

Dataset

The dataset collected consisted of 11 variables: 3 variables (age, oxygen saturation, and respiratory rate) have continuous values, 5 variables are ordinal (acuity, billing codes, and the 3 medication variables), 1 variable (chief complaint) is categorical, and 2 variables (history of asthma and prediction variable of asthma eligibility) are binary. The dataset was discretized following the values given in Table 1 for use in constructing the expert BN.

BN learning algorithms, including the one used in this study, generally require complete datasets, i.e., no missing values. The frequency of missing data in several variables (shown in table 1) limited the choice of methods to handle the missing values (e.g., removal of cases, removal of variables, imputation). The "non-asthma" column includes all patients who were not diagnosed with asthma (n=3,638) and the "asthma" column includes patients diagnosed with asthma (n=385). The missing elements were encoded as an additional value a variable may take. The dataset was randomly split into a training set (n=3017) and testing set (n=1006) with equal proportions of asthma cases.

The ANN, SVM, and GP methods may take continuous values, therefore an alternative dataset was constructed for their use from the original data (i.e. prior to discretization). For this dataset missing data for the continuous variables were imputed using a k-nearest neighbors imputation method. For the medication variables, a missing value was treated as having no history of that medication found in the medical record. Missing acuity values were treated as an additional "0" value. All ordinal and categorical variables were encoded using 1-of-m encoding. Finally, the dataset was scaled to [-1,1] for the SVMs and [0,1] for the ANN technique. The dataset was scaled using Gaussian normalization for the GP.

Table 1 – Missing values

Variable	Values	Non-asthma (%)	Asthma (%)
History of Asthma	Present, Absent	0.0	0.0
Billing Codes	Number: 0, 1, >1	0.0	0.0
Meds: -agonists	Number: 0, 1, >1	51.3	45.2
Meds: Steroids	Number: 0, 1, >1	51.3	45.2
Meds: Other	Number: 0, 1, >1	51.3	45.2
Chief Complaint	141 unique	0.0	0.0
Acuity Level	ESI level: 1-5	1.21	0.26
Age Category	2-3, 4-6, 7-11, 12-18	0.0	0.0
Respiratory Rate	<20, 20-24, 25-29, 30-34, 35-39, >40	4.15	3.38
Oxygen Saturation	<91, 91-92, 93-94, 95-96, 97-98, >98	3.13	0.26

Reference standard

The reference standard for an asthma diagnosis was a free-text diagnosis of "asthma exacerbation," "status asthmaticus," "wheezing," or "reactive airway disease" [5]. A board-certified internal medicine physician determined the asthma diagnosis through manual chart review. Electronic and paper charts were searched for a diagnosis for every ED visit during the study period. Patients without a discharge diagnosis were not included in the study.

Bayesian network

A BN is formalism consisting of a directed acyclic graph with nodes representing variables and a joint probability distribution over the variables. BNs can be created by hand using expert knowledge. The network parameters for the BN are set by the training set and predictions are made on the testing set. BNs are advantageous in that the prediction inference algorithms handle missing data which is prevalent in clinical systems, and they allow an investigator to choose an optimal detection threshold balancing sensitivity and specificity. BNs have been used for disease detection and diagnosis [9].

Max-Min Hill-Climbing

As an alternative to creating the structure of a BN by hand using expert knowledge, a BN was constructed automatically from the data using machine learning techniques. Many algorithms exist for learning BNs; the MMHC algorithm has been shown to outperform on average a number of other prototypical and state-of-the-art BN learning algorithms in an extensive empirical evaluation and was therefore chosen for use in the analysis [10]. MMHC learns the structure of a BN given a dataset, after which the

network parameters are then estimated directly from the data. The MMHC algorithm is available in the Causal Explorer library[1] [11]. The training dataset was used to learn the network structure and estimate the parameters. The Norsys Netica™ API probabilistic inference algorithm was used to predict asthma in the test dataset; the same that was used in the expert BN. This test dataset was used to calculate an AUC for the method.

Support Vector Machine

SVM's, when used in a binary classification problem such as in this study, construct a maximal margin separating hyperplane to discriminate between the two classes of data [12]. SVMs make use of a kernel function to map the input data to a new space typically of much higher dimensions, where an optimization procedure is run to find the linear separating hyperplane. Several different kernel functions are often considered for classification tasks; the full polynomial and radial-basis kernel functions (RBF) were both used in this task. The full polynomial kernel takes two parameters: the degree d of the function, and C, a cost parameter. The RBF kernel also takes two parameters: the cost parameter C, and which determines the width of the function. SVMs have been used to classify clinical data and in clinical data analysis [13, 14].

The SVM classifiers were implemented using LibSVM [15]. The choice of parameters for the classifiers was optimized with stratified 10-fold nested cross validation over the training dataset using empirical area under the receiver operating characteristic curve as a performance measure. The values for each parameter was selected from the following ranges: $C - \{10^{-6}, 10^{-4}, 10^{-2}, 1, 10^{2}, 10^{4}\}$, $d - \{1, 2, 3, 4, 6, 8\}$, and $ - \{10^{-6}, 10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 1\}$. The best classifier was selected and trained using the entire training dataset, and evaluated on the test data set.

Artificial Neural Network

ANNs are modeled after the brain's interconnecting neurons [16]. They use a non-linear approach to create statistical models. ANNs can be used to discover patterns in a dataset, and have been applied to identify asthma patients using responses to a questionnaire [6]. "Learning" occurs through adjusting the connection weights between nodes, finding a result, and then re-adjusting the connection weights.

Our network was developed using the Netlab [16]. A nested 10-fold cross-validation was employed to select the network architecture. The networks were trained and tested using gradient descent with adaptive learning back-propagation with mean squared error as the fitness measure. The best network was selected and trained on the entire training set. An independent test set was then used to calculate a receiver operating characteristic curve (ROC) and asthma probabilities.

Gaussian Process

A GP applies Bayesian techniques to an ANN to create a probabilistic structure that can be used to calculate probabilities. Bayesian methods have been applied to ANNs by placing prior probabilities on the weights of the network. Using Bayesian methods with ANNs elevates the need for a monitoring data set and allows parameters to be determined on the network being trained [17]. GPs are an extension to BNs, but they place prior probabilities on the function.

We used the Gaussian Processes for Machine Learning (GPML) toolbox in Matlab [17, 18] to develop a GP for asthma prediction. The hyperparameters were optimized using the supplied function in the GPML toolbox. We applied a Laplace's Approximation for a binary Gaussian process and selected the commonly used squared exponential covariance function. The hyperparameters are related to the squared exponential covariance function to determine the amplitude and shift of the function in space. The hyperparameter lengths are associated with each variable and determine how much the variable is can vary in its dimension.

Analysis

Performance was evaluated using ROC curves [19]. The ROC curve measures overall test performance and is obtained by plotting sensitivity versus 1-specificity. The area under the ROC curve (AUC) was the primary outcome measure [20] for all techniques with the exception of the SVM. We determined standard operational characteristics for each method including sensitivity, specificity, predictive value, and likelihood. For probabilistic systems, sensitivity was varied from 80%, 85%, 90%, and 95% to determine standard operational characteristics. To compare the methods with the expert-built BN, sensitivity was fixed at 90% as in the previous paper [5].

Results

There were 4,115 patient visits during the study period. Of these, 92 visits were excluded, and 385 (9.6%) had a final diagnosis of asthma. The patient demographics were reported in the previous study. The Bayesian network structure displayed in the left hand side of figure 1 was constructed manually using expert knowledge. The right hand side of figure 1 displays the network structure learned from the data using MMHC.

1 Causal Explorer containing MMHC can be downloaded at http://www.dsl-lab.org.

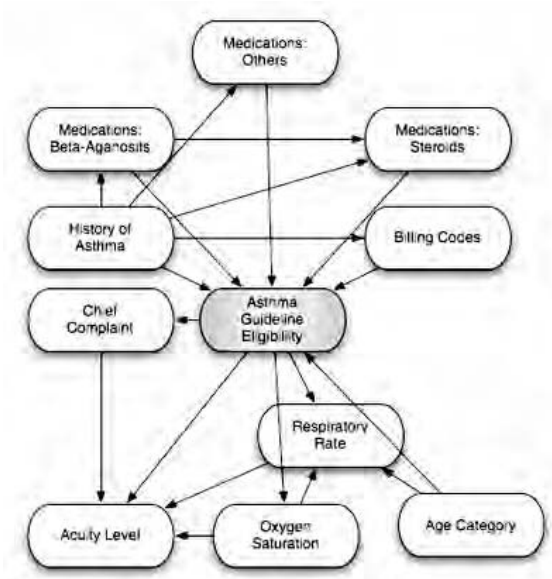
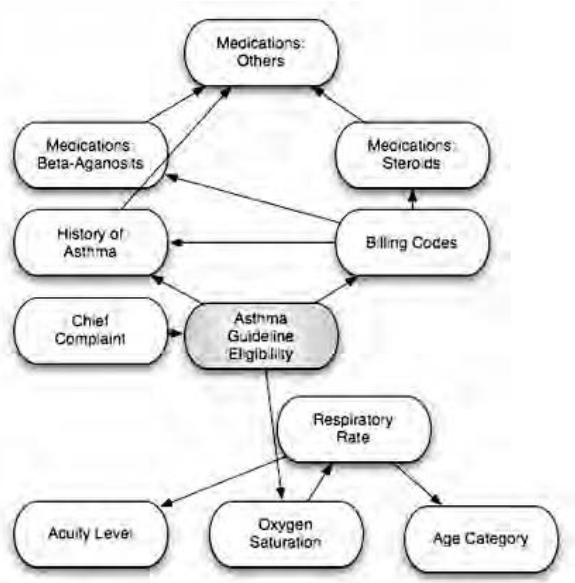


Figure 1 - Expert-built Bayesian Network



MMHC Bayesian Network (structure learned form data)

The AUC for the expert BN was 0.959 (95% CI: 0.933 – 0.977). The MMHC created network AUC was 0.962 (95% CI: 0.935 – 0.980). The ANN AUC with 160 hidden nodes and 160 hidden layers was 0.936 (95% CI: 0.902 – 0.961), and the AUC for the GP was 0.956 (95% CI: 0.923 – 0.976). The ROC curves for the original BN, the BN learned with MMHC, the ANN, and the GP are shown in figure 2. The SVM produces binary predictions with the threshold implicit in the SVM formalism therefore, an SVM ROC curve is not included in this portion of the analysis.

Table 2 shows operational characteristics of the methods with the fixed 90% sensitivity (for methods where the threshold can be varied). The results are reported directly on the predictions made by the SVM since sensitivity cannot be varied. The final parameters selected and used by the SVM was the full polynomial kernel function with a degree of 1 and cost parameter of 1. Operational characteristics for multiple sensitivities were calculated, as in the previous study, for the MMHC network are shown in table 3 for comparison.

Table 2 - Operational characteristics

	SEN (%)	SPEC (%)	PPV (%)	NPV (%)	PLR	NLR
BN	90.0	88.3	44.7	98.9	7.69	0.11
MMHC	90.0	90.1	49.2	98.9	9.16	0.10
ANN	90.0	84.4	38.0	98.8	5.81	0.11
GP	90.0	90.3	49.7	98.9	9.37	0.10
SVM	71.9	98.7	85.2	97.1	54.5	0.29

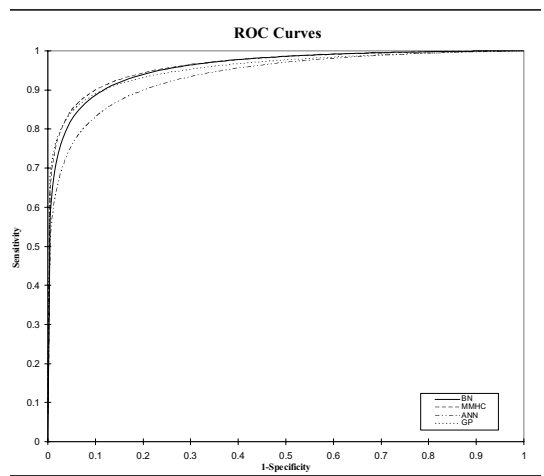


Figure 2 – ROC Curves

The test set had 96 asthma cases and 910 non-asthma cases. At 90% sensitivity, the expert-built BN had 86 true positive predictions and 105 false positive predictions, 805 true negative predictions and 10 false negative predictions. The MMHC network also had 87 true positive predictions and 90 false negative predictions, 820 true negative predictions and 9 false negative predictions. The ANN had 87 true positive predictions and 142 false positive predictions, 768 true negative predictions and 9 false negative predictions. The GP had 87 true positive predictions and 88 false positive predictions, 822 true negative predictions and 9 false negative predictions. At the resulting 72% sensitivity, the SVM had 69 true positive predictions and 12 false positive predictions, 898 true negative predictions and 27 false negative predictions.

Table 3 - MMHC operational characteristics with fixed sensitivity values

SEN (%)	SPEC (%)	PPV (%)	NPV (%)	PLR	NLR
80	97.1	74.8	97.9	28.1	0.20
85	93.8	59.1	98.3	13.7	0.17
90	90.1	49.2	98.9	9.16	0.10
95	86.3	42.1	99.4	6.90	0.06

SEN: Sensitivity, SPEC: Specificity, PPV: Positive Predictive Value, NPV: Negative Predictive Value, PLR: Positive Likelihood Ratio, NLR: Negative Likelihood Ratio

Discussion

The accuracy of the MMHC discovered BN, the SVM, and the GP were comparable to the expert-built BN, however, the ANN did not perform as well as the expert-built BN.

Sparse data may have caused problems in some of the techniques. In our dataset, asthma prevalence was 9.6% of the cases. With this few cases, the SVM and ANN, without adjustment, may not have been able to properly detect the asthma exacerbations. We did not perform any procedures for handling imbalanced data in this study (e.g., over-sampling, under-sampling, one class SVMs, etc). However, such adjustment may lead to better performance of the SVM and ANN.

The MMHC algorithm depends on tests of conditional independence. Accurate estimates for those tests depend on the number of samples and the domain of the variables involved in the testing. For this dataset the Chief Complaint variable can take one of 141 possible values. With the amount of sample provided to the learning algorithm caution must be taken when considering the network produced.

In summary, we believe these methods could be applied to create a real-time detection system for identifying asthma patients using commonly available clinical data.

Acknowledgments

The first author was supported by NLM T15 007450-03.

References

[1] Allergy & Asthma Advocate. Quarterly patient newsletter of the American Academy of Allergy, Asthma and immunology. 2004

[2] National Heart, Lung, and Blood Institute, National Asthma Education and Prevention Program. Expert Panel Report 2: Guidelines for the diagnosis and management of asthma. 1997

[3] Scribano PV, Lerer T, Kennedy D, Cloutier MM. Provider adherence to a clinical practice guideline for acute asthma in a pediatric emergency department. Acad Emerg Med. 2001;8:1147-1152.

[4] Grimshaw JM, Eccles MP, Walker AE, Thomas RE. Changing physicians' behavior: what works and thoughts on getting more things to work. J Contin Educ Health Prof. 2002;22:237-243.

[5] Sanders DL, Aronsky D. Detecting Asthma Exacerbations in a Pediatric Emergency Department Using a Bayesian Network. Proc AMIA Symp. 2006;:684-688.

[6] Sanders DL, Aronsky D. Biomedical informatics applications for asthma care: a systematic review. J Am Med Inform Assoc. 2006. [E-pub].

[7] Hirsch S, Shapiro JL, Turega MA, Frank TL, Niven RM, Frank PI. Using a Neural Network to Screen a Population for Asthma. Ann Epidemiol. 2001;11:369-376.

[8] Sanders DL, Gregg W, Aronsky D. Identifying asthma exacerbations in a pediatric emergency department: A feasibility study. Int J Med Inform. 2006. [E-pub].

[9] Aronsky D, Haug PJ. Diagnosing community-acquired pneumonia with a Bayesian network. Proc AMIA Symp. 1998;;632-636.

[10] Tsamardinos I, Brown LE, Aliferis CF. The max-min hill-climbing Bayesian network structure learning algorithm. Machine Learning. 2006;65:31-78.

[11] Aliferis CF, Tsamardinos I, Statnikov A, Brown LE. Causal Explorer: A Probabilistic Network Learning Toolkit for Biomedical Discovery. In International Conference on Mathematics and Engineering Techniques in Medicine and Biological Sciences (METMBS). 2003;:371-376.

[12] Vapnik V. Statistical Learning Theory. Wiley-Interscience; 1998

[13] Lagor C, Aronsky D, Fiszman M, Haug PJ. Automatic identification of patients eligible for a pneumonia guideline: comparing the diagnostic accuracy of two decision support models. Medinfo. 2001;10:493-497.

[14] Kote-Jarai Z, Matthews L, Osorio A, Shanley S, Giddings I, Moreews F, Locke I, Evans DG, Eccles D, Williams RD, Girolami M, Campbell C, Eeles R. Accurate prediction of BRCA1 and BRCA2 heterozygous genotype using expression profiling after induced DNA damage. Clin Cancer Res. 2006;12:3896-3901.

[15] Chang CC and Lin CJ. LIBSVM: a library for support vector machines. 2001

[16] Nabney IT. Netlab. Springer; 2004

[17] MacKay DJC. Information Theory, Inference & Learning Algorithms. Cambridge University Press; 2002.

[18] Matlab® v 7.3. http://www.mathworks.com

[19] Hanley JA and McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology. 1982;143:29-36.

[20] Hand DJ and Till RJ. A Simple Generalisation of the Area Under the ROC Curve for Multiple Class Classification Problems. Machine Learning. 2001;45:171-186.

Address for correspondence

Judith W Dexheimer, 2209 Garland Ave, Nashville, TN, 37232-8340, judith.dexheimer@vanderbilt.edu

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Analysis and Redesign of a Knowledge Database for a Drug-Drug Interactions Alert System

Daniel Luna, Victoria Otero, Daniela Canosa, Sergio Montenegro, Paula Otero, Fernan Gonzalez Bernaldo de Quirós

Department of Medical Informatics, Hospital Italiano de Buenos Aires, Argentina

Abstract

Physicians tend to ignore drug-drug interactions alerts, this is due to the large amount of irrelevant interactions displayed and the interface in which these alerts are shown. The high rate of clinically inadequate alerts produce "alerts fatigue". This high number of incorrect alerts predisposes physicians to underestimate the electronic prescription systems as useful tools in their practice. We decided to analyze and redesign our drug-drug interactions alerting system knowledge database. In order to do so, we cleaned our knowledge database according to the clinical significance of drug-drug interactions. New drug interactions taxonomy was created in four levels based on clinical significance and the recommendations given in each single monograph of interaction. We proceeded to recategorize the alerts as Active, which present themselves to the physician interrupting the prescribing process, or Passive, which allow physicians to accept the recommendations, and adopt some action in order of minimizing the interaction risks.

Keywords:

drug interactions, medical order entry systems,
clinical decision support systems

Introduction

Errors in medicine, as well as in other human activities, occur frequently. Although most of them do not have harmful consequences, some cause injuries of varying degrees, and may even cause death. It has been reported that a fourth of those errors are related to medications [1]. Errors in medication lead to the so called preventable adverse drug events [2]. Considering the different steps in the medication cycle (prescribing, transcription, dispensing, administrating, and monitoring), it is during prescribing that almost half of the errors occur [3]. The most common mistake during this step is related to a lack of knowledge of the drug and patient information [4]. Some of the errors during the prescription step are labeled **drug-drug interactions** (DDI) [5]. A DDI occurs when one drug affects the metabolism of a second drug, thereby producing adverse effects [6]. DDI occur frequently, but most of them do not lead to adverse events. It has been estimated that only 10 to 15% of these interactions have clinical significance [7]. The occurrence of these interactions varies according to the clinical setting (inpatient, emergency, outpatient) [8]. Evidence shows that physicians do not recognize these interactions 50% of the time, and one-third of the time in serious interactions [9, 10]. Published studies and reviews indicate that **Computerized Physician Order Entry** (CPOE) that provide contextual help at the point of prescription, help in the prevention of prescribing errors and drug adverse events [11-13]. When CPOE are implemented, the clinical workflow can be affected, and generate diverse responses among the physicians using them. When asked about the potential help of CPOE in the prescription process, more than half of the practitioners agreed that they are useful [9, 14], however some studies report that doctors ignore such alerts in 57 to 95% of the time [15, 16]. One of the most important reasons for this high rate of alert overriding is evident in the literature, and is, without a doubt, due to the high rate of false positives (clinically inadequate alerts) which give rise to "alerts fatigue." This high number of incorrect alerts predisposes physicians to underestimate the CPOE as useful tools in their practice. Among the causes of the high rate of manual override of the alerts we found [8, 17]:

- Problems related with the design of the knowledge database of such systems that generate a high rate of false positives.
- Issues related to the utility of the alerts interface.
- The lack of permanent inspection of the interaction between the system and the users, in order to create cycles of continuous improvement.

The Hospital Italiano de Buenos Aires has implemented a CPOE, in the context of an electronic medical record [18], that includes an **alert system for drug-drug interactions**. This work is motivated by the fact that we have not evaluated the rate of overriding alerts by our physicians and on the problem described in the literature about these clinical decision support systems. The objective of our present work is to describe the analysis of the knowledge database of our drug-drug interaction alert system. This analysis includes annotation and purging of the knowledge database according to clinical significance, and proposes changes in its classification of recommendations and alerts visualization.

Selected for best paper award.

Table 1 - Description of fields contained in individual monographs of each DDI of the knowledge database

Drug-Drug Interaction Monographs	
Title	Both drugs (or family of drugs) participating in the interaction
Summary	Referred to the global effect of the interaction
Recommendations	Useful facts for the acting physician to resolve the case when interactions are present. Could recommend to avoid the joint prescription of the related elements, utilize an alternative drug, require complementary testing for monitoring the therapy if adopted, advise, and alert the patient, etc.
Related drugs	Extends the range of the monograph and includes pharmacological, pharmacokinetics, or chemical agents related to the drugs in question
Routes of administration	For each interaction, only the routes of administration for each particular drug are included in order to avoid the appearance of interactions alerts when the drugs are administered thru routes that do not generate an interaction (false positive)
Mechanism of interaction	Details the proposed or postulated mechanism of interaction
Significance	Code assigned to each interaction. It is based on three parameters: potential damage to the patient, frequency and predictability of occurrence, and quality of the documentation that sustains the interaction. They are classified into four levels: **Level 1:** High significance, interaction with great potential to cause damage to the patient, predictable, and frequent, and that it is well documented **Level 2:** Moderate significance, potentially damaging interaction, less predictable, less frequent, or with incomplete documentation **Level 3**: Minimal significance, interaction with low potential for damage to the patient, of variable predictability, of infrequent appearance, or that is based on little documentation **Level 4:** Without clinical significance, even thou this type of interaction can occur, the documentation is based on theoretical considerations or is not clinically significant. Also adverse effects can not be predicted
References	Includes the bibliographic references of published works sustaining the presented information

Methods

The knowledge database of the alert system was created using different sources, including clinical pharmacology textbooks, monographs on products, consultation with specialist in our institution, specific literature searching and a publication on pharmacological interactions named "Evaluation of Drug Interactions (EDI)" [19] maintained by First DataBank Inc. [20]. To build the knowledge database, the information contained in Table 1 was applied.

The EDI is organized in' 18 chapters, according to drug groups (antihypertensive drugs, narcotics, etc.) and according to its index arround 43,000 potential DDI exist. These DDI are generated from the related drugs included in each of the monographs in the EDI [20]. Our knowledge database was created mainly using this index as a guideline. Due to the smaller number of individual drug monographs of a drug-drug or drug-family interaction in comparison with the potential DDI contained in the index (1,201 vs. 43,000), the first step (done by a clinical pharmacologist), was annotating and purging each potential pair according to other sources. In addition, rounds with experienced professionals with the drug in question were also conducted. The objective was that only those interactions with a clear bibliographical and clinical background would remain in the knowledge database. Once the purging was completed, each DDI was analyzed, and a new classification of alerts was defined according to the recommendations for an action contained in the monograph. Finally all interactions were categorized either as active or passive, depending on whether they would manifest themselves actively, interrupting the prescriptive workflow or not.

Results

The first step was the **purging** of the unsubstantiated interactions based on the EDI index. Each of the 1,201 monographs in the EDI were individually analyzed, particularly the section named "Related drugs", where the 43,000 DDI included in the EDI index originated from. As a result of this analysis by a clinical pharmacologist, consultations with other sources, and revision rounds with specialists with daily experience in the use of the drugs, 39,191 DDI were discarded from the knowledge database (originally created from the EDI index as a guideline), and only 2,608 DDI were kept, each related to an individual DDI monograph. Therefore, our knowledge database was formed now with 3,809 DDI (Figure 1).

After this purging, the entries were reclassified according to their clinical significance:

- **Level 1:** 600 (High significance, interaction with great potential to cause damage to the patient, predictable, and frequent, and that it is well documented)
- **Level 2:** 1494 (Moderate significance, potentially damaging interaction, less predictable, less frequent, or with incomplete documentation)

- **Level 3:** 1653 (Minimal significance, interaction with low potential for damage to the patient, of variable predictability, of infrequent appearance, or that is based on little documentation)
- **Level 4:** 62 (Without clinical significance, even thou this type of interaction can occur, the documentation is based on theoretical considerations or is not clinically significant. Also adverse effects can not be predicted)

In addition, during the revision process, each DDI was analyzed in detail for its recommendations given to the prescribing physician. Based on this analysis, a five-domain taxonomy was created. All the recommendations contained in the alert system were then grouped according to this taxonomy (Figure 2).

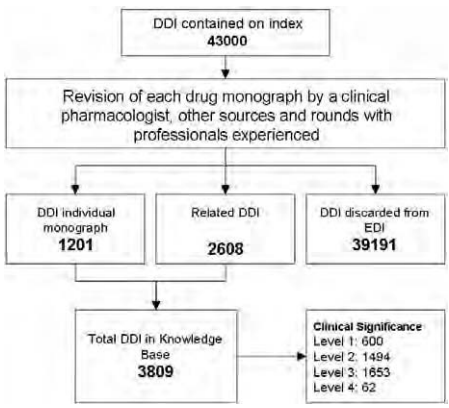


Figure 1 - Purging of the knowledge database

Recommendations taxonomy

Avoid joint use

This domain refers to those interactions that do not provide other options to the physician but to avoid both drugs in combination. This domain is the most important of all because it does not provide alternatives to the prescription. The physician must therefore justify his/her action if he/she decides to proceed and prescribe both drugs simultaneously.

Monitoring Required

This domain includes two sub domains: clinical monitoring and monitoring by complementary tests.

Clinical Monitoring includes all recommendations related to signs or symptoms that physician should inquire about during consultation and follow up with the patient to verify the effectiveness of the drug, control adverse events, or check for drug toxicity.

Some examples include:

- Clinical monitoring of desired effects:
  - Neuromuscular blockage
  - Heart frequency
  - Blood pressure
- Clinical monitoring of adverse effects:
  - Hyperglycemia
  - Gastric ulcers
  - Skin rashes
- Clinical monitoring to detect toxicity:
  - Neurotoxicity
  - Mielotoxicity
  - Cardiotoxicity

Clinical Monitoring with Complementary Tests includes laboratory tests and other studies in this category:

- Monitoring with laboratory tests:
  - Liver function tests
  - Complete Blood Count (CBC)
  - Drugs blood level
- Monitoring with other studies:
  - Electrocardiogram (EKG)
  - Electromyogram (EMG)
  - Central venous pressure

Evaluate alternative drugs

This domain includes a recommendation to search for other drugs as possible substitutions for one of the interacting pair; it also includes a support system that would provide substitution alternatives, assuring identical or similar therapeutic efficacy as the drug being replaced, and also make sure that an alternative drug will not interact with the second drug in the pair. Some examples of alternative options are:

- Acetyl salicylic acid: Acetaminophen
- Cimetidine: Famotidine/Nizatidine
- Guanetidine: Methildopa
- Erythromycin: Azytromicine

Modify administration

This recommendation does not avoid the joint administration of both drugs; however, it suggests the mode of administration, for instance:

- Space the administration of both drugs as far apart as possible in time
- Modify the dosage of one or both drugs
- Select alternative routes for administering one or both drugs
- Choose alternative pharmaceutical formats

Inform the patient

Faced with the decision of administrate an interacting pair of drugs, physicians can inform the patient about signs of alarm and other additional recommendations in order to minimize possible consequences of the interaction. Among others, they are:

- Signs of alarm related to hepatotoxicity
- Potential decrease in the contraceptive effect (evaluate alternative contraceptive methods)

- Modifications in the diet
- Signs of alarm of myolisis

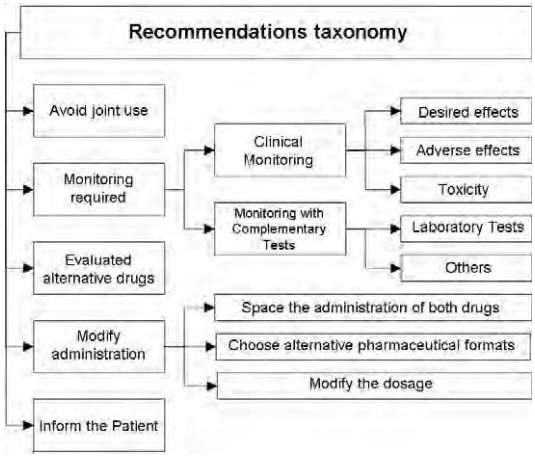


Figure 2 - Recommendations taxonomy contained in the alert system knowledge database

Alerts characteristics

Alerts are a type of clinical decision support systems and they must be intrusive, stop the workflow, and get the user's attention. Due to the fact that physicians prefer this type of alerts to be selectively used and shown in the right clinical context [21], we decided to re-categorized all DDI in the knowledge database as **active** or **passive** DDI alerts:

- **Active** DDI alerts**:** these alerts would present themselves to the physician interrupting the prescribing process, and forcing physician to explain his/her action if he/she decides to go ahead with prescription of the interacting drugs. Only 695 interactions (Table 2) are included in this category, coming from two types of DDI:
  - DDI for which the monograph recommends the avoidance of simultaneous administration. So, these DDI fall in the first domain described above ("Avoid joint use").
  - DDI that have been classified as belonging to Level 1 or High Significance are DDI with great potential to produce harm to the patient. They are frequent, and they are well documented.
- **Passive** DDI alerts: these include the remaining 3,114 interactions which allow physicians to accept the recommendations, and adopt some action in order of minimizing the interaction risks. These DDI are in the remaining domains in the DDI categories. These alerts would appear in the background, and would not interrupt the prescribing process (Table 2).

Significance	Avoid joint use Domain	Other domains
Level 1	39	561
Level 2	80	1414
Level 3	14	1639
Level 4	1	61
	Actives = 695	**Passives** = 3114

Table 2 - Re-categorization of the DDI in actives (Significance level 1 + avoid joint use domain) or passives (Significance level 2, 3 y 4 + other domains)

Discussion

The knowledge database of the support systems for decision making in the field of pharmacological prescription is generally commercially acquired, or is developed based on periodical publications. Such knowledge databases are frequently quite inclusive, putting more emphasis on the domain coverage than in the clinical relevance or the severity of adverse effects that interactions may provoke [22]. Due to this limitation, and based on the information supplied by different studies [23, 24] in which problems with the EDI are presented, we decided to undergo an analysis and subsequent purging of the knowledge database of our alert system (the EDI being an important source as guideline of our alert system as indicated above).

Clear recommendations are available as to what characteristics pharmacological alerts must meet [17, 21, 25], Therefore, the objective of the redesign of the knowledge database of our alert system was undertaken to improve the acceptance of these alerts by physicians, and minimize interruptions in the prescribing process, only leaving the most serious DDI in this group. Already there are reports that confirm that the redesign of this knowledge bases (only leaving a reduced and very specific set of alerts) has increased acceptance by physicians in general [26] and reduces the number of manual overrides of the alerts [27]. We also believe that the creation of taxonomy of recommendations related to DDI allows the physician to accept such recommendations more readily, without having his/her actions being considered as ignored alerts.

Before we implement the changes in the knowledge base of our alert system, we will conduct a study to evaluate the usability of new alerts with a group of physicians from our institution. Based on the results of this evaluation, future implementation will be decided.

References

[1] Leape LL, Brennan TA, Laird N, Lawthers AG, Localio AR, Barnes BA, et al. The nature of adverse events in hospitalized patients. Results of the Harvard Medical Practice Study II. N Engl J Med. 1991 Feb 7;324(6):377-84.

[2] Nebeker JR, Barach P, Samore MH. Clarifying adverse drug events: a clinician's guide to terminology, documentation, and reporting. Ann Intern Med. 2004 May 18; 140(10):795-801.

[3] Bates DW, Cullen DJ, Laird N, Petersen LA, Small SD, Servi D, et al. Incidence of adverse drug events and potential adverse drug events. Implications for prevention. ADE Prevention Study Group. Jama. 1995 Jul 5; 274(1):29-34.
[4] Leape LL, Bates DW, Cullen DJ, Cooper J, Demonaco HJ, Gallivan T, et al. Systems analysis of adverse drug events. ADE Prevention Study Group. Jama. 1995 Jul 5; 274(1):35-43.
[5] Allard J, Carthey J, Cope J, Pitt M, Woodward S. Medication errors: causes, prevention and reduction. Br J Haematol. 2002 Feb;116(2):255-65.
[6] Goodman LS, Hardman JG, Limbird LE, Gilman AG. Goodman & Gilman's the pharmacological basis of therapeutics. 10th ed. New York: McGraw-Hill 2001.
[7] Seymour RM, Routledge PA. Important drug-drug interactions in the elderly. Drugs Aging. 1998 Jun;12(6):485-94.
[8] Luna D, Gonzalez Bernaldo de Quiros F, Martinez M, Gambarte ML, Morchón A, Costa G, et al. ¿Porque se ignoran los alertas sobre interacciones farmacológicas? 3er Virtual Congress of Medical Informatics - Informedica; 2004.
[9] Glassman PA, Simon B, Belperio P, Lanto A. Improving recognition of drug interactions: benefits and barriers to using automated drug alerts. Med Care. 2002 Dec;40(12):1161-71.
[10] Luna D, Hares D, Schpilberg M, Hernandez G, Soriano E, Martinez M, et al. Validación de la base de conocimiento de un sistema notificador de interacciones farmacológicas. 5to Simposio de Informática en Salud - 31 JAIIO; 2002; Santa Fe, Argentina.
[11] Bates DW, Teich JM, Lee J, Seger D, Kuperman GJ, Ma'Luf N, et al. The impact of computerized physician order entry on medication error prevention. J Am Med Inform Assoc. 1999 Jul-Aug;6(4):313-21.
[12] Poikonen J, JM L. Medication-Management Issues at the Point of Care. J Healthc Inf Manag. 1999;13(2):43-52.
[13] Teich JM, Merchia PR, Schmiz JL, Kuperman GJ, Spurr CD, Bates DW. Effects of computerized physician order entry on prescribing practices. Arch Intern Med. 2000;160(18):2741-7.
[14] Magnus D, Rodgers S, Avery AJ. GPs' views on computerized drug interaction alerts: questionnaire survey. J Clin Pharm Ther. 2002 Oct;27(5):377-82.
[15] Payne TH, Nichol WP, Hoey P, Savarino J. Characteristics and override rates of order checks in a practitioner order entry system. Proc AMIA Symp. 2002:602-6.
[16] Weingart SN, Toth M, Sands DZ, Aronson MD, Davis RB, Phillips RS. Physicians' decisions to override computerized drug alerts in primary care. Arch Intern Med. 2003 Nov 24;163(21):2625-31.
[17] van der Sijs H, Aarts J, Vulto A, Berg M. Overriding of drug safety alerts in computerized physician order entry. J Am Med Inform Assoc. 2006 Mar-Apr;13(2):138-47.
[18] Luna D, Otero P, Gomez A, Martinez M, García Martí S, Schpilberg M, et al. Implementación de una Historia Clínica Electrónica Ambulatoria: "Proyecto ITALICA". 6to Simposio de Informática en Salud - 32 JAIIO; Buenos Aires, Argentina; 2003.
[19] Zucchero F, Hogan M, Sommer C, eds. Evaluations of drug interactions. St. Louis, Missouri, USA: First DataBank Inc. 2005.
[20] First DataBank Inc. Evaluations of Drug Interactions (EDI). 2005 [accessed 10 Oct 2006]; Available from:http://www.firstdatabank.com/products/evaluations/
[21] Krall MA, Sittig DF. Clinician's assessments of outpatient electronic medical record alert and reminder usability and usefulness requirements. Proc AMIA Symp. 2002:400-4.
[22] Reichley RM, Seaton TL, Resetar E, Micek ST, Scott KL, Fraser VJ, et al. Implementing a commercial rule base as a medication order safety net. J Am Med Inform Assoc. 2005 Jul-Aug;12(4):383-9.
[23] Abarca J, Malone DC, Armstrong EP, Grizzle AJ, Hansten PD, Van Bergen RC, et al. Concordance of severity ratings provided in four drug interaction compendia. J Am Pharm Assoc (Wash DC). 2004 Mar-Apr;44(2):136-41.
[24] Malone DC, Abarca J, Hansten PD, Grizzle AJ, Armstrong EP, Van Bergen RC, et al. Identification of serious drug-drug interactions: results of the partnership to prevent drug-drug interactions. J Am Pharm Assoc (Wash DC). 2004 Mar-Apr;44(2):142-51.
[25] Feldstein A, Simon SR, Schneider J, Krall M, Laferriere D, Smith DH, et al. How to design computerized alerts to safe prescribing practices. Jt Comm J Qual Saf. 2004 Nov;30(11):602-13.
[26] Spina JR, Glassman PA, Belperio P, Cader R, Asch S. Clinical relevance of automated drug alerts from the perspective of medical providers. Am J Med Qual. 2005 Jan-Feb;20(1):7-14.
[27] Shah NR, Seger AC, Seger DL, Fiskio JM, Kuperman GJ, Blumenfeld B, et al. Improving acceptance of computerized prescribing alerts in ambulatory care. J Am Med Inform Assoc. 2006 Jan-Feb;13(1):5-11.

Address for correspondence

Dr. Daniel Luna: daniel.luna@hospitalitaliano.org.ar Área de Informática Médica. Hospital Italiano de Buenos Aires. Gascón 450. Ciudad Autónoma de Buenos Aires. Argentina. (C1181ACH). Tel/Fax:+54-11-4959-0507

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Closing the Loop : Bringing Decision Support Clinical Data at the Clinician Desktop

Claudine Bréant[a], Francois Borst[a], René Nkoulou[b], Olivier Irion[c], Antoine Geissbuhler[a]

[a]*Service of Medical Informatics, University Hospitals of Geneva, Geneva, Switzerland*
[b]*Division of Cardiology, University Hospitals of Geneva, Geneva, Switzerland*
[c]*Division of Obstetrics, University Hospitals of Geneva, Geneva, Switzerland*

Abstract

We describe the development of an inquiry office to bridge the gap between clinician needs for decision support systems and readily available large quantities of integrated clinical data. With this link, an information feedback mechanism is implemented that closes the loop of information flow by bringing decision support information from the data warehouse at the clinician desktop. As a result, and as a new DRG cost reimbursement system has been introduced, we have provided the heads of over 30 medical services with an intranet web-based application to access patient encoding of diagnoses, procedures, and DRGs of their respective service. The inquiry office has also developed a query service to process specific requests. It has implemented the automatic screening of patient clinical data of past and current hospitalizations in order to select cases for multiple studies, research, and teaching projects. The purpose of this clinical data warehouse and its information feedback process is to offer a coherent, comprehensive, and reliable return of information to improve decision making, to enable research projects, and to facilitate statistical outputs.

Keywords:

data warehouse, hospital information system, feedback, decision support system

Introduction

With vast amounts of high quality clinical data stored in various databases, the challenge today resides in extracting dynamic information from static data stores. Ultimately, the goal is to provide intelligent and relevant guidance at the point of decision making towards the delivery of better clinical health services. This translates into the integration of information feedback into clinical practice, medical management, research and teaching projects.

Although, it is frequently discussed that electronic patient record (EPR) systems have a tremendous potential to improve care and performance measurements, few developments have been made in the attempt of bringing practice feedback at the clinician desktop [1].

Until now, the prime focus has been to design data warehouses to tackle heterogeneity of medical databases and lack of connection between clinical systems [2][3][4]. The data warehouse technology has become widely spread in the clinical area and health institutions report on the value of integrating their data into a single repository [5][6]. A data warehouse can be simply defined as a copy of transactional data specifically structured, integrated, and organized for complex querying, data analysis, and decision support applications [7].

However, clinicians, medical managers, and researchers understand well their problematic but do not necessarily master the system query languages and the data structures. Today, the full exploitation of clinical data warehouses is impeded by the lack of user-friendly interfaces integrated at the clinician desktop. Furthermore, truly effective methods of feedback from a clinical data warehouse have yet to be found [1]. Needs in terms of access to a clinical data warehouse have been evaluated in the setting of an academic healthcare center and are three folds:

1. automatic patient screening with notification of cases matching a pre-defined criteria (i.e. similar cases retrieval). This mechanism is necessary to both accurately filter patient cases and avoid the tedious task of manually reviewing patient records for a particular study, research, or teaching project [8].
2. on-line dynamic navigation to explore views displaying both indicators (i.e. aggregated data) and detailed patient data. In a multi-dimensional data warehouse, indicators can be calculated across patient records and medical specialties along several axes of analysis [7]. To build a strong analysis, the end-users need to access these indicators as well as the underlying detailed data.
3. data mining and statistics functionalities for automatically searching large volumes of clinical data to extract new patterns representing useful information [9][10][11].

An additional important requirement is the patient data confidentiality that must be enforced according to the institution regulations.

In this study, we describe the development of an efficient inquiry office to bridge the gap between clinicians' needs

and readily available large quantities of integrated clinical data. The paper is organized as follows; materials and methods are described first with a brief review of the overall data workflow architecture, including more specifically the data warehouse content and data model. The focus is then on the information feedback process that was lately added to bring decision support information from the data warehouse to the clinical desktop. With this new link, an information feedback mechanism is implemented that closes the loop of information flow as depicted in Figure 1. Finally, results are presented for : 1) an intranet web-based application for medical activity analysis linked to a new cost reimbursement system, and 2) for a query service for a customized access to the clinical data warehouse.

Materials and methods

The data workflow architecture

The electronic patient record (EPR) organizes the acquisition and the visualization of patient clinical data during the care production workflow. Furthermore, this data is managed by the various independent clinical care systems such as for instance the radiology information system (RIS), the admission/discharge/transfer (ADT) system, and the laboratory system. In this setting, the University Hospitals of Geneva (HUG) have been accumulating large amounts of clinical data for various domains of clinical activity, embedding knowledge and experience produced during the patient care processes. However, to improve and evaluate clinical practice, information processing must go beyond basic automation (i.e. patient oriented access only) and convert EPR data into aggregated, multidimensional information. Hence, following the tendency in other industries, it requires the integration of patient clinical data into a single repository.

Table 1 – Content of the Archimed data warehouse.It currently includes seven main data sources from the HIS.

Archimed Clinical Data Marts	**HIS components (*data source names*) and years covered**	**Facts (millions)**
ADT	1990 (*Impact*) ->2005 (*DPA*)	10M
Diagnoses (ICD codes)	1990 (*PDP*)	2.8M
Procedures (CHOP codes)	1990 (*PDP*)	1M
DRG codes	2005 (*APDRG*)	0.5M
Laboratory	1993 (*Unilab*)->2005 (*Unilab2*)	75M
Clinical data questionnaires	2002 (*DPI-Form*)	10M
Radiology exams	1990 (*Unimage*)->2004 (*Xplore*)	2M

The Archimed clinical data warehouse database has succeeded in procuring an integrated and coherent view of a wide variety of medical and administrative data, all centered on patient care processes, acquired from multiple

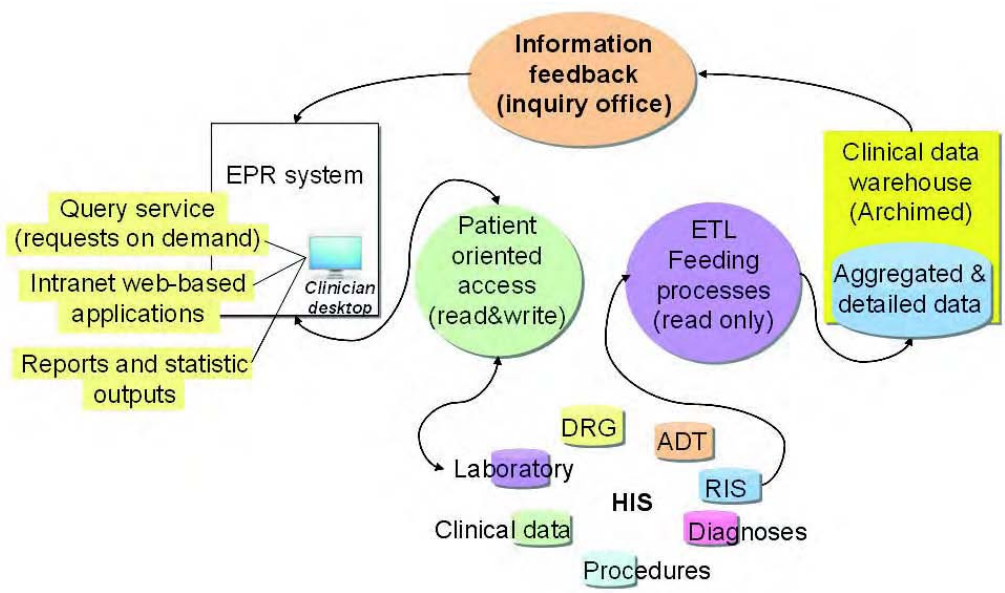


Figure 1 – The data workflow architecture. The EPR system provides a read/write access to clinical data that is patient oriented only. ETL processes are run daily to extract, transform, and integrate HIS data into a single multi-dimensional database (Archimed). The information feedback process allows the return of decision support information towards the clinician desktop.

independent and heterogeneous systems part of the HIS [12]. Extract/Transform/Load (ETL) processes are run daily to refresh data. It currently stores patient clinical data since the early 90's for a total of over 750'000 patients, 1.7M hospitalization stays, and 2M out-patient visits. More specifically, the Archimed database includes around 100 millions patient facts spread over 7 data marts, as listed in Table 1. As an example, the ADT data mart stores detailed patient stay administrative information consisting in 10 millions of patient facts. It was fed by the *Impact* ADT system until the end of 2004, and then by the upgraded *DPA* system from 2005.

Archimed is designed according to the data warehouse bus (DWB) architecture and the multi-dimensional model advocated by Kimball [7]. In this architecture, the data warehouse is composed of a set of data marts connected to each other through some specific dimensions called conformed dimension tables. The conformed dimension tables store key data on which each care process relies and on which the whole institution agrees and complies. In the case of Archimed, the conformed dimension tables have been identified at an early stage of the design and model the axes of analysis of the data warehouse. These are the patient administrative data, the episode of care description, the medical services, the care units of the hospital, as well as the care centers. This architectural design builds the foundation for a simple and powerful querying of the patient medical data as if it were initially part of the same database [12]. Figure 2 illustrates the Archimed data model. Black dots show how the fact tables are linked to the conformed dimensions tables.

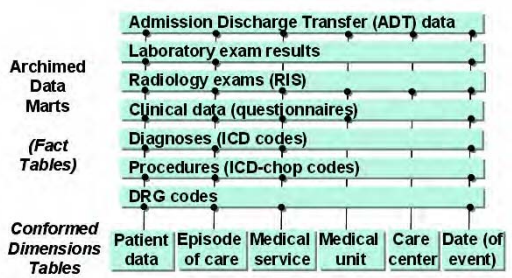


Figure 2 – The Archimed clinical data warehouse bus architecture

Based on the DWB architecture, each data mart or domain of activity has been progressively integrated into the Archimed system, breaking the overall implementation task to manageable proportions. This architecture also provides scalability to the system, facilitating the future addition of new data marts.

The information feedback process

A new inquiry office has been created to go beyond the production of simple static reports and to fully leverage the value of the Archimed data warehouse. The inquiry office is responsible for implementing the information feedback process to bring decision support data at the clinician desktop. It fulfills the following tasks:

- Act as a mediator between the data warehouse system and the clinicians needs,
- Develop both computerized tools and their integration into the clinician workflow,
- Enforce procedures regarding patient data confidentiality, according to our institution regulations.

The restitution of information is currently performed according to the two following methods: 1) intranet web-based applications, 2) a query service for a customized access to the data warehouse.

On-line access through intranet web-based applications

We have utilized the intranet web-based technology to facilitate the navigation into a pre-defined subset of data at different levels of granularity. Views of clinical information, possibly combining the data from several data marts are prepared from the data warehouse. These are available on-line at the desktop of the end-user through intranet web-based applications. The user can define and parameterize views by, for instance, choosing a period of interest or by specifying a patient population. Different levels of granularity can be presented to convey the information. These range from aggregates to summarize data at various levels of interest, to detailed patient clinical data. *Drill-down* and *roll-up* links allow to navigate between views to explore more detailed data, or on the contrary to analyze more synthetic information.

Furthermore, the intranet web-based applications managed by the inquiry office are connected to the hospital access right management system hence providing the level of data confidentiality required by the institution.

A query service for customized access to the data warehouse

We have defined a new environment to automatically screen patient data according to a user-defined criterion. As described above, our data model gives us the ability to easily build complex criterions involving data from as many clinical activities as needed. Such queries take only a few seconds to a few minutes to run against the whole database. The basic result obtained is a list of patient cases. According to users needs, this list can be enriched in a second step with an extraction of partial or total clinical data for the patients of the initial list. Finally, results are communicated either simply by e-mail, or through intranet web-based views made available within the EPR environment.

The processing of a new request is as follows. The clinician first contacts the inquiry office to communicate a new request. Proper authorization to access the data is checked according to the HUG patient confidentiality regulations. Then, the inquiry office builds the queries that access the data warehouse, and results are carefully validated with the clinician. Some parameters are set such as the frequency and the duration of the hospital patient population screening.

Results

Statmed : a desktop tool to understand the implications the new DRG reimbursement system at the HUG

As in many countries, the reimbursement system in Switzerland for acute inpatient care has recently transferred from a standard daily reimbursement to a complete diagnosis-related group (DRG) reimbursement system. The DRG system classifies hospital cases into one of approximately 900 groups, expected to have similar hospital resource use. It is now part of the HUG payment system since 2006. DRGs are assigned by a grouper program mainly based on ICD diagnoses, procedures, age, sex, and the presence of comorbidities. Reimbursement is calculated from a relative weight of a procedure multiplied by a base rate.

Figure 3 - View of the procedures performed by the division of visceral surgery and corresponding number of cases. The patient icon allows to drill down to more detailed patient medical information.

This is a major change for the HUG and some of the implications are unknown. In that case, information feedback is of paramount to accompany the change and help users to get insights and understanding of the new reimbursement system. Consequently, we have provided the heads of over 30 medical services with an intranet web-based application available at their desktop to access both patient detailed and summarized data related to the encoding of diagnoses, procedures, and DRGs of their respective service.

The user first selects a period of interest and a medical service (i.e. division of visceral surgery). Then, he can choose to view the number of diagnoses, procedures (Figure 3), or DRGs (Figure 4) coded for patients who completed their stay within the period of interest in the selected medical service.

For further analysis, the user can select a DRG code and drill-down to the encoding of the underlying diagnoses and procedures to study the link between the medical activity and the financial outcome. Finally, from any dashboard displaying codes of diagnoses, procedures, or DRGs, the user can retrieve the corresponding list of patient identification numbers and connect to the EPR system to have a view of the entire record file to explore other details regarding the patient hospitalization stay. For instance, when encountering an intriguing anomaly in the data, it is then necessary to access the full patient record to determine the exact cause.

In addition to the clinicians, this program is used by the administrative staff of about 15 members, responsible for the ICD encoding of the patient diagnoses and procedures. They regularly use the program to perform quality control verifications, to detect possible errors in code attribution, and to monitor their activity.

Figure 4 - View of the DRG codes for patients hospitalized in the division of visceral surgery. Icons on the left part of the screen can be used to drill down to a more detailed description of underlying patient cases, diagnostic and procedure codes.

The current release of Statmed merges data from four different data marts: Diagnostic and Procedure ICD Encoding, DRG codes, and ADT administrative data. It is a Java based program that automatically generates Structured Query Language (SQL) queries to retrieve data, to build aggregate and then to display dashboards presenting the data.

Automatic patient screening and notification of selected cases

The inquiry office is regularly called for support by clinicians for implementing the automatic screening of patient clinical data of past and current hospitalizations in order to select cases for multiple studies, research, or teaching projects. As an illustration, we have listed below some examples of patient screening requests among about eighty inquiries processed in 2006 at the HUG:

- The european multicenter Micado randomized controlled trial which enrolls women with preterm premature rupture of membranes between 28 and 32 weeks (division of obstetrics).
- The multicenter AMIS registry and the MIDAS study collect data on patients with acute coronary syndrome in Switzerland (division of cardiology).

- Study regarding guideline and management of adult asthma in Switzerland (division of general internal medicine).
- Study of children with cystinuria (division of nephrology).
- Extraction of clinical data for the calculus of the APACHE and SAPS scores for the patients of the intensive care unit (division of intensive care).
- Epidemiological study of hip fractures among the elderly population in Geneva's county (Switzerland) (division of geriatrics).
- Two-years follow-up of intensive care patients (division of intensive care).
- Study of the obstetrical outcome after treatment of cervical dysplasia (division of obstetrics).

Regarding qualitative results, and more particularly in the case of the AMIS registry, it has been established that a retrospective database can be used to analyze clinical practice and furthermore participate in the modification of the therapeutic behavior of the physicians [13].

Conclusion

In this work, we have demonstrated the feasibility of a mechanism to bring dynamic data prepared from a hospital-wide clinical data warehouse to the clinician desktop. The results obtained so far pave the way for the development of more client interfaces depending of specific needs. The web-based technology allows these applications to be quickly developed and required no deployment effort. Hence, the data warehouse is not any longer a stand-alone system but becomes fully integrated into the clinician routine workflow.

To fully leverage the data warehouse towards a usable information system, we have created an inquiry office including methodology and tools in order to assist users access data. Graphical user interfaces and web-based technology are used to provide on-line meaningful views of clinical information. As of today, clinicians are using the system for analysis of the new HUG cost reimbursement scheme and for monitoring their medical activities. This system necessitates very few training. Specific requests regarding case studies, the retrieval of similar cases, and statistics are processed on demand by the inquiry office. About eighty requests have been processed in 2006. We have defined a new environment to automatically screen patient data according to a user-defined criterion, and to bring results at the clinician desktop.

The long-term objective of this work is to have this system becoming an integrated part of the clinician workflow for quantitative measurable results. Existing tools will need to evolve, and new developments will be initiated to follow the users needs. For instance, we have tested the coupling of Archimed with data mining tools such as the Weka open source toolkit and Clementine®. A few data mining projects have been initiated in the area of detection and surveillance of nosocomial infections, and to study patient readmissions.

References

[1] Grant A, Moshyk A, Diab H, Integrating feedback from a clinical data warehouse into practice organization. Int J Med Informatics, 75:232-239, 2006.

[2] Ledbetter CS, Morgan MW, Toward best practice: leveraging the electronic patient record as a clinical data warehouse J Healthc Inf Manag. 2001 Summer;15(2):119-31.

[3] Ebidia A, Mulder C, Tripp B, Morgan MW. Getting data out of the electronic patient record: critical steps in building a data warehouse for decision support. Proc AMIA Symp. 1999;:745-9.

[4] Einnbinder JS, Scully KW, Pates RD, Schubart JR, Reynolds RE. Case study: a data warehouse for an academic medical center. J Health Inf Manage 15 (2001) 165-175.

[5] Wisniewski MF, Kieszkowski P, Zagorski BM, Trick WE, Sommers M, Weinstein RA. Development of a clinical data warehouse for hospital infection control. J Am Med Inform Assoc. 2003 Sep-Oct:10(5):454-62.

[6] McGregor C,Bryan G, Curry J, Tracy M. The E-baby data warehouse: a case study. 35th Annu Hawaii Intl. Conf. on System Sciences, 2002.

[7] Kimball R, Reeves L, Ross M, Thornthwaite W (1998) The data warehouse lifecycle toolkit.Wiley, New York.

[8] Stephen R, Boxwala A, Gertman P. Feasibility of using a large Clinical Data Warehouse to automate the selection of diagnostic cohorts. AMIA Annu Symp Proc. 2003;:1019.

[9] Cao H, Markatou M, Melton GB, Chiang MF, Hripcsak G. Mining a clinical data warehouse to discover disease-finding associations using co-occurence statistics. AMIA Annu Symp Proc. 2005;:106-10.

[10] Wan TT. Healthcare informatics research: from data to evidence-based management. J Med Syst. 2006 Feb;30(1): 3-7.

[11] Silver M, Sakata T, Su HC, Herman C, Dolins SB, O'Shea MJ. Case study: how to apply data mining techniques in a healthcare data warehouse. J Healthc Inf Manag. 2001 Summer;15(2):155-64.

[12] Breant C, Thurler G, Borst F, Geissbuhler A. Design of a Multi Dimensional Database for the Archimed Data Warehouse.Stud Health Technol Inform. 2005;116:169-74.

[13] Fassa A-A, Urban P, Radovanovic D, Duvoisin N, Gaspoz JM, Stauffer J-C, Erne P; Trends in reperfusion therapy of ST segment elevation myocardial infarction in Switzerland: six year results from a nationwide registry. Heart 2005;91: 882-888

Address for correspondence

Claudine Bréant PhD,
Service of Medical Informatics,
University Hospitals of Geneva (HUG),
1211 Geneva, Switzerland.
Email: claudine.breant@sim.hcuge.ch

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Are Clinicians' Information Needs and Decision Support Affected by Different Models of Care? Experimental Study

Vitali Sintchenko[a,b], Tim Usherwood[b], Enrico Coiera[a]

[a]*Centre for Health Informatics, University of New South Wales,*
[b]*Western Clinical School, University of Sydney, Sydney, Australia*

Abstract

This study explores task- and healthcare model-specific differences in clinicians' information needs which can affect the uptake of decision support. Results of a web experiment involving 104 general practitioners are presented. Respondents indicated that guidelines were the most important source of information with almost equal weighting for acute, chronic and preventive care. A patient's quality of life was identified as the most important determinant of decision-making in all three models of care. Risk assessment tools and information about outcomes were more valuable (P<0.05) for chronic and preventive care than for acute cases. The participants accessed electronic risk assessment tools in 54%, 45% and 81% of acute, chronic and preventive care scenarios, respectively. Participants estimated that the electronic decision support would have a significantly higher impact in preventive care than in chronic or acute care settings (P=0.01). The differences in the information needs of clinicians related to different care models have to be considered in the design of clinical decision support systems. Systems that target preventive model decisions may have higher adoption and impact.

Keywords:

information needs; information systems; clinical decision-making

Introduction

Lack of knowledge about the information needs of clinicians has been identified as one of the major reasons for the slow uptake of decision support tools in clinical practice [1,2,3]. Evidence suggests that the type of decision task affects the effectiveness of electronic decision support systems (EDSS). The provision of computerised decision support improves the quality of decision-making and outcomes of patients with acute illnesses in a hospital setting. However, the impact of EDSS on care decisions for patients with chronic conditions or genetic risks is less certain [4,5]. Despite the increasing availability of genetic testing and the emergence of new programs for shared genetic risk assessment [6,7], knowledge of clinician's predictive testing and EDSS use remains limited [8]. A large variety of attitudes and beliefs influence clinical decision-making by clinicians [5]. Our previous research points to task- and healthcare model-specific differences in their information needs, which can affect the uptake of EDSS [9]. To test this hypothesis, three modules of decision support were developed and a web-based experiment was designed to aid clinical risk assessment for acute, chronic and preventive care models. The objectives of this experiment were to explore (a) task specific information needs related to acute, chronic and preventive care, and (b) the free-willed use of computerised decision support tools by clinicians performing the above tasks.

Methods

Study population and design

We surveyed a convenience sample of Australian general practitioners to examine the differences in their information needs and beliefs about the role of EDSS in acute, chronic and preventive care. Participants were recruited among Fellows and Trainees of the Royal Australian College of General Practitioners through the Quality Assurance & Continuing Professional Development Programme. Participation was voluntary and no monetary incentives were offered. Participants were randomised to the EDSS group and the unaided decision-making group and asked to review three clinical scenarios in a random order to avoid learning effect as well as to answer case-specific questions (Figure 1). The first group was provided with access to decision support modules at the time of answering the questions. The modules provided probabilistic patient-specific information about future clinical outcomes. Clinicians were not specifically informed that their use of EDSS modules would be logged. The second group only saw the decision support modules after answering case-specific questions at the end of the survey. The questions were designed using a 5-point Likert scale to identify potential factors affecting EDSS acceptance as well as specific usage patterns on the part of clinicians. At the end of the survey each participant was provided with feedback regarding (a) the distribution of answers from the group of participants, including their own answers, and (b) the impact of the use of decision support tools on this process.

The experiment was conducted between April and August 2006 as an online exercise with interactive decision sup-

port modules accessible from a personal computer over a secure Internet connection to the University of NSW server. This survey can be accessed from: http://129.94.108.23/DSSSurvey/index.jsp

Decision support modules

Three risk assessment aids were built using published scores based on demographic, clinical and behavioural information that is routinely collected when taking a history (Copas, 2002). Community-acquired pneumonia decision support provided risk assessment based on the Pneumonia Severity Index [10] and the genital herpes risk assessment tools was based on local Australian data. The Breast Cancer Risk Assessment module was based on probabilities obtained from two recently validated models: the BOADICEA model of genetic susceptibility to breast cancer from the University of Cambridge [11] and BRCARPO model of genetic testing of BRCA1/BRCA2 and prevalence of breast cancer [12] (Figures 2 and 3). Scenarios of community-acquired pneumonia, genital herpes and breast cancer were designed specifically to represent, respectively, acute, chronic and preventive models of care because of their relative differences in the acuity and complexity of clinical decision-making, as well as the utility of shared decision-making and decision outcomes.

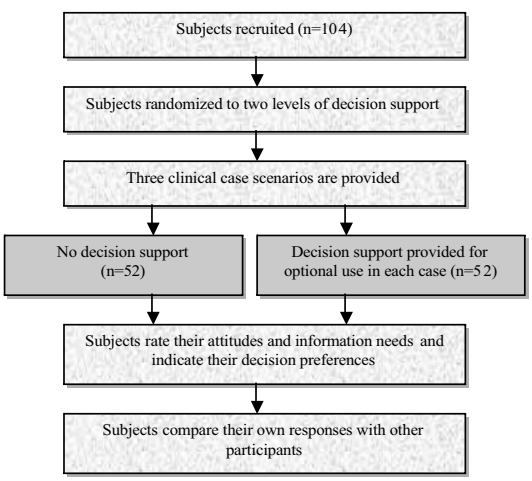


Figure 1 - An online experiment design

Outcome measures

Using a log-file, the participants' choice to use a decision support tool was monitored. The attitudes and self-reported information needs of subjects were recorded. The impact will be assessed using one-way analysis of covariance (ANOVA) and chi-square statistics. Statistical significance was set at $P<0.05$.

Results

Characteristics of participants

104 general practitioners completed the experiment. Twenty-five of respondents (25%) also treated residents of aged care facilities and 15 (14%) served as sessional hospital medical officers. All respondents were accredited practitioners representing all States and Territories of Australia. Half of the participants (n=53) indicated that they use a computer in for keeping medical records and electronic prescribing. General practitioners familiar with computers were equally distributed in the two study groups.

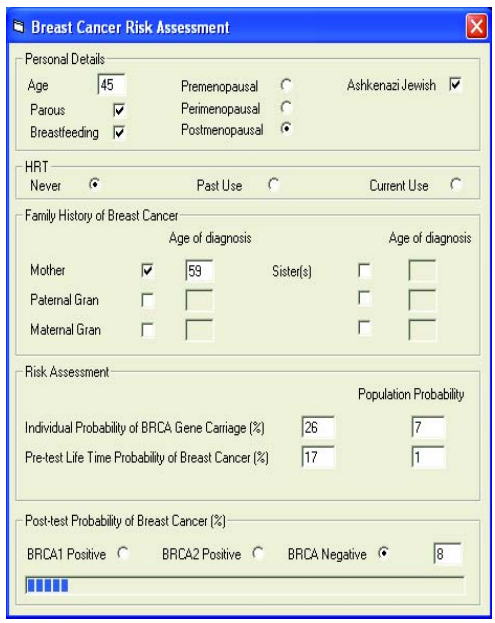


Figure 2 - Snapshot of the decision support module for the risk assessment of breast cancer

Information needs for acute, chronic and preventive care

Overall, participants were positive about the value of probabilistic EDSS in primary care. All clinicians rated clinical guidelines, condition-specific risk calculators and computerised decision support as important sources of evidence (Table 1). The potential impact on a patient's quality of life and his or her satisfaction were the two most important determinants of clinicians' decision-making in all three models of care. Respondents indicated that clinical guidelines were the most important source of information with almost equal weighting for acute, chronic and preventive care.

Condition-specific risk calculators and information about outcomes of care for previous patients with similar problems were more valuable ($P<0.05$) for chronic and preventive care situations than for acute cases (Table 1). When two arms of the study were compared, the participants in the decision support arm of the experiment accessed electronic risk assessment tools in 54%, 45% and 81% of acute, chronic and preventive care scenarios,

*Table 1 - Attributes affecting the quality of clinical decision-making for different models of care, mean scores**

Statement	Acute care	Chronic care	Preventive care	P**
The importance in decision-making				
Compliance with clinical guidelines	1.793	1.526	1.758	NS**
Cost to taxpayer	2.414	2.069	2.379	NS
Impact on quality of life	1.448	1.345	1.276	NS
Patient satisfaction	1.448	1.241	1.310	NS
Access to clinical guidelines	1.862	1.757	1.758	NS
Access to condition-specific risk calculators	2.345	2.034	1.862	0.04
Information about outcomes of care of previous patients with similar problems	2.276	1.793	1.861	0.05
Information essential for optimal decision-making				
Document patient history	1.621	1.413	1.552	NS
Clinical protocols	2.034	1.758	1.621	NS
Cost of care to a patient	2.793	1.897	2.931	0.001
Information most often lacking				
Document patient history	2.414	2.414	2.931	0.03
Clinical protocols	2.655	3.310	1.828	0.002
Cost of care to a patient	2.517	2.655	1.931	0.02

* Items rated on scale of 1 (Strongly agree) to 5 (Strongly disagree). Statistics were calculated by ANOVA.
** NS, not significant.

respectively. The majority of general practitioners who used decision support tools also felt that EDSS could have a high impact on the quality of decision-making in different models of care (Figure 3).

Information gaps in acute, chronic and preventive care

When specifically asked about information types essential for optimal decision-making, participants agreed that a documented patient history, relevant test results and clinical protocols were the most relevant pieces of information.

Respondents felt that the cost of care to a patient was also a contributor to chronic care decisions and was more important for chronic care than for other models P=0.001). When asked to identify the specific types of information most often lacking at the time of clinical decision-making, respondents indicated that a relevant patient history and test results were more often lacking in acute and chronic care. In contrast, clinical protocols (P=0.002) and cost information (P=0.02) were more often lacking in preventive care (Table 1).

Participants estimated that the electronic decision support would have a significantly higher impact in preventive care than in chronic or acute care settings (P=0.01) (Figure 4).

Discussion

This study showed that perceived information needs between acute, chronic and preventive care models in general practice were similar. The main differences in information needs in these settings were related either to the availability of information or to access to decision support.

Figure 3 - Snapshot of the decision support module for the risk assessment of community-acquired pneumonia

While the lack of documented patient history and test results has been judged as more important in acute and chronic care, finding appropriate clinical protocols and information about the cost of care was of higher priority in preventive care. Furthermore, medical professionals indicated that the successful information searching has a more significant impact on clinical decision-making in preventive care than in other setting. It could reflect differences in the goals of management or variations in risk preferences for different scenarios shown to influence clinical decisions [5]. Although physician uncertainty has been cited as a cause of practice pattern variations [5,9], differences in physicians' objectives concerning the goals of treatment and management in different models of care have received less attention [13]. However, these differences imply that clinicians may use different information seeking strategies for acute, chronic and preventive care tasks.

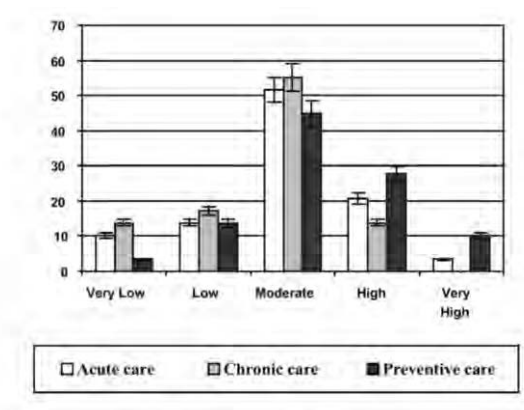


Figure 4 - Distribution of the perceived impact of information provided by decision support tools on the quality of clinical decision making in different models of care. Estimates: very low=0-20%; low=21%-40%; moderate=41%-60%; high=61%-80%; very high=81%-100%.

Many decision support systems are based on electronic guidelines or protocols. However, this experiment gathered further evidence to support the suggestion that the impact of computerised decision support exceeds that of guidelines [9]. Our results also suggest that the uptake of decision support may be affected by differences in acute, chronic and preventive care tasks. Our participants more frequently used a decision support tool for the breast cancer risk assessment than for acute or chronic care conditions. Our findings about the use of condition-specific risk calculators using pathology testing data are of particular significance as diagnostic tests account for about 25% of ambulatory health care costs, with 80% of all health expenditures directed by physicians [14].

Previously undocumented facts have emerged from our experiment: a higher utility and uptake of decision support for preventive decision tasks than for acute care decisions; a gap in decision support for preventive medicine; significant differences in the perceived impact of information on the clinical decision-making process for acute, chronic and preventive models of care.

These conclusions should be interpreted in light of the limitations in the study design. First, the experiment relied on self-reported behaviour without verification that clinicians actually practice in the manner described. Our study was experimental and did not take into account operational issues concerning the application of EDSS in a clinical setting. It is also possible that clinicians in the intervention arm used EDSS relatively often because they felt they were participating in a new study. Second, we used a convenience sample of primary care practitioners. Participation was voluntary and clearly the study may be biased towards those who felt more comfortable with electronic decision support. However, clinical vignettes as a method for measuring the competence of physicians and the quality of their actual practice have been validated [15] and interactivity in the web-based questionnaires increased compliance [16,17]. Lastly, we surveyed general practitioners. Specialist clinicians are likely to differ from primary care providers in decision-making styles and information needs [18]. However, this has the advantage of representing the point of view of professionals who see the whole spectrum of problems at the front end of health care delivery and do not necessarily represent "early adopters" of new concepts.

In conclusion, the differences in the information needs of clinicians related to different care models have to be considered in the design and implementation of clinical decision support systems. Systems that target preventive model decisions may have higher adoption and impact. These findings discover the relative value of different types of information needed for the optimisation of clinical decision-making in primary care and identify key strate-

gies for the design and implementation of successful EDSS for genomic medicine.

Acknowledgments

VS was supported by the Australian National Institute of Clinical Studies. Authors are indebted to Ken Nguyen for his help in the design of the web questionnaire and also thank the Quality Assurance & Continuing Professional Development Programme of the Royal Australian College of General Practitioners and the GPs who participated in the survey.

References

[1] Sim I, Gorman P, Greenes RA, Haynes RB et al. Clinical decision support systems for the practice of evidence-based medicine. J Am Med Inform Assoc 2001;8:527-34.

[2] Pluye P, Grad RM, Dunikowski LG, Stephenson R. Impact of clinical information-retrieval technology on physicians: A literature review of quantitative, qualitative and mixed methods studies. Intern J Med Inform 2005;74(9):745-68.

[3] Dreiseitl S, Binder M. Do physicians value decision support? A look at the effect of decision support systems on physician opinion. Artif Intell Med 2005;33:25-30.

[4] Bates D, Kuperman G, Wang S, Ganghi T, Kittler A, Volk L, et al. Ten commandments for effective clinical decision support: making the practice of evidence-based medicine a reality. J Am Med Inform Assoc 2003;10:523-30.

[5] D'Alessandro DM, Kreiter CD, Peterson MW. An evaluation of information-seeking behaviours of general practitioners. Pediatr 2004;113:64-9.

[6] Ravdin PM, Siminoff LA, Davis GJ, Mercer MB, et al. Computer program to assist in making decisions about adjuvant therapy for women with early breast cancer. J Clin Oncol 2001;19:980-91.

[7] Baum M, Ravdin PM. Decision-making in early breast cancer: guidelines and decision tools. Eur J Cancer 2002;38:745-9.

[8] Knottnerus JA. Community genetics and community medicine. Fam Pract 2003;20:601-6.

[9] Sintchenko V, Coiera E. Decision complexity affects the extent and type of decision support use. In: Bates D (Ed). Proc Annual AMIA Symp, Washington DC, 11-15 Nov 2006; pp.724-8.

[10] Dean NC, Suchyta MR, Bateman KA, Aronsky D, Hadlock CJ. Implementation of admission decision support for community-acquired pneumonia: a pilot study. Chest 2000;117:1368-77.

[11] Antoniou AC, Pharoah PPD, Smith P, Easton DF. The BOADICEA model of genetic susceptibility to breast and ovarian cancer. Br J Cancer 2004;91:1580-90.

[12] Berry DA, Iversen ES, Gudbjartsson DF, Hiller EH, et al. BRCARPO validation, sensitivity of genetic testing of BRCA1/BRCA2, and prevalence of other breast cancer susceptibility genes. J Clin Oncol 2002;20:2701-12.

[13] Flach SD, Longenecker JG, Tape TG, Bryan TJ, Parenti C, Wigton RS. The relationship between treatment objectives and practice patterns in the management of urinary tract infections. Med Dec Making 2003;23;131-9.

[14] Birbeck GL, Gifford DR, Song J, Belin TR, Mittman BS, Vickrey BG. Do malpractice concerns, payment mechanisms, and attitudes influence test-ordering decisions? Neurol 2004;62;119-21.

[15] Peabody JW, Luck J, Glassman P, Dresselhaus TR, Lee M. Comparison of vignettes, standardized patients and chart abstractions. J Am Med Assoc 2000;283:1715-22.

[16] Emery J, Walton R, Murphy M, Austoker J, et al. Computer support for interpreting family histories of breast and ovarian cancer in primary care: comparative study with simulated cases. Brit Med J 2000;321:28-32.]

[17] Balter KA, Balter O, Fondell E, Lagerros YT. Web-based and mail questionnaiers: A comparison of response rates and compliance. Epidemiol 2005;16:577-9.

[18] Green MJ, Peterson SK, Baker MW, Harper GR, et al. Effect of a computer-based decision aid on knowledge, perceptions, and intentions about genetic testing for breast cancer suscpetibility. A randomised controlled trial. J Am Med Aassoc 2004:292:442-52.

Address for correspondence

Dr Vitali Sintchenko,
Centre for Health Informatics,
University of NSW
UNSW Sydney 2052 Australia
v.sintchenko@unsw.edu.au

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Modeling and Acquisition of Drug-Drug Interaction Knowledge

Frédéric Mille[a,b], Patrice Degoulet[a], Marie-Christine Jaulent[a]

[a] INSERM,UMR°S 872, Eq 20, Les Cordeliers, Paris, F-75006 France ; Université Pierre et Marie Curie-Paris6, UMR°S 872, Paris, F-75006 France ; Université René Descartes, UMR°S 872, Paris, F-75006 France;
[b] AP – HP, Hôpital Robert Debré, Pharmacie, Paris, 75019 France, France

Abstract

Objectives*: The effectiveness of computerized clinical decision support systems (CDSS) depends on the quality of the knowledge they refer to. In this article, we are interested in the acquisition, modeling and representation of the knowledge embedded in the "national reference framework of drug-drug interaction" published by the French Health Products Safety Agency.*

Methods*: A model of drug-drug interactions has been designed using bottom-up and top-down approaches. This model is the basis for the design of an XML format to represent and extract information on drug interactions from the reference framework.*

Results*: A specific tool has been developed to extract the information from a corpus of 1053 drug monographs using a methodology similar to the one used by the GEM-Cutter tool to extract information from clinical guidelines. Strategies to integrate the XML files produced into CDDSSs are discussed.*

Discussion-conclusion: *Modeling and acquisition of drug-drug interaction knowledge from a corpus of drug monographs is a potential approach to foster the development of CDSS and improve their specificity.*

Keywords:

drug interactions, knowledge modeling, XML

Introduction

Technology–based interventions have been recommended for reducing the likelihood of medication errors. Computerized physician order entry (CPOE) has been cited as one of the most effective ways to avoid medication errors[1-3] caused by misinterpretation of handwritten orders, incorrect doses, wrong dose forms and inappropriate administration times. The next step is to integrate effective Clinical Decision Support Systems (CDSS) into CPOE systems.

CDSS are information systems designed to improve clinical decision making, for example detecting drug-drug interactions (DDIs) during the order entry process. Indeed, DDI detection systems were among the first CDSS developed. When a potential problem is identified, these systems can provide real-time alerts, allowing the clinician to make appropriate changes before the prescription is finalized.

The effectiveness of CDSS depends on the quality of the knowledge base which they use. Obviously, for detection of DDIs, any CDSS requires knowledge of these DDIs. The main issue addressed here is the acquisition, modeling and representation of the knowledge in this field. In France, the main source of knowledge for DDIs is the "national reference framework of the drug-drug interaction" published by French Health Products Safety Agency "Agence Française de Sécurité Sanitaire des Produits de Santé" (AFSSAPS) [4]. DDIs are described in free text in these monographs and are not directly accessible by a computerized system.

A simplest way to represent knowledge in monographs would be to build manually two-column tables describing the interactions between pairs of drugs. This approach is difficult to maintain as knowledge progresses. The most advanced approach would be automatic free text understanding by Natural Language Processing (NLP) techniques. These techniques could automatically identify concepts in DDI monographs and semantic relations between these concepts. However, the NLP approach is complex and beyond the scope of this work. Between these two extremes, an affordable and plausible approach is to propose tools to structure the knowledge contained in the monographs so as to make this knowledge available.

The aim of the study is to acquire DDI knowledge in a structured knowledge base that could be used by a computerized system. The paper presents a method and a tool for building a knowledge resource usable in the context of prescription systems.

Our work is based on the hypothesis that DDI knowledge formulation has similarities with clinical guidelines knowledge formulation. Indeed, DDI monographs contain recommendations and risk factors that can be considered in the same way as guidelines. Clinical guidelines have become an important medium for the standardization and dissemination of medical knowledge. The "document-centric approach" has been introduced to facilitate the use of guideline knowledge. In the document-centric approach, the original text of the guidelines is systematically marked-up with respect to the model and kept in the form of a structured document.

Selected for best paper award.

The best known instance of "document-centric approach" is the Guideline Elements Model (GEM) which is an XML (Extensible Markup Language) framework based on a hierarchy of concepts describing guideline contents [5].

The encoding of a clinical guidelines using the GEM framework consists of structuring the textual document using the set of XML markups provided by the framework.

The encoding of a monograph consists of structuring the textual document using the set of XML markups. This can be a complex process, as it requires in-depth analysis of the text content. Similarly, for clinical guidelines, substantial variation is observed in the encoding by different users [6]. This would suggest, as concerns our study, that the complexity of manual analysis can affect the structure of the encoded monographs.

To tackle these problems and support the process of manual marking up of monographs, we developed a specific interface, inspired by the GEM-Cutter application [5].

We propose to use the same methodology to structure the information within monographs.

The first step was to identify the relevant information in the monographs so as to build an XML schema related to DDI information. This schema is implemented in a specific environment to acquire descriptions of DDIs directly from text. The second part of our work was to use and evaluate this tool to develop a knowledge base concerning DDI information from 1000 monographs. The practical result is the knowledge base itself. We present the advantages and limitations of our approach and consider the exploitation of the resulting knowledge base.

Materials and methods

Methods

We used a two-step approach to identify the information elements contained in the "national reference framework of the drug-drug interaction" published by AFSSAPS. First, we identified the main concepts of DDI. Second, we used natural language processing tools to analyze the content of DDI texts.

Then we built a XML schema of DDI information, which was evaluated for its ability to represent the initial text information and maintain its meaning.

Finally, we built specific software to write XML files.

Top-down approach knowledge identification

The identification of large classes relied on the manual study of structure in the monographs. It was done based on the reference frame of AFSSAPS.

Bottom-up approach knowledge identification

We applied a data-driven knowledge approach to identify the concepts of the domain present in the monographs. We used a syntax analyzer module based on the hypothesis of similar dependencies between terms which have similar meanings. Then, a pharmacist reviewed the different terms on the basis of shared syntactical contexts and exploited all the network data to cluster and organize the terms.

The lexicon built with this software made possible a second manual step of semantic analysis registering the principal semantic structures contained in the DDI texts, and their frequencies.

Establishing a XML schema

The creation of an XML language to describe monographs requires conceptual modeling of DDIs. We built an XML schema of DDI knowledge from existing knowledge and terminological analysis. This hierarchy of concepts constitutes the skeleton of the XML schema.

Development of an XML Editor

We built X-DIE (XML Drug Interaction Editor), an XML editor with added functionalities to facilitate the editing of XML. X-DIE is a graphic editor which hides the code in the background and presents the content to the user in a more readable format, similar to the version which must be ultimately posted. The interface is based on a multiple windowing system that displays the original monograph, a XML browser and the corresponding XML-encoded text.

Material

The French drug database Theriaque® developed by the Centre National Hospitalier d'Information sur le Medicament (CNHIM) is responsible for the dissemination of independent information about all drugs available in France [7]. The database contains complete information about the drugs: the pharmaco-therapeutic group, the active component, the excipient, the commercial presentation, indications, contra-indications and AFSSAPS reference concerning DDIs. The set of all 1053 monographs of DDIs available in the Theriaque® database constitutes the initial corpus.

The syntax analyser module is SYNTEX, a natural language processing tool that is widely used to build ontologies from texts. SYNTEX performs a syntactic analysis of the sentences of the corpus, and yields a network of the dependence of words and phrases [8, 9].

The PROTEGE editor was used to build and browse the hierarchy of concepts.

The XML schema was built using an open-source XML file editor, JAXE.

The interface for the marking up of monographs was developed in JAVA with the API DOM (Application Programming interface Document Object Model) dedicated to the production and management of XML files.

Results

Knowledge acquisition results

Top-down approach

In the frame of reference of AFSSAPS, a DDI is defined by a pair of protagonists "A + B" who can be an active substance, indicated by their international non-proprietary names, or a therapeutic class, itself being the subject of

interactions "of class". The wording of a DDI is structured in 4 paragraphs: ❶ nature of the risk, ❷ brief mechanism of action (if known), ❸ action to be taken and ❹ degree of constraint.

For the last type of paragraphs four degrees of constraint are possible: ❶ Contra-indication: This is an absolute character and should not be transgressed, ❷ Not advised: The association should generally be avoided, except if rigorous examination of the risk/ benefit ratio suggests otherwise, and imposes close monitoring of the patient, ❸ Cautioned: The association is possible if, in particular at the beginning of treatment, simple recommendations are respected making it possible to avoid undesirable DDI (adaptation of doses, reinforcement of monitoring, be it clinical, biological, ECG, or other), ❹ Take into account: The risk of DDI exists, generally in the form of accumulation of adverse effects; no practical recommendation can be proposed. It is left to the physician to evaluate the appropriateness of association.

In case of a DDI being classified as a contra-indication or not advised, the action to be taken is generally summarized only as a constraint. Cautioned DDI are often associated with recommendations that are simple to implement to avoid the interaction (adaptation of doses, biological controls, etc.). The classification "take into account" is not associated with any practical recommendation because it announces the addition of adverse effects that can only be avoided by using other therapeutic strategies.

This first stage of the analysis generated various categories of knowledge found in the DDI monographs. These categories are considered to be the main concepts to be identified in DDI texts and are used to group terms in the following analysis.

Several classes are distinguished but there is no organization of knowledge in each class.

The level of organization of the knowledge was insufficient for application in CDSS.

Bottom-up approach

The initial corpus contained about 87,707 words. Once processed with SYNTEX, the corpus gave 9393 candidate terms appearing at least twice. The expert selected 2657 candidate terms : ❶ 1349 noun phrases out of 3150, ❷ 1071 nouns out of 3775, 3) ❸ and 237 adjectives out of 336.

These terms have been classified into 5 categories: ❶ pharmacology and pharmacokinetics (931 terms), ❷ galenical, active principle, dose and regimen (936 terms, this category includes the various characteristics of the associations causing DDIs), ❸ physiopathology (282 terms, this group includes all diseases or symptoms secondary to DDIs), ❹ physiology (188 terms), ❺ others (320 terms).

These selected candidate terms were normalized within each category and this gave a final total of 888 concepts.

The combination of bottom-up and top-down approaches enabled us to build a DDI model with 6 main classes (figure 1): risk association, consequence, mechanism, limitation, precaution for use, risk factors.

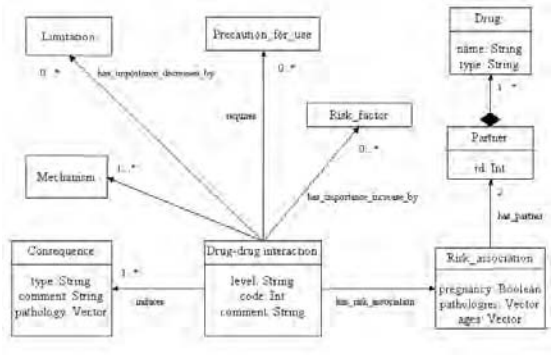


Figure 1 - DDI model

The XML representation schema

The JAXE software allows an XML schema to be built, compatible with the W3C norm.

XML files are constructed as a hierarchy of 72 discrete tags with 6 major branches intended to capture the information in the 6 main sections of a monograph. The 6 branches are:

- *Risk association*: the sub-tree describes the drug association responsible for the DDI, and the clinical situations (pathologies, ages) in which the DDI occurs.
- *Mechanism*: the sub-tree describes the DDI mechanisms (pharmacology, pharmacokinetics, etc.).
- *Risk factors*: the sub-tree describes any physiology (age, pregnancy) and/or diseases that may increase the risk of adverse effects of a given DDI.
- *Consequences*: the sub-tree describes the diseases and symptoms resulting from the DDI.
- *Precaution for use*: the sub-tree describes the precautions to take to reduce the effects of the DDI (substitution, alternative drugs) and/or to reduce its consequences (clinical or biological exams).
- *Limitations*: the sub-tree describes any disease or condition where the benefit for the patient outweighs the risk of an adverse event due to the DDI.

Other types of elements are represented in the XML schema. They include: drug, pathology, biological examinations.

XML elements can have attributes in the start tag. Attributes provide additional information about elements. The attributes of our elements are found in published classifications particularly ATC (Anatomical Therapeutic Chemical), ICD10 (International Classification of Disease – tenth edition) and MedDRA (Medical Dictionary for Regulatory Activities). With these attributes, DDI detection system (DIS) will be able to perform various types of inference between DDI information and the clinical context of the patient. In addition, the XML schema includes

elements to allow quantitative information (for example duration, dose, etc.) to be coded.

The X-DIE environment for the encoding of monographs

The X-DIE environment has been developed to encode monographs. The main interface is a three-windowing system (figure 2), inspired by the GEM-Cutter interface: ❶ window A displays the DDI monograph in full text, ❷ window B displays the hierarchy of the elements for the XML file under construction and ❸ window C displays the information to provide for each element of the file.

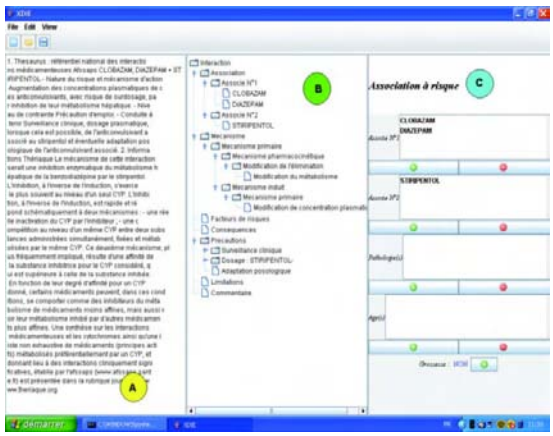


Figure 2 - X-DIE

Our environment allows the user to look for drugs (active principles, therapeutic class and chemical class) for which there is a link, in the data base Theriaque®, with the displayed monograph. It also allows the user to search classifications including ICD10 and MedDra for pathologies similar to those cited in the visualized monograph. Other information is otained by the user.

Knowledge base

One thousand and six of the total of 1053 monographs present in the Theriaque® database could be successfully entered: the coverage was thus 96%. The monographs which were not successfully encoded had at least one of the following two characteristics: ❶ The monograph describes an DDI involving a drug which is no longer marketed; in these cases we chose not to enter the corresponding XML file. ❷ The monograph describes a DDI whose mechanism is a modification of plasma protein binding. This mechanism was omitted by the expert.

Discussion and conclusion

We developed a DDI model capable of representing full DDI information and we developed the X-DIE environment to capture this structured data from full text monographs. We used both general knowledge about DDI contained in the "national reference framework of the drug-drug interaction" published by AFSSAPS, and knowledge extraction with NLP techniques from the same database.

General knowledge helped us to define the generic structure of the DDI model. This top-down approach, although performed manually, rapidly gave an organization of the main concepts, because the domain is delimited. To refine this generic structure, we extracted the knowledge present in the "national reference framework of the drug-drug interaction". For this bottom-up approach, we combined natural language processing results using manual semantic analysis of the relevant DDI candidate terms specifying generic concepts identified by the top-down analysis. The bottom-up and top-down approaches are standard methods in knowledge engineering.

Their implementation in this way ensures that the final structure covers a significant proportion of the information contained in the DDI monographs. However, the percentage of coverage of all the monographs has not yet been established. Moreover, the use of monographs from only one knowledge source potentially imposes limitations: the interoperability of the XML files is reduced in theory because the source for the knowledge and the data to encode is the same. An improvement would be to design the model from different sources of DDI data.

It became clear, during the encoding process, that the monographs were poor in recommendations. A few XML markups were rarely used, for example "clinical examination". We also noted that some recommendations are vague; biological examinations are cited under the generic term "biological monitoring" and clinic examinations are quoted under the generic term "clinical monitoring". This was a surprising finding because monographs may contain: ❶ monitoring options (blood tests, clinical and extra clinical examinations), ❷ guidelines to manage DDIs (modification of the times or sequence of administration, routes of administration, compensation by administration of a third compound or substitution of depleted endogenous substances, adjusting doses, discontinuation as symptoms appear, and suggestions of non interacting alternatives), ❸ factors increasing the risks of DDI (age, sex, genetics, predisposing diseases, and use of alcohol and tobacco).

This work clearly illustrated the paucity of the content of some of the monographs as concerns recommendations. Thus, our model could be used as a tool to enrich and standardize the DDI monographs.

X-DIE tool facilitates the encoding of monographs. However, the data capturing is still largely manual. The encoding process could be considerably simplified if X–DIE could incorporate automatic text processing functions, such as the identification of linguistic markers. Previous work by our group on guidelines demonstrates the value of the identification of linguistic markers [10]. Such extension will exploit the already built ontology to highlight he pertinent information in the text.

Some studies have indicated that the weak specificity and irrelevance of alerts in CDSS lead to loss of confidence in

these systems, a phenomenon known as the "Cry wolf syndrome" [11, 12]. Perceived poor specificity of drug alerts may be a major obstacle to efficient utilization of information and may prevent such alerts contributing fully to improving safety. If there is a false alert, these systems substantially increase the time required to carry out a task, due to the disruption in workflow. High signal-to-noise ratios may also produce alert fatigue and result in physicians skipping past alerts without considering or even reading them. One major limitation of existing DDI detection systems is associated with the use of static knowledge, most often embedded in fixed mapping tables: they proceed by comparing between drugs according to a hard-coded interaction mapping table, without any consideration of the clinical context of the patient or of all the information contained in monographs. These systems are often highly inclusive, placing more emphasis on breadth of coverage than on clinical relevancy or severity of adverse events. In published studies, practitioners explained that CDSS should integrate context relevance information, guidelines or evidence-base medicine. In fact, the majority of DDI can be compensated by dose adjustment or prevented by a well-considered sequence of administration and represents therefore a manageable risk [13]. Physicians wish informative support on DDI, concerning management. A distinction between clinically relevant and negligible DDI is essential [14].

For all these reasons we suggest that the reference framework should be integrated in the form of XML files in CDSS. The current XML-Schema has to be augmented in order to include all information for each DDI in the reference framework: some instructions for laboratory testing to monitor side-effects, replacement of ordered drug with another drug, change in dose, or additional drug. It also contains clinical situations where the use of an association is acceptable in spite of the potential consequences of the DDI; in these clinical situations, it is not necessary to announce the DDI or post an alert. Indeed, in these situations, the clinician can deviate from recommendations for good clinical reasons and the benefits of the drug association outweigh the disadvantages of the potential DDI.

To construct a more accurate and evidence-based CDSS for DDI detection, knowledge from DDI databases as provided by AFSSAPS must be arranged and integrated, but it is also essential to construct algorithms that can exploit this knowledge base appropriately. The next steps include the development of algorithms able to use XML files. Also, the impact of this combination (algorithms and knowledge base) on CDSS specificity will have to be evaluated.

References

[1] Oren E, Shaffer E, and Guglielmo B. Impact of emerging technologies on medication errors and adverse drug events. *American Journal of Hospital Pharmacy*. 2003; 60: 1447-1458.

[2] Bates D, Leape L, Cullen D, et al. Effect of computerized physician order entry and a team intervention on prevention of serious medication errors. *JAMA*. 1998; 280(15): 1311-1316.

[3] Teich J, Merchia P, Schmiz J, Kuperman G, Spurr C, and Bates D. Effects of computerized physician order entry on prescribing practices. *Arch intern med*. 2000; 160: 2741-2747.

[4] Thésaurus des interactions médicamenteuses. 2006, Agence française de sécurité sanitaire des produits de santé: Paris. p.165.

[5] Shiffman R, Karras B, Agrawal A, and Chen R. GEM : A proposal for a more comprehensive guideline document model using XML. *J Am Med Inform Assoc*. 2000; 7(5): 488-498.

[6] Karras B T, Nath S D, and Shiffman R N. A preliminary evaluation of guideline content mark-up using GEM--an XML guideline elements model. *Proc AMIA Symp. 2000*: 413-7.

[7] Amiel M, and Husson M. The THERIAQUE database and information on side effects of drugs. *Therapie.* 1994; 49(5): 455-458.

[8] Fabre C, and Bourigault D. Linguistic clues for corpus-based acquisition of lexical dependencies. *in Corpus Linguistics 2001 Conference. 2001*: UCREL Technical Papers.

[9] Charlet J, Bachimont B, and Jaulent M C. Building medical ontologies by terminology extraction from texts: An experiment for the intensive care units. *Comput Biol Med*. 2005; 36(7-8): 857-70.

[10] Georg G, Colombet I, and Jaulent M C. Structuring Clinical Guidelines through the Recognition of Deontic Operators. *Stud Health Technol Inform*. 2005; 116: 151-6.

[11] Glassman P, Simon B, Belperio P, and Lanto A. Improving recognition of drug interactions : benefits and barriers to using automated alerts. *Medical care*. 2002; 40(12): 1161-1171.

[12] Magnus D, Rodgers S, and Avery A. GPs' views on computerized drug interaction alerts : questionnaire survey. *Journal of clinical pharmacy and therapeutics*. 2002; 27(2): 377-382.

[13] Bergk V, Gasse C, Rothenbacher D, Loew M, Brenner H, and Haefeli W. Drug interactions in primary care: impact of a new algorithm on risk determination. *Clin Pharmacol Ther.* 2004; 76: 85-96.

[14] Bergk V, Gasse C, Schnell R, and Haefeli W. Requirements for a successful implementation of drug interaction information systems in general practice: results of a questionnaire survey in Germany. *Eur J Clin Pharmacol*. 2004; 60(8): 595-602.

Address for correspondence

Dr. Frédéric Mille. UMR_S 872, 15 rue de l'école de médecine, 75006 Paris, France. frederic.mille@rdb.aphp.fr

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

A Systems Development Life Cycle Approach to Patient Journey Modeling Projects

Joanne M. Curry[a], Carolyn McGregor[a], Sally Tracy[b],

[a] *Health Informatics Research, University of Western Sydney, Australia*
[b] *AIHW National Perinatal Statistics Unit, University of New South Wales, Sydney, Australia*

Abstract

Patient Journey Modeling, a relatively recent innovation in healthcare quality improvement, models the patient's movement through a Health Care Organisation (HCO) by viewing it from a patient centric perspective. A Systems Development Life Cycle (SDLC) provides a standard project management framework that can improve the quality of information systems. The concept of following a consistent project management framework to boost quality outcomes can be applied equally to healthcare improvement. This paper describes a SDLC designed specifically for the health care domain and in particular patient journey modeling projects. It goes on to suggest that such a framework can be used to compliment the dominant healthcare improvement method, the Model for Improvement. The key contribution of this paper is the introduction of a project management framework in the form of an SDLC that can be used by non-professional computer developers (ie: health care staff), to improve the consistency and quality of outcomes for patient journey redesign projects. Experiences of applying the SDLC in a midwife-led primary-care maternity services environment are discussed. The project team found the steps logical and easy to follow and produced demonstrable improvement results along with ongoing goal-focused action plans.

Keywords:

medical informatics, SDLC, health care Improvement, patient journey modeling

Introduction

Systems development life cycles (SDLC) were developed to provide a formal structure for the development of quality information systems. An overarching concept of a SDLC is the inclusion of a project management framework for planning, managing and controlling the people, development process and problem solution from the projects inception to the delivery of the required system [1, 2]. This theoretical construct can be applied in a similar manner in a variety of domains including health care improvement.

For a development project to be successful, the people involved in the project must have a detailed plan to follow. Attainment of the required goals depends heavily on having a plan that includes an organized, methodical sequence of tasks and activities that culminate in the delivery of a system that meets the clients' needs for reliability and efficiency. This is a specific goal of a SDLC.

Such concepts can also be aligned to the goals of the dominant health care improvement method, the Model for Improvement (MFI) [3] and it is suggested that by integrating the two approaches, patient journey modeling projects can be conducted in a more consistent manner, delivering higher quality process improvements. In this paper we propose a SDLC approach to Patient Journey Modeling projects that compliments the Model for Improvement via the introduction of a project management framework. Key benefits of the proposed SDLC approach are the provision of a planning, monitoring and control structure that can be used by both IT and non-IT staff with little or no previous process improvement experience to improve the consistency and quality of outcomes for patient journey redesign initiatives.

The paper begins with a background on patient journey modeling and systems development life cycles. Methods and research motivations are then presented, followed by a discussion on the proposed SDLC for patient journey modeling. The paper finishes with highlights of the SDLCs application and a discussion on how the SDLC integrates with the Model for Improvement.

Background

Patient journey modeling

Patient journey modeling is a patient-centric activity that details a patient's progress through a healthcare system for a given service [4, 5]. The goal of Patient Journey Modeling (PJM) is to improve health care quality by reducing variability in the care process. Specifically this includes evidence-based best practice, collecting required information only once, reducing the number of times a patient is moved, eliminating excessive activities, reducing duplicate communications and providing clear and concise information to the patient.

Several terms are used in the literature to refer to the concept of patient journey modeling (PJM) including: clinical pathways, patient flow redesign, clinical practice improvement and redesigning healthcare [6-8].

The most prominent method being used to reengineer health care processes, the 'Model for Improvement' (MFI), provides a framework for developing, testing and implementing changes that lead to improvement. As the method of choice for the Institute of Healthcare Improvement, it has been used extensively in the US [3] and by the NHS in England [4, 5].

Systems development life cycle

A Systems Development Life Cycle (SDLC) is a project management framework that organizes activities into phases [1]. If problem-solving activities are to be productive, the work conducted must be structured and goal-oriented. Computing professionals achieve these results by organizing the work into projects. A *project* is a planned activity (or set of activities) that has a definite beginning and end and that produces a desired result. *Project Management* deals with the planning, monitoring and control of all aspects of a project including the people involved, the problem solution and development process itself.

An SDLC was first introduced to the computing field in the 1960's with the goal of providing guidelines to improve the quality of computer developments [1, 2].

Many systems being developed today follow a development path consisting of 3 core elements: *Analysis, Design* and *Implementation*. *Analysis* activities provide an understanding of the business information system requirements. *Design* activities define the technical architecture and structure of the new systems to satisfy the business requirements. *Implementation* is the actual construction, testing and installation of a functioning information system.

These 3 phases address the core activities required to develop an information system but two additional phases are also required when developing quality systems. A *Project Planning/Initiation* phase involves those activities required to initiate, plan and obtain approval for the project. Once the new system is completed and installed, the development team must perform activities to determine if the project satisfied the original business needs or whether the system needs amendment or enhancement. This is known as the *Post Implementation Review* or *Evaluation* phase.

Current issues

Present patient journey modeling approaches lack any type of project management structure for planning the improvement project or monitoring and controlling its progress. The MFI also lies at a level of abstraction above step-by-step procedures and does not adequately address the integration of technology to assist the change process. The introduction of a SDLC to patient journey modeling projects provides a mechanism for overcoming these issues and provides staff, both IT and non-IT, with a logical guide to achieving improvement results.

Methods and research motivation

The research described uses a constructive research process [9] enriched by aspects of a participatory action research environment [10]. This has involved patient journey modeling sessions with management and staff at Ryde Hospital's Maternity department. Ryde provides midwife-led primary care maternity services for identified low-risk women through the public healthcare system [11, 12].

The work described in this paper originated as part of a Quality Review conducted at Ryde Hospital in 2006. Preliminary analysis of the areas under review indicated that although patient satisfaction was consistently high there were some significant patient assessment, information duplication and system administration issues. An SDLC was designed specifically for the PJM project. This included step-by-step activities and expected deliverables.

A Systems Development Life Cycle for patient journey modeling projects

The proposed Systems Development Life Cycle for patient journey modeling as shown in ***figure 1***, follows the standard SDLC format.

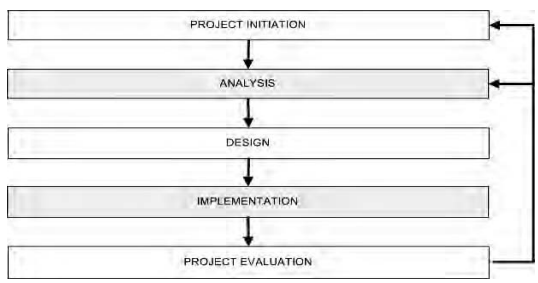


Figure 1 - A systems development life cycle for patient journey modeling

There are 5 phases, namely: Project Initiation, Analysis, Design, Implementation and Project Evaluation. Outputs from the Project Evaluation phase are fed back into the Life Cycle and either trigger new process improvement projects or lead to further enhancements of the resulting patient journey. Figures 2-6 and their description, outline each of the phases in more detail.

Typically senior management will assign a process improvement team leader. In SDLC terms this person is known as the project manager. This role is responsible for reporting progress, leading the team and ensuring that goals and deadlines are met.

Project initiation phase

The primary purpose of the *Project Initiation* phase is to set the scope of the patient journey modeling project and to inform those who will be affected by the results what will occur during the project. ***Figure 2*** shows the 6 major activities that make up this phase with all outputs being stored in a Project Repository.

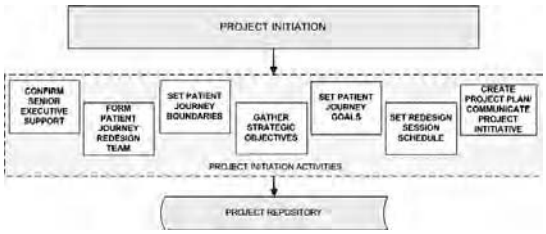


Figure 2 - Project initiation phase activities

The primary contributor to process improvement success is active, strong and visible executive sponsorship throughout the project [13, 14], thus the first activity must be to *secure a sponsor at senior executive level*. This will involve identifying a high-level project 'champion' who will 'talk-up' the project, is authorized to make resource decisions and can report project results to the executive team. Failure to secure a senior executive sponsor may see the project without required resources (both staff and physical) or lead to the project being cancelled if difficulties arise or budgets are tightened.

Once an executive sponsor is confirmed, the project manager sets about *staffing the project team*. In healthcare the inclusion of a representative from all areas affected by the results must be attempted. This includes clinicians, management, IT and administrative staff, and in the case of PJM projects, patient representatives as well. This helps to create a sense of ownership of resulting changes and promotes a culture of on-going process improvement. Ideally staff will be allocated to the project full-time but typically resources can only be released on a part-time basis.

Following formation of the project team the *scope of the patient journey modeling project* is discussed and agreed. The scope sets out what areas are to be included in the project analysis and what areas will be explicitly excluded. This information forms the basis for the first formal documentation created by the project team. It should be recorded in a retrievable medium and stored in the *project repository.* A *project repository* is a central storage area for all project information. Ideally this should be in electronic form as this allows future project teams to easily review past projects, identify successful attributes and reuse them where possible.

Following setting of the project boundaries, strategic objectives relating to the patient journey are gathered. *Strategic Objectives* are set by the Executive and are high level goals and measurements that drive the organisation's overall direction. These may not be clearly defined in some organizations but must be clarified before the project can continue. Specific *patient journey redesign goals* are then defined based on the strategic objectives. Each goal should address a particular area of the problem domain and have expected measurements assigned. These measurements relate to the redesigned patient journey and will determine the degree of improvement attained, post implementation.

Once the team knows what it is trying to improve and how changes will be measured, a *schedule for the running of the redesign sessions* can be created. This aligns with the final activity in the Project Initiation phase, creating the Project Plan. The *Project Plan* lists all of the activities and tasks that will be carried out during the project, estimates their duration and assigns them to a project team member. This plan then guides the progress of the project determining what activities will be conducted, when they must be completed by and who is responsible for their completion. The project manager is responsible for monitoring and controlling the plan and identifying problems. Key information regarding the project initiation phase and the future plan should then be communicated to all areas. This can be done via internal news distributions or as part of in-service sessions. Areas that will be directly affected by resulting changes must be continually kept 'in-the-loop' so that they develop a sense of ownership and there are no surprises during implementation. All documentation created during the phase is added to the project repository for use in the next phase, *Analysis*.

Analysis phase

The *Analysis phase* focus' on creating a graphical representation of the current situation and analyzing how this journey can be improved from the patient's perspective (***Figure 3***).

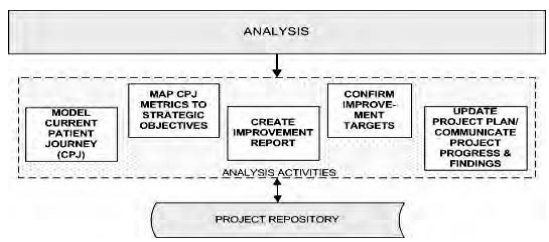


Figure 3 - Analysis phase activities

The first activity is to *create a model of the current patient journey.* This can be done using any process modeling technique including established techniques such as Lean Thinking or emerging techniques such as the 'multi-layered patient flow' communication tool [15, 16].

Existing *measurement criteria* should also be gathered during this exercise. The current patient journey model is created during facilitated group sessions. It is critical during this activity that key stakeholders are involved in the facilitation sessions including administrative, volunteers and patients as appropriate.

Measurements are then *mapped to the strategic objectives* documented in the project initiation phase. This is quite often an enlightening activity, as it will highlight where work is not adding value to the organization, staff or patients.

Following completion of both activities, a comprehensive *Improvement report* is created for management. This details what inefficiencies have been uncovered and what plans of action are possible to improve the current situation. This report should cover areas such as poor use of

human and physical resources, duplicated information collection and communications, unnecessary patient movements, excessive workflow activities and issues of confusion or lack of information for the patient. The report should also recommend priority action areas and activities. The report is submitted to management for discussion and approval.

Once approval is received the *target improvement areas and actions* should be reviewed and confirmed by the project team. It is now time to update the *project plan* based on any new information or scheduling issues and inform the rest of the organization on the outcomes of the analysis phase and what the next steps will be. All documentation created during *Analysis* is stored in the *project repository.*

Design phase

The *Design* phase (***Figure 4***) uses the outputs from the Analysis phase to redesign the patient journey aiming to improve the quality of care and reduce the level of variability for patients experiencing the same journey. It is in this phase that the team will start to work with existing technology constraints and systems and the requirement for new or integrated IT solutions.

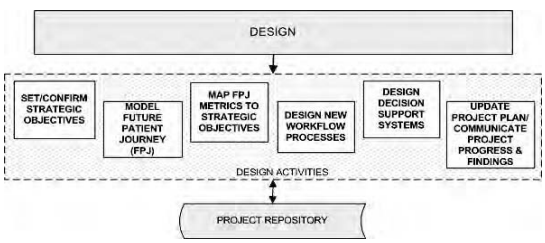


Figure 4 - Design phase activities

The first activity is to *confirm the strategic objectives* that the redesigned patient journey supports.

Once the direction is clear, facilitated group sessions are again used to create a visual representation of the *future patient journey.* This will include new or improved process flows, more efficient use of human and physical resources, streamlined information collection and dissemination, reduced patient movements, improved patient interactions and measurement criteria for all areas.

The *measurement* items are again mapped to the *strategic objectives* to ensure that the new journey is adding value to the organization and its future direction. Some adjustments may be necessary to the measurements defined in the previous activity and as with all other phase documentation these will be updated in the *project repository.*

The completion of the future patient journey and agreed metrics leads to the *design of new or enhanced workflows*. These may be automated or manual workflows. To enable the defined measurements to be captured and analysed, a *decision support system design* is required. This will identify what measurements must be captured and at what stage of the workflow enactment they are required. This is typically the domain of the IT section and will be derived directly from the patient journey redesign work already conducted. The *project plan* is again updated and *project progress and findings are communicated* to the organization.

Implementation phase

The *Implementation* phase (***Figure 5***) is mainly concerned with the development and implementation of the designs output from the Design phase. This will primarily involve the IT section but will still require input from the project team. This will be in the form of system *testing of new/ integrated workflows and the decision support system*.

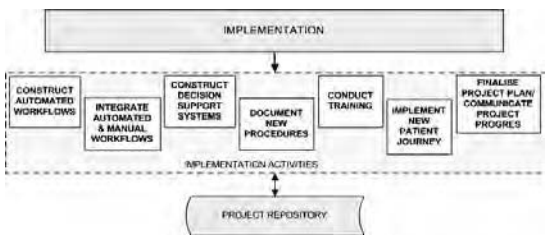


Figure 5 - Implementation phase activities

In parallel to this, the project team will update or *create documentation detailing the new patient journey* along with updated daily workflows. *Training* will need to be conducted on this material as well as any new systems that are to be implemented. The new patient journey goes 'live', with the *implementation of new/enhanced systems and workflows*. The project plan is finalised and details of the project's implementation is communicated. Staff are also advised that following implementation, further improvement is encouraged and can be communicated to the project manager for inclusion in the Evaluation phase.

Project evaluation phase

The *Project Evaluation* phase (***Figure 6***) should be commenced within 3 months of implementation. The main goal of this phase is to revisit the new procedures and determine if they are delivering the expected benefits.

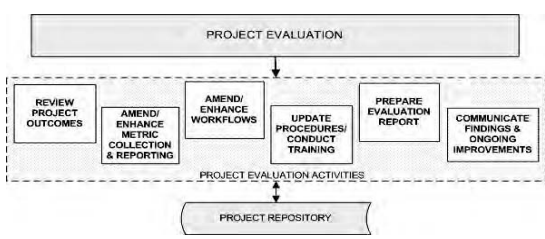


Figure 6 - Project evaluation phase activities

The first activity involves *reviewing the actual results* delivered by the new patient journey and measuring these against expectations. Decision Support System output should be analysed to determine if process metric collection and reporting is meeting set targets *Amendments or enhancements* to the actual measurements and the way they are gathered and reported may be required. This may also lead to *refinement of the newly implemented workflows*.

Any changes to metric collection and reporting or workflows require updating of the documentation in the *project*

repository and possibly republication of updated patient journey procedures. Additional *training* may also be required.

An *evaluation report* is prepared for management outlining evaluation findings and actions taken to further improve the patient journey. Findings may lead to recommendations to revisit the *analysis phase* to conduct a major revision of the (now) current patient journey or may trigger recommendations of completely new patient journey modeling project/s.

Results of the evaluation phase are again communicated to the organization including any new changes or projects.

Results

The Ryde exercise has almost completed the first 3 phases of the proposed patient journey modeling SDLC. Work is still progressing, although some new workflows and procedures have already been tested. This includes the introduction of a new patient assessment form that can be completed and submitted online reducing the number of times a woman is required to attend the hospital prior to her first antenatal appointment. Patient permission paperwork has also been improved reducing the number of forms completed from 3 to 1. Explicit action plans are also in place to complete identified improvements. Health care staff found the SDLC easy to follow and confirmed that when integrated with the MFI, the SDLC had given them a solid direction and set of activities to complete. Specific mention was made of the fact that it could be followed by staff inexperienced in health care redesign projects and made the interactions with the IT section more seamless.

Discussion

Integrating the SDLC with the MFI. This approach is seen as complementary to the Model for Improvement (MFI) not as a replacement. Specific activities within the SDLC align with MFI tasks (ie: set patient journey goals and setting aims) and others can be integrated as part of the MFI (storing gathered information in a project repository). The MFI also gives further information on some of the SDLC activities such as *'Forming the Team', 'Setting Aims'* and *'Defining Measurements'*. An important point to note is how *Plan-Do-Study-Act* (PDSA) cycles are integrated into the SDLC. Once the Design phase is complete, specific areas of the future patient journey model can be selected for implementation on a trial group. Once this trial is complete, the improved procedures can be expanded until management and staff is confident that full-scale implementation should proceed. This means that the Implementation phase can be conducted in an iterative manner, with several iterations leading to implementation of the full future patient journey model over time. The SDLC approach also extends the MFI by including activities for the design and development of technology solutions to support process change. Most importantly the inclusion of a project management framework that supports the planning, management and control of improvement project work provides inexperienced project staff with a solid basis for delivering improvement results within required timeframes.

Conclusion

The integration of a SDLC approach with the MFI for patient journey modeling is achievable and can be understood by all improvement team members, both IT and non-IT. The project team found the SDLC steps logical and easy to follow and produced demonstrable improvement results in the required timeframe, along with ongoing goal-focused action plans.

Acknowledgments

The authors would like to thank the staff of the Ryde Maternity Department at Ryde Hospital, NSW, Australia.

References

[1] Satzinger JW, Jackson RB, Burd SD. Systems Analysis and Design in a Changing World. 3rd ed. Boston: Thomson Course Technology; 2004.
[2] Pressman RS. Software Engineering: A Practitioner's Approach. 4th ed. New York: McGraw-Hill; 1997.
[3] Institute of Healthcare Improvement, The Breakthrough Series: IHI's Collaborative Model for Achieving Breakthrough Improvement. White Paper. Boston: Institute of Healthcare Improvement; 2003, pp. 20.
[4] NHS Modernisation Agency, Improvement Leaders' Guide: Process mapping, analysis and redesign. Best Practice Guidelines. London: Department of Health; April 2005, pp. 22.
[5] NHS Modernisation Agency, Improvement Leaders' Guide: Improving flow. Best Practice Guidelines. London: Department of Health; April 2005, pp. 22.
[6] UK Department of Health, Patient Pathways. Organisation Policy: UK Department of Health; 2005
[7] NSW Department of Health, Easy Guide to Clinical Practice Improvement - A Guide for Healthcare Professionals. Sydney; October 2002, pp. 45.
[8] Szwarcbord M. Redesigning care: improving flow across the hospital [Conference Presentation]. Glenelg: 1st Australasian Redesigning Health Care Summit, Glenelg South Australia, 8-9 March 2005; 2005.
[9] Kasanen E, Lukka K. The Constructive Approach in Management Accounting Research. Journal of Management Accounting Research 1993;5:pp.243-265.
[10] Avison D, Lau F, Neilsen PA, Myers M. Action Research. Communications of the ACM 1999;42(1):pp.94-97.
[11] Tracy S. Proposal for a primary health care model of maternity services at Ryde Hospital - Introduction of the Ryde Midwifery Group Practice. Unpublished Report. Sydney: Midwifery Research & Practice Development Unit; 15th March 2004, pp. 44.
[12] Australian College of Midwives, National Midwifery Guidelines for Consultation and Referral. Guideline. Canberra: Australian College of Midwives; January 2004
[13] Change Management Learning Centre, 2005 Best Practices in Change Management. Benchmark Report. Loveland, CO: Prosci, pp. 51.
[14] Kotter JP. Leading Change. Boston: Harvard Business School Press; 1996.
[15] Institute of Healthcare Improvement, Going Lean in Health Care. Innovation Series Report. Cambridge, MA,: Institute of Healthcare Improvement; 2005, pp. p.24.
[16] Curry JM, McGregor C, Tracy S. A Communication Tool to Improve the Patient Journey Modeling Process. 28th IEEE International Conference of the Engineering in Medicine and Biology Society (EMBC2006); New York, USA.

Address for correspondence

Joanne Curry
email: jm.curry@uws.edu.au)

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

The Nurse—Patient Trajectory Framework

Gregory L. Alexander[a]

[a]*Sinclair School of Nursing, University of Missouri—Columbia, USA*

Abstract

The development of nursing knowledge should give structure and form to the practice of nursing. The development of Nursing Process Theory resulted from early nursing observations and inferences from nursing practice that resulted in formal data accumulation processes, mutual correspondence between nurses and patients, and exchange of information. The development of the nursing process discipline helped to substantiate the need for professional nursing services. The shifts towards examining the links between processes and outcomes, professional accountability, and classification of distinct nursing functions have influenced the development of information systems. The Nurse—Patient Trajectory Framework described in this paper may be used to show the relationships between the virtual information system and the real world that it affects. The framework is visualized along two separate and distinct nurse and patient trajectories.

Keywords:

nursing informatics, information systems, nursing process

Introduction

Nurses and patients interact with computers on many levels. Staff nurses collect information from their clients and, in some cases, simultaneously store client information in information systems. Nursing administration or researchers can use the information contained in this warehouse of data to evaluate utilization, financial impact, and risk management activities of nursing services. The information can also be used to maintain an ongoing record of events, actions, behaviors, perceptions, and progress made during healthcare encounters. Effectiveness of nursing information systems to understand and predict nursing and patient outcomes depends on the usability of the system, organizational workflow, and satisfaction during nursing interactions with computer information systems. Therefore, understanding interactions between nurses, patients, computers, and other elements of information technology has the potential to improve nursing care processes and, subsequently, patient outcomes impacted by nursing services. The following paper describes a framework for nursing informatics (NI) called the Nurse—Patient Trajectory Framework. The framework utilizes nursing process theory, human computer interaction, nursing and patient trajectories as components of a framework that can be used to evaluate patient care systems.

Background

Nursing process theory

Nursing, as a practice discipline, should be concerned with the development of nursing knowledge and nursing knowledge should be used to give structure and form to the practice of nursing [1]. From the roots of the beginning of professional nursing education the nursing process was taught as a means to structure nursing care. Original components of Nursing Process Theory were developed through extensive clinical observations and evaluations of nurses performing the nursing process [2-3]. Emphasis on these clinical observations and inferences made from nursing practice was seen as a forerunner to the nursing process, with its respective components being precursors to a formal data accumulation process [3]. In early anecdotal observations emphasis was placed on the importance of reciprocal relationships between patients and nurses and the importance of the process of nursing care. The nursing process represented the first attempts to develop reciprocity between patients and staff. Reciprocity was garnered through mutual correspondence between patient and staff, through mutual exchange of privilege, and through the mutual dependence, action and influence the patient and staff have on one another.

The four practices basic to the nursing process, as recognized by Orlando [2], were observation, actions, reporting, and recording (Figure 1). Observations included direct or indirect information obtained about a patient while on duty. Direct information was derived from the nurse's own experience of patient behavior. Indirect information included reports of actions, records, or reports of other nurses or allied health professionals. Actions, such as the ability to make decisions or planning care, occur within a context and environment. These contexts are highly influenced by organizational design, area of application, characteristics of the decision maker, maturity of the setting, and importance of the decision [4]. Finally, recording and reporting of nursing information regarding observations and actions is a fundamental function of nursing practice. The effectiveness of nursing process is dependent upon the clinical inferences made from information captured. Every clinical inference made involves an element of risk for the nurse, patient, and the relationship between them [3].

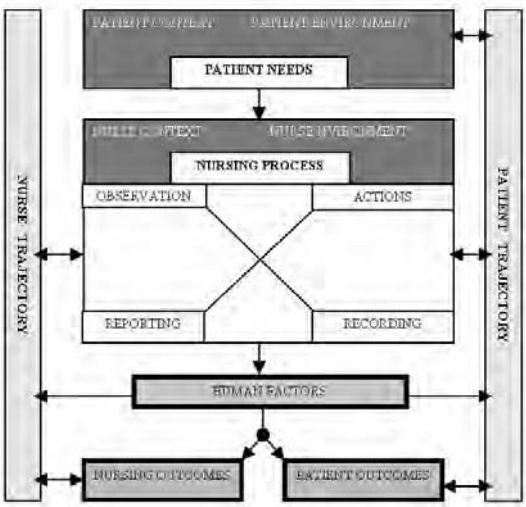


Figure 1 - The Nurse-Patient Trajectory Framework

Nursing informatics: creating the interface between nursing process and technology

Werley and Grier, two pioneers in nursing informatics, advocated for the development of nursing information systems(IS)[5]. Their work led to the identification of nursing data elements essential to diagnosing nursing problems, choosing nursing actions, and evaluating the nursing process in electronic health records (EHR). Furthermore, they suggested research directions to promote the development of technology in caregiving activities. Finally, Werley and Grier published one of the first nursing models that established a link between IS and the development of nursing knowledge.

Their model integrated community data, institutional data, interpersonal data, and patient data into a hierarchical framework. It was suggested that these information sets were needed to assist nurses in making decisions at various levels of functioning[6]. Sets of information thought to be important to nursing decisions included aggregate data on communities including population risk, cultural status, institutional data on finances and structure of facilities, interpersonal elements of caregiving including nursing interventions, orders, and outcomes, and finally, patient data including diagnoses, psychosocial factors, and patients perceptions and goals. Facets of this systems model approach can be seen in subsequent NI frameworks [7;8].

Schwirian developed the NI pyramid as a model for Nursing Informatics [9]. Schwirian defined NI as, "the use of information technology in relation to any of the functions which are within the purview of nursing and which are carried out by nurses."(p. 134). NI activity was depicted as an interface between the computer hardware and software, raw nursing related information, and the user within the context of their profession or organization. All of these elements led to a common goal or objective. The model is described as being flexible and multidimensional allowing the researcher to enter into the model at various points depending on the research questions and hypotheses posed.

Graves and Corcoran [10] defined NI as a, ". . .combination of computer science, information science, and nursing science designed to assist in the management and processing of nursing data, information and knowledge to support the practice of nursing and delivery of nursing care" (p.15). These authors emphasized the processing of nursing information as it progressed from data to information, and finally, to nursing knowledge. In a related article, Goosen [11] extended Graves and Corcoran's NI model to include decisions made in clinical practice, activities that follow nursing decisions, and in the final evaluation, consideration of patient outcomes. These models emphasized the importance of understanding how nurses utilize information to develop knowledge. Goosen went one step further to include pragmatic aspects of information or how information leads to nursing actions.

Turley [12] described a model for NI based on three themes derived from past definitions of nursing informatics. The themes regarded the use and the position of the computer and computer science in informatics, conceptual issues, and functional performance in informatics. These themes underscored the important role computer technologies play in the daily functions of nurses. The model utilized nursing science as a base of knowledge to promote the advancement of nursing informatics as a discipline. Computer science, information science, and cognitive science were represented as spheres in the model that overlay nursing science. The juncture between the three spheres represented the informatics domain. Computer science represented the development of hardware and software to facilitate new understanding and new ways of representing knowledge. Information science facilitated knowledge of organizational structure and informational flow through the organization. Finally, Turley indicated that cognitive science helps to clarify information technology by improving information retrieval, perception of information encountered, and understanding of information processing. Staggers and Parks [13] developed a model called the Nurse—Computer Interaction Framework which has been used to help understand interactions between nurses, computers, and enabling elements that optimize the ability of nurses to process information via computerized systems. The authors identified five elements commonly included in human—computer interaction (HCI) frameworks including the user, computer, tasks, interfaces, and environmental elements. After a review of several frameworks the authors reached several conclusions regarding previous NI frameworks including: 1) most frameworks lack environmental and task oriented elements that are essential to understanding computer interactions, 2) elements of frameworks are conceptualized differently across different frameworks, and 3) most frameworks do not include a dimension of time.

Staggers and Parks [13] included a developmental trajectory for NI including time dimensions not previously developed in NI models. According to the model the NI trajectory has important implications because: a) nurse—computer inter-

actions can change over time and b) the location of phenomena along the trajectory has important implications for outcomes related to nurse—computer interactions.

The most recent model of NI described by Effken [8], the informatics research organizing model (IRO), extends Donabedian's structure—process—outcomes model and emphasizes elements of Nursings' metaparadigm including the system, nurse, patient, and health. Effken described the IRO model as being highly abstract and as being able to accommodate various middle range theories and conceptual frameworks. Effken indicated that all organizing frameworks for NI must address and represent two essential components including context and components of nursings' metaparadigm. Based upon these criteria, previous frameworks for NI were found to exclude specific elements of these essential components. According to Effken, some of the current NI models do not explicitly make the patient part of the model, while other models do not define the context or include all elements of nursings' metaparadigm.

Nurse-patient trajectory framework

Previous NI models have been criticized for not explicitly including aspects of patient care but, being more about nursing management than patients [8]. As discussed previously, reciprocal relationships between patient—nurse, nurse—nurse, and the nurse—significant other are integral components that need to be included in the clinical decisions made by nurses. Evidence of these relationships is represented in the knowledge gained through interactions between these individuals. These interactions facilitate an exchange of communication between patient and nurse that lead to better understanding of the contextual and environmental factors attributed to each person. Crucial factors that must be recognized in shared information are cultural, social, economic, and physical characteristics; excluding this information interferes with the ability to fully understand potential outcomes of a patient [8]. Including this information can facilitate more effective nursing actions that can lead to better individual outcomes along nurse and patient trajectories.

Defining nurse and patient trajectories

The term trajectory in health care can be defined as the assembling, scheduling, monitoring, and coordinating of all steps necessary to complete the work of patient care. The term trajectory refers not only to the pathophysiological process of a patients disease state, but also refers to the total organization of work done throughout all nurse and patient interactions and refers to the impact of patient care processes on those interactions and the organization[14;15]. Trajectories involve different medical and nursing actions by people with different types of skills and resources, trajectories lead to a separation of tasks between workers, including kinfolk and the patient, and trajectories must consider the different relationships between all workers[15].

Two separate trajectories, the nurse trajectory and patient trajectory, are identified in the proposed framework (Figure 1). While appearing to be in parallel with each other these trajectories could be viewed as quite independent of each other. Associated with each trajectory is a trajectory scheme that can be imagined as a sequence of potential events or anticipated events along the trajectory[15]. The beginnings of the trajectory may have two different dimensions for the nurse or patient. The nurse may characterize the diagnosis or chief complaint as the beginning of the trajectory. A patient's trajectory may begin when a symptom or a need appears before coming in contact with a health care professional.

The patient's context and environmental characteristics are also seen as separate and different from nursing context and nursing environment. Context has been described as a multilayered construct that has cultural, economic, social, and physical implications for understanding potential and actual outcomes[8;13]. These actual and potential outcomes are associated along two trajectories, one for nursing and one for patients. The potential and actual outcomes are affected by how technology is integrated into the environment and by the users ability to interact with technology.

Nurse trajectories

Nurse trajectories begin when the diagnoses or chief complaint is determined. Nursing contexts are described in the observations, actions, reports, and records of nursing information. Patient behaviors and perceptions of the nurse that is described in the context of the nursing data influence clinical decision making. Decisions may be influenced by accessibility of information, how information is classified and stored, how it is communicated, how technology is used, and design of workspace including both physical and virtual environments (Salvendy, 2005). Finally, a set of nursing outcomes is identified on the nursing trajectory. While these nursing outcomes may certainly overlap with patient outcomes (i.e. patient safety) the implications for nursing will be different than for the patient. For example, if a nursing process is changed related to medication administration practices the nurse might require education of new policy changes and possibly competency evaluation while the patient only knows that they have received the right medication, just in time and at the right dose.

In contrast to those who attempt to define the nursing process through more descriptive measures, other evaluations of nursing processes using EHR center on quality of documentation, patient satisfaction, and nurse perceptions. Evidence of the benefits and the lack of benefit of IS that incorporate nursing documentation and case management strategies have been reported [16-18]. Nahm and Poston evaluated an integrated point of care systems effect on nursing documentation. The authors identified several attributes of computerized IS that contribute to quality documentation including: a) prompts or reminders within assessments and interventions to alert nurses to required documentation, b) ability to collect real time nursing data, c) standardized, streamlined assessments and interventions in menus and interfaces, d) mandatory fields requiring nursing attention before the nurse can proceed, e) information retrieval from past visits, and f) incorporated work tools that sequence and consolidate tasks and provide reminders when part of the nursing process is missed [16].

The authors also found that computerized documentation did not interfere with patient satisfaction.

In another study evaluating nursing documentation pre- and post-implementation of IS the authors found that IS did not significantly improve documentation within the first 6 months of the study [17]. However, with re-education of nurses on the use of IS documentation of assessments of outcomes, goals, and nursing interventions performed did improve by the end of the 18 month post-implementation phase[17]. Nurse perceptions of clinical information systems were evaluated to determine different views between computer users and non-users about how IS affected their practice [19]. Interestingly, this study indicated that there was a significant difference between the two groups when asked about satisfaction and professional status with computerization. Nonusers were less satisfied because they felt the computer interrupted their thought processes, they felt they could not trust the computer, and they felt the computer thought to much for them resulting in a reduced professional status[19].

Patient trajectories

Patient trajectories may begin with an identified need or symptom and are dependent on a separate set of contextual and environmental factors than healthcare workers. Patient trajectory schemes may be well developed and thought out before they even have any contact with healthcare providers. At the onset of the identified need or symptom patients may begin accessing healthcare information via the World Wide Web or other sources so that they are armed with information for the healthcare worker by the time the diagnosis or chief complaint is made.

There are cultural, economic, social, and physical considerations within the context of the patient environment that shape the patient trajectory[7]. For example, a patient's physical location can have implications for the availability of medical technology or other sources of health care such as information on the Internet. Patient trajectories may also be influenced by past personal experiences or by relationships with other people with similar needs.

The use of technology to evaluate nurse-patient trajectories

Technology can influence trajectories by producing an entirely new trajectory or by lengthening trajectories[15]. New trajectories are created when medical information, that was previously difficult to find, is found with a simple keystroke, with an automated computer alerting system, or by creating color changes in critical text fields holding vital clinical information. The new information may lead to different clinical decisions or judgments regarding treatments, potential outcomes, diagnoses, or utilization of health care resources.

The lengthening of trajectories creates new medical, organizational, and personal problems for patients who are living longer than expected and thus have more complex illnesses and trajectory schemes. Lengthened trajectories may lead to increased specialization, costs, and oftentimes uncertainty of outcomes. Although, research on nurse and patient trajectories is limited some research exists that can be used to describe trajectory evaluation[20-22].

Nurse-patient perceptions and human-computer interaction

The complexities of organizing therapeutic actions are derived from the multiple trajectories, the range and number of complex tasks, which affect the course of patient care and the organization of those tasks [14-15]. The ability of nurses to perceive and organize their work in IS depends on how nurses interact with the computer systems.

Human computer interaction (HCI), oftentimes used interchangeably with usability or human factors, addresses specific issues of human performance during computer interactions including ease of learning, use, remembrance, satisfaction, efficiency, error-forgiving interactions, and seamlessness of fit to tasks [23]. Previous research has shown that IS can provide benefits by seamlessly linking homebound persons with Alzheimer's disease and caregivers with information resources, online health related support, and training [7]. The information system was found to provide strong interpersonal support [7]. This type of early intervention via a computer linking patient, caregiver, and healthcare provider will allow for earlier intervention by nurses in the patient trajectory scheme and may improve patient outcomes.

In other studies, research has shown that understanding the interface between users and the information system plays a role in nurse-patient trajectories. Healthcare providers are challenged by the availability of information at the point of care. In a study, designed to discover and implement design principles to facilitate healthcare practitioners access to healthcare information, a strong correlation was found between total time, navigational ability, and perceived functionality within a computer interface[24]. Ability to find accurate information on which to base decisions can affect the ability of nurses to provide care that is evidence based. Further studies have shown that by reducing the barriers to the use of clinical reminders, such as usefulness, workflow, and efficiency, quality of care of inpatients may improve [25].

Conclusion

The purpose of a framework or a model is to show varying degrees of relevance between virtual or imagined systems and the real world it represents [1]. The purpose of this paper is to describe a framework that incorporates the fundamental components of nursing, the nursing process, with principles of human computer interaction. The framework can be visualized along two separate and distinct trajectories, nurse or patient, which ultimately, depending on the design of the information system, may impact nursing and patient outcomes.

Acknowledgement

Supported by the National Institutes of Health T-15-LM07089-12 National Library of Medicine Fellowships.

References

[1] Young A, Taylor SG, McLaughlin-Renpenning K. Connections: Nursing Research, Theory, and Practice. St. Louis: Harcourt; 2001.

[2] Orlando I. The Dynamic Nurse Patient Relationship. New York: National League for Nursing; 1961.

[3] Kelly K. Clinical inference in nursing. Nursing Research 1966;15(1):23-6.

[4] Abbott PA, Zytkowski ME. Supporting clinical decision making. In: Englebardt SP, Nelson R, editors. Health Care Informatics: An Interdisciplinary Approach.St. Louis: Mosby; 2002. p. 115-28.

[5] Nursing Information Systems. New York, NY: Springer; 1981.

[6] Newcomb BJ. Issues related to identifying and systematizing data--group discussions. In: Werley HH, Grier MR, editors. Nursing Information Systems.New York, NY: Springer; 1981. p. 278-96.

[7] Brennan PF. Characterizing the use of health care services delivered via computer networks. Journal of the American Medical Informatics Association 1995;2(3):160-8.

[8] Effken E. An organizing framework for nursing informatics research. Computers Informatics Nursing 2003;21(6):316-25.

[9] Schwirian PM. The NI pyramid: A model for research in Nursing Informatics. Computers in Nursing 1986 May;4(3): 134-6.

[10] Graves JR, Corcoran S. The study of nursing informatics. Image Journal of Nursing Scholarship 1989;21(4):227-31.

[11] Goossen WTF. Nursing information management and processing: A framework and definition for systems analysis, design, and evaluation. International Journal of Bio-Medical Computing 1996;40:187-95.

[12] Turley J. Toward a model for Nursing Informatics. Image Journal of Nursing Scholarship 1996;28(4):309-13.

[13] Staggers N, Parks PL. Description and initial applications of the Staggers & Parks Nurse--Computer Interaction Framework. Computers in Nursing 1993;11(6):282-90.

[14] Goorman E., Berg M. Modelling nursing activities: Electronic patient records and their discontents. Nursing Inquiry 2000;7(1):3-9.

[15] Strauss A, Fagerhaugh S, Suczek B, and Wiener C. Social Organization of Work. Chicago, IL: University of Chicago Press; 1985.

[16] Nahm R, Poston I. Measurement of the effects of an integrated point of care computer system on quality of nursing documentation and patient satisfaction. Computers in Nursing 2000;18(5):220-9.

[17] Larrabee JH, Boldreghini S, Elder-Sorrells K, Turner Z, Wender RG, Hart JM, et al. Evaluation of documentation before and after implementation of a nursing information system in an acute care hospital. Computers in Nursing 2001;19(2):56-65.

[18] Williams SU. Computerized documentation of case management: From diagnosis to outcomes. Nursing Case Management 1998;3(6):247-54.

[19] Axford R, Carter B. Impact of clinical information systems on nursing practice: Nurses' perspectives. Computers in Nursing 1996;14(3):156-63.

[20] Hawthorne MH. Using the trajectory framework: Reconceptualizing cardiac illness. Scholarly Inquiry for Nursing Practice 1991;5(3):185-95.

[21] Corbin JM. The Corbin and Strauss Chronic Illness Trajectory model: An update. Scholarly Inquiry for Nursing Practice 1998;12(1):33-41.

[22] Alverzo J. The use of aesthetic knowledge in the management of brain injury patients. Rehabilitation Nursing 2004;29(3):85-9.

[23] Staggers N. Usability Concepts and Clinical Computing. In: Ball MJ, Hannah KJ, Newbold SK, Douglas JV, editors. Nursing Informatics: Where Caring and Technology Meet. 3rd ed. New York: Springer; 2000. p. 95-109.

[24] Alexander GL, Hauser S, Steely K, Ford G, Demner-Fushman D. A Usability Study of the PubMed on Tap User Interface for PDAs. Proceedings MEDINFO 2004. 2004.

[25] Patterson ES, Nguyen AD, Halloran JP, Asch S.M. Human factors barriers to the effective use of ten HIV clinical reminders. Journal of the American Medical Informatics Association 2004;11:50-9.

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

System Analysis and Improvement in the Process of Transplant Patient Care

Catherine J Staes[a], R Scott Evans[a,b], Scott P Narus[a], Stanley M Huff[a,b], John B Sorensen[c]

[a]*Department of Biomedical Informatics, University of Utah School of Medicine, Utah, USA*
[b]*Intermountain Healthcare, Utah, USA*
[c]*Department of Surgery, University of Utah School of Medicine, Utah, USA*

Abstract

Clinical information c oncerning transplant patients is voluminous and difficult to manage using paper records. A system analysis was performed to assess information system needs of the liver, kidney, and pancreas transplant program at LDS Hospital in Salt Lake City, Utah. After evaluating workflow, decision support needs, and requirements, we designed and implemented an extendable information system to support care following liver transplantation. We developed and implemented a standardized operative note, forms to enter external laboratory results and transplant-related information into the electronic health record, and computerized alerts to notify the transplant nurses when liver transplant patients had new, abnormal, or overdue laboratory results. The information system has improved the quality of clinical data available in the EHR, clinician satisfaction, and efficiency with management of laboratory results. The components developed for this project can be extended to meet other transplant program needs.

Keywords:

system analysis, ambulatory care, transplantation

Introduction

In the United States during 2005, 28,107 persons underwent solid organ transplantation for end-stage disease.[1] Long-term survival depends on the patient's clinical status, the surgical procedure, the donated organ, and the management of immunosuppressive therapy and complications. The process of evaluating patients, matching them with donors, and monitoring them after transplantation, generates large volumes of information from multiple sources. Physicians, nurses, and support staff that work with transplant programs need to manage and access this large volume of laboratory, clinical and other data to make decisions.

LDS Hospital, in Salt Lake City, Utah, has a liver, kidney, and pancreas transplant program that selects and monitors adult patients that reside in eight western states in the United States. LDS Hospital is 1 of 21 hospitals and numerous outpatient facilities included in an enterprise called Intermountain Healthcare (IHC). In 2004, 159 kidney, liver or pancreas transplant surgeries were performed at LDS Hospital, and 1,216 transplant patients required outpatient monitoring. The volume of information managed by the transplant program increases each year as the annual number of transplantations increases and survival rates improve.[2] In 2000, the director of the transplant program at LDS Hospital requested assistance with developing "a database". Their information system was almost completely based on paper records. The transplant program could purchase or build a stand-alone information system, or they could integrate their needs with electronic record systems available at LDS Hospital.

The objectives of this project were to define the requirements for a transplant program information system, to identify a transplant process that could be improved with computerized information technology, to initiate system development, and to make a positive impact on patient care.

Methods

Using a systematic approach [3], we addressed each software development phase and used our findings to inform the next phase. We performed a system analysis, requirements analysis, and defined a feasible project scope. The system was designed and developed using knowledge engineering methods, vocabulary development, and application programming. We performed usability testing and quality assessment on the new computerized components, implemented the system, trained the transplant team, and evaluated the impact and user satisfaction. Selected methods and results are reported.

System analysis

System analysis is important for understanding problems, directives, and opportunities; establishing priorities; and assessing feasibility.[4] Before expending resources, the following questions needed answers. What information was needed to manage transplant patients? Who needed information and from where will it be accessed? What processes would benefit from computerization? Workflow and information flow were assessed by attending weekly transplant meetings and observing processes in the office and clinic. Existing data forms, reports, and record systems were assessed. The medical, nursing, and support staff were interviewed to determine priorities and clarify the process of transplant patient care. We reviewed clinical

transplant systems described in the literature[4-6] and stand-alone systems available from software vendors.

Application development

The IHC information system had functionality that could meet some transplant program needs.[7] The transplant clinicians had access to the LDS Hospital clinical information system "HELP" and the IHC enterprise-wide, longitudinal electronic health record (EHR).[7] The EHR included laboratory results from 16 IHC hospitals and five IHC clinic laboratories. Within the EHR, clinicians could view IHC laboratory results and enter and view medications, allergies, problems, and clinical notes. A patient list function in the EHR was used to specify patients whose data should be used to trigger alerts and create reports. An application was available to build data entry forms to enter structured laboratory and transplant data. An integrated decision support system was used to create a logic module for transplant related alerts. An integrated messaging application was used by clinicians to view alerts and laboratory data, and navigate within the EHR.

Assessment of data quality on the paper flowchart

An oversized paper flowchart (11 by 25.5 inches) was the primary record used by the transplant team for tracking patients after transplantation. The paper flowchart was maintained and continued to be the primary clinical longitudinal record throughout the entire project. The quality of laboratory data on the paper flowchart is an indicator of information flow and the quality of information being used for decision- making. The completeness of creatinine, tacrolimus, and cyclosporin A results transcribed onto the paper flowcharts was assessed before and after the alerts were implemented. On the day of a patient's chart review, the previous four months of computerized IHC laboratory results for the patient were printed. Each computerized result was compared to the result recorded on the paper flowchart and classified as a "perfect match," "missing," or "different" (incorrect value or specimen collection date). We excluded results that were collected before the transplant program managed the patient. We included only one result each day because only one result is charted on the paper flowchart each day. Both inpatient and outpatient IHC results were included.

Results

System analysis

Information required for transplant patient care was located in the LDS Hospital "HELP" system, the IHC enterprise EHR, and the inpatient and outpatient paper record systems at LDS Hospital. Additional transplant patient information was stored in at least 10 other paper and electronic record systems. . During the study, the flowchart was found to be useful for integrating IHC and external laboratory results in chronological order so clinicians could view trends in laboratory results and immunosuppression drug dose and levels side by side. In addition, the flowchart included patient demographics, interventions, complications, and some nursing documentation. The paper flowchart was not always accessible, was time-consuming to maintain, difficult to reproduce if misplaced, and not amenable to computerized decision support or data analysis.

Six major processes associated with the management of liver, kidney, and pancreas transplant patients were identified (Figure 1). During the final process, clinicians monitored patients after transplantation to manage immunosuppression therapy and to prevent organ rejection, infections, and medication toxicity. Each process involved a unique set of patients, records, and activities; however, common features were identified. All six processes required the reporting of information to the national organization United Network for Organ Sharing (UNOS), and communication with external entities.

The system analysis generated the following findings:

- Among all patients managed by the transplant nurses, more patients are managed after transplantation (77%) than prior to transplantation (23%) when they are being evaluated or waiting for transplant. The population of post-transplant patients is increasing (an average of 4% each year between 2000 and 2004).
- The transplant clinicians closely manage liver transplant patients for their lifetime. Kidney and pancreas patients are transferred to community physicians a few months after transplantation.
- Approximately 70% of the laboratory results for liver transplant patients are available in the EHR. Clinicians report that laboratory results are the most important clinical data used for managing patients after transplantation.
- When managing patients after transplantation, nurses repeatedly performed a task amenable to computerization and decision support: they perform surveillance on laboratory results and observe occurrences, trends, and omissions. In contrast, prior to transplantation, clinicians gathered a heterogeneous set of information to make one decision (e.g., Is the person eligible for transplantation?). Also, prior to transplantation, there is a greater need to exchange information and establish interfaces with outside entities.
- The nurses were concerned about losing patients to follow-up. Transplant patients need periodic testing for immunosuppression drug levels and kidney toxicity to recognize complications and monitor medication tolerance. Laboratory testing is performed three times a week soon after transplantation, then decreased over time to every three months. The nurses requested a system for tracking laboratory results to identify new, critical, and overdue laboratory results.
- Transplant management information systems available from vendors did not meet our needs for several reasons. Interfaces would need to be developed to download laboratory and other clinical data from the EHR. Information charted in the transplant information system would not be accessible for viewing and decision support from the EHR and would be yet another electronic data storage silo. The system would duplicate functionality that already existed in the EHR.

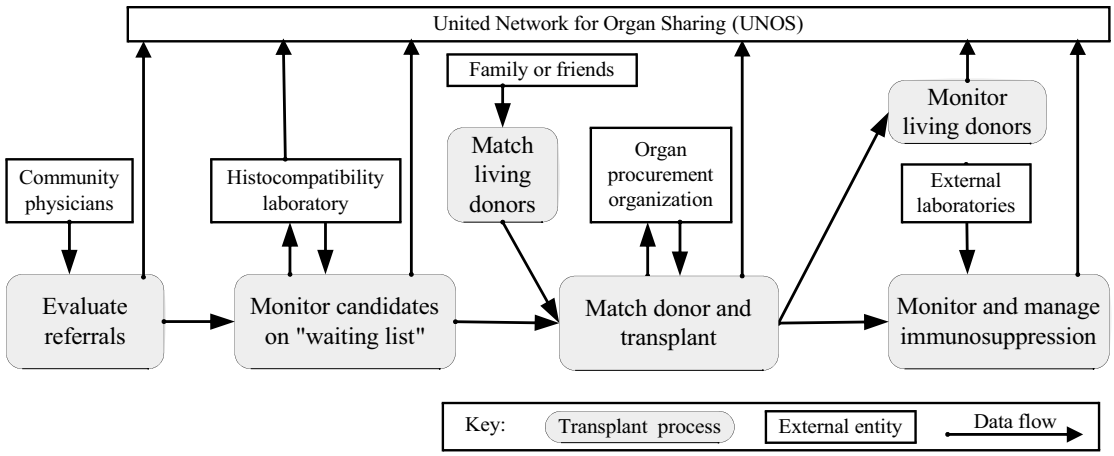


Figure 1 - Major processes associated with the management of solid organ transplant patients

Finally, external laboratory results would still need to be hand entered into the system to integrate external and IHC results.

- Information required for UNOS reporting was not always available in the current free-text operative note. The methods used for calculating ischemia times varied between clinicians.

Project scope

A project targeting the management of patients after liver transplantation was selected for the first system improvement project.

Requirements

The following major requirements were identified:

1. Identify the patient population and each patient's status with the program.
2. Use models that can be extended to other organs and processes of transplant care.
3. Use vocabulary useful for clinical care, but consider concepts reported to UNOS when creating lists of "reasons for transplant" and data collection forms.
4. Improve the quality of documentation about the transplant event, the donor, and risk factors so information is available for decision support, reporting to UNOS, and analysis of outcomes.
5. Integrate external and IHC lab data in the EHR.
6. Create a side-by-side view of laboratory results and medication dosage.
7. Meet regulatory and accreditation standards.
8. Use and do not duplicate functionality that exists in the IHC EHR.

System Design

Improved data collection

A new paper operative note was developed with input from the surgeons. This note standardized documentation about the transplantation and included a graphic to calculate organ ischemia times. This information is important for both clinical management and accurate reporting to UNOS.

Model development and data input

After defining core data fields and improving documentation, data models and entry forms were developed to store information in the EHR. Forms were created to enter the patient's status with the program, selected laboratory results reported by laboratories external to IHC[8], and core information about the transplantation, the donor, and risk factors present at the time of transplantation. External laboratory results in the EHR were critical for automating surveillance of overdue laboratory results and avoiding false overdue alerts. Donor and risk information is important for decision making (e.g., CMV prophylaxis) and program evaluation.

Data views and reports

A "Transplant" view of laboratory results was created so clinicians can see external and IHC results in chronological order with analytes displayed as they are ordered on the paper flowchart.[8] Reports were developed to support workflow and summarize transplant patient information.

Computerized alerts

The transplant physicians and nurses defined logic used to generate alerts. Alerts were received for all new creatinine results (Figure 2). An additional message was included when creatinine results showed an acute or trending increase of 0.3 mg/dL or more. The creatinine alerts prompted the nurses to view all new laboratory results in the EHR. Alerts were also received for all new immunosuppression drug levels and for abnormal potassium and magnesium values. Each alert message included the date of transplantation and the time since transplantation, and indicated if the patient was hospitalized. Alerts were also received when patients were overdue for creatinine or immunosuppression drug level testing. For example,

patients that underwent transplantation during the previous three to six months are expected to get their blood tested every two weeks. If a patient had no creatinine or immunosuppression drug level for 21 days, then an ìoverdueî alert was triggered. When reviewing and acknowledging alerts, the nurses documented their actions in the EHR.

Implementation

In February 2003, the "Transplant" view of laboratory data was implemented. In spring 2004, the surgeons started using the new operative note; the medical assistant started entering external laboratory results; the nurses started entering information about transplantations; the patient list was generated and information about each patient's date of transplantation and program status was entered into the system; and the nurses, pharmacists, and support staff started using the messaging application. The office staff supported this implementation by sending all phone messages electronically instead of leaving paper messages. On May 28, 2004, the alerts were implemented. By fall 2004, the three liver transplant nurses and their assistant were using the alerts to review laboratory results and identify patients overdue for testing. The medical assistant transcribed laboratory results to the paper flowchart (previously, the nurses performed this task).

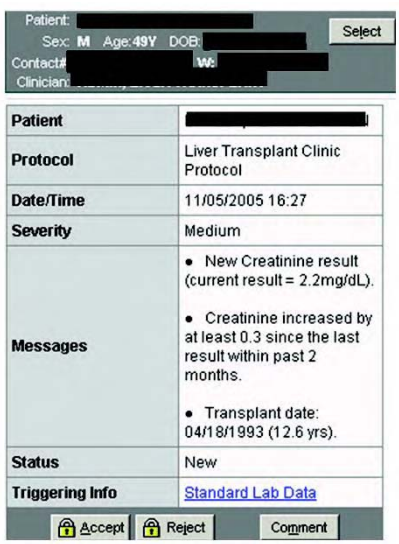


Figure 2 - Alert for a liver transplant patient

Assessment of data quality on the paper flowchart

Among results collected within 7 days prior to the chart review, the proportion of results missing on the paper flowchart decreased from 33 to 16%, before and after implementing alerts, respectively. Among results collected 14 days to 4 months prior to the chart review, the proportion of missing results decreased from 12 to 5% after implementing alerts. The change was significant for both time intervals (Pearson chi-square 0.02). The missing results may not have been transcribed to the flowchart for a variety of reasons.

Impact on workload and workflow

The computerized alerts increased workload but improved workflow. During 2005, the nurses received approximately 130 new creatinine alerts each week. In addition, each week the nurses received 12 alerts for patients newly recognized as overdue for testing. Nurses could quickly find patients with abnormal values and trends and could review laboratory results online and respond from the hospital setting where physicians were accessible for consultation.

Previously, the nurses could only manage laboratory reports while physically in the transplant office. The alerts facilitated cross-coverage among the staff. The nurses managed one list of alerts and could view all new alerts and the actions taken by others. The overdue alerts consistently identified patients that needed follow-up and led to a new system for sending letters to overdue patients. In addition, clinicians were notified when their patients became hospitalized in an IHC facility. As of December 2006, the liver transplant team continues to use the system daily to manage their patients and receive 50-60 alerts each day. They want to expand the system to manage kidney patients.

Discussion

To our knowledge, there is no published literature describing a system analysis of transplant patient care or the design of alerts for the outpatient care of liver transplant patients. The alerts were successful because we identified a problem important to the clinicians that was amenable to information technology. We fit the solution into the workflow, and created a simple intervention.[9] The clinicians have an improved system for tracking patients and responding quickly to laboratory data. In addition, the program administrator has a tool for communicating needs (Figure 1) and was able to get resources for a pre-transplant application.

The transplant patient record continues to be a hybrid of electronic and paper systems. The paper flowchart is still used because clinicians need to view laboratory results and medication dose side by side. Medications are not stored in the EHR in a manner to support this view. The alerts improved the completeness of the record, but the process of transcribing lab results creates a record that remains less than 100% complete.

Conclusion

The objectives defined for this project were met, largely because the system analysis informed us about how to integrate transplant program needs with existing tools and clinical systems. The information system has improved the quality of data available for decision- making. The requirements and systems developed for this project can be extended to meet other transplant program needs.

Acknowledgments

We thank the LDS Hospital transplant team, IHC, and the National Library of Medicine (training grant #2T15LM07124) for their support.

References

[1] UNOS. Based on organ procurement transplantation network (OPTN) data as of November 24, 2006. Accessed November 27, 2006. Available from http://www.optn.org/data/.

[2] HHS/HRSA/HSB/DOT. 2005 OPTN/SRTR Annual Report 1995-2004 (http://www.hrsa.gov/)

[3] Whitten J, Bentley L, Dittman K. System analysis and design methods. fifth ed. Boston: Irwin / McGraw-Hill 2001.

[4] Sarkar MS SD, Giuse D, Ing, Shultz E, Miller RA. Supporting longitudinal care for transplant patients with an external laboratory data entry application (poster). Proc AMIA Annu Fall Symp; 2001 Nov 3-7, 2001; Washington, DC; 2001.

[5] Gordon RD, Markus B, Mitchell S. A liver transplant center information management system. Gastroenterol Clin North Am. 1988 Mar;17(1):61-9.

[6] Kurtz M, Bennett T, Garvin P, Manuel F, Williams M, Langreder S. Demonstration of SLUMIS: a clinical database and management information system for a multi organ transplant program. Proc Annu Symp Comput Appl Med Care. 1991:889-90.

[7] Clayton PD, Narus SP, Huff SM, Pryor TA, Haug PJ, Larkin T, et al. Building a comprehensive clinical information system from components. The approach at Intermountain Health Care. Methods Inf Med. 2003;42(1):1-7.

[8] Staes CJ, Bennett ST , Evans RS, Narus SP, Huff SM, Sorensen JB. A case for manual entry of structured, coded laboratory data from multiple sources into an ambulatory electronic health record. J Am Med Inform Assoc. 2006 Jan/Feb;13(1):12-5.

[9] Bates DW, Kuperman GJ, Wang S, Gandhi T, Kittler A, Volk L, et al. Ten commandments for effective clinical decision support: making the practice of evidence-based medicine a reality. J Am Med Inform Assoc. 2003 Nov-Dec;10(6):523-30.

Address for correspondence

Catherine Staes, BSN, MPH, PhD;
Department of Biomedical Informatics;
University of Utah
Email: Catherine.staes@hsc.utah.edu

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

St Elsewhere's or St Everywhere's: Improving Patient Throughput in the Private Hospital Sector

Jennifer A Laffey[a] and Moran Wasson[b]

[a] *Sydney Adventist Hospital Information Services Department Consultant, EpiSoft Pty Ltd*
[b] *Sydney Adventist Hospital Director of Clinical Services, Acute Care and Perioperative*

Abstract

Communication errors have been found to be most common root cause of medical errors by the US-based Agency for Healthcare Research and Quality [1]. Although elective admissions to hospital involves a high volume of important healthcare communications where incorrect, missing or illegible information could result in a serious medical error, there is little published research on the impact of improving pre-admission communication flow between admitting doctors and hospitals.

The Sydney Adventist Hospital (the San) is a 341-bed private hospital in Sydney's northern suburbs that provides a comprehensive range of health services. A process improvement program began in early 2005 to streamline preadmission communications. The objectives of this ongoing program are broadly to improve patient safety and to increase operating efficiency. The first major initiative within this program was to implement a standardised method for inpatient booking / referral with over three hundred admitting doctors. Eighteen months on, the hospital has been able to demonstrate a significant shift in the timeliness of patient bookings from specialists' rooms, more comprehensive provision of clinical indicators that can facilitate resource planning in operating theatres and on the wards, and reduction in the ratio of bookings made in areas other than the hospital bookings department. The program continues with focus on improving accuracy of data entry, rationalising patient forms, making more effective use of information received and automation of pre-admission information flows.

Keywords:

patient admission, operating rooms, perioperative care, patient throughput, preadmission communication, inpatient booking

Introduction

The case for improving efficiency

The Australian private hospital sector comprises a substantial proportion of the Australian acute care sector with nearly 40% of all inpatient admissions and 56% of surgical admissions being to private hospitals [2]. The number of separations in the private hospital sector has grown by nearly 4% in the period 2003-4 to 2004-5 compared with a 2.6% increase in public hospital separations in the same period. There has been a 6% increase in the number of same-day separations from private hospitals in 2004-5 compared with the previous year [2].

Private hospitals have a need to more effectively manage patient throughput to meet this growing demand while optimising patient safety and satisfaction with the services provided.

The case for improving safety

Communication errors have been found to be the most common root cause of medical errors by the US-based Agency for Healthcare Research and Quality [1].

A study of 197 anaesthetic-related incidents by Kluger et al [3] found that poor communication was a major contributing factor to patient death, comorbidity or unplanned admission to ICU or HDU following surgery. Similarly, inadequate pre-admission communication was found to be a contributing factor in the murder of a woman by a psychiatric patient who was incorrectly diagnosed and discharged from a facility in the Hunter region of NSW [4].

Recent examples of poor pre-admission communication in the Sydney Adventist Hospital (the San) have resulted in:

- a medical secretary providing the wrong side information on a surgical list. This was subsequently entered to the theatre management system as received i.e. incorrectly. No written documentation was received from the doctor. This was picked up by the surgeon prior to the operation;
- no notification of a patient's recent admission to another hospital that would have alerted the San to the patient being at greater risk of having a multi-resistant organism (which they were later found to have). The patient was subsequently admitted to a shared room on the ward placing another patient at increased risk of infection

There were over 4 million elective inpatient admissions in Australia in 2004-5 (AIHW 2004-5). Elective admissions to hospital involve a high volume of important healthcare communications where incorrect, missing or illegible information could involve a sentinel event (such as death from anaesthetic complications or operation on the wrong body part).

The purpose of this research was to compare the quality of preadmission information before and after implementation of the new booking process. While a number of studies have reported on quality improvement projects in the preadmission area [5,6,7], there is little published research about the impact of improving preadmission communication flows between admitting doctors and hospitals.

About this facility

The San is a 341-bed not for profit private hospital in Sydney's northern suburbs that provides a very comprehensive range of health services including medical and surgical acute care, diagnostic services and emergency care. It runs twelveoperating theatres, two cardiac catheter laboratories, two endoscopy rooms and inpatient radiology procedure rooms. It admits about 43,000 patients per annum inclusive of 27,000 surgical, endoscopic and cardiac cathteterisation cases.

There are doctors on staff at the hospital but most of the medical care is provided by Visiting Medical Officers. Prior to this process improvement program being implemented, a wide range of preadmission referral documentation was provided by VMOs ranging from none at all (phone-call from secretary) to a comprehensive letter documenting comorbidities and current medications as well as planned treatment details. This was provided via the patients. Most typically, for surgical patients, a surgical list was written or typed by the doctor's secretary that included patient demographics, the planned procedure(s) and occasionally other details. This list was faxed to the hospital 2-3 days prior to the admission date of those patients. A written referral from the admitting doctor was also received with the patient paperwork in many cases.

About this study

A process improvement program was commenced in late 2004 to enhance the flow of information prior to hospital admission (Patient Throughput program). The objectives of this ongoing program are:

- to improve patient safety;
- have better and more timely information available for theatre and ward resource planning
- centralise bookings processes hence achieve efficiency gains
- streamline patient throughput on the day of admission and
- improve patients' satisfaction in administrative aspects of their hospital episode

The first major initiative within this program was to implement a standardised method for inpatient booking / referral with over three hundred admitting doctors (the "Hospital Booking Letter" project).

This paper describes the results of this initiative. It also describes some of the preliminary findings of other patient throughput initiatives.

Materials and methods

Agreement on standardised content for the Hospital Booking Letter involved extensive consultation with hospital staff, both clinical and administrative, as well as the Medical Advisory Committee.

The booking letter includes the patient's basic demographic information, provisional diagnosis, comorbidities, allergies and infections (with the key clinical indicators affecting order of surgical list specified), admission details (date, whether transfer from other facility), surgical procedure details if applicable, special equipment or resources required, pre-operative consults, surgical implants and diagnostic orders (pathology, radiology, ECG). The back page of the document includes a chart for medication orders on admission. Some of the Hospital Booking Letter is structured (tick-boxes) to reduce the time to fill by the doctor, to draw attention to clinical information of particular importance and to facilitate data entry.

The new process involved the admitting doctor completing this referral during the patient consultation and the doctor's secretary faxing it to a designated Toll Free number as soon as the decision was made to admit the patient and the date of admission was known. This involved a significant change of process in the doctors' rooms, not only by the doctors who had formerly provided no written documentation but also by their secretaries who were used to faxing one list of patients through shortly before the admission date. It also involved a major change of process in the hospital because there was a significantly larger volume of faxes being received. Faxes to the Toll Free number were printed in two different locations, one in the Hospital Bookings department and one in another administrative area to be sorted for use by the Pre-Admission Clinic staff to further process pre-operative consultations and diagnostic test requests.

Implementation with the admitting doctors involved site visits to a large number (over 100) admitting specialists to explain the benefits of the new process to themselves and their administrative staff, to provide them with information packs and to walk them through the key information that was useful to the hospital prior to admission. Where possible, both specialist and secretary were met with and where not possible, just the secretary.

For doctors admitting comparatively fewer patients per annum, the process was implemented via mail-out and explanatory telephone call to the secretaries and / or doctors.

In scope were all medical and surgical patients with the exception of Day Chemotherapy (who have additional information needs), Maternity patients (who attend a booking appointment with a midwife in first or second trimester) repeat visit renal dialysis patients (who are admitted by staff in the unit) and patients admitted via the Emergency Department. Gastroenterologists (Endoscopy bookings) were excluded initially but are currently being implemented on the new process.

Results

More timely information

Figure 1 below shows the notification profile (days between booking and admission) in the twelve months since February 2005 when only some doctors were implemented on the new process compared with February 2006 when most were using the new process. The notification profile of bookings in August 2006 reflects additional specialties (cardiologists) implemented on the new process since February 2006.

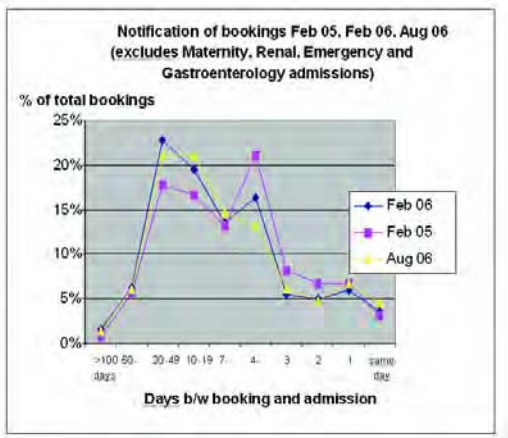


Figure 1 - Days between booking and admission

While there have been no changes in the 'tails' of the curve with booking notification greater than 50 days or on the day prior or same day of admission (or in the 7-9 day notice range), there has been a general shift towards more advance notification in the other ranges with fewer notifications being received 2-6 days prior to admission and more notifications being received 10-50 days prior to admission.

Better quality information

A random sample audit of 80 bookings (representing approximately 3% of all bookings for the period) was undertaken in the six week period from mid Sept 2006 to end Nov 2006. Bookings received via the Hospital Booking Letter contained provisional diagnosis in 86% of cases, comorbidities in 55% of cases (diabetes in 6%), allergy information (including 'nil known') in 41% of cases and infection information (including 'none') in 3% of cases. Pre-Admission Clinic (tests and / or consults) were ordered in 40% of cases.

In contrast, a sample audit of 79 bookings received in the same period via the typed surgical list method included provisional diagnosis in no cases, comorbidities in 4% of cases (diabetes in 1%), allergy information in 1% of cases and infection information in no cases.

In a smaller sample audit of 22 bookings in the same six week period, data entry accuracy was also measured. Accuracy was found to be exceptionally high for the basic surgical booking data of patient demographics, planned procedure, admission date, Medicare item numbers and surgical implants. Of the data entered, information was entered accurately albeit with a few minor spelling errors. In contrast, an unacceptably high number of errors of omission of additional data was found: provisional diagnosis and special notes/instructions were not entered at all; comorbidities were only entered in 2/6 bookings containing comorbidities (33% accuracy only) and allergies entered in 4/7 cases (57% accuracy).

Since senior staff members of the bookings department cross-check data entered on the day prior to admission, this may not have been reflective of the final level of data entry quality. However, the hospital's stated direction is that data are entered accurately and comprehensively the first time hence the audit was carried out on the initial booking entry.

Doctors' handwriting was found to be an issue in 7/22 cases (32%) in the same smaller audit sample which may have accounted for some missing data (not legible to data entry clerk).

Improved patient safety

A definitive measurement plan for this objective of the program is still being put in place making use of data from the hospital's risk management system. The key confounder to overcome is an increased number of risk incidents related to pre-admission communication arising from improved compliance in incident recording.

Centralising booking processes

The Hospital Booking Letter initiative resulted in a 6% increase in the proportion of bookings flowing to the hospital bookings department as opposed to other departments (an increase from 74% in Feb 2005 to 80% in Aug 2006).

Other objectives of the program

The Hospital Booking Letter initiative was not expected to have any effect on streamlining processes on the day of admission but nevertheless some minor benefits were noted in this area. The Pre-Admission Clinic uses the Hospital Booking Letter to identify patients who may require consultation with a case manager prior to admission. This is in addition to those patients where pre-operative tests or consults have been specifically ordered by the doctor. Attendance at the Pre-Admission Clinic reduces the time of both clerical and nursing admission on the day of admission.

With respect to improving administrative aspects of the patient's visit to the San, the more timely receipt of bookings has allowed the hospital to post out, for a larger proportion of admissions, an estimate of out of pocket expenses as opposed to providing this information on the day of admission. However, the main gains in this area are likely to arise from future patient throughput initiatives, especially the project to rationalise patient forms, implement an online pre-admission process and the project to refine the algorithm used for theatre time prediction.

Discussion

The Hospital Booking Letter project achieved a number of the stated objectives of the Patient Throughput program, particularly with respect to more timely and comprehensive receipt of pre-admission information from the admitting doctor. The main issues that arose during implementation are discussed below.

Process change

While the majority of doctors and secretaries took the process change in their stride, there were a number who were resistant to a change that they perceived as more work, both from a documentation point of view and from a process (fax per patient booking) point of view.

The project steering committee regularly reviewed a list of doctors who were not conforming, discussed the likely barriers and agreed methods to overcome these. In some cases, there were practical reasons for non-compliance (the doctor's rooms simply ran out of Booking Letter forms). In some cases, doctors with several staff in different locations did not effectively pass on details about the new process to their colleagues. In a small number of cases, there was complete refusal to conform by either the admitting doctor or their secretary or both. In general, most admitting doctors and their staff who originally did not implement the process have responded well to the rationale of improving patient safety.

Similarly the staff in the hospital took some while to get used to effectively handling and managing the additional paperwork. Reliability of the fax server was found to be of paramount importance. Given the volume of bookings received daily, a few hours' downtime of the fax server on several occasions in late 2005-early 2006 resulted in chaos in the hospital bookings department when the system came back online. The fax server was subsequently upgraded and no issues have arisen now for more than eight months.

The process is now operating efficiently with the Hospital Bookings department sorting bookings into priority order based on date of admission and entering bookings into the system within 24 hours of receipt. Similarly the administrative area located adjacent to the Pre-Admission Clinic is effectively sorting paperwork for follow-up by the Pre-Admission Clinic.

Data quality

The theatre management system routinely now contains information well in advance of patient admission about the patient's comorbidities, allergies and infections, all of which can potentially affect the order of surgical list. The Theatre Manager has been able to more effectively manage potential equipment conflicts (such as navigator probes). The Radiology Department now has increased notice of where Image Intensifiers have been ordered to plan equipment and radiographer time in theatre and ICU staff can see where an ICU bed has been ordered post-operatively. The hospital's manual handling co-ordinator has visibility of some overweight patient admissions for resource planning on the wards but it is believed that doctors under-report this comorbidity and an education program about weight information is currently underway with VMOs.

Following on from the data entry audit of late 2006, an ongoing process of monitoring and feedback to bookings staff has been put in place and a follow-up audit is planned for early 2007 to measure improvement. This process is designed to be non-threatening to individuals but to incrementally improve the accuracy of the whole department over time by focusing on problem areas in the data. The system audit trail allows for individual follow-up if required by the department manager.

Making effective use of data received

While some of the senior Nursing Unit Managers such as in day surgery and in theatres are making very good use of the information received to plan resources, others are not and require ongoing support and education about the IT system and how to more effectively use the information within it. This is partly an issue with the IT system not making key data sufficiently visible for busy clinicians to view easily and partly an issue related to nursing work practice. Both issues are currently being addressed with IT system enhancements and nurse education and training.

Current and future patient throughput initiatives

Notwithstanding some of the issues described above, the patient throughput program at the San has already achieved a number of early gains against its stated objectives with the Hospital Booking Letter initiative. Ongoing process improvement continues to more fully realise the objectives.

Patient admission forms rationalisation

During site visits to doctors' rooms for the Hospital Booking Letter project, a number of medical secretaries provided feedback that the San's patient forms were very confusing for patients. This patient paperwork pack is currently being revised into a single booklet inclusive of only essential information and forms combined. At the same time, nursing staff are working on revising the patient history form to include additional questions that would normally be asked on admission with a view to reducing the time for the nursing admission.

Pre-Admission Clinic process improvement

This service to patients continues to grow relative to overall admissions. Until recently, the scheduling process in the Pre-Admission Clinic involved a number of inefficiencies for clerical staff, especially with respect to communicating with other departments involved in the patient's visit. A scheduling system designed in house was put in place in Nov 2006. The project brief was to eliminate duplicate processes (paper and electronic diary), facilitate departmental communication and save clerical staff time. All these objectives have been demonstrably achieved within three weeks of going live with the new scheduling system.

Improving the predictive model for operating theatre times

Predicting time in the operating theatre as accurately as possible is essential for more effectively managing over- and underutilised theatres (hence streamlining patient throughput) and also improving patient satisfaction by minimising waiting time in the hospital prior to their operation.

The San currently has a model for predicting theatre times based on the procedure item number (or item number combination) and the surgeon. Basic benchmarking against other technologies in the sector has found that the San's algorithms are more sophisticated than many but they still result in a systematic underestimation of actual time in both theatre and Endoscopy procedure rooms.

It is hypothesised that other patient variables routinely captured by the hospital bookings system could also affect theatre times. The San has recently engaged a statistician to review these data and advise on whether more accurate predictions can be achieved.

Scanning received paperwork

This is being investigated as a means of improving workflow and interdepartmental communications and ensuring that essential patient paperwork (such as the consent to treat, signed by the patient and their admitting doctor) does not go missing.

Electronic automation of pre-admission

This involves a tripartite communication flow between doctor, hospital and patient and the hospital is exploring methods to facilitate online electronic booking by the doctor and online entry of admission and history information by the patient.

Conclusion

The patient throughput program at the San is a program of continuous improvement made up of a number of initiatives involving people, process and technology.

Incremental improvements in pre-admission communications have been found to have a positive impact on operating efficiencies and inpatient resource planning. Improving efficiency in the pre-admission process is essential to sustain quality of service delivery in the face of increasing demand for inpatient services.

Although improved data quality is expected to improve patient safety, the effective use of available information by clinicians is also a factor. Further research is needed to determine whether such communication improvement initiatives have a positive impact on patient safety.

Acknowledgments

The authors gratefully acknowledge Sydney Adventist Hospital for permission to publish these findings and the work of the cross-functional team involved in the Patient Throughput program especially staff in Information Services, Patient Services, Business Development, Nursing Administration, Medical Administration and Medical Records.

References

[1] AHRQ's Patient Safety Initiative: Building Foundations, Reducing Risk. Interim Report to the Senate Committee on Appropriations. AHRQ Publication No. 04-RG005. 2003 (Dec). Agency for Healthcare Research and Quality, Rockville, MD. http://www.ahrq.gov/qual/pscongrpt/

[2] Australian Insitute of Health & Welfare. Australian Hospital Statistics 2004-5. Health Services Series No 26, 2006. http://www.aihw.gov.au/publications/index.cfm/title/10305

[3] Kluger MT, Tham EJ, Coleman, NA, Runciman WB, Bullock MFM. Inadequate pre-operative evaluation and preparation: a review of 197 reports from the Australian Incident Monitoring Study. Anaesthesia. 2000; 55(12): 1173.

[4] Peterson K. Where is the line to be drawn? Medical Negligence and Insanity in Hunter Area Health Service v Presland, Sydney Law Review. 2006; 28(1): 181. http://www.austlii.edu.au/au/journals/SydLRev/2006/9.html

[5] Shearer W, Monagle J, Michaels M. A model of community based, preadmission management for elective surgical patients. Can J Anaesth.1997; 44(12): 1311-1314.

[6] Hoeksema J and Munski J. Development of a Patient-Centre Preprocedure Program. AORN Journal. 1997; 65(2): 388-395.

[7] MacLeod AM. Review of a hospital preadmission program: applying the principles of continuous quality improvement. Leadership. 1997; 4(5): 11-14.

Address for correspondence

Jennifer Laffey
c/- Information Services Department,
Sydney Adventist Hospital,
185 Fox Valley Road,
Wahoonga NSW 2076
Email: jennifer.laffey@sah.org.au.

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

A Meta Schema for Evidence Information in Clinical Practice Guidelines as a Basis for Decision-Making

Katharina Kaiser[a], Patrick Martini[a], Silvia Miksch[a,b], Alime Öztürk[a]

[a] *Institute of Software Technology and Interactive Systems, Vienna University of Technology, Austria*
[b] *Department of Information and Knowledge Engineering, Danube University Krems, Austria*

Abstract

Clinical practice guidelines are an important instrument to aid physicians during medical diagnosis and treatment. Currently, different guideline developing organizations try to define and integrate evidence information into such guidelines. However, the coding schemas and taxonomies used for the evidence information differ widely, which makes the use cumbersome and demanding. We explored these various schemas and developed a meta schema for the evidence information, which covers the most important components of the existing ones, is comprehensible, and easy to understand for the users. We developed and assessed the usefulness and applicability of our meta schema with guideline developers and physicians.

Keywords:

clinical practice guidelines, evidence-based medicine, recommendations, study characteristics, clinical decision support, otolaryngology

Introduction

Evidence-Based Medicine (EBM) is defined as "the integration of best research evidence with clinical expertise and patient value" [1]. EBM advocates the use of up-to-date best scientific evidence from health care research as the basis for making medical decisions. One means to communicate research evidence is to integrate it into clinical practice guidelines (CPGs), which are "systematically developed statements to assist practitioner and patient decisions about appropriate health care for specific clinical circumstances" [2]. Evidence-based CPGs involve a comprehensive search of the literature, an evaluation of the quality of individual studies, and recommendations that are graded to reflect the quality of the supporting evidence.

Evidence-based recommendations are mostly classified in particular grading schemas to provide a unique format at least for guidelines of the developing organization. Various organizations popularized taxonomy systems for grading the quality of evidence and the strength of recommendations (SoR) (e.g., [3-5]) and developed methodologies to categorize recommendations according to their systems.

In this paper we discuss the importance of a meta schema for levels of evidence (LoEs) and SoRs as a means for comparing, handling, and connecting LoEs and SoRs of different taxonomy systems for supporting the medical decision-making process. Due to the profusion of grading schemas users are often puzzled by the message a grade conveys. The different application of codes (e.g., I, II, III, …; A, B, C, …; 1, 2, 3, …; Ia, Ib, IIa, …) and the different definitions of the levels are not only confusing to users, but also aggravates a comparison and decreases the transparency of the schemas [6]. Table 1 and Table 2 give two examples of LoEs and SoRs from two different guideline-developing organizations.

Table 1 - Levels of Evidence used by the University of Michigan Health System

Level	Definition
A	Randomized controlled trials
B	Controlled trials, no randomization
C	Observational trials
D	Opinion or expert panel

The overall objective of this work is to facilitate the decision-making process on the basis of a systematic representation of the evidence information. A systematic representation is required to handle evidence information in computer-interpretable guideline representation languages (see [7] and [8] for a comprehensible overview). To achieve this objective we meet the following more specific objectives:

1. Development of a meta schema for grading evidence information. This schema should cover the most important components of various rating schemas for LoEs and SoRs.
2. Mapping of different LoEs and SoRs used by different organizations into this meta schema.

For our research we used 21 evidence-based CPGs from the clinical specialty otolaryngology. We selected the clinical specialty otolaryngology, because there are many, well structured guidelines available for our purpose. Based on the different LoEs and SoRs in these CPGs we have developed a meta schema for both graded and ungraded evidence information and SoRs.

Table 2 - Strength of Recommendations defined by the Scottish Intercollegiate Guidelines Network (SIGN)

Strength	Definition
A	At least one meta analysis, systematic review, or RCT* rated as 1++, and directly applicable to the target population; or A systematic review of RCTs or a body of evidence consisting principally of studies rated as 1+, directly applicable to the target population, and demonstrating overall consistency of results
B	A body of evidence including studies rated as 2++, directly applicable to the target population, and demonstrating overall consistency of results; or Extrapolated evidence from studies rated as 1++ or 1+
C	A body of evidence including studies rated as 2+, directly applicable to the target population and demonstrating overall consistency of results; or Extrapolated evidence from studies rated as 2++
D	Evidence level 3 or 4; or Extrapolated evidence from studies rated as 2+

* RCT … Randomized Controlled Trial

The meta schema connects existing grading systems to provide a means to increase the transparency among the various schemas and to appraise ungraded recommendations. By the direct comparability of various grading systems the communication of the underlying information is quickly and concisely possible.

Furthermore, decision-support is a crucial topic in computer-supported guideline's research. The meta schema will thereby facilitate the integration of evidence information and form a basis to handle the multitude of grading systems on an equal level.

In the following section we describe the process of developing our meta schema and we show the results in the subsequent section. Furthermore, we discuss the outcomes and cease with concluding remarks and future work.

Methods

Guyatt et al. [9] defined several criteria for developing an optimal grading schema. As our intention is not to develop a new system, but a meta schema that connects the existing schemas, other requirements apply. These are to have sufficient categories to cover a good portion of systems, to be consistent with existing systems, and to be simple and transparent to users.

We selected 21 CPGs from the National Guideline Clearinghouse[1], which were developed by nine different organizations (see Table 3). These CPGs cover eight different representations of LoEs and three different representations of SoRs.

Table 3 - Guideline development organizations

Organization/Cooperation	Number of Guidelines
American Academy of Family Physicians; American Academy of Otolaryngology – Head and Neck Surgery; American Academy of Pediatrics	1
American Academy of Pediatrics	2
Allergic Rhinitis and its Impact on Asthma Workshop Group	1
Cincinnati Children's Hospital Medical Center	3
Finnish Medical Society Duodecim	1
Institute for Clinical Systems Improvement	6
Practice Guidelines Initiative	1
Scottish Intercollegiate Guidelines Network	3
University of Michigan Health System – Academic Institution	3

After analyzing the different grading schemas of these organizations we decided to use the definitions of LoEs and SoRs of the Scottish Intercollegiate Guidelines Network[2] (SIGN) [10] as a basis for our meta schema. One reason for this decision was that SIGN's representation of evidence information is systematically evaluated and clearly structured and defined. As the SIGN approach does not cover all required information to represent the different LoEs and SoRs, we expanded the definitions for other organizations to cover their representations of the evidence information, too.

The most relevant attributes for developing the meta schema are:

- Code schema of guideline developing organization
- Study design and quality
- Strength of recommendations
- Benefits and harms

During the development process we conducted interviews with various guideline developers and physicians. We discussed the correctness, sensibility, availability, and understandability of the hierarchical structure, the quality of the LoEs and SoRs, the mapping tables, and the balance between benefits and harms. Furthermore, we surveyed the

1 http://www.guidelines.gov (last assessed December 3, 2006)
2 http://www.sign.ac.uk (last assessed December 1, 2006)

availability of required information and the facilitation of the decision-making process. Moreover, the covering of the meta schema with existing grading systems was verified during the entire process. The remarks and comments were incorporated in our schema altogether.

Code schema of the guideline developing organization

This attribute is essential to differentiate between various grading schemas, because a symbol or code communicating a grade can represent different meanings (see also [6]). For example, the *University of Michigan Health System* uses the symbols "*A", "B", "C", "D*" for LoEs, whereas *SIGN* uses these symbols for SoRs. Thus, it is not possible to extract the evidence information from the guidelines and map them to the meta schema, without the information about the developing organization.

Study design and quality

The quality of evidence is described by LoEs. They are mostly explicitly represented in the guidelines but different symbols are used to refer to them (see for instance Table 1).

The attribute of the study design covers all study types (i.e., Meta-Analysis, Systematic Reviews, Randomized Controlled Trials (RCTs), Cohort Studies, Case Control Studies, Expert Opinion) used in CPGs.

We represent the LoE on the basis of the study design's attribute, because in that way we get an ordered structure, where *meta-analysis* is on the top of the hierarchy and *no study design* is at the bottom.

The study design attribute plays a significant role by assigning a grade to ungraded evidence information. Often CPGs include information about the study design upon which the recommendations are based, but they do not provide any explicit grades for their evidence (e.g., "*The recommendations are supported by randomized controlled trials. Adverse parasympathetic events were reported by participants in randomized controlled trials, the most frequent and troublesome being increases sweating which occurred in about one-quarter of patients taking 5 mg three times per day and about one-half of patients taking 10 mg*" [11]).

Another attribute to be considered for establishing the LoEs is the study quality. It refers to the detailed study methods and execution. The study quality is thereby the degree to which a study employs measures to minimise biases, focussing on internal validity [12].

Our representation has to address both the study design and the study quality. The levels have to be clearly distinguishable and easily and clearly interpretable.

Strength of recommendations

For SoRs also different symbols or names are used but they are not always explicitly mentioned in our CPGs. Three out of nine organizations have defined SoRs and only six of the 21 CPGs include explicitly defined SoRs. In 15 CPGs no information about SoRs is included.

The representation of SoRs has to be representable to the different existing SoRs. The requirements for our SoR taxonomy are that

- The strengths have to be clearly distinguishable from each other
- The names of the grades have to be meaningful
- The strengths have to be easily and clearly interpretable
- The number of grades should be limited to ensure an easy understanding and application

The GRADE Working Group[3] developed a system for defining the recommendations based on four factors [4]:

1. The trade-offs, taking into account the estimated size of the effect for the main outcomes, the confidence limits around those estimates, and the relative value placed on each outcome
2. The quality of the evidence
3. Translation of the evidence into practice in a specific setting, taking into consideration important factors that could be expected to modify the size of the expected effects, such as proximity to a hospital or availability of necessary expertise
4. Uncertainty about baseline risk for the population of interest.

Recently, medical associations and organizations adapt the GRADE approach for their needs [9,13]. However, the publication of guidelines using the new grading systems will take time. In our CPGs, this new approach is not implemented yet. The strength of a recommendation is only based on the underlying quality of evidence.

Benefits and harms

Information about benefits and harms of a particular treatment plays a significant role in the decision-making process. But in our guidelines they are described very briefly and limited. They do not contain information about the trade-off between the benefits and harms either. For embedding information about benefits and harms into the decision-making process, we need them to be represented explicitly. Thus, a schema for the trade-off between benefits and harms is necessary, because this information is essential for the medical staff to assess benefits and harms of a treatment recommendation. For example:

1. In patients with peptic ulcer, drug A reduces acidity. This recommendation is based on RCTs.
2. In patients with cardiac problems, drug A may cause heart attacks and hence is contraindicated. This recommendation is based on case reports.

The second argument is based on lower quality evidence, but defeats the first argument, because of the more important claim (heart attack is worse than having acidity reduced).

Results

The meta schema

Based on the considerations described in the previous section we developed a meta schema for representations of the quality of evidence, the strength of recommendations, and benefits and harms.

3 http://www.gradeworkinggroup.org (last assessed November 30, 2006)

Definition of levels of evidence

Our definition of LoEs is based on the *study design* and the *study quality*. The *study design* is essential to get a hierarchical representation and to assign a level to ungraded evidence information in CPGs. The LoEs consist of symbols that cover information about the study design and the quality of the studies. We introduced our own symbols (e.g., I_1, I_2, … II_1, II_2, …) that represent both the study design and quality. The first character describes the study design whereas the number describes the quality. Table 4 shows a part of our LoEs schema.

Table 4 - Part of the meta schema representing Levels of Evidence

Study Design	Evidence Level	Definition
Meta Analysis	I_1	Meta-analysis of RCTs
	I_2	High quality meta-analysis
	I_3	Well-conducted meta-analysis
	I_4	Meta-analysis
Systematic Reviews	II_1	High quality systematic reviews of RCTs with large sample
	II_2	High quality systematic reviews of RCTs with small sample
	II_3	High quality systematic reviews of RCTs with very low risk of bias
	II_4	Systematic reviews of RCTs
	II_5	High quality systematic reviews of cohort studies
	II_6	High quality systematic reviews of case-control studies
	II_7	Systematic reviews
…	…	…

Definition of strengths of recommendations

Based on the requirements for our SoRs taxonomy we defined four different strengths, because more than four hierarchical levels are hardly distinguishable, but less do not adequately cover existing systems. The four grades are:

1. Strong Recommendation
2. Recommendation
3. Weak Recommendation
4. No Recommendation

They have a unique definition and are easy to differentiate from each other. For example, *Strong Recommendation* is directly applicable to the target population and bases on at least one meta-analysis, systematic review of RCTs, RCTs with very low risk of bias, high quality meta-analysis of observational studies, or high quality systematic reviews of observational studies.

Our aim with these definitions of SoRs is to provide guideline users a proposed recommendation that should only be a direction if there is no explicit representation of SoRs in the CPGs.

Definition of trade-off between benefits and harms

Table 5 shows the definitions of the trade-off between the benefits and harms. Our definitions are based on the descriptions used by the GRADE working group [4], because they have a well defined categorization of the trade-off between the benefits and harms in their grading schema.

Table 5 - Schema for trade-off between benefits and harms

Classification	Benefits and Harms
Clear Benefit	The benefits of the recommended approach clearly exceed the harms.
Benefit	The recommended intervention explicitly does more good than harm or the benefits outweigh the harms.
Unclear Balance	It is unclear whether the recommended intervention does more good than harm. The trade-off between benefits and harms is quite unclear.
No Clear Benefit	The recommended intervention clearly does not do more good than harm.

Mapping the evidence information of CPGs

The meta schema should provide a general representation of different classifications of LoEs and SoRs used in CPGs. We connected each individual LoE and SoR taxonomy to our meta schema. Figure 1 shows a mapping between SIGN and our meta schema. The mapping tables provide guideline users a better handling and understanding of the evidence information in CPGs. With this representation the users have a means to compare different LoEs and SoRs.

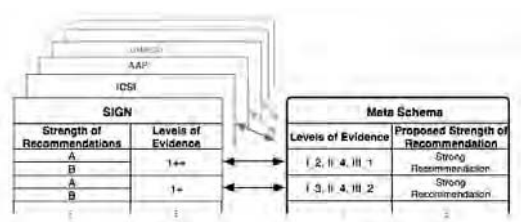


Figure 1-– Mapping table for the meta schema showing mappings from and to the SIGN schema

Discussion

The most significant factor for decision-making is the strength of recommendations. In most taxonomies the following aspects are taken into account:

- The level of evidence of individual studies
- The type of outcomes measured by these studies (patient-oriented or disease-oriented)
- The number, consistency, and coherence of the evidence as a whole
- The relationship between benefits, harms, and costs

In most guidelines this information is not entirely available. Thus, it is only possible to assign a constricted and intermediate SoR. Furthermore, benefits and harms for each recommendation are needed to incorporate them in the constitution of the SoR. But often, benefits and harms are only given for the entire guideline and not individual recommendations.

Conclusion

Our meta schema is an instrument to connect different schemas of LoEs and SoRs. The meta schema is representable to eight different systems defining LoEs and three different systems defining SoRs and incorporate the ideas and concepts of the GRADE Working Group. Furthermore, it is possible to assign a LoE to an ungraded evidence recommendation based on the study design and quality if available. It covers also information about the trade-off between benefits and harms, which are mostly not included in the existing grading schemas. We used the attributes study design and study quality (defined in LoEs), SoRs, the organization's code schema, and benefits and harms, which were significant for the development process.

Furthermore, we think that our meta schema can also support instruments for guideline appraisal (e.g., AGREE [14], GLIA) in terms of providing means to better understand and compare the various existing grading schemes for evidence information.

Acknowledgments

This work is supported by "Fonds zur Förderung der wissenschaftlichen Forschung - FWF" (Austrian Science Fund), grants P15467-N04 and L290-N04.

References

[1] Sackett D. Evidence Based Medicine: What it is and what it is not. BMJ, 312:71, 1996.

[2] Field M and Lohr K. Clinical Practice Guidelines: Directions for a New Program. National Academy Press, 1990.

[3] Harbour R and Miller J. A new system for grading recommendations in evidence based guidelines. BMJ, 323:334–336, 2001.

[4] GRADE Working Group. Grading Quality of Evidence and Strength of Recommendations. BMJ, 328(7454):1490-1498, 2004.

[5] Ebell MH, Siwek J, Weiss BD, Woole SH, Susman J, Ewigman B, and Bowman M. Strength of Recommendation Taxonomy (SORT): A Patient-Centered Approach to Grading Evidence in the Medical Literature, American Family Physician, 69(3): 548–556, 2004.

[6] Schünemann HJ, Best D, Vist GE, and Oxman AD for the GRADE Working Group. Letters, numbers, symbols and words: how to communicate grades of evidence and recommendations. Canadian Medical Association Journal (CMAJ), 169(7): 677–680, 2003.

[7] Peleg M, Tu SW, Bury J, Ciccarese P, Fox J, Greenes RA, Hall R, Johnson PD, Jones N, Kumar A, Miksch S, Quaglini S, Seyfang A, Shortliff E, and Stefanelli M. Comparing Computer-Interpretable Guideline Models: A Case-Study Approach, JAMIA, 10(1):52–68, 2003.

[8] de Clercq PA, Blom JA, Korsten HHM, and Hasman A. Approaches for Creating Computer-interpretable Guidelines that Facilitate Decision Support, Artificial Intelligence in Medicine, 31(1):1–27, 2004.

[9] Guyatt GH, Guttermann D, Baumann MH, Addrizzo-Harris D, Hylek E, Philips B, Raskob G, Zelman Lewis S., and Schünemann HJ. Grading strength of recommendations and quality of evidence in clinical guidelines: report from an American college of chest physicians task force. Chest, 129:174–181, 2006.

[10] Scottish Intercollegiate Guidelines Network. SIGN 50: A guideline developers' handbook. SIGN, 2004.

[11] Practice Guideline Initiative. Symptomatic treatment of radiation induced xerostomia in head and neck cancer patients, 2004.

[12] Moher D, Jadad AR, Nichol G, Penman M, Tugwell P, and Walsh S. Assessing the quality of randomized controlled trials: an annotated bibliography of scales and checklists. Control Clin Trials, 16:62-73, 1995.

[13] Schünemann HJ, Jaeschke R, Cook DJ, Bria WF, El-Solh AA, Ernst A, Fahy BF, Gould MK, Horan KL, Krishnan JA, Manthous CA, Maurer JR, McNicholas WT, Oxman AD, Rubenfeld G, and Turino GM. An Official ATS Statement: Grading the Quality of Evidence and Strength of Recommendations in ATS Guidelines and Recommendations. American Journal of Respiratory and Critical Care Medicine, 174:605-614, 2006.

[14] The AGREE Collaboration. Writing Group: Cluzeau FA, Burgers JS, Brouwers M, Grol R, Mäkelä M, Littlejohns P, Grimshaw J, Hunt C. Development and validation of an international appraisal instrument for assessing the quality of clinical practice guidelines: the AGREE project. Quality and Safety in Health Care 2003; 12(1): 18–23.

[15] Shiffman RN, Dixon J, Brandt C, Essaihi A, Hsiao A, Michel G, and O'Connell, R. The guideline implementability appraisal (GLIA): development of an instrument to identify obstacles to guideline implementation. BMC Med Inform Decis Mak, 5, July 2005.

Address for correspondence

Katharina Kaiser
Vienna University of Technology
Institute of Software Technology & Interactive Systems
Favoritenstrasse 9-11/188, A-1040 Vienna, Austria
Email: kaiser@ifs.tuwien.ac.at

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Creating Interoperable Guidelines: Requirements of Vocabulary Standards in Immunization Decision Support

Karen M. Hrabak MSN RNC[a], James R. Campbell MD[b], Samson W. Tu MS[c], Robert McClure MD[d], Robert (Tony) Weida PhD[d]

[a] *The Nebraska Medical Center, United States*
[b] *University of Nebraska Medical Center, United States*
[c] *Stanford Medical Informatics, Stanford University, United States*
[d] *Apelon, Inc., United States*

Abstract

Interoperable support of electronic health records and clinical decision support technology are central to the vision of sustainable information infrastructure. Efforts to implement interoperable clinical guidelines for immunization practice have been sparse. We used the SAGE knowledge workbench to develop a knowledge base to provide immunization decision support in primary care. We translated the written clinical guideline into a structured decision logic format. The semantic content to completely capture CDC clinical decision logic required 197 separate concepts but was completely captured with SNOMED CT and LOINC. Although 88% of concepts employed pre-coordinated codes, 6% of guideline concepts required expanded vocabulary services employing Boolean logical definition using two or more SNOMED concepts. Post-coordination requirements were modest, representing just 6% of guideline semantic concepts. We conclude that creation of interoperable knowledge bases employing clinical vocabulary standards is achievable and realistic. Employment of information model (HL7 RIM) and vocabulary (SNOMED CT, LOINC) standards is a necessary and feasible requirement to achieve interoperability in clinical decision support.

Keywords:

interoperability, standard vocabulary, clinical practice guidelines, clinical decision support

Introduction

An expert system employs an inference method linked to a domain ontology [1], ideally evaluates patient state data from the electronic health record, and issues recommendations for care. Advances in clinical vocabulary development [2] and interest multi-nationally [3] has promulgated a core set of reference terminologies which offer to evolve into the comprehensive clinical ontologies needed for electronic health record (EHR) technology and decision support. SNOMED CT is a core reference terminology within these recommendations. NDF-RT and laboratory LOINC are controlled vocabularies with reference features also within the core of standards recommended within the US. At this time however, implementation of these standards is infrequent, in part related to confusion regarding best deployment and difficulties with conversion of legacy data. While a single study report [4] has argued that these terminologies may be insufficient to support guidelines, the functionality of these terminologies relative to representation of domain knowledge for a guideline expert engine poses complex issues not yet resolved.

Clinical practice guidelines seek to standardize care and facilitate the provision of evidence-based care. Historically published in free text formats, efforts to encode and implement guidelines within the EHR for clinical decision support face many challenges. Interpretation and translation of the written guideline text is necessary. Ambiguity within the source publication requires clinical expertise to precisely formulate the guideline logic [5]. Concept modeling problems exposed by guideline encoding include differences in granularity and definition between the guideline and the domain ontology and interactions with the vendor information model.

The National Vaccine Advisory Committee guides best practices surrounding immunization administration in the United States [6]. Included in their recommendations is an emphasis on accurate vaccine recording practices and a method to send clinical reminders to patient and practitioners. Other investigators have recognized the importance of this domain and have worked to create decision support for immunization practice. These past efforts have successfully modeled immunization clinical decision support forecasting and reminder systems [7-8]. The IMM/SERV system supported childhood immunization forecasting and maintained a web based knowledge maintenance and testing environment [9]. The immunization reminder recall system [8] provided immunization decision support utilizing a modular architecture. Concepts in its knowledge base have been mapped to medical entities dictionary (MED) employed at the New York Presbyterian Hospital to provide integration with clinical records at that facility. Representation and maintenance of the knowledge

domains in these systems employed tabular, rules based and procedural approach. Neither of these past implementations has employed clinical vocabulary standards. The resulting decision models employed were unique to the environments in which they were developed.

The SAGE [10] consortium is a collaboration of academic and private sector interests with the shared goal of creating interoperable guideline decision support. In order to support interoperability, the knowledge bases created for SAGE employ only the core

vocabulary resources recommended by the US National Committee on Vital and Health Statistics (NCVHS). Knowledge modeling occurs on an open source workbench created with the Protégé [11] knowledge tool.

The program interface between the SAGE inference engine and the vendor clinical information system communicates via an HL7 reference information model (RIM) compliant query engine termed the virtual medical record (vMR). For purposes of inference support the SAGE engine employs a suite of vocabulary services which bind the decision support software to the SNOMED CT structures. This binding provides ontologic features of subsumption and concept definition. Concept definition beyond the pre-coordinated scope of SNOMED CT is handled with vocabulary service extensions and post-coordination within the SAGE SNOMED CT extension. The end result is a knowledge construction and domain ontology (extension of SNOMED) which can freely interoperate with any other system compliant with SAGE and SNOMED CT standards.

Given the clinical importance of immunization practices and the historical efforts of other decision support scientists, achieving interoperability through the use of standard terminologies is critical. We therefore organized, enumerated and characterized the vocabulary and knowledge services of the SAGE immunization guideline in order to inform the concerns of EHR vendors and emphasize the benefits of vocabulary standards compliance.

Methods

Guideline clarification and logic modeling

The US Center for Disease Control (CDC) through the Advisory Committee on Immunization Practices issues guidelines for the US with specific recommendations for vaccination of child and adult populations [12]. We obtained the immunization recommendations (Fall 2005) for children and adults for this analysis. Our initial task was to translate by hand the written guideline content into a structured representation. We compiled indications, contraindications and deferral criteria for each vaccine. Age specific criteria for eligibility, dosing intervals, and catch up rules for missed vaccine doses were identified. These criteria were used to formulate a knowledge base specification document containing logical IF-THEN statements which formalized the CDC logic while employing the source guideline concept statements. Figure 1 represents an example of a logic statement derived from the pediatric immunization guideline.

From the logic base we compiled an inventory of concept references employing methods we have described [13] and worked with local clinical experts to disambiguate guideline statements which were unclear. For example, the guideline source concept ‘progressive neurologic disorder’ required clinical domain expert definition, resulting in a logical union of the concepts ‘Lennox Gastaut’, ‘Tuberous sclerosis’ and the nested union of concepts ‘Developmental delay’ and ‘Encephalopathy’.

We then compared our concept inventory against pre-coordinated concepts from SNOMED CT and LOINC. When source concepts were not pre-coordinated but could be formulated correctly with logical constructions of pre-coordinated concepts, we did so. Remaining concepts that were clearly outside of the scope of pre-coordinated SNOMED

CT were modeled into an extension namespace for the guideline following editorial principles published by the College of American Pathologists [14].

Rule 1: First dose MMR pediatric	
IF	NO CONTRAINDICATION TO MMR
	AND
	NO REASON FOR DEFERRAL
	AND
	AGE >= 12 MONTHS
	AND
	NUMBER OF MMR DOSES = 0
	AND
	NO VARICELLA VACCINE ADMINISTERED WITHIN 28 DAYS
THEN	
	ADVISE ADMINISTER MMR VACCINE

Figure 1 – Structured decision logic statement

All immunization logic rules and vocabulary concepts were then linked to a set of EHR queries employing the idealized record structure (vMR) which we have developed with the HL7 Clinical Decision Support Technical Committee in compliance with the HL7 Reference Information Model (RIM) [10]. This link involved an explicit assertion of the expected EHR records required and the

attributes required to bind the decision logic to the clinical patient record. Detailed data models of guideline concepts identify how patient data associated with the concepts are represented by a vMR class. Table 2 lists the 13 primary classes of the vMR in the left hand column.

Vocabulary services

Recognizing that guideline statements sometimes requested unique concepts for query from the record, while other statements implied retrieval from within a set of concepts (all instances of diabetes), we reviewed the guideline to clarify the vocabulary services required. We categorized all vocabulary service requirements on a scale reflecting the complexity of the concept relative to pre-coordinated NCVHS vocabularies, and the expected query function in the run-time environment.

Categorization of concept query requirements

Category 1: Concept instance only is directly referenced by the guideline logic. This category includes instances such as gender code, qualifier values (e.g. contraindicated or true) and lab codes (e.g. Hepatitis B surface antigen, Measles virus IGG antibody).

Category 2: Concept, along with all specializations, are implicitly referenced by the guideline. This category includes concepts that may have many conceptual variations within the EHR (such as "Diabetes mellitus" or "Hemoglobinopathy") and the guideline expects all more specialized children of the concept to be included at run-time query.

Category 3: Boolean constructions

- 3a: Guideline concept is represented by the logical Boolean 'OR' (Set union) of two or more category 2 references. For example "Functional or anatomic asplenia" is logically defined by the union of the sets of concepts: "Splenectomy", "Functional splenectomy", "Congenital asplenia", "Sickle cell disease", "Asplenia syndrome" and "Hyposplenism" - including children - within the SNOMED CT ontology.
- 3b: Concept requires the logical Boolean 'AND' (Set intersection). An example of a category 3b concept is "bisexual male" which is the intersection of the concept sets "Patient is male" and "Bisexual".
- 3c: Concept expression includes a Boolean 'NOT' (Set complement). Concepts in this category include concept expressions that exclude descendant concepts within a hierarchy. For example, the guideline concept of "Chronic respiratory disease" when clinically reviewed, was defined to exclude the SNOMED specialization concepts of "Chronic rhinitis" and "Chronic sinusitis" at run-time.

Category 4: Concept post-coordination required. This category includes concepts requiring extension development for SNOMED CT. Since they are not pre-coordinated and cannot be defined from logical statements employing pre-coordinated concepts, they are defined employing SNOMED formalisms as extension vocabulary. An example for the concept "Maternal hepatitis B surface antigen positive" is included in Figure 2.

SAGE knowledge modeling

Employing the immunization rule logic and the concept inventory, we then proceeded to model the complete guideline using the SAGE guideline ontology [10]. Context of care, clinical workflow and organizational resources are elements of the SAGE ontology. All decision logic rules and vocabulary queries were bound and modeled employing SAGE formal criteria and SAGE actions linked to the vendor EHR. For comparison with previous immunization guideline work, we counted and summarized these execution elements.

The SAGE guideline workbench produces an XML knowledge base that can be shared between clinical systems. We validated the immunization knowledge base with a series of experiments including syntax checking of the XML and simulated run-time assessment employing test cases. We are now validating the knowledge base against actual clinical records in two separate enterprise systems.

Results

Vocabulary inventory

This "birth-to-death" immunization knowledge module was a complex construction. The 45 pages of clinical guideline publication were distilled into 75 separate "IF-THEN" statements in support of three clinical implementation scenarios proposed by the clinical team. The scenarios included vaccination advice at birth, a primary care office visit, and a survey scenario for population based reminders. An inventory of the source utterances from the guideline statements yielded 147 conceptual references. Disambiguation and expert clinical opinion was required with 7 concepts which were then defined within the SNOMED CT extension ontology. Table 1 provides a summarization of the pre-coordinated vocabulary concepts that were ultimately required to support guideline logic. These concepts were installed in vMR queries as data type restrictions which defined the value sets for retrieval of information from the EHR by the SAGE decision engine. Each query employed one or several coded concepts from the distinct concepts tallied for the guideline.

```
<concept conceptID= "12110001000004100"
  conceptStatus="0"  fullySpecifiedName="Maternal
   hepatitis B surface antigen  positive (situation)"
   isPrimitive="0">
  <descriptions>
    <description term="Maternal hepatitis B surface
   antigen
     positive(situation)"   descriptionStatus="0"
     descriptionType="3"  languageCode="en-US"
     initialCapitalStatus="0"
  </descriptions>
  <relationshipSet>
    <relationship relationshipType="116680003 IS-A"
       conceptId2="4164710007 Family history of
       clinical finding"  characteristicType="0"
       refinability="0"/>
    <relationshipGroup>
```

```
    <relationship relationshipType="246090004
   Associated
       finding" conceptId2="165806002 Hepatitis B
       surface
       antigen positive" characteristicType="0"
        refinability="0"/>
    <relationship relationshipType="408729009 Finding
      context" conceptId2="410515003 Known present"
       characteristicType="0" refinability="0"/>
    <relationship relationshipType"408731000 Temporal
       context" conceptId2="410512000 Current or
       specified"
       characteristicType="0" refinability="0"/>
     <relationship relationshipType"408732007 Subject
        relationship context" conceptId2="72705000
       Mother" characteristicType="0" refinability"0"/>
     </relationshipGroup>
  </relationshipSet>
</concept>
```

Figure 2 – Post-coordinated concept definition

Table 2 summarizes the final analysis of vMR queries required to support rule logic with a total of the vocabulary concepts. Not all guideline concepts could be accurately modeled employing pre-coordinated SNOMED CT. Table 3 summarizes the complexity of the vocabulary model and services required to support immunization guidelines. This reflects run-time management requirements (only category 1 concept references do not require retrieval of data sets which include all children of the concept) as well as the requirements for post-coordinated vocabulary development (category 4 concepts).

Table 1- Pre-coordinated concepts by semantic type (Category 1 and 2 concept complexity)

SNOMED domain	***n***
Context-dependent category	3
Disorder	51
Finding	23
Observable entity	10
Occupation	4
Organism	1
Person	1
Procedure	9
Product (clinical drug)	50
Qualifier	16
Racial group	1
Substance	5
Total	*n* = 174

Table 2- Concept inventory by vMR query class

vMR query	**SNOMED**	**LOINC**
Adverse reaction	42	
Agent	0	
Alert	0	
Appointment	0	
Encounter	0	
Goal	0	
Observation	39	5
Order	6	
Medication Order	24	
Problem	78	
Procedure	4	
Referral	0	
Substance administration	32	

Characteristics of knowledge model

SAGE employs the frame-based knowledge modeling of the Protégé environment. Immunization logic criteria are formulated into frames which enforce a set of constraints on data query from the EHR. Criteria are employed within decision models which reproduce the source guideline logic and communicate with the vendor record via action specifications. The full immunization knowledge model required 236 Boolean criteria, 207 presence criteria and 161 comparison criteria. These were employed in 88 decision models which employed 82 action specifications. The full immunization guideline model, including SAGE workbench and vocabulary coding, is available for public use and evaluation [15].

Table 3- Concept inventory by complexity

Category	***n***
Category 1 (Concept entity)	35 (17.8%)

Category	***n***
Category 2 (Subsumption)	139 (70.5%)
Category 3a (Boolean with OR)	8 (4.1%)
Category 3b (Boolean with AND)	2 (1.0%)
Category 3c (Boolean with negation)	2 (1.0%)
Category 4 (Post coordination)	11 (5.6%)
Total	$n = 197$

Discussion

The development of reference terminologies for clinical vocabulary standards has created utility but also poses new challenges for the knowledge information specialist. Previous studies [16] have documented the limitations of pre-coordinated terminologies, but a commitment to compositional forms means that procedures and methods for management of post-coordination must be developed.

In contrast to a previous report [4], we found that comprehensive coding in support of our guideline was feasible, but that vocabulary services for the guideline engine had to be extended to include support for two services: 1) Boolean definitions of complex concepts and 2) integration of post-coordination within a SNOMED extension. Since management of large extension vocabulary sets requires new skills and software functionality such as description logic classifiers, this is a matter of developing understanding within the informatics community.

Run-time support provided by reference terminologies such as SNOMED CT is also important to decision support engines. Our experience clearly documents the significance of support for aggregation within record query activity. 80-90% of queries into the EHR were searching not just for a single concept, but for one within a related set. By providing for identification of all specializations of a concept with hierarchical relationships, SNOMED CT supplies knowledge structures which replace the need for exhaustive code list generation in knowledge bases. This defines a clear benefit resulting from standard reference terminology deployment, as well as an important use case for evaluating evolution of these vocabulary systems.

Conclusions

It is feasible to implement guideline decision support within a knowledge engine employing international standard vocabularies. Effective use of these reference terminologies requires new procedures for vocabulary management and deployment. Benefits to the knowledge engineer include savings in domain knowledge development and true semantic interoperability.Acknowledgement

The SAGE project has been supported by grant 70NANB1H3049 of the NIST Advanced Technology Program.

References

[1] Musen MA. Scalable architectures for clinical decision support. *Methods Inf Med.* 1999;38:229-38.
[2] Wang AY, Sable JH, Spackman K. The SNOMED clinical terms development process: refinement and analysis of content. *Proc AMIA Symp*. 2002;: 845-9.
[3] SNOMED Standards Development Organization [Online]. 2006 Nov 28 [cited 2006 Dec 1]; Available from: URL: http://www,snomed.org/news/documents/snomed_meeting_statement_nov2006_update.pdf
[4] Dykes PC, Currie LM, Cimino JJ. Adequacy of evolving national standardized terminologies for interdisciplinary coded concepts in an automated clinical pathway. *J Biomed Inform* 2003. 36:313-25.
[5] Tierney WM. Improving clinical decisions and outcomes with information: a review *Int J Med Inform.* 2001 Jun;62(1):1-9.
[6] Poland G, Shefer A, McCauley M, Webster P, Whitley-Williams P, Peter G. Standards for adult immunization practices. *Am J Prev Med.* 2003;25(2):144-150.
[7] Miller PL, Frawley SJ, Sayward FG. Issues in computer-based decision support in public health illustrated using projects involving childhood immunization. *J Public Health Manag Pract.* 2001;7(6), 75-86.
[8] Wang D, Jenders R, Dasgupta B. Translating national childhood immunization guidelines to a computer-based reminder based reminder recall system within an immunization registry. *Proc AMIA Symp*. 1999;:994-8.
[9] Miller PL, Frawley SJ, Brandt C, Sayward FG. A prototype website for immunization maintenance website. *Proc AMIA Symp*.1997;:293-7.
[10] Tu SW, Campbell J, Musen MA. The SAGE guideline modeling: motivation and methodology. *Stud Health Technol Inform.* 2004;101:167-71.
[11] Gennari JH, Musen MA, Fergerson RW, Grosso WE, Crubezy M, Eriksson H, Noy NF, Tu SW. The evolution of protégé: An environment for knowledge-based systems development. *Int J Hum Comput Stud.* 2003;58(1):89-123.
[12] Center for Disease Control, Advisory Committee on Immunization Practice [Online]. 2006 [cited 2006 Nov 28]; Available from: URL: http://www.cdc.gov/nip/ACIP/default.htm
[13] Tu S, Musen MA, Shankar R, Campbell J, Hrabak K, McClay J, Huff S, McClure R, Parker C, Rocha R, Abarbanel R, Beard N, Glasgow J, Mansfield G, Ram P, Ye Q, Mays E, Weida T, Chute C, McDonald K, Mohr D, Nyman M, Scheitel S, Solbrig H, Zill D, Goldstein M. Modeling guidelines for integration into clinical workflow. *Medinfo.* 2004; 11(Part 1):174-8.
[14] SNOMED Clinical terms Concept modeling/Style authoring guide. College of American Pathologists. October, 2004.
[15] SAGE Project Overview [Online]. 2006 Dec 3 [cited 2006 Dec 3]; Available from: URL: http://sage.wherever.org/projectoverview/projectoverview.html
[16] Campbell JR, Carpenter P, Sneiderman C, Cohn S, Chute CG, Warren J. Phase II evaluation of clinical coding schemes: completeness, taxonomy, mapping, definitions, and clarity. *J Am Med Inform Assoc*. 1997 May-Jun;4(3):238-51.

Address for Correspondence:

Karen Hrabak MSN, RNC
khrabak@nebraskamed.com

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Automatic Treatment of Temporal Issues in Clinical Guidelines in the GLARE System

Luca Anselma[a], Paolo Terenziani[b], Stefania Montani[b], Alessio Bottrighi[b]

[a] *Dipartimento di Informatica, Università di Torino, Torino, Italy*
[b] *Dipartimento di Informatica, Università del Piemonte Orientale "Amedeo Avogadro", Alessandria, Italy*

Abstract

Temporal constraints play a fundamental role in clinical guidelines. For example, temporal indeterminacy, constraints about duration, delays between actions and periodic repetitions of actions are essential in order to cope with clinical therapies. This paper proposes a computer-based approach in order to deal with temporal constraints in clinical guidelines. Specifically, it provides the possibility to represent such constraints and reason with them (i.e., perform inferences in the form of constraint propagation). We first propose a temporal representation formalism and two constraint propagation algorithms operating on it, and then we show how they can be exploited in order to provide clinical guideline systems with different temporal facilities. Our approach offers several advantages: for example, during the guideline acquisition phase, it enables to represent temporal constraints and to check their consistency; during the execution phase, it allows the physician to check the consistency between action execution-times and the constraints in the guidelines, and to provide query-answering and temporal simulation facilities (e.g., when choosing among alternative paths in a guideline).

Keywords:

artificial intelligence, clinical guidelines, temporal constraint representation, temporal reasoning, repeated/ periodic actions.

Introduction

Clinical guidelines are a means for specifying the "best" clinical procedures and for standardizing them. In recent years, the medical community has started to recognize that computer-based systems dealing with clinical guidelines provide relevant advantages, since, e.g., they can be used to support physicians in the diagnosis and treatment of diseases, or for education, critical review and evaluation aims [1]. Thus, many different approaches and projects have been developed to create domain-independent computer-assisted tools for managing clinical guidelines (see e.g., [2, 3]). Most of these approaches distinguish between an *acquisition phase*, in which expert-physicians (usually in cooperation with knowledge engineers) introduce clinical guidelines into the computer-based system, and an *execution phase*, when user-physicians execute a given guideline on a specific patient (i.e., on the basis of the patient's data). Moreover, recently, several approaches have started to focus also on the treatment of temporal aspects [4, 5, 6, 7]. As a matter of fact, in most therapies, actions have to be performed according to a set of temporal constraints concerning their relative order, their duration and the delays between them. Additionally, in many cases, actions must be repeated at regular (i.e., periodic) times. Furthermore, it is also necessary to carefully take into account the (implicit) temporal constraints derived from the hierarchical decomposition of actions into their components and from the control flow of actions in the guideline. A complete automatic treatment of temporal constraints involves – besides the design of an expressive representation formalism – also the development of suitable temporal reasoning algorithms operating on them, to be applied both at acquisition and at execution time. However, subtle issues such as the trade-off between the expressiveness of the representation formalism and the efficiency of correct and complete temporal reasoning algorithms have to be faced in order to deal with temporal constraints in a principled and well-founded way; few works in the area of computerized guidelines have deeply analyzed this topic so far.

In this paper, we first provide a brief overview of the state of the art and we discuss the advantages of a principled approach. Then, we introduce our representation formalism and our reasoning mechanisms to perform temporal reasoning in the acquisition and in the execution phase. Finally, we describe how to exploit our formalism and algorithms to provide clinical guidelines systems with temporal reasoning facilities, and we address comparisons and conclusions.

Background

State of the art

In the area of clinical guidelines, despite the large amount of work devoted to the *representation* of temporal constraints, little attention has been paid to temporal *reasoning*. Notable exceptions are represented by the approaches by Shahar [6] and by Duftschmid et al. [4]. In Shahar's approach, the goal of temporal reasoning is not to

deal with temporal constraints (e.g., to check their consistency), but to find out proper temporal abstractions to data and properties. Therefore, temporal reasoning is not based on constraint propagation techniques, in fact interpolation-based techniques and knowledge-based reasoning are used. Duftschmid et al. have proposed a comprehensive approach based on the notion of temporal constraint propagation [4]. In particular, in Duftschmid et al.'s approach, different types of temporal constraints – deriving from the scheduling constraints in the guideline, from the hierarchical decomposition of actions into their components and from the control flow of actions in the guideline – are supported. Temporal constraint propagation is used in order to (1) detect inconsistencies, and (2) provide the minimal constraints between actions. In [4], there is also the claim that (3) such a method can be used by the guideline interpreter to assemble feasible time intervals for the execution of each guideline activity.

New challenges and open problems

Despite the large amount of work, there still seems to be a gap between the range of phenomena covered by Artificial Intelligence temporal reasoning approaches and the needs arising from clinical guidelines management. In particular, in clinical guidelines the following features have to be supported:

1. qualitative and quantitative constraints, as well as repeated/periodic events; all types of constraints may be imprecise and/or partially defined;
2. a structured representation of complex events (in terms of part-of relations), to deal with structured descriptions of the domain knowledge;
3. the distinction between classes of actions (e.g. an action in a general guideline) and instances of such actions (e.g., the specific execution of an action in a guideline);
4. the consistency of the temporal constraints between classes and instances. This involves dealing with the inheritance of constraints (from classes to instances) and with the predictive role of constraints between classes[1].

Obviously, the interplay between issues (1)-(4) needs to be dealt with, too. For example, the interaction between composite and periodic events might be complex to represent and manage. In fact, in the case of a composite periodic event, the temporal pattern regards the components, which may, recursively, be composite and/or periodic events, as in Ex.1.

(Ex. 1) *The therapy for multiple myeloma is made by six cycles of 5-day treatment, each one followed by a delay of 23 days (for a total time of 24 weeks). Within each cycle of 5 days, 2 inner cycles can be distinguished: the melphalan treatment, to be provided twice a day, for each of the 5 days, and the prednisone treatment, to be provided once a day, for each of the 5 days. These two treatments must be performed in parallel.*

In Ex. 1, the instances of the melphalan treatment must respect the temporal pattern "twice a day, for 5 days", but such a pattern must be repeated for six cycles, each one followed by a delay of 23 days, since the melphalan treatment is part of the general therapy for multiple myeloma. Unfortunately, no current approach in Artificial Intelligence (henceforth AI) and in guideline literature proposes a comprehensive approach in which all the above phenomena can be represented, and correct, complete and tractable temporal reasoning can be performed. In this paper, we introduce an approach addressing all the above-mentioned issues.

Methods

Representing temporal constraints in clinical guidelines

Regarding temporal constraints concerning non repeated actions in the guidelines, we have chosen to model them using a well-known AI framework, STP (Simple Temporal Problem) [8]. This framework takes into account conjunctions (sets) of bounds on the distance between two time points (of the form $c \leq P1 - P2 \leq d$), and correct and complete temporal reasoning (e.g., for consistency checking) can be performed in cubic time by a classical all pairs shortest paths algorithm (such as Floyd-Warshall's one) [8]. STP allows to model precise and imprecise temporal locations (dates), durations, delays between points and qualitative temporal constraints (such as "A is before B").

Let us introduce first the constructs to model the repetition/periodicity constraints.

In our approach, the constraints on repetitions and periodicities are temporal constraints of the form

Repetition(A, RepSpec)

where *A* is a (possibly composite) action repeated according to the parameter *RepSpec*.

RepSpec is a recursive structure of arbitrary depth of the form

$$\mathbf{RepSpec = <R_1, R_2, \ldots, R_n>}$$

where each level R_i states that the actions described in the next level (i.e., R_{i+1}, or – by convention – the action *A*, if $i=n$) must be repeated a certain number of times in a certain time span. To be more specific, R_i consists of a quadruple

$$\mathbf{R_i = <nRepetitions_i, I\text{-}Time_i, repConstraints_i, conditions_i>}$$

where the first term represents the number of times that R_{i+1} must be repeated, the second one represents the time span in which the repetitions must be included, the third one may impose a pattern that the repetitions must follow, and the last one allows to express conditions that must

1 For example, given a guideline stating that action A_2 must be executed between one and two days after A_1, and given an execution of the action A_1 on a given patient at day d_1, one expects to have an instance of A_2 within day d_1+1 and day d_1+2.

hold so that the repetition can take place. Informally, we can roughly describe the semantics of a quadruple R_i as the natural language sentence "*repeat* R_{i+1} $nRepetitions_i$ *times in exactly* $I\text{-}Time_i$, *if* $conditions_i$ *hold*".

repConstraints$_i$ is a (possibly empty) set of pattern constraints, representing possibly imprecise repetition patterns. Pattern constraints may be of type:

- **fromStart(min, max)**, representing a delay between the start of the I-Time and the beginning of the first repetition;
- **toEnd(min, max)**, representing a delay between the end of the last repetition and the end of the I-Time;
- **inBetweenAll(min, max)**, representing the delay between the end of each repetition and the start of the subsequent one;
- **inBetween((min_1,max_1),…,($min_{nRepetitions_{i-1}}$,$max_{n\text{-}Repetitions_{i-1}}$))** representing the delays between each repetition and the subsequent one. Note that any couple (min_j, max_j) may be missing, to indicate that the constraint does not impose any temporal constraint between the j^{th} repetition and the $(j+1)^{th}$ one.

Let us see an example to illustrate the use of *repConstraints*:

(Ex. 2) Intrathecal methotrexate must be administered 7 times during 88 weeks, never less than 10 weeks apart or more then 14 weeks apart.

Ex. 2 may be represented with a one-level specification:

Repetition(Intrathecal_methotrexate,
< <7,88wk,{inBetweenAll(10wk, 14wk)}, ∅>>).

It is worth noting that $repConstraints_i$, $nRepetitions_i$ and $conditions_i$ are not mandatory.

conditions$_i$ is a (possibly empty) set of conditions that influence the repetitions. The conditions may be of type:

- **while(B)**, where B is a Boolean predicate. It states that, as soon as B becomes false, a break from the repetitions is forced, i.e., the repetitions must immediately interrupt. As an example, we may consider the following:

(Ex. 3) Give acetaminophen twice a day until the fever has gone.

This may be represented as:

Repetition(acetaminophen,
< <_, _, ∅, {while(fever)}>, <2, 1d, ∅, ∅> >.

- **onlyIf(B)**, where B is a Boolean predicate. It states that, if B is true, the repetition may be performed and, if B is false, the repetition must not be performed and we can pass to the next repetition. This construct allows to skip single repetitions. As an example, we may consider the following:

(Ex. 4) Give acetylsalicylic acid twice a day for a maximum of 15 days, only if there is migraine.

This may be represented as:

Repetition(acetylsalicylic_acid,
<<_, 15d, ∅, ∅>, <2, 1, ∅, onlyIf(Migraine)>>.

The formalism we are introducing allows one to manage different kinds of imprecision; in fact:

- there may be arbitrary delays between the repetitions;
- the *(min, max)* specifications in $repConstraints_i$ make it possible to specify variable delays between the repetitions.

Dealing with imprecise temporal constraints is very important for the practical applicability of the approach.

Moreover, the repetitions may be nested at arbitrary depth, representing simple cases with fewer levels as in Ex. 2 and more complex cases with more levels as in Ex. 5, an excerpt from a clinical guideline for the treatment of Childhood Acute Lymphoblastic Leukaemia:

(Ex. 5) The therapy lasts 88 weeks and it is repeated twice in four years. In the therapy, cotrimoxazole must be given twice daily on two consecutive days every week.

Ex. 5 may be represented in the following way:

Repetition(Cotrimoxazole, < <2, 4y, ∅, ∅>, <_, 88wk, ∅, ∅>, <2, 1wk, {inBetweenAll(0,0)}, ∅>, <2, 1d, ∅, ∅>>), where the pattern constraint *inBetweenAll(0,0)* in the third triple imposes that the days must be consecutive.

In order to make our approach to temporal constraints more user-friendly, a (possibly graphical) interface could be used to acquire and represent temporal constraints (concerning both (i) dates, durations, delays and qualitative relations between non-repeated events and (ii) repetition/ periodicity constraints).

Reasoning with temporal constraints in clinical guidelines

Internal representation

STP provides a suitable representation for temporal constraints on non-repeated actions. In fact, the constraints can be modeled as graphs on which the well-known Floyd-Warshall's algorithm operates to check consistency. However, the STP framework is not expressive enough to cope with repeated/periodic actions. Thus, we have chosen to model the constraints regarding repeated actions into separate STPs, one for each repeated action. Thus, in our approach, the overall set of constraints between actions in the guideline is represented by a tree of STPs (STP-tree henceforth). The root of the tree (node N1 in the example in figure 1) is the STP which homogeneously represents the constraints (including the ones derived from the control flow of actions in the guideline) between all the actions in the guideline (e.g., in N1, the fact that the duration of the chemotherapy is 168 days), except repeated actions. Each node in the tree is an STP, and has as many children as the number of repeated actions it contains. Each edge in the tree connects a pair of endpoints in an STP (the starting and ending point of a repeated action) to the STP containing the constraints between its subactions and it is labeled with the list of properties describing the temporal constraints on the repetitions (i.e., *RepSpec*). For

example, in Fig. 1, we show the STP-tree representing the temporal constraints involved by Ex. 1.

Additionally, an independent STP must be used in order to represent the temporal constraints about the specific *instances* of the actions of the guidelines, as emerging from executions of the guidelines on specific patients.

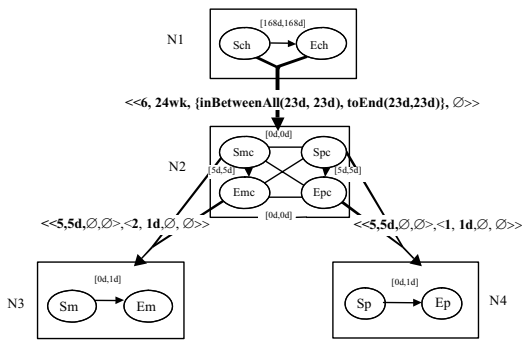


Figure 1 - STP-tree for the multiple myeloma chemotherapy guideline in Ex. 1. Thin lines and arcs between nodes in a STP represent bound on differences constraints. Arcs from a pair of nodes to a child STP represent repetitions. Arcs between any two nodes X and Y in a STP of the STP-tree are labeled by a pair [n,m] representing the minimum and maximum distance between X and Y. Sch, Ech, Smc, Emc, Spc, Epc, Sm, Em, Sp and Ep stand for the starting (S) and ending (E) points of chemotherapy, melphalan cycle, prednisone cycle, melphalan treatment and prednisone treatment, respectively.

Checking the consistency of a guideline

In order to check the consistency of the STP-tree, it is not sufficient to check the consistency of each node separately. In such a case, in fact, we would neglect the repetition/periodicity information. Temporal consistency checking, thus, proceeds in a top-down fashion, starting from the root of the STP-tree. Basically, the root contains a "standard" STP, so that the Floyd-Warshall's algorithm can be applied to check its consistency. Thereafter, we proceed top down towards the leaves of the tree. For each node *X* in the STP-tree (except the root), we progress as in the *STP_tree_consistency* algorithm:

1. temporal constraints on the incoming arcs are inherited by X;
2. STP+inherited temporal constraints are propagated by means of Floyd-Warshall's algorithm;
3. check that the propagated temporal constraints are consistent.

Property. The *STP_tree_consistency* algorithm is correct, complete and tractable.

Reasoning with the executions of the guideline

In the following, we report an algorithm for checking the consistency of the execution of a guideline with respect to its related guideline. In our work, as in most approaches to clinical guidelines, we suppose that one has *full observability* of instances (i.e., all the instances of actions which have been executed have been observed and inserted into the knowledge base), and that, for each instance, one knows the corresponding class of actions and/or repetition in the guidelines.

Algorithm *integratedConsistency*:

1. check that all and only the instances predicted by the guideline are present;
2. inherit the repetition/periodicity constraints and the temporal (non-periodic) constraints from the guideline to the instances;
3. propagate the temporal constraints via the Floyd-Warshall's algorithm;
4. check that the propagated temporal constraints are consistent.

Property. The *integratedConsistency* algorithm is correct, complete and tractable.

Architecture of GLARE

In order to enhance the generality of the temporal reasoning approach, the temporal reasoning algorithms are provided by a modular approach, in which a layered Temporal Server (TS) is loosely coupled with a guideline system (see figure 2). The clinical guideline system delegates temporal-related problems to the TS module. The core of TS is the *temporal reasoner* (TR), which implements the temporal reasoning algorithms and the related data structures. The *facilities layer* uses the two consistency-checking algorithms to provide the facilities (a)-(d) described below. Moreover, for acquiring and representing temporal information the *interface layer* may make use of visualization techniques.

Results

Exploiting temporal reasoning in clinical guidelines systems

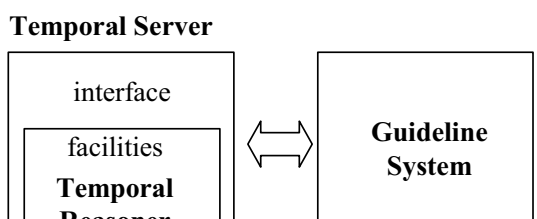


Figure 2 - Loosely coupled architecture to extend a guideline system with temporal reasoning facilities

In the previous sections, we have proposed a principled approach coping with issues (1)-(4). The adoption of our approach can provide computer-based guideline systems with crucial advances. In the following, we discuss several facilities that can be designed on the basis of our representation formalism and reasoning algorithms, both during guideline acquisition and execution. Although the

approach we propose is system-independent, in some cases we will exemplify it by sketching how we are planning to implement it in GLARE (GuideLine Acquisition, Representation and Execution) [9, 10, 11]. GLARE is a prototypical system to acquire and execute clinical guidelines, developed by the Computer Science Department of the Università del Piemonte Orientale of Alessandria (Italy) in cooperation with Azienda Ospedaliera San Giovanni Battista of Torino (the second hospital in Italy).

During **acquisition**, expert physicians (usually with the help of a knowledge engineer) represent a clinical guideline in a computer-based formalism, considering also the temporal constraints it contains. An automated ***consistency-checking*** facility is required to grant the temporal consistency of the guideline in a principled way. Such a facility can be provided through a call to the *STP_tree_consistency* algorithm, which can be advocated at any stage during the acquisition of a clinical guideline, so that incremental consistency checking is also possible. By default, consistency checking can also be executed at the end of each acquisition working session.

During **execution**, a guideline (e.g., the guideline for asthma) is applied to a specific patient, i.e., specific instances of the classes of actions in the guideline are executed. Our *integratedConsistency* algorithm can be exploited as the core of a user-oriented approach providing some crucial temporal facilities to user-physicians. Several of these facilities rely on the fact that our algorithm also provides the *minimal network* of the constraints between instances (considering also the constraints inherited from the guideline).

a) First of all, for scheduling purposes, it is important to provide a facility to assess when the next actions have to be performed, given the constraints in the whole guideline and given the time when the last actions in the guideline have been executed. The execution time of the next action(s) can be obtained through the steps 1-3:

1. retrieve the set of candidate next actions through a navigation of the control flow relations in the guideline;
2. apply the *integratedConsistency* algorithm to obtain the minimal network of temporal constraints;
3. retrieve the actions' possible execution-times from the minimal network (in the form of distances from the last-executed action, or from the origin of time).

By combining this facility with the query answering one (see facility (c) below), temporal reasoning can also be used in an interactive way to determine schedules which are consistent with the temporal constraints. For example, given a pattern $A_1, \ldots, A_n$ of actions in a guideline, temporal reasoning can be used in order to answer queries such as "*If I perform action A_1 today at 11 am, when will I have to perform $A_2, \ldots, A_n$?*", or "*Is it OK if I perform A_1 today at 12 am, A_2 at 6 pm and A_3 at 8 pm, and, if so, when will I have to perform A_4?*";

b) From the point of view of quality evaluation/assessment, it is important to provide a facility to check whether the temporal constraints in the guideline have been respected or not by the instances of actions that have been executed (considering also partial – i.e., ongoing – executions). Such a facility is directly provided by the *integratedConsistency* algorithm;

c) One of the main goals of guideline computer-based systems is to support decision making. In such a context, providing a (temporal) query-answering facility is a crucial task. Such a facility can be efficiently implemented on the basis of the minimal network provided by the *integratedConsistency* algorithm (along the lines discussed in [12]0), both to answer yes/no queries (e.g., "*is it correct to execute action A now, and action B within the next two hours?*") and/or to have in output the minimal distance between the instances of actions;

d) Still considering decision making, temporal reasoning can be profitably coupled with "simulation" computer-based facilities to see the temporal consequences of choosing among different alternative paths in a guideline. In particular, GLARE provides the "what if?" facility allowing physicians to discriminate among different alternatives of a decision by simulating the consequences of each choice, i.e., by visiting the paths in the guideline stemming from each one of the alternatives (see, e.g., [10]). Taking advantage of the *integratedConsistency* algorithm, our approach provides physicians with a way of comparing paths from the temporal point of view (i.e., in order to find the maximal and minimal temporal duration of each path). This facility can be obtained as follows:

1. for each path P_i to be compared
2. hypothesize the existence of an instance of each action in P_i which has not yet been executed;
3. apply the *integratedConsistency* algorithm;
4. retrieve the minimal and maximal duration of P_i.

Discussion and conclusion

In this paper, we propose a principled domain- and system-independent approach to the treatment of temporal constraints in clinical guidelines. We introduce a new representation formalism, coping with different types of temporal constraints, including constraints on (possibly periodic) repeated events. We devise two reasoning mechanisms to check the consistency of temporal constraints in clinical guidelines. Finally, we show how they can be exploited to provide clinical guideline systems with temporal reasoning facilities.

The approach by Duftschmid et al. [4]0 is the closest one to ours in the literature. With respect to such an approach, we propose an extended language to deal with repetitions (e.g., we cope with conditioned repetitions, through the '*while'* and '*onlyIf'* constructs) and we extend the basic STP framework, via the definition of the STP-tree and of the related constraint propagation algorithms. Finally, from the point of view of end-users, we also provide the (a)–(d) facilities discussed in the previous section. In particular, the treatment of the "a-posteriori" consistency between the temporal constraints in the guideline and of

the execution times requires several extensions both from the representation point of view (since a separate STP needs to be used) and from the algorithmic point of view (since new constraint-propagation-based temporal reasoning algorithms have to be devised).

Acknowledgments

We are very grateful to Prof. Gianpaolo Molino, Azienda Ospedaliera San Giovanni Battista, Turin, Italy and his group for their continuous advise and cooperation in developing and testing the GLARE system, and, specifically, the temporal approach described in this paper.

References

[1] Gordon C and Christensen JP, eds., Health Telematics for Clinical Guidelines and Protocols (IOS Press, Amsterdam), 1995.

[2] JAMIA, Focus on Clinical Guidelines and Patient Preferences, JAMIA 15(3), 1998.

[3] Peleg M, Tu S, Bury J, Ciccarese P, Fox J, Greenes RA, Hall R, Johnson PD, Jones N, Kumar A, Miksch S, Quaglini S, Seyfang A, Shortliffe EH, Stefanelli M., Comparing computer-interpretable guideline models: a case-study approach, J Am Med Inform Assoc.,10(1):52-68, 2003.

[4] Duftschmid G, Miksch S, Gall W, Verification of temporal scheduling constraints in clinical practice guidelines, Artificial Intelligence in Medicine 25(2): 93-121, 2002.

[5] Keravnou ET, Special issue: Temporal Reasoning in Medicine, Artificial Intelligence 8(3), 1996.

[6] Shahar Y, A Framework for Knowledge-Based Temporal Abstraction in Clinical Domains, Artificial Intelligence 90(1):79-133, 1997.

[7] Shahar Y, Musen MA, Knowledge-based temporal abstraction in clinical domains, Artificial Intelligence in Medicine 8(3):267-298, 1996.

[8] Dechter R, Meiri I, Pearl J, Temporal Constraint Networks, Artificial Intelligence 49:61-95, 1991.

[9] Terenziani P, Molino G, Torchio M, A Modular Approach for Representing and Executing Clinical Guidelines, Artificial Intelligence in Medicine 23:249-276, 2001.

[10] Terenziani P, Montani S, Bottrighi A, Molino G, Torchio M, Supporting physicians in taking decisions in Clinical Guidelines: the GLARE's "what if" facility, Proc. AMIA 2002, 772-776, 2002.

[11] Terenziani P, Montani S, Torchio M, Molino G, and Anselma L, Temporal Consistency Checking in Clinical Guidelines Acquisition and Execution: the GLARE's Approach, JAMIA, Special issue on the Fall AMIA Symposium, pp. 659-663, 2003.

[12] Brusoni V, Console L, Pernici B, Terenziani P, LaTeR: managing temporal information efficiently, IEEE Expert 12(4):56-64, 1997.

Address for correspondence

Address: Corso Svizzera 185, 10149 Torino, Italy
E-mail: anselma@di.unito.it

Chapter 6.

Improving Quality

This page intentionally left blank

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Information and Communication Processes in the Microbiology Laboratory – Implications for Computerised Provider Order Entry

Andrew Georgiou[a], Joanne Callen[b], Johanna Westbrook[a], Mirela Prgomet[b], George Toouli[c]

[a] *Health Informatics Research & Evaluation Unit, University of Sydney*
[b] *Faculty of Health Sciences, University of Sydney*
[c] *South Western Area Pathology Services, Liverpool, Sydney*

Abstract

The aim of this multi-method study based at a microbiology department in a major Sydney metropolitan teaching hospital was to: i) identify the role that information and communication processes play in a paper-based test request system, and ii) examine how these processes may affect the implementation and design of Computerised Provider Order Entry (CPOE) systems. Participants in this study reported that clinical information can impact on the urgency and type of tests undertaken and affect the interpretation of test results. An audit of 1051 microbiology test request forms collected over a three-day period showed that 47% of request forms included clinical notes which provide a variety of information often vital to the test analysis and reporting process. This transfer of information plays an important role in the communication relationship between the ward and the laboratory. The introduction of new CPOE systems can help to increase the efficiency of this process but for that to be achieved research attention needs to be given to enhancing the provision and communication of clinical information.

Keywords:

Evaluation studies, hospital information systems, laboratories, microbiology, pathology, qualitative research

Introduction

Computerised Provider Order Entry (CPOE) systems pose major challenges for hospital pathology laboratories [1], with important implications for a range of laboratory processes including inter-department functions, work organisation and laboratory effectiveness [2-5]. CPOE systems provide clinicians with the ability to enter orders directly into a computer [6]. The incorporation of functions such as clinical decision support and patient database linkage provide the potential to significantly impact on the quality of health care delivery leading to improved patient outcomes [7-9].

However, within the research and medical literature there has been relatively little attention given to the effect of CPOE on pathology laboratories [10]. These services play a crucial role in overall patient safety and outcome, accounting for an estimated 70% of all information used in decision making for admission, treatment and discharge [11, 12]. Pathology services are information intense units reliant on the efficient management and timely communication of relevant information to maximize the delivery of health care [13]. Moreover, the pathology department is comprised of a number of organisational structures and bodies each with its own unique and highly complex way of performing tasks and interacting with other departments. The aim of this study was to: i) identify the role that information and communication processes play in a paper-based test request system in the microbiology department; and ii) examine how these processes may affect the implementation and design of CPOE systems.

Methods

Design and research setting

A multi-method study (using quantitative and qualitative data collection techniques) was conducted in a microbiology department based at an Australian metropolitan teaching hospital. The department receives 131,000 microbiology test requests and specimens annually and employs 55 staff. It is divided into bacteriology, molecular biology, serology, virology, mycobacteriology and parasitology sections. The department is part of a pathology service involved in the introduction of Cerner Millennium PowerChart (Version 2004.01). The pathology service is responsible for a large metropolitan area and provides diagnostic services to a number of hospitals (including teaching hospitals) and clinics. This study forms part of a large multi-hospital research project evaluating the implementation of CPOE with data collection occurring from sites pre- and post- implementation.

Qualitative data relating to existing information and communication processes connected with test ordering and reporting within the microbiology laboratory were collected by observations, interviews and a focus group. Quantitative data collected from the microbiology department consisted of the volume of tests ordered and measured the presence of additional clinical information provided by doctors. The results provided a baseline indication of the existence of clinical information on test request forms by the requesting doctor.

Selection and sampling logic

The site was chosen because it was about to introduce an electronic ordering system that was mandatory for all inpatients. Qualitative data collection began with a focus group consisting of four laboratory scientists and one laboratory manager (n=5). These participants were chosen for their suitability (i.e., the department manager attested to their experience and knowledge of microbiology laboratory processes). The aim of the focus group was to discuss participants' views and expectations of the impact of a new CPOE system that was due to be introduced within the next three months. A set of semi-structured questions were used to gather impressions about the current system of paper-based test requests and what changes participants thought the new system would introduce. Participants were asked to raise both positive and negative features of the current laboratory processes. The participant base was extended using purposive sampling techniques [14], whereby participants directed us to other key informants. This increased the number of participants to eleven. Overall it included two senior laboratory scientists, one laboratory business manager, three technical officers and five laboratory scientists. Interviews were repeated with participants for clarification and further exploration of issues raised, with a total of 20 interview sessions conducted. This process provided a valuable feedback mechanism which enhanced our confidence in the validity of emerging themes [15]. One researcher conducted all the interviews and the focus group. Eight observation sessions were conducted by two researchers with each session lasting on average 1.5 hours (total of 12 hours of observations). One researcher conducted five of the observation sessions and the other conducted the remaining three.

Data collection

In the course of our analysis of the qualitative data, the research team undertook the collection of microbiology test request data. This provided the study with an important triangulation technique to investigate emergent themes [16]. Request forms were audited by one researcher over a three-day period between 29 June 2006 and 1 July 2006. No details related to patient identification were collected. The data collected included the number of test request forms received with and without the inclusion of clinical notes. For the purposes of this study clinical notes refers to patient specific clinical information written on the laboratory request form by the doctor requesting the test. Clinical notes therefore, may include: signs and symptoms; the site from which the specimen was obtained; medical history; physical examination; medications and the reason for the test request. One of the above pieces of clinical information was needed for the test request form to satisfy the criteria of "test request contained clinical notes."

The observations, interviews and focus group were conducted between August 2005 and October 2006. A letter outlining the study, its voluntary nature, the confidentiality of findings and participants, and a consent form, were provided to all participants. The research was approved by the University of New South Wales Human Research Ethics Committee and the relevant Area Health Service Research Ethics Committee.

Data analysis

The quantitative data were entered into SPSS (Version 12.0.1 for Windows 2004) and analysed using descriptive statistics. The focus group and one interview were recorded and transcribed. The remaining interviews and observations were recorded by the researchers in note form. The qualitative data were interpreted using a grounded theory approach [16] to derive themes that explained the information and communication processes within the microbiology department. Triangulation of analysis involved a number of iterative sessions involving a total of five researchers discussing and analysing the data: two who had collected the data and three others from the research team [14].

Results

The results are presented in two parts: firstly the qualitative data about laboratory information and communication processes related to the test ordering process, and secondly the volume, type and inclusion of clinical notes on microbiology test requests.

Laboratory information and communication processes related to the test ordering process – qualitative analysis

Three themes relating to information and communication processes surrounding test requests were identified:

- Theme 1: The context of the microbiology department
- Theme 2: Communication of clinical information
- Theme 3: Expectations of the new electronic ordering system

The context of the microbiology department

Participants explained that microbiology departments have their own specific requirements and needs that are not always applicable to other departments. For instance, the issue of timeliness has a particular context-dependent meaning for microbiology that is not identical to other pathology departments (e.g. biochemistry) for whom the optimisation of turnaround times for the processing and issue of results is an important performance indicator. Microbiology deals predominantly with diseases caused by infectious agents (e.g. bacteria, viruses, fungi and parasites) requiring time to grow before an appropriate test result is available.

The communication of clinical information

Participants highlighted the provision of relevant and appropriate clinical information by doctors ordering tests as a key area that directly impacts on their efficiency. In hospitals where electronic ordering has not been implemented this means the provision of a hand written test request form, including clinical notes, from the requesting doctor. If clinical information is not included the request may be judged to be incomplete or inadequate and in need of some form of follow up, often through direct telephone contact with the requesting doctor. This point was described by one participant in the following way:

> "As a whole the request that we receive, we need to know what the specimen is. We need to know what they want us to do with it, and it needs to be legible, so it really is an error, because we have to use our time to verify what they actually want." (Focus group participant)

Clinical notes are very important to laboratory staff. This is because they play an important role in setting the context for the test. Laboratory managers explained that this contextual information improves the laboratory's input. For instance, it may help a pathologist detect the need for more tests, or perhaps identify when a doctor may have asked for an inappropriate test.

> "They don't tell us what they want and we process what we think. If we didn't get the correct clinical details we may not necessarily make it up for the right thing…. (Focus group participant)

A salient example of this is for the disease tuberculosis, which the laboratory may not routinely test for unless it is either specifically requested, or when relevant clinical information is provided.

> "There are times when we process a specimen, then they [clinicians] ring up and say: have you done TB [tuberculosis] on this? We say – well you didn't ask for it. They should have given us the clinical details that would have allowed us to do that." (Focus group participant)

Expectations of new electronic ordering

The introduction of electronic ordering was expected to alter the way the department communicates with clinicians. In particular, laboratory personnel would not be required to decipher hand written notes anymore, which should eliminate instances of unclear or illegible requests. Participants described the potential of more effective exchange of valuable and relevant clinical information.

> "There should be some benefits to the laboratory, in that there will be less data entry, I guess. The patients' demographics etc, will come across. There should be less confusion, as to what tests are requested by the medical staff. We are hoping to get a lot more clinical details…" (Focus group participant)

The volume and inclusion of clinical notes on test request forms – quantitative analysis

The total number of microbiology specific tests requested within the 1,051 test request forms received were 1,078 as some request forms contained multiple test requests (Table 1). A large proportion of test request forms (47%) contained clinical notes documented by the clinician on the request form. The average number of tests requested per day, over the 3 day period, was 359 (range 338 to 374).

Table 1 – Number of microbiology specific tests requested with and without clinical notes

	No of tests (n=1078)	**With Clinical notes (n=506)**	
Day	n	n	%
1	374	186	49.7
2	338	146	43.2
3	366	174	46.9
Average	359.3	168.7	46.9

Table 2 highlights the results of the most frequently ordered tests. The most requested tests were urine cultures (35%) followed by blood cultures (21%) and specific site swab cultures (8%). The majority of urine culture requests (n=233[62%]) and blood culture requests (n= 142[62%]) did not contain clinical notes. However most wound culture requests (n=42[75%]) did contain clinical notes.

Table 2 – Frequency of a selection of the most ordered microbiology requested tests

Test categories	**No of tests (n=1078)**	**%**
Non-specific site swab cultures	27	3
Stool cultures	35	3
Infection control cultures	38	4
Fluid cultures	44	4
Sputum cultures	52	5
Wound cultures	56	5
Genital cultures	74	7
Specific site swab cultures	81	8
Blood cultures	230	21
Urine cultures	379	35
Others	62	6

Discussion

The comparison of results collected from the audit of microbiology test requests with the themes identified from the interviews and focus group session provides a means to triangulate different types of data, and encourages a better understanding of the meaning and significance of different findings. The results showed an important proportion (47%) of microbiology test requests received by the microbiology department contained some clinical notes

provided by clinicians. This indicates that clinicians often need to communicate further information to the microbiology department beyond simply identifying the test to be performed. There are cases, as in most blood and urine cultures, (which make up the bulk of tests requested), that do not contain any clinical notes. In some instances, (as suggested in the interviews and focus groups) this may require laboratory staff to follow up the missing information using telephone communication.

The translation of data into clinically meaningful information

The results highlight the role that the supply and processing of clinical information plays in the microbiology laboratory. The traditional format through which this information is communicated has been the hand written request form. Aside from their obvious clerical function, these forms also represent an important link between doctors and the laboratory [17] through which contextual patient data are communicated. This information can impact on the urgency, choice and even interpretation of pathology tests and results. The laboratory process involves the translation of data into clinically meaningful information. This role can be described as a core function of the laboratory service [18] and represents an important contribution to the patient care process [19]. For some commentators, such as Marques and MacDonald the absence of clinical information in certain situations can be misleading and even potentially dangerous [20].

Communication and the transfer of information

This study also demonstrates that the exchange of information across the hospital is a two-way process through which clinicians not only provide clinical information to laboratories, but also receive it back in an enhanced form. This relationship demonstrates the importance that communication plays in this process. The ordering process can be conceptualised as part of a collaborative effort of multidisciplinary groups [21]. For Toussaint and Coiera every information exchange is a communication act including a simple exchange between two people or even two machines [22]. Communication systems are important parts of the information structure [23] and can play a major role in the decision making process.

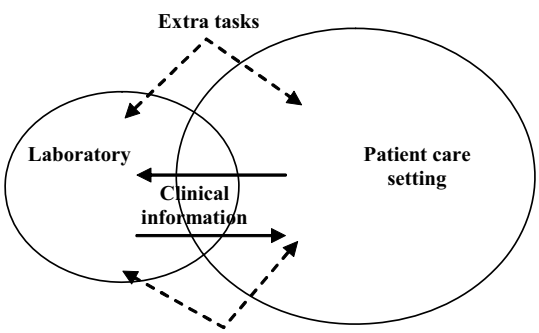


Figure 2 – The laboratory/ward information and communication relationship

Most information transactions within health services occur without the involvement of electronic data or systems [23], usually in conversations or through paper exchange. In reality, hospital communication systems form a collection of differing components and types [24]. Electronic systems like CPOE will impact significantly on existing channels and relationships [2]. The results of this study suggest that CPOE systems can be expected to enhance communication ensuring legibility and clarity in the ordering process and contribute to improvements in the clinical decision making process [4, 7]. However, there is also evidence that CPOE systems may disrupt previous channels of communication and replace them with inadequate or unsuitable alternatives often involving workarounds and extra tasks [25]. As Gorman et al. have stressed, incomplete or incorrect models of the process can lead to problems in the uptake and operation of CPOE [21]. Figure 2 depicts the importance of clinical information for the communication exchange between the laboratory and the clinical setting. The broken lines highlight the potential for this flow to be disturbed by design inconsistencies and barriers. It is important therefore that information and communication processes (at times unique to each hospital) are clearly identified as a means of maximizing the benefits of CPOE systems.

Limitations of this study

This study was undertaken at one site, using a microbiology laboratory department during the lead up to the implementation of a new CPOE system. The experiences of this one site will not be identical to other laboratory departments in other hospitals. Nevertheless the issues outlined will have wider resonance. The multi-method design adopted by this study has the advantage of providing rich contextual qualitative data about how the department's information and communication requirements are perceived along with descriptive data summarising the existing arrangements using hand written requests. This multi-method approach helped to enhance the findings and inform the discussion with participants. The results provide a useful evaluative framework with which to approach the question of CPOE implementation. But it also suggests the need to closely examine and quantify the impact different types of clinical information provided for different test requests can have on the laboratory process and their subsequent communication with doctors. While this task was beyond the scope of this study, it remains a natural area for follow up.

Conclusion

This study underscores the important role that the provision, processing and exchange of clinical information plays in microbiology laboratory processes. Clinical information helps to inform the laboratory of the type and urgency of tests required as well as assisting pathology staff to add interpretative value to the information provided back to medical staff. The exchange and transfer of clinical information is underpinned by a complex variety of communication channels within the hospital. New CPOE systems can increase the efficiency of this process and enhance the richness of the information exchange. To date, little attention has been provided to this issue. We recommend that more research into this area be undertaken so as to make these channels of communication and information exchange more explicit, and as a means of

providing information to enhance the design and implementation of CPOE systems.

Acknowledgements

This study was part of an Australian Research Council Linkage Grant project that funded the evaluation of the impact of information and communication technologies on organisational processes and outcomes. The project was undertaken with the collaboration of the New South Wales Health Department. The authors also acknowledge laboratory personnel for their participation in this study.

References

[1] Davidson EJ and Chismar W. Examining the organizational implications of IT use in hospital-based health care: a case study of computerized order entry. in Proceedings of the 32nd Hawaii International Conference on System Sciences. 1999. Hawaii.

[2] Stablein D, Welebob E, Johnson E, Metzger J, Burgess R, and Classen DC, Understanding hospital readiness for computerized physician order entry. Jt. Comm J Qual Saf 2003;29(7); 336-44.

[3] Massaro TA, Introducing physician order entry at a major academic medical center: I. Impact on organizational culture and behavior. Acad Med 1993;68(1); 20-25.

[4] Mekhjian HS, Kumar RR, Kuehn L, Bentley TD, Teater P, Thomas A, Payne B, and Ahmad A, Immediate Benefits Realized Following Implementation of Physician Order Entry at an Academic Medical Center. J Am Med Inform Assoc 2002;9(5); 529-539.

[5] Kuperman GJ, Boyle D, Jha A, Rittenberg E, Ma'Luf N, Tanasijevic MJ, Teich JM, Winkelman J, and Bates DW, How Promptly Are Inpatients Treated for Critical Laboratory Results? J Am Med Inform Assoc 1998;5(1); 112-119.

[6] Doolan DF and Bates DW, Computerized physician order entry systems in hospitals: mandates and incentives. Health Aff 2002;21(4); 180-8.

[7] Sittig D and Stead W, Computer-based physician order entry: the state of the art. J Am Med Inform Assoc 1994;1(2); 108-123.

[8] Abelson D, Connelly D, Klee G, Maag T, and Smith J, Technology Assessment: Computerized Physician Order Entry. 2001, Institute for Clinical Systems Improvement.

[9] The Leapfrog Group for Patient Safety, Fact Sheet: Computerized Physician Order Entry (CPOE) (2003) Available at: http://www.leapfroggroup.org/media/file/Leapfrog-Computer_Physician_Order_Entry_Fact_Sheet.pdf Accessed 2 March 2006.

[10] Georgiou A, Williamson M, Westbrook J, and Ray S, The impact of computerised physician order entry systems on pathology services: a systematic review (doi:10.1016/j.imedinf.2006.02.04). (In press) Int J Med Inf 2007.

[11] Forsman RW, Why is the laboratory an afterthought for managed care organizations? Clin Chem 1996;42(5); 813-6.

[12] Becich MJ, Information management: moving from test results to clinical information. Clin Leadersh Manag Rev 2000;14(6); 296-300.

[13] Review of NHS Pathology Services in England, Report of the Review of NHS Pathology Services in England 2006 Available at: www.dh.gov.uk Accessed 10 August 2006.

[14] Bowling A, Research methods in health. Open University Press: Buckingham; 1997.

[15] Guba E and Lincoln Y, Fourth generation evaluation. Sage: Newbury Park, CA; 1989.

[16] Glaser B and Strauss A, The Discovery of Grounded Theory: Strategies for Qualitative Research. Aldine Publishing Company: New York; 1967.

[17] Deeble J and Lewis-Hughes P, Directions for pathology. National Health Strategy Background Paper No 6. National Health Strategy: Melbourne; 1991.

[18] Travers E, Management of Information and its Economic Impact in Clinical Laboratory Management, E. Travers, Editor. 1997, Williams & Wilkins: Baltimore.

[19] Hardwick DF, "Directoring" and managing in a professional system. Modern Pathology 1998;11(6); 585-92.

[20] Marques MB and McDonald JM, Defining/measuring the value of clinical information. Clin Leadersh Manag Rev 2000;14(6); 275-9.

[21] Gorman PN, Lavelle MB, and Ash JS, Order creation and communication in healthcare. Methods Inf Med 2003;42(4); 376-84.

[22] Toussaint PJ and Coiera E, Supporting communication in health care. Int J Med Inf 2005;74(10); 779-81.

[23] Coiera E, When conversation is better than computation. J Am Med Inform Assoc 2000;7(3); 277-86.

[24] Coiera E, Communication Systems in Healthcare. The Clinical Biochemist Reviews 2006;27; 89-98.

[25] Georgiou A, Westbrook J, Braithwaite J, Iedema R, Ray S, Forsyth R, Dimos A, and Germanos T, When requests become orders - A formative investigation into the impact of a computerized physician order entry system on a pathology laboratory service (In press). (doi:10.1016/j.medin.2006.04.0020) Int J Med Inf 2007.

Address for correspondence

Health Informatics Research & Evaluation Unit
Cumberland Campus, University of Sydney,
PO Box 170 Lidcombe NSW 1825, Australia
Andrew Georgiou (Senior Research Fellow)
Telephone: + 61 2 9036 7331 Email: a.georgiou@usyd.edu.au

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Using an Accident Model to Design Safe Electronic Medication Management Systems

Farah Magrabi[1], Geoff McDonnell[1], Johanna I. Westbrook[2], Enrico Coiera[1]

[1]Centre for Health Informatics, University of New South Wales, Australia
[2]Health Informatics Research & Evaluation Unit, Faculty of Health Sciences, University of Sydney

Abstract

Large-scale implementation of electronic prescribing systems (e-PS) is likely to introduce at least some machine-related errors that will harm patients. We present a dynamic systems modeling approach to developing a comprehensive multilevel accident model of the process, context and task interaction variables which give rise to human error and system failure when e-PS are used in routine care. System dynamics methods are used to represent interactions between medication management processes and the context that is relevant to error generation, interception and transmission, agent-based methods represent task interactions. Capturing the patterns of failure within an accident model will facilitate an evidence-based approach to hazard analysis and design of e-PS features to improve patient safety. The model will have broad potential to guide the design, implementation and regulation of e-PS.

Keywords:

safety management, medical order entry systems, medication errors, computer simulation, systems theory.

Introduction

Medication errors are one of the most significant causes of iatrogenic harm and death within health care systems internationally [1]. Electronic medication management systems or prescribing systems (e-PS) are strongly recommended by health care bodies as an effective patient safety intervention to reduce medication errors [1-3]. Because of the widely reported potential benefits of e-PS in improving patient safety, quality of care and efficiency in healthcare delivery, these systems are central to health information and communications technology (ICT) implementation strategies worldwide [2, 3].

Safety of e-Prescribing Systems – an emerging problem in healthcare

Alongside their potential benefits, ICT systems can also be a source of harm to patients [4], and this has sparked urgent debate in the international Health Informatics community about the risks of harm associated with introducing this new technology. Several recent studies have identified the potential of these systems to have an adverse impact on patient safety [5]. For example, one study generating much controversy following its publication in December 2005 reported a significant increase in patient deaths following the introduction of electronic orders at the University of Pittsburgh Children's hospital (mortality rate increased from 2.80% to 6.57% post-implementation) [6, 7]. In another US study, Koppel et. al. found that computerization within a Pennsylvania hospital facilitated 22 types of medication errors [8]. The US Pharmacopoeia, a drug industry standards group which monitors patient safety, reported a steady growth in medication errors associated with e-PS [9]. Nearly 20% of these errors reported in 2003 were automation related.

Need for safety interventions

Evaluations of e-PS are increasingly identifying the need for patient safety interventions to minimize the impact of human error and system failure associated with ICT use in routine clinical tasks. For example, an evaluation of an e-PS within a Veterans Administration Medical Centre in Salt Lake City found that 93% of adverse events could be prevented through the implementation of a safety feature that would individualize safe dosage information for each patient [10]. In another investigation Horsky et. al. found that the absence of multiple system safeguards to check for the type of drug and the dose at successive stages of the medication process contributed to a serious error [11]. In primary health care, which has significantly higher rates of computerization than hospitals, Gandhi et. al. estimated that two out of three prescribing errors could be prevented with safety features that would provide dosage information [12]. These recent studies underline an urgent need to proactively engineer safer systems as the roll out of e-PS accelerates worldwide [13]. Safety is a neglected area as these systems are currently not regulated and there is little formal guidance to inform policy for regulating ICT in healthcare [14-16]. Efforts to investigate the safety of such systems currently rely on an ad hoc combination of methods to retrospectively reconstruct events which led to accidents [6, 8, 11].

In this paper we present a systems approach to studying the complex interactions which give rise to medication errors when e-PS are used in patient care. Developing an accident model, which describes the processes that may lead to failure when an e-PS is used in routine clinical work, should facilitate safer systems design. Such in silico

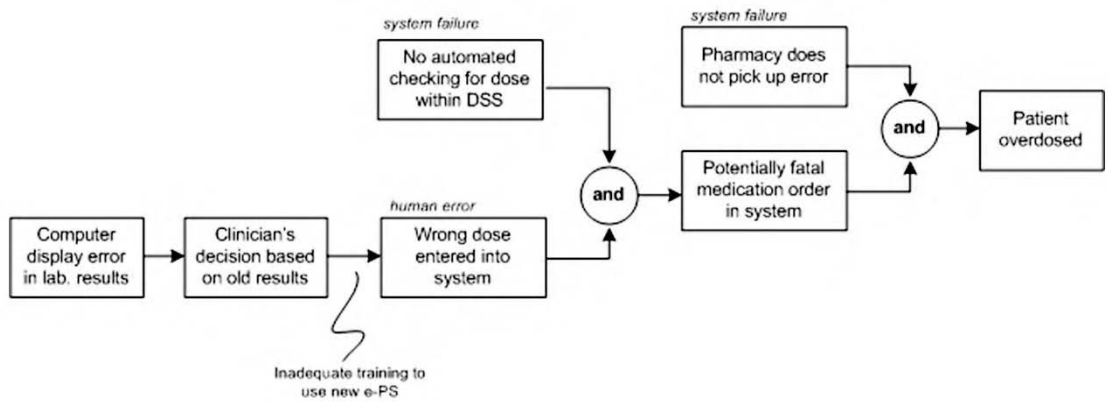


Figure 1 - An accident model for an e-prescribing system (based on case study findings in Horsky et.al. [18])

models provide a safe test bed for exploring complex feedback interactions, and allow system designers to simulate the impact of alternative designs in the clinical environment. Firstly we describe the nature of e-PS failures and review current methodologies to engineering safer systems. We then present an approach to developing an accident model to describe e-PS failures in silico.

The nature of e-PS failures

Efforts to engineer safer e-PS hinge on understanding how accidents occur. Electronic medication management systems support clinicians in complex tasks related to the treatment of disease in individual patients. For example, in the management of chronic disease such as asthma, medications are tailored to individual patients depending on stage of disease, allergies and side effects of drugs. Clinicians' decision-making styles vary within the different specialties and by professional cultures. A significant proportion of the fine-scale decision-making tasks undertaken by clinicians to treat individual patients are complex and cannot be fully standardized [17]. Consequently there is a high potential for human error when using computerized systems, which impose highly standardized human-computer interactions. In the case of medications management, pre-existing *human-human* interactions which impact *human-computer* interactions are particularly significant [18], therefore any accident model must capture task interactions critical to the process. Additionally, the context within which work is undertaken also influences user interactions with e-PS. A number of variables including professional culture and organizational factors give rise to variability in work practices when e-PS are used in a highly socio-technical [19, 20] environment such as healthcare, contributing to system failure. Therefore any effort to analyze medication errors must capture the process, context and task interactions associated with e-PS use.

Approaches to engineering safe systems

An accident model is a representation of the processes which give rise to failures within a system [21] (Figure 1). Such models are used to examine accident causing variables and their associated interactions. Capturing the underlying causes and patterns of failures within an accident model provides a basis upon which to identify hazards and to design interventions to prevent or reduce the impact of accidents. There are no accident models that specifically account for the failure of ICT systems in healthcare. Current efforts to improve safety rely on accident models developed in aviation, defense, nuclear power and the manufacturing sectors, which have pioneered research in systems safety. The Safety Science literature distinguishes three main types of accident models, sequential, epidemiological and systemic accident models [21].

Sequential accident models

The majority of traditional approaches to engineering safer systems are based on sequential models in which accidents are considered as the result of a sequence of events occurring in a specific order. Failure Modes and Effects Analysis is the most popular sequential approach for investigating large technical failures [22]. Although sequential models predominantly focus on technical failures, accidents resulting from human error are considered as any deviation from a standardized operation procedure, much the same way as a discrete technical failure. Probabilistic Risk Assessment [22] which examines the combinations of risk associated technical failures and human errors is being applied to model high-impact, low frequency iatrogenic injury events in medical care e.g. anesthesia patient risk. However, sequential models are inadequate for describing human errors in complex tasks and system failure where direct causal relationships between variables cannot be established e.g. organizational culture and policy.

Epidemiological accident models

Recognizing the difficulty in establishing direct causal relation-ships between events, epidemiological models describe accidents as the outcome of a combination of variables. In this approach accidents occur when a sufficient number of variables, some manifest and some latent, come together in space and time. Reason's Swiss Cheese Model is the best known epidemiological model in which system failures are identified within a causal sequence of events [23]. In com-

parison with sequential approaches, epidemiological approaches such as the Root Cause Analysis [22] consider more than just the proximate events that preceded an accident, therefore they are useful for investigating major accidents but are less useful for smaller adverse events in which system failure is difficult to isolate.

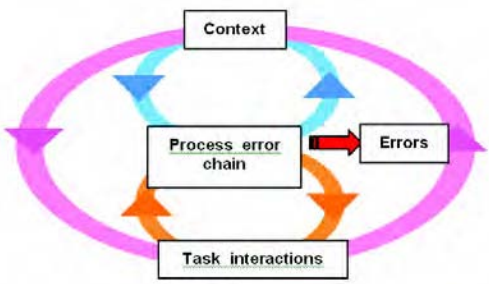


Figure 2 - A multilevel model of the process, context and task interactions which give rise to medications errors

Sequential and epidemiological accident models which examine discrete failure events and focus on causal relationships between variables have limited applicability in engineering safe e-PS. These techniques were designed to assess risks associated with discrete component failures in industrial systems such as a large process plant. In such tightly coupled technical sys-tems, human interaction with machines is broken down into well-defined sequence of events where operational procedures can be standardized. Human error is then considered in terms of deviations from standardized procedures. Healthcare work in contrast is complex, individualized, highly contextual and dynamic [17]. A significant proportion of transactions cannot be completely standardized and therefore human error cannot be fully represented within sequential and epidemiological models.

New systemic approaches to accident modeling

In comparison with sequential and epidemiological accident models, which focus on deterministic causal mechanisms of failure events, systemic models attempt to capture the dynamic behavior of a system. Accidents are viewed as emergent phenomena within complex systems. Hazard analysis in this approach consists of identifying and monitoring the sources of variability that give rise to human error and system failure resulting in accidents. The Functional Resonance Accident Model presented by Hollnagel et. al. is a simple systemic model based on stochastic resonance [21]. In this model, accidents arise from the functional couplings and their inherent variability within complex socio-technical systems. Highly variable hazardous interactions are identified so that barriers can be designed to control their variability and monitor performance. Although this model was specifically developed for highly technical aviation systems in which many of the control actions to maintain safety are automated, the availability of software tools to simulate the model make it a candidate for further evaluation.

Rasmussen's behavioral model is another systemic approach that could be useful for describing the inherent variability in work practices associated with e-PS [25]. This model describes a safety envelope within which a system must operate, defined by four key boundaries to (i) economic failure (ii) unacceptable workload; (iii) operational procedures and (iv) safety boundary beyond which accidents occur. It is essentially an overarching framework to integrate sequential and epidemiological models. Its strengths lie in the fact that it explicitly considers interactions among all levels of the organizational hierarchy to monitor and maintain safety.

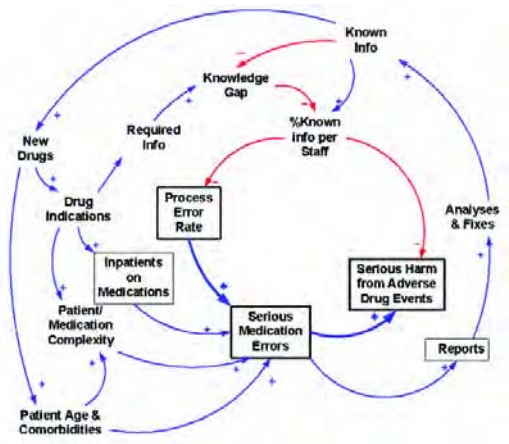


Figure 3 - A causal loop diagram of the process and context interactions associated with medication errors [24]

Perhaps the most sophisticated recent systemic model is that of Leveson which considers safety as a control problem [26]. In this model accidents result from inadequate control and enforcement of safety constraints on a system because (i) hazards are not known, (ii) control action is not adequate or the wrong action is performed, (iii) inconsistencies between process models used by the automation or human (mental models) and the actual process, (iv) missing or inadequate feedback. Similar to Rasmussen's model, this approach views accidents as consequences of socio-technical interactions among all levels of the organizational hierarchy that violate system safety constraints. But it is superior in terms of its formalization based on control theory, a very well developed discipline within Electrical Engineering. When applied to a complex aircraft collision avoidance system this method was found to be more complete than a fault tree analysis [27]. It was also effective in an investigation of a Canadian public water safety system revealing complex couplings between different levels of organization that were responsible for a negative safety culture and ignorance of basic science [27]. Systemic accident models that account for failures in highly automated industrial systems are not directly transferable to healthcare where automation largely supports humans in making complex decisions. However they provide a framework to examine complex relationships at multiple levels of the process, context and task interactions that contribute to e-PS errors.

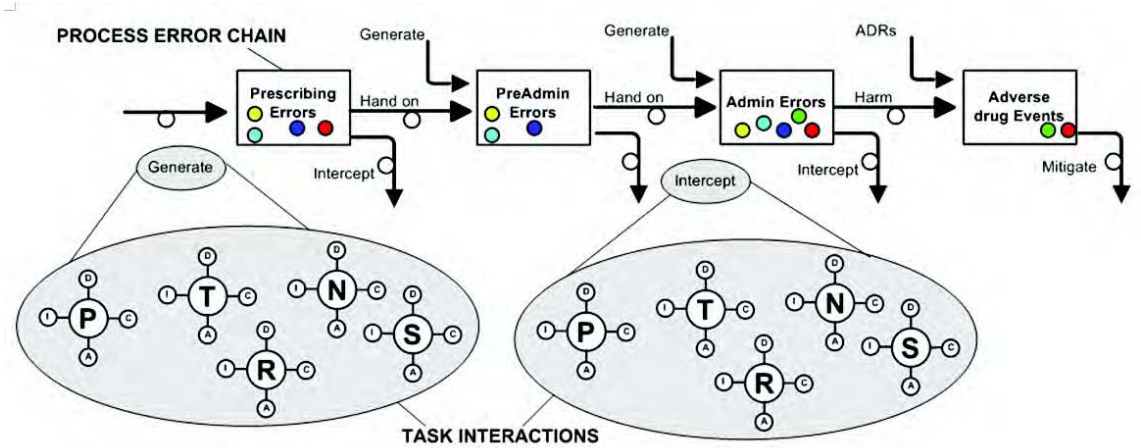


Figure 4 - Task interactions critical to error generation and interception, represented as agents and linked to a process error propagation and mitigation chain. Individual patient type errors can be tracked using patients

A multi-method multilevel accident model for e-PS failures

We propose to use dynamic systems modeling [28] to bring together a comprehensive multilevel model of the process, context and task interaction variables which give rise to human error and system failure when e-PS are used routinely in hospital inpatient care (Figure 2).

Process error chain & context interactions

Based on a systemic approach to accident modeling [26], difference equations and feedback interactions are used at an aggregated level to describe the pattern of interaction between medication management processes and the context that is rele-vant to error generation, interception and transmission. We employ a process chain of error events and harm, based on the framework developed by Bell et. al. [29] for e-PS impacts, and by Anderson et. al. [30] to represent medication error generation, interception and propagation within key prescribing activities e.g. order transmission dispensing, supply and ad-ministration. Bell's framework of e-PS impacts explicitly describes the inputs, outputs, resources and information required at each of the five steps of prescribing, transmission, dispensing, administration and monitoring within the process chain. Anderson's model specifically illustrates the sources of error within this chain. We capture the long-term organizational context that produces errors using a model developed by McDonnell et. al. [24] (Figure 3) which describes the complex interactions of patients and clinicians, information, medications, work practices and infrastructure and policies within the medications management process.

Process & task interactions

Agent based methods can represent task interactions critical to error generation and interception at the disaggregated indi-vidual level. Agents, such as clinicians and patients, are linked to the process error chain (Figure 4). Explicit rules of interaction guide the overall behavior of agents within the model, and errors become an emergent property of the modeled system, rather than being explicitly identified in the model. Key agent actions include informing, deciding, acting and communicating. For example, errors could be generated within prescribing tasks when (i) a specialist is not adequately informed and their decision is based on an old laboratory result, or (ii) an interruptive environment imposes cognitive loads resulting in a wrong decision, or (iii) from ambiguous computer screens that do not allow residents to enter orders correctly (action), or (iv) medications changes are not communicated in a timely manner to nurses. Individual patient errors are tracked by including patients within the model. The approach serves to identify error modes due to misperceptions of the situational context (informing role), cognitive errors (deciding role), task execution (acting role) and communication especially during transmission to downstream processes.

We will undertake a process to systematically assess validity of the model (a) conceptual validity using expert opinion, research evidence and study data (b) structural validation through formal qualitative inspections to confirm structure, parameters, extreme conditions and dimensional consistency. The model will also be examined quantitatively under extreme conditions to assess behaviour sensitivity, boundary adequacy and phase relationship (c) behavioural validity using behaviour pattern tests (d) simulation verification and (e) replication. The validation and verification of the simulation model will be based on Sterman's evaluation framework [28]. The model will be calibrated using data from the controlled experiments and observational studies.

Discussion & conclusion

There are no empirical accident models to provide a basis for engineering safe ICT systems for complex socio-technical systems such as healthcare. Sequential, epidemiological and systemic models developed within

defense, aviation, nuclear power and industrial applications are not directly transferable to healthcare, which is characterized by complex interactions that are not easily standardized. We have presented a framework for a systems approach to developing a novel accident model to describe e-PS accidents in healthcare. By bringing together a multi-level multi-method model of process, context and task interaction variables we will develop an accident model that accounts for human error and system failure mechanisms associated with complex fine-scale tasks unique to decision-making in dynamic environments such as healthcare. The novelty of this approach lies in combining the strengths of multiple systems and agent-based modeling methods within a simulation environment to examine complex interactions between multiple parameters otherwise not possible with traditional techniques. The next stages of this work will involve identification of variables to calibrate the model. We will then experiment with safety interventions to minimize the impact of human error and system failure on medication errors.

Acknowledgments

FM is supported by the Faculty of Medicine, University of New South Wales; JW is supported by an Australian National Health and Medical Research Council (NHMRC) CD grant.

References

[1] US Institute of Medicine: Crossing the quality chasm: A new health system for the 21st Century. USA: National Academy Sciences Press; 2000.
[2] National Health Service: Reports by the chief pharmaceutical officer. Building a safer NHS: Improving safety medication. London: National Health Service.
[3] Leap Frog Group. Computer assisted order entry fact sheet. Washington: Leap Frog Group; 2001.
[4] Ash JS, Berg M, Coiera E. Some unintended consequences of information technology in health care: the nature of patient care information system-related errors. J Am Med Inform Assoc 2004;11(2):104-12.
[5] EFMI. Bad Health Informatics Can Kill: Working Group for Assessment of Health Information Systems of the European Federation for Medical Informatics(EFMI): iig.umit.at/efmi/ [14 Nov 2006].
[6] Han YY, Carcillo JA, Venkataraman ST, Clark RS, Watson RS, Nguyen TC, Bayir H, Orr RA. Unexpected increased mortality after implementation of a commercially sold computerized physician order entry system. Pediatrics 2005;116(6):1506-12.
[7] Gesteland PH, Nebeker JR, Gardner RM. These are the technologies that try men's souls: common-sense health information technology. Pediatrics 2006;117(1):216-7.
[8] Koppel R, Metlay JP, Cohen A, Abaluck B, Localio AR, Kimmel SE, Strom BL. Role of computerized physician order entry systems in facilitating medication errors. JAMA 2005;293(10):1197-203.
[9] US Pharmacopeia: MEDMARX 5th Anniversary Data Report: A Chartbook of 2003 Findings and Trends 1999ñ2003; 2004.
[10] Nebeker JR, Hoffman JM, Weir CR, Bennett CL, Hurdle JF. High rates of adverse drug events in a highly computerized hospital. Arch Intern Med 2005;165(10):1111-6.
[11] Horsky J, Kuperman GJ, Patel VL. Comprehensive analysis of a medication dosing error related to CPOE. J Am Med Inform Assoc 2005;12(4):377-82.
[12] Gandhi TK, Weingart SN, Seger AC, Borus J, Burdick E, Poon EG, Leape LL, Bates DW. Outpatient prescribing errors and the impact of computerized prescribing. J Gen Intern Med 2005;20(9):837-41.
[13] Fox J, Das S. Safe and sound: artificial intelligence in hazardous applications /. California: AAAI Press/MIT Press; 2000.
[14] Coiera EW, Westbrook JI. Should clinical software be regulated? Med J Aust 2006;184(12):601-2.
[15] Australian Government Department of Health and Ageing Therapeutic Drugs Administration: www.tga.gov.au/devices/devices.htm. [cited Oct 24]
[16] US Food and Drug Administration: Regulatory Requirements for Medical Devices: www.fda.gov/cdrh/manual/510kprt1.html. [cited Oct 24];
[17] Bar-Yam Y. Making things work: solving complex problems in a complex world. Newton: NECSI-Knowledge Press; 2004.
[18] Coiera E. Interaction design theory. Int J Med Inform 2003;69(2-3):205-22.
[19] Berg M. Patient care information systems and health care work: a sociotechnical approach. Int J Med Inform 1999;55(2):87-101.
[20] Coiera E. Four rules for the reinvention of health care. BMJ 2004;328(7449):1197-9.
[21] Hollnagel E. Barriers and accident prevention. Aldershot: Ashgate; 2004.
[22] Dhillon BS. Human reliability and error in medical system. Singapore: World Scintific Publishing; 2003.
[23] Reason J. Human Error. Cambridge: Cambridge University Press; 1990.
[24] McDonnell G, Heffernan M. The Dynamics of Hospital Medication Errors: A Systems Simulator Testbed for Patient Safety Interventions. In: 23rd International Conference of The System Dynamics Society July 17 - 21, Boston.; 2005.
[25] Rasmussen J. Risk management in a dynamic society: a modelling problem. Safety Science 1997;27(2/3):183.
[26] Leveson N. A new accident model for engineering safer systems. Safety Science 2004;42(4):237-270.
[27] Leveson N. Model-Based Analysis of Socio-Technical Risk: Technical Report, Engineering Systems Division, Massachusetts Institute of Technology.; 2002 June.
[28] Sterman JD. Business Dynamics: Systems Thinking and Modeling for a Complex World. Irwin: McGraw-Hill; 2000.
[29] Bell DS, Cretin S, Marken RS, Landman AB. A conceptual framework for evaluating outpatient electronic prescribing systems based on their functional capabilities. J Am Med Inform Assoc 2004;11(1):60-70.
[30] Anderson JG, Jay SJ, Anderson M, Hunt TJ. Evaluating the capability of information technology to prevent adverse drug events: a computer simulation approach. J Am Med Inform Assoc 2002;9(5):479-90.

Address for correspondence

Centre for Health Informatics,
University of New South Wales,
Sydney 2052, Australia.
Email: f.magrabi@unsw.edu.au

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Securing Chemotherapies: Fabrication, Prescription, Administration and Complete Traceability

Stéphane Spahni[a], Christian Lovis[a], Monique Ackermann[b], Nicolas Mach[c], Pascal Bonnabry[b], Antoine Geissbuhler[a]

Service of Medical Informatics, Department of Radiology and Medical Informatics[a] Pharmacy[b,] Service of Oncology, Department of Internal Medicine[c] University Hospitals of Geneva, Switzerland

Abstract

Decision support, order entry, drug and care administration with their respective documentation cannot be seen as independent actions, especially in term of medical approach and patient safety. Chemotherapy treatment is a good example to illustrate the various implication of technology information in these multifaceted and intricate processes. Chemotherapy administration can be a highly complex process. It can take place over a variable period of time, from hours to months. Chemotherapies can be produced specifically for a given patient and can have dramatic effects in case of errors. Chemotherapy treatment will depend from various information including patient specific data such as temperature, weight or laboratory or drug specific knowledge such as side effects or administration directives. At the University Hospitals of Geneva (HUG), processes reengineering accompanied with new applications covering the whole chain of the processes involved in chemotherapy treatment (prescription preparation, administration, control) have been developed. This paper presents the overall approach leading the computerization of these processes and the first evaluations about the potential benefits of the computer-aided controls during the administration phase.

Keywords:

hospital information systems (HIS), GS1, RFID, pharmacy, chemotherapy.

Introduction

The need for increased safety and efficiency in care production is an important goal in healthcare [1]. Optimization of care production and a better efficiency of care logistics while improving quality and safety is however a challenging goal. Information technologies allow to meet both goals, increasing both efficiency and safety [2, 3]. However, reaching these goals implies to embed information technologies at all steps of the process, to identify clearly all actors and objects involved and, usually, to reengineer deeply these processes, the associated procedures and the way the are (often historically) done. To illustrate this, we present the developments made around chemotherapy treatments at the HUG.

The HUG is a consortium of hospitals in four campuses and more than 30 ambulatory facilities in the state, comprising more than 2,000 beds, 5,000 care providers, over 45,000 admissions and 750,000 outpatients' visits each year. It covers the whole range of in- and outpatient care, from primary to tertiary facilities. The HUG is the major public healthcare facility in the Geneva region and the near France. The in-house developed computerized patient record (CPR) is used in all facilities and runs on more than 4,500 PC's. Over than 20,000 records are open every day by more than 4,000 care providers from all functions [4].

Chemotherapeutic agents are prescribed for various oncological and haematological disease states. Although they are considered to be the treatment of choice for many cancers, these medications are associated with serious and potentially life-threatening side effects. The toxicities of these anti-cancer drugs, the multidisciplinary actors involved in the whole treatment process, create a very high risk of devastating medical errors. In an effort to minimize the potential for chemotherapy-related errors, the HUG have spent the last years performing important developments and prospective risk analysis to lower the criticality of the whole process [5]. As part of the results, the preparation of all chemotherapies administered at any of the facilities of the HUG has been centralized at the pharmacy. The centralization guarantees the processes, the quality and the traceability of the preparations and their components and, very important, the safety of the operators. This first phase allowed justifying the need and the pertinence of a unique description of substances used and their associated protocols. One major task was the creation of a global database containing all substances, materials and description of procedures that can be used for in-house preparations. This could not be done without reaching consensus between all actors involved. Building the protocols is one of the very complex tasks that have been made in close collaboration with and under the supervision of specialists. This database can then be used by the applications developed for requesting new preparations, managing the pending requests, organizing the concrete preparations and managing the traceability before, during and after the production process.

The prescription side

At the prescription side, we developed a family of tools including components to help the creation and management of specific protocols, to use them for given patients, but also to help the follow-up of all patients getting chemotherapies. With the correct right accesses, users can see the list of patients and their respective chemotherapy protocols. For each patient, one can see various information, such as the protocol type, the number of cycles involved, the first day and the next day of chemotherapy. These features help oncologists to follow their patients, the current status of all running protocols, but it is also a powerful tool for evaluation of the treatment, the prevention of side effects and the efficiency of this prevention.

Figure 1 - chemotherapy protocol for a given patient

At the time of prescription, physicians can choose the appropriate protocol and use it with the data pertaining to the correct patient. The system will help adjusting doses and chronology of procedures, as well as all elements pertaining to the selected chemotherapy regimen, including the fabrication of the chemotherapy if required (see Figure 1).

The system can produce alerts in some situations, for example when overlapping dates and regimens. Other alerts might be added in the future, such as warning if the renal function is worsening, or the white blood cells counts too low.

The pharmacy side

In many cases, the regimen associated with a protocol does not exist per se commercially, but must be produced specifically by the HUG's central pharmacy.

Figure 2 - the main page of the pharmacy system for drug production

The tools developed for the pharmacy allow the management of most of the logistics needed to produce drugs from raw substances. This includes reception of the raw substances and their identification/validation, tagging, stocking and localization, lot follow-up and usage. They also ensure a complete traceability of all actions on these substances or their derivates up to the final product preparation and its distribution. The production of products must follow strict production protocols - distinct from prescription protocols – that includes materials, procedures, validation steps and safety behaviours. This production includes specific chemotherapy treatments, but also a wide range of in-house made products, such as anti-cough syrup or disinfectants, amongst others. The computerized production protocols can be adapted according numerous parameters, ranging from patient specific data to mass production variables. Drugs validation and laboratory analyses are often made both at the level of raw substances and on the final product (Figure 2). These analyses ensure the quality of the raw components and of the final product used for patient care. As long as a product, raw or final, has not been validated, it is kept in "quarantine" and cannot be used. The system allows following of all lots, production and expiration dates, suppliers, end users, remaining stocks, and the position of the product in the production workflow (Figure 3). It ensures standards in the production chain and increases operators' safety. In parallel, a computerized high precision balance aiming at reducing the risk of errors during the fabrication is used by a specific tool for supporting the fabrication of specific preparations (Ca[1]to®). This tool automatically gets the various volumes of the preparations to be realized. These

data are used to ease the validation of each step by controlling the plausibility of the result.

Figure 3 - Example of workflow and stock management

Drug administration – the nursing side

One of the complex tasks of computerization of this process involves the administration part. Before the chemotherapy is administrated, complete instruction will be printed in the ward for nurses, including information about drugs, sides-effects and their prevention such as nausea, and precautions for the administration itself. These precautions will cover many aspects, such as solutions incompatibilities, administration temperature, administration route or protection against light (Figure 4).

The last step in the process is the bedside traceability of the administration of the drug to the patient. In order to be able to properly trace all steps of administration, it has been necessary to tag all infusions and actors, the nurses involved and the patient. This phase is currently ready for its initial deployment in a pilot ward, but is delayed due to logistics problems with the labels. A solution based on Pocket PC is being realized. The goals of this module are the following:

- Verification that it is the right preparation for the right patient;
- Automatic control that the preparation is still administrable (used-by date);
- Verification at the bedside that no stopping condition have been issued meanwhile;
- Complete traceability up to the “last mile”.

These functions require the correct identification of the three “partners” involved: the patient, the nurse and the preparation. While there are many ways to identify the partners, proprietary or non proprietary, we decided to start using international numbering of objects and selected GS1’s coding schema (GS1 is the result of the merge of the two organizations EAN – European Article Numbers and UCC – Uniform Code Council). All three partners have therefore their own code that can be read by the Pocket PC and validated on-line according to the information stored in the Hospital Information System (HIS): the patient and the nurse have their own permanent GS1 codes and each preparation produced by the pharmacy receives an unique identification code enabling the tracking in the institution. All GS1 codes are using the GS1-128 encoding scheme.

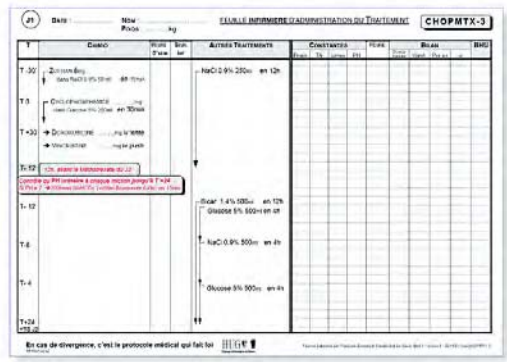


Figure 4 - Administration directives

The GS1 encoding can be used on various transporters, such as linear barcodes, data matrix, radio-frequency tags, etc. The problem with the linear barcodes comes from the size of the printed code: 1 dimensional formats are rather large, and data matrix are not commonly read by classical readers. The labels on infusions and other preparations are devoted to human-readable information, therefore reinforcing the ability for human to cross-check information (Figure 5). For this reason, we decided not to use printable codes. To achieve this, the use of labels with integrated RFID chips has been chosen for the electronic identification of the preparations as well as the patients (using a wristband with a RFID sticker) and the nurses.

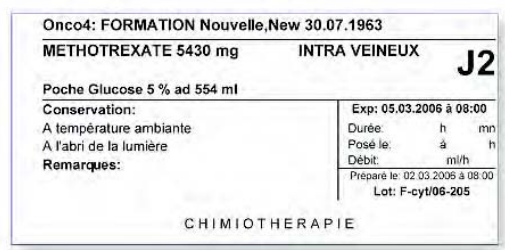


Figure 5 - Self adhesive label with a radio-frequency tag (RFID)

The labels for preparations are printed at the pharmacy before production and joined to the raw material required for the fabrication, while the labels for patients will be produced at the admission desks (they are currently produced at the ward until the trial is complete). This solution enables the use of a unique reader at the bedside for safely getting the identities of the patient, the nurse and the preparation to be administered. The RFID reader is used with a PDA (HP iPaq) with wireless LAN connection to the CPR. Therefore, validation check can be made in real-time and at the bedside. The same device is also used after the start of the administration for e.g. controlling the current administration status after the staff rotation, for indicating

1 Computer Aided Therapy for Oncology (cato). http://www.cato.at

a possible suspension of the treatment and finally for registering the end of the administration.

Figure 6 - The overall administration process at the bedside

Initial evaluation

Evaluation of the impact of the introduction of such a project is important in many respects – benefits in patient safety, acceptance by the users, impact on the procedures and workload, etc. Recent work of Bates [6] shown that the origin of errors in drug prescription and administration are distributed as follows: prescription 49%, re-transcription 11%, preparation 14% and administration 26%. For patient safety, considering that if about 50% of the errors during the prescription, preparation and dispensing phase are caught only 2% of the administering and documenting errors are detected [5, 7], having a reliable evaluation of the impact is crucial. A study has therefore been conducted in order to evaluate the safety benefits and the acceptance of the use of PDAs during the administration phase [8, 9], addressing thus specifically the 26% of drug dispensation errors. This study has yet only been realized "off-site" with 62 users (specialized and non-specialized nurses as well as assistants at the pharmacy). The study was oriented in order to measure the following aspects:

- Benefits of the computer-assisted controls against a fully manual control with or without a paper-based check-list (7 checkpoints: identity of the patient, name of the substance to administer, dose, administration path, date and day of administration use-by date taking into account administration duration, conservation);
- Acceptance of the system by the future users;
- Getting a feed-back regarding the ergonomics of the application.

In order to evaluate the respective benefits of the two helps (checklist, PDA), participants had to check two sets of protocols containing randomized errors (2% and 16% of errors respectively). The results were the following [9]:

Non-specialized persons:

	No assistance	Checklist	PDA
Detection [%]	83.5 %	97.5 %	100 %
Trust interval	78.9 – 87.4	95.1 – 89.9	
Test duration [min]	32	18	8

Specialized persons:

	No assistance	Checklist	PDA
Detection [%]	90.1 %	100 %	100 %
Trust interval	85.7 – 93.6	87.2 – 100	
Test duration [min]*	32	18	8

* No significant differences in the mean test duration were noted between the two categories of persons.

As one can see, the results of the study are very promising for the future use of the new application:

- While the use of the check-list already significantly reduces the number of undetected errors, it is only with the PDA that non-specialized staff detected 100% of the problems;
- The time spent for doing the controls is reduced by a factor of up to 4 when using the PDA.
- Between 73% and 80% of the persons preferred to use the PDA in various situations while they are only 13% to 20% preferring to use the check-list.

We are therefore very confident that the new system will bring a valuable contribution to patient's safety and daily work of nurses. Further evaluations will of course be performed after the whole system is in production.

Discussion and conclusion

Implementing such a system is a challenging goal, merging virtual and real worlds. It involves numerous actors and cannot be achieved without managing carefully organisational impact. Analyzing the complete process of chemotherapy treatment, from the design of a protocol, to the prescription, the production, the administration and the follow-up as a coherent and shared workflow is an important paradigm in improving both safety and efficiency of the process.

On the clinician side, this process has initiated a global formalization and sharing of protocols and guidelines. Except special cases, the prescription of drugs regimens not agreed within protocols is not more possible. Only adjustments pertaining to patient characteristics are allowed; doses are then validated according to several variables that cannot be over

passed. The prescription has gained in transparency and readability. In the pharmacy, benefits are numerous, from better management of raw substances and traceability to increased safety and standardization of operators work. For nurses, validation of the administration and confidence in the overall process should be increased when the final phase will be introduced. By now, there is already a clear benefit due to the standardization, the completeness and the readability of treatment directives.

The system has been well accepted by all actors involved. The importance of the quality of the information is well recognized, and the tools do offer a significant improvement. However, it must be emphasized that the formalization and the validation of all processes, including each protocol, require a significant amount of time, especially from oncologists. But we now have reached the state where it is the oncologists who do not have the new applications who are strongly pushing for getting it (at the cost of formalizing and standardizing their protocols).

The initial project for the support of the prescription and production of chemotherapies has been successfully deployed and is well accepted by its users. The second phase, the bedside validation of the administration of the medications to the patients is in development and should be used in a pilot ward within a few months. We expect to have a significant improvement to the overall security of the care of the patients thanks to several achievements: better documentation in the patient record, suppression of the hand-written orders, controls at the prescription, suppression of multiple retranscription, production and administration levels, etc. The use of Pocket PCs and radio-frequency technologies at bedside is expected to grow as real-time application and complete traceability are being progressively installed.

The patient will be the final beneficiary of the system, getting a globally improved process with increased safety and traceability. Further studies will be held once the bedside administration validation system will be in production.

References

[1] ITAC. Report to the President, Revolutionizing Health Care through Information technology. President's Information Technology advisory Committee June 2004.

[2] Bates DW, Cohen M, Leape LL, Overhage JM, Shabot MM, Sheridan T. Reducing the fre-quency of errors in medicine using information technology. J Am Med Inform Assoc 2001;8(4):299-308.

[3] Ruland CM. Improving patient safety through in-formatics tools for shared decision making and risk communication. Int J Med Inform 2004;73(7-8):551-7.

[4] Geissbuhler A, Lovis C, Lamb A, Sphani S. Ex-perience with an XML/HTTP-based federative approach to develop a hospital-wide clinical in-formation system. Medinfo 2001;10(Pt 1):735-9.

[5] Bonnabry P, Cingria L, Ackermann M, Sadeghi-pour F, Bigler L, Mach N. Use of a prospective risk analysis method to improve the safety of the cancer chemotherapy process. Int J Qual Health Care 2006;18(1):9-16.

[6] Bates DW, Cullen DJ, Laird N et al. Incidence of adverse drug events and potential adverse drug events. JAMA 1995; 274: 29-34.

[7] Bates DW. Using information technology to re-duce rates of medication errors in hospitals. BMJ 2000; 320: 788-91.

[8] Bates DW, Nathan S, Cullen DJ et al. The costs of adverse drug events in hospitalized patients. JAMA 1997; 277: 307-11.

[9] Baalbaki R. Mesure de la fiabilité du contrôle ul-time avant l'administration des médicaments aux patients. Diploma work presented at the Science Faculty of the University of Geneva, Switzer-land, July 2006.

Address for correspondence

Stéphane Spahni
Medical Informatics Service
Radiology and Medical Informatics Department
University Hospitals of Geneva
24 rue Micheli-du-Crest
CH-1211 Geneva 14
Email: Stephane.Spahni@hcuge.ch

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Multitasking by Clinicians in the Context of CPOE and CIS Use

[a]Sarah Collins, [a,b,c] Leanne Currie, [c]Vimla Patel,
[a,c]Suzanne Bakken, [c]James J. Cimino

[a] *Columbia University School of Nursing, New York, New York, USA*
[b] *New York Presbyterian Hospital, New York, New York, USA*
[c] *Department of Biomedical Informatics, Columbia University, New York, New York, USA*

Abstract

Interest in studying distractions and interruptions in the context of clinician workflow has increased in light of evidence that these events may negatively impact patient safety. Additionally, many recent informatics-based studies analyze computer provider order entry (CPOE) and clinical information system (CIS) implementation and its effects on clinician workflow. This study expands the development and use of a taxonomy to characterize distractions to clinicians and their subsequent actions in the context of CPOE/CIS use. We found a total of 75 distracting events in 406 minutes of observational data of which 32 led to interruptions and 30 led to continued multitasking. The above primary actions led to 5 tasks not completed and 4 episodes of clinician's lack of recall. Differences in the distribution of the source of distractions and primary action of the distracted clinicians may be a function of clinical setting and clinician type using the CPOE/CIS. Nine secondary actions, potentially resulting in a slip or a mistake, suggest that CPOE may necessitate different forms of safety nets than traditional clinician communication.

Keywords:

communication, distraction, interruption, computer provider order entry, clinical information system

Introduction

Multitasking is a valued skill in the clinical setting allowing for the efficient execution of daily activities, yet, may fail to be an effective mechanism in clinical practice when it leads to cognitive overload [1]. Clark and Brennan [2] noted that the properties of a communication medium impose constraints on the communication process; therefore, the design of communication modalities should account for these constraints [3].

During coordinated activities, such as communication amongst the health care team, responsible clinicians must establish common ground. Common ground is defined as similar experiences, beliefs, and knowledge and is necessary to ensure that clinicians' mental models reflect each others' needs within the context of the task and current situation [4]. Coordinated communication may serve as a "rescue" to account for the fact that the introduction of each new interaction can detract from the clinician's finite cognitive resources [4]. When cognitive resources are exhausted, the average amount of attentional resources, also known as working memory, available to any single interaction may be reduced [5].

Computerized Provider Order Entry (CPOE) systems attempt to alleviate clinician cognitive overload through providing an organized electronic format for work activities. CPOE systems can have several types of decision support functions such as automated alerts and reminders; however, excessive alerts have been shown to cause cognitive overload and alert fatigue [6]. Alert fatigue occurs when the number of reminders and alerts are perceived too be excessive. This number often varies from clinician to clinician and may cause clinicians to override both critical and non-critical alerts, compromising the desired safety effect of integrating decision support into CPOE [6].

The benefits of CPOE implementation likely outweigh the unintended consequences [7]. However, due to the asynchronous channel's inability to form a mental model of the cognitive needs of the clinician, new ways to detect distractions, interruptions and multitasking in the setting of CPOE and general clinical information system (CIS) usage are needed. Furthermore, other clinicians within the same workspace may be equally vulnerable to similarly reduced attentional resources. This study describes the use of a taxonomy to characterize and analyze distractions and subsequent actions in the setting of CPOE/CIS usage.

Background

Studies characterizing interruptions and distractions during clinician workflow continue to demonstrate their prevalence and significance to the health care work environment's culture [8,9]. Interruption rates consistently approach 30% of all clinical communication, in many settings. On the other hand, interruptions serve as the most frequently used method of communication in the health care environment and are thus considered a beneficial activity [8]. Of great concern however, up to 43% of medication errors have been attributed to distractions such as interruptions [10]. CPOE systems show great promise in their potential to increase patient safety and have been

Selected for best paper award.

shown to reduce medication errors up to 81% [11]. However, unexpected "silent errors", i.e., latent errors resulting from mismatches between clinician workflow and CPOE or CIS development and deployment, have emerged as potentially contributing to medical errors [7].

Table 1 – Taxonomy of distractions during CPOE use

Distraction with multitasking (Event): Period when a distraction causes a subject to interact in two or more concurrent communication events [9].

Primary Actions	Definition	Examples				
		User	**Source**	**Original Task**	**Distraction Quote or Episode**	Secondary **Action**
Interruption (I)	Cessation of productive activity before the current task was completed for an externally imposed reason [13].	MD	*C*	Ordering urine electrolytes	Difficulty navigating CPOE order process; sought assistance from second clinician but was unsuccessful	*IT*: User stopped after 1 minute and 5 seconds of unsuccessful attempts.
Continued Multitasking (CM)	Continued interaction in two or more concurrent communication events or tasks.	MD	*SC*	Ordering Mucomyst medication	"Is that the patient's chart you have?"	*LR*: Difficulty recalling Mucomyst dose previously stated by another clinician.
Deferred Task (DT)	Acknowledged distraction stimulus from external source that was not followed by cessation of original task or continued multitasking.	MD	*SC*	Ordering patient laboratory draws	"Did patient X ever get the CT and biopsy?" (different patient than active patient record on CPOE screen)	*IT*: User acknowledged distraction with indication to follow-up, yet follow-up was not observed.

Secondary Actions	**Definition**
Lack of Recall (LR)	Inability to quickly recall previously verbalized information about the task (could result in slip or mistake [16]).
Incomplete Task (IT)	Either original task and/or interruption task not completed during the observed session (could result in slip or mistake [16]).
Change in Plan of Care (CPC)	Information of a new plan interrupts a task and causes the interrupted task to be irrelevant and discarded (could result in slip or mistake [16]).
Source of Distraction Event	**Definition**
Information Need (IN)	Required data to answer a question is not known.
Synchronous Communication (SC)	Two parties exchange messages across a communication channel at the same time (e.g. face-to-face, telephone) [17].
Computer (C)	Distraction caused by technical problems (e.g. frozen user interface) or CPOE/CIS usability difficulties.

In an earlier study, we developed a taxonomy to characterize distractions and interruptions during the use of CPOE in a medical intensive care unit (MICU) [12]. The taxonomy extended the work of Coiera et al.'s definition of distraction with multitasking [9] and Flynn et al.'s definition of an interruption [13]. Our previous study found that information needs accounted for 55% of the distraction events detected and clinical communication accounted for 40% [12]. Distraction events may be attributed to more than one source. For example, a clinical communication may be related to an information need. Additionally, 14% of distractions were attributed to computer problems. For example, while the observed MICU resident was changing a CPOE oxygen order, another clinician distracted him with an information need. In this instance, the second clinician asked for information that the first clinician needed to obtain from the CIS. The distraction interrupted the resident, the resident answered the question, and then lack of recall of the original intended oxygen order occurred for 1 minute and 6 seconds [12].

In a second study, as part of the Infobuttons project at Columbia University, 51% of clinician information needs went unmet [14]. Consistent with Coiera et al.'s findings that clinicians use synchronous channels of communication more frequently, Currie and colleagues found that for

76% of information needs concerning domain-related questions, the individual in need of the information utilized a person to answer the question (rather than a paper or computer-based resource) [14]. The health care setting, rich with information needs and with an apparent preference for face-to-face communication sets up an interrupt-driven environment [15] which has previously been shown to both compromise patient safety [10] and to provide an opportunity for coordinated activity "rescues" [4]. However, to date, these phenomena have been studied in the absence of CPOE/CIS, therefore this study characterizes and analyzes interruptions and distractions to clinicians in the context of CPOE/CIS use.

Methods

Morae ™ software served as the portable usability lab to collect all observational data during randomly selected periods of time [14]. We used a Taxonomy of Distractions During CPOE use which was developed for earlier work on this topic to characterize distraction and interruption events during observational data of medical resident rounds in the MICU (See Table 1) [12]. We developed the taxonomy by iteratively using deductive and inductive methods to characterize our observational data. This was congruent with the hybrid method to categorize interruptions and activities as described by Brixey et al. in the HyMCIA study [18]. For this study, we used the taxonomy and extended our sample to residents and other clinician's who were using CPOE/CIS on inpatient medical/surgical floors collected for the Infobuttons project at Columbia University Medical Center (CUMC).

Table 2 – Time and events by type of clinician for combined rounds and non-rounds data

Clinician	Minutes (%)	#Distraction	Events/Hr	Interruption (%)	Continued	Multitask-Deferred Task (%)
MD MICU Rounds	93 (23)	22 (29)	8	10 (31)	3 (10)	9 (69)
MD Non-rounds	158 (39)	41 (55)	28	13 (41)	25 (83)	3 (23)
RN Non-Rounds	94 (23)	8 (11)	5	5 (16)	2 (7)	1 (8)
PT/OT Non-Rounds	45 (11)	2 (2.5)	2.5	2 (6)	0	0
Student Non-Rounds	16 (4)	2 (2.5)	7.5	2 (6)	0	0
Total	406	75	Avg 10	32	30	13

The coding schema, as shown in Table 1, which was used to characterize distraction events, includes the initial event of Distraction with Multitasking when the clinician engages in another interaction in addition to the current use of CPOE. The initial event is followed by a primary action of the distracted clinician: an Interruption, a Deferred Task, or Continued Multitasking. The primary action may or may not be followed by a secondary action: Lack of Recall, Incomplete Task, or Change in Plan of Care.

Results

We observed 38 clinicians from a combined data set of MICU rounds and CPOE/CIS use on a medical/surgical unit at the Columbia-Presbyterian campus of New York Presbyterian Hospital. A total of 75 Distraction with Multitasking events were detected in 406 minutes of observational data. Table 2 shows the breakdown of total minutes and events per type of clinician observed.

Distractions during MICU rounds

During 93 minutes of MICU rounds, observational data of a medical resident using CPOE found that a total of 22 distractions occurred, with one distraction occurring on average every 4.2 minutes. Ten of the events were categorized as Interruptions, three as Continued Multitasking, and nine as a Deferred Task. An Interruption preceded one Incomplete Task and two Lack of Recall episodes and one Change in Plan of Care. One Deferred Task preceded one Incomplete Task. Three sources of distraction were identified. It is important to note that these sources are not mutually exclusive; therefore, the sum of the percentages does not equal 100 percent. These sources were as follows: information need events [14] accounted for 12 out of 22 events (55%), clinician communication accounted for 9 out of 22 events (40%), and frozen CPOE user interface screens accounted for 3 out of 22 events (14%).

Distraction during MedSurg non-rounds

The second observational data set consisted of clinicians (physicians, nurses, physical therapists, occupational therapists, and medical students) while using a CPOE/CIS system. The observational data ranged in length of time from one minute to 37 minutes. The recording time represents time when the user was actively using the CPOE/CIS system.

The analysis of 313 minutes of data that was collected identified a total of 53 Distraction with Multitasking events. When examining overall data, on average a distraction event occurred every six minutes. Of these 53 events: 27 events were characterized as a Continued Multitasking primary action with two secondary actions of Lack of Recall; 22 events were characterized as an Interruption primary action with two secondary actions of Incomplete Tasks; four events were characterized as a Deferred Task primary action with one secondary action of Incomplete Task.

Distractions by user type

Sixteen physicians were observed using CPOE/CIS for a total of 158 minutes with 41 identified Distraction with Multitasking events. Of the 41 events, 25 (60%) resulted in a Continued Multitasking primary action and two of those primary actions led to a Lack of Recall secondary action. Thirteen of the 41 events, or 31%, produced an Interruption primary action with one of those actions leading to an Incomplete Task secondary action. Of the remaining three

Deferred Task primary actions, only one led to an Incomplete Task secondary action. An event occurred every 3.8 minutes while a physician was using CPOE/CIS.

Observational data of 13 nurses using CPOE/CIS totaled 94 minutes and identified 8 Distraction with Multitasking events. 62%, or 5 of the 8 events, were followed by an Interruption primary action, with one of those actions leading to an Incomplete Task secondary event. In contrast, 2 (25%) of the 8 events led to Continued Multitasking primary actions. The remaining event was identified as a Deferred Task primary action. An event occurred every 11.75 minutes while a nurse was using CPOE/CIS.

Four physical therapists were observed using CPOE/CIS for a total of 31 minutes and encountered one Distraction with Multitasking event. Observational data collected on one occupational therapist using CPOE/CIS for a total of 14 minutes identified one Distraction with Multitasking event. The two events that occurred during physical therapist and occupational therapist CPOE/CIS use both resulted in an Interruption primary action with no occurrence of a secondary action. On average a physical therapist or occupational therapist encountered an event every 22.5 minutes.

Three medical students were observed using CPOE/CIS for 16 minutes with two Distraction with Multitasking events identified. Each of the two events encountered by medical students was followed by an Interruption primary action and no secondary action. The medical students, on average, experienced an event every eight minutes.

Of the 75 events in the combined data set, 10 resulted in a secondary action (See Table 3).

Table 3 – Total count of primary and secondary actions

Primary Actions		Secondary Actions		
		Lack of Recall	Incomplete Task	Change in Plan of Care
Interruption	N = 32	2	3	1
Deferred Task	N = 13	0	2	N/A
Continued Multitasking	N = 30	2	0	N/A
Total	N =75	4	5	1

An information need [14] accounted for 21 of the 53 non-rounds events (40%). In these cases, 13 of the 22 information needs caused an Interruption, two of which the information need failed to be met and an Incomplete Task resulted. Six of the 22 information needs caused Continued Multitasking, two of which caused Lack of Recall episodes. Two Deferred Tasks resulted from an information need, one of which the information need was deferred and resulted in an Incomplete Task. Of note, 17 of the 21 events resulting from information needs utilized a human resource [14].

Discussion

We observed 6 hours and 46 minutes of clinicians using a CPOE/CIS system in the MICU and in MedSurg. Distractions occurred 75 times during this time period. The MICU rounds data resulted in a greater proportion of Deferred Tasks while the medical/surgical unit non-rounds data resulted in a greater proportion of Continued Multitasking. However, Interruptions as a primary action from a distraction event were found to have high proportions in both the rounds and the non-rounds data. Possibly the nature of the structure, pace and coordinated activity [4] of MICU rounds leads the resident using CPOE/CIS to be less likely to engage in Continued Multitasking. The less structured work of non-rounds CPOE/CIS use may increase the likelihood of engaging in conversation with a colleague while using the CPOE/CIS system, a form of Continued Multitasking.

The emergent phenomenon that physicians opted for the primary action Continued Multitasking 60% of the time versus nurses who opted for the primary action Interruption 62% of the time possibly relates to the type of distractions encountered by each type of clinician or the nature of the CPOE/CIS tasks required by the two types of clinicians. Though data samples were small, physical therapists, occupational therapists, and medical students followed a distraction with an Interruption primary action in all cases. This phenomenon may influence the design of clinician tailored interfaces of CPOE/CIS systems. Further investigation is necessary to determine if the nature of the informatics task, CPOE specific versus CIS specific, influences the secondary action of the clinician and if the type of clinician plays a role in determining the chosen secondary action when encountered with a distracting event.

The taxonomy was able to characterize a consistent rate of Distraction with Multitasking events for physicians using CPOE/CIS during MICU rounds and during non-rounds usage. The nature of the events appeared to differ; yet the frequency of events did not. Analysis of the content of distraction events shows that the rounds events were more structured and clinically focused. The events during non-rounds, in addition to clinically focused communication events, included “casual, polite and social” conversation.

The combined average event rate for the physician observational periods was one event every 3.98 minutes. This event rate shows more frequent occurrence of distraction events than compared to the non-CPOE/CIS specific Coiera et al.’s 2002 study of physicians in an emergency department that detected a distraction event every 11.1 minutes [9].

A similar comparison is shown for the distraction events experienced by nurses in this study to interruptive events experienced by nurses in a level one trauma center in 2005 [19]. Brixey et al.’s definition of an interruption included distracting events and recipient blocked tasks [19] (deferred tasks), allowing for comparison of total events between the studies [12]. In the context of nurse CPOE/CIS use, our study detected an event every 11.75 minutes

compared to the non-CPOE/CIS specific Brixey et al.'s detection of an event every 15 to 24 minutes [19].

The above comparisons to other studies show that general workflow distraction events rates may vary from distraction event rates in the context of CPOE/CIS use. One previous study of distractions in general clinician workflow did detect distraction event rates at 1 every 4 minutes [20], comparable to the CPOE/CIS distraction event rate. However, the internally consistent rates among rounds and non-rounds physician data shows a possible trend of increased distraction event rates in the context of CPOE/CIS use. Our small sample size, yet rich data set, indicates that CPOE/CIS use lends a clinician vulnerable to at least a similar rate and possibly a higher rate of distraction events than are found in general clinician workflow.

Due to the nature of some clinical tasks (i.e. looking up laboratory orders) there exists some difficulty determining the end point of an intended task. Additionally, a secondary action might result in a slip or mistake [16], but we were unable to ascertain if a slip or mistake occurred because we did not have follow-up data [12]. However, the detection of secondary actions, such as Lack of Recall, though a small count, is concerning if the context of CPOE/CIS use does not allow for the recognition and "rescue" of cognitive overload by colleagues at the same level seen in coordinated activity amongst clinicians [4].

Conclusions

Distractions per hour of CPOE/CIS use are as prevalent, and possibly more prevalent, than distractions per hour in general clinician workflow. Evidence of the existence of secondary actions indicates an opportunity for a slip or mistake to occur [17]. The taxonomy is comprehensive enough to capture the distraction events, primary actions and secondary actions that occur in the context of clinicians' use of CPOE/CIS systems.

The interrupt-driven nature of the clinical work environment impacts the cognition of a clinician while using CPOE/CIS. Health care providers rarely work at private workstations in secluded areas; clinical information needs, as well as social engagement, both addressed through clinician communication, contribute to distraction events and possible slips or mistakes [17] in patient care. Given the prevalence of distraction events in the context of CPOE/CIS use, and previous work indicating the relationship between distractions and potential for patient harm, the results of this study indicate an area ripe for further analysis.

Acknowledgments

This research is sponsored by a grant from the National Institute for Nursing Research (T32NR007969) and a grant from the National Library of Medicine (1R01LM07593).

References

[1] Laxmisan A, Hakimzada F, Sayan O, Green R, Zhang J, Patel V. The Multitasking clinician: Decision-making and cognitive demand during and after team handoffs in emergency care. International Journal of Medical Informatics. (in press, 2006.)

[2] Clark H, Brennan S. Grounding in communication. Perspectives on Socially Shared Cognition, Resnick L, Levine JM, Teasley SD, Eds. Washington, DC: Amer. Psychol. Asssoc., 1991, 127-149.

[3] Patterson ES, Cook RI, Woods DD, Render ML. Examining the complexity behind a medication error: generic patterns in communication. IEEE SMC 2004; Part A, 34(6):749 – 756.

[4] Klein G, Woods D, Bradshaw J, Hoffman R, Feltovich P. Ten Challenges for Making Automation a "Team Player" in Joint Human-Agent Activity. IEEE Intelligent Systems. 2004; 19(6):91-95.

[5] Coiera E. Interaction Design Theory. International Journal of Medical Informatics 2003; 69: 205-222.

[6] Van der Sijs H, Aarts J, Vulto A, Berg M. Overriding of drug safety alerts in computerized physician order entry. J Am Med Inform Assoc. 2006; 13(2):138-47.

[7] Ash JS, Berg M, Coiera E. Some unintended consequences of information technology in health care: the nature of patient care information system-related errors. JAMIA 2004; 11:104-112.

[8] Alvarez G, Coiera E. Interruptive communication patterns in the intensive care unit ward round. International Journal of Medical Informatics 2005; 74: 791-796.

[9] Coiera E, Jayasuria R, Hardy J, Bannan A, Thorpe M. Communication loads on clinical staff in the emergency department. MJA 2002; 176(9): 415-418.

[10] MEDMARX Data Report 2002.

[11] Bates DW, Teich JM, Lee J, Seger D, Kuperman GJ, Ma'Luf N, Boyle D, Leape L. The impact of computerized physician order entry on medication error prevention. J Am Med Inform Assoc. 1999 Jul-Aug;6(4):313-21.

[12] Collins S, Currie LM, Bakken S, Cimino JJ. Interruptions During the Use of a CPOE System for MICU Rounds. AMIA Symp Proc. 2006.

[13] Flynn EA, Barker KN, Gibson JT, Pearson RE, Berger BA, Smith LA. Impact of interruptions and distractions in dispensing errors in an ambulatory care pharmacy. American Journal of Health-System Pharmacy. 1999; 56(13):1319-25.

[14] Currie LM, Graham M, Allen M, Bakken S, Patel V, Cimino JJ. Clinical information needs in context: an observational study of clinicians while using a clinical information system. AMIA Symp Proc. 2003;190-194.

[15] Chisholm CD, Collison EK, Nelson DR, Cordell WH. Emergency department workplace interruptions: are emergency physicians "interrupt-driven" and "multitasking"? Acad Emerg Med. 2000; 7(11):1239-43.

[16] Zhang J, Patel V, Johnson TR, Shortliffe EH. A cognitive taxonomy of medical errors. Journal of Biomedical Informatics 2004; 37:193–204.

[17] Coiera, E. Communication Systems in Healthcare. Clinical Biochem Review. 2006; 27:89-98.

[18] Brixey J, Robinson D, Johnson C, Johnson T, Turley J, Patel V, Zhang J. Towards a hybrid method to categorize interruptions and activities in healthcare. International Journal of Medical Informatics (in press 2006.)

[19] Brixey J, Robinson D, Tang Z, Johnson T, Zhang J, Turley J. Interruptions in Workflow for RNs in a Level One Trauma Center. AMIA Symp Proc. 2005;86-90.

[20] Spencer R, Logan P. Role-based communication patterns within an emergency department setting. In: Proceedings HIC 2002; 2002; Melbourne; 2002.

Address for correspondence

Sarah A. Collins, RN, BSN
Mail Code 6, 630 West 168th Street, New York, NY 10032
Telephone: 1-781-801-9211, e-mail sac2125@columbia.edu

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Diffusion of Electronic Health Records – Six Years of Empirical Data

Christian Nøhr [a], Stig Kjær Andersen [b], Knut Bernstein [c], Morten Bruun-Rasmussen [c], Søren Vingtoft [c]

[a] *Virtual Centre for Health Informatics, Department of Development and Planning, Aalborg University, Denmark*
[b] *Virtual Centre for Health Informatics, Department of Health Science and Technology, Aalborg University, Denmark*
[c] *MEDIQ, Copenhagen, Denmark*

Abstract

The Danish EHR-Observatory has monitored Danish EHR projects since 2000 with respect to a number of parameters such as diffusion, diffusion rate and the hospital owners expectations. On the basis of the data and a model for technology diffusion three scenarios for future diffusion are built. The results show that the national goal to have EHRs fully implemented in hospitals by 2008 will not be reached. The scenarios built from empirical data provide a useful benchmarking tool for planning and evaluating the EHR implementation programs in hospitals.

Keywords:

Medical Records Systems, Medical Order Entry Systems, state government, diffusion of innovation, technology transfer

Introduction

Several countries follow up on policies to implement electronic health record (EHR) systems for managing and communicating patient information, and a number of them have developed national strategies for the implementation [1-5].

The US does not have a similar national strategy, but on April 27, 2004 President Bush issued an executive order establishing the Office of the National Coordinator for Health Information Technology (ONCHIT) [6].

All the strategies and plans have time estimates for when the EHR systems should be implemented. In the Australian plan they expect that "by 2008, Australia will be well in advance in achieving the goal of electronic connectivity between all major health institutions and health care providers". The Canada Health Infoway goal is "to have an interoperable electronic health record in place across 50per cent of Canada, by population, by the end of 2009". In the US plan the mission is to implement EHR systems nationwide within ten years.

The diffusion of EHR systems among general practitioners in Denmark has been reported earlier [7]. Today practically all GPs (98%) have EHRs in their office, and they communicate 3.5 mill messages every month, where the most common are: 100% of discharge letters, 80% of prescriptions, 98% of lab-reports, and 63% of referrals to hospitals or practicing specialists. The messages comply with the EDI standard and are communicated on the Danish health care network which is developed and administered by MEDCOM. Detailed statistics on current adoption and usage can be found on their homepage [8].

In the Danish revised strategy the goal is to have EHRs fully implemented in hospitals by 2008 [3].

There are however few studies which monitor the implementation status or the actual diffusion of EHR systems. Ford et.al. [9] gathered and synthesized data from six studies on EHR adoption rates. Applying technology diffusion theory they designed a model to project estimated future EHR adoption trends and timelines in three future scenarios, optimistic, best estimate, and conservative. Finally they determined the likelihood of achieving universal EHR adoption. The focus is on physicians in small practices. Their results show that under current conditions, EHR adoption will reach its maximum in 2024, and they conclude that the EHR products now available are unlikely to achieve full diffusion within the 2014 timeframe being targeted by policy makers.

Jha et.al. identified surveys on EHR adoption and addressed their quality [10]. They based their estimates on studies of high or medium quality and found that through 2005, approximately 23.9 percent of physicians used an EHR in the ambulatory settings, while 5 percent of hospitals used computerized order entry systems.

The present study construct a model that project likely EHR adoption patterns based – not on historic estimates as in [9;10] – but on empirical data for the Danish hospitals obtained by the Danish EHR-Observatory through national surveys from 2000 to 2006 [11-16].

The purpose of the EHR-Observatory is to support the realization of the national strategy by monitoring and assess the development, implementation and application patterns of EHR systems in Danish hospitals.

The study makes two new contributions to the EHR discussions and the EHR policy making. First it confirms and substantiates the gap between the optimistic implementation plans seen in several countries and the actual development of useable and accepted systems. Second it provides a benchmarking tool for planning and evaluating EHR implementation programs in hospitals.

Selected for best paper award.

Methods

Identification of what to count

An important limitation in surveying the adoption and diffusion of EHR systems is the definitions of systems handling patient data: Electronic medical record (EMR), electronic patient record (EPR) computerized medical record (CMR), computer-based patient record (CPR), and electronic health record (EHR). There are only minor differences in the meaning depending on the defining country of origin, health sector, professional discipline, and period of time.

There have been many formal definitions of these terms, but they display more similarities than differences with respect to the purpose, functions and goals of electronic records [17].

The International Standard Organization ISO has developed an internationally agreed definition of the EHR: "a repository of information regarding the health of a subject of care, in computer processable form". Here the EHR is defined in terms of its structure, but it is internationally broad and encompasses most of the EHR systems currently used.

These very broad definitions make it difficult to determine the exact object of a diffusion survey. However EHR systems are in general subdivided into components or modules each handling one or more functionalities [18].

The most common components are:

Clinical documentation

Handles progress notes either as free text directly entered into the system or via predefined structured notes. Voice recognition systems are also seen as data entry method.

Physician order entry (POE)

Used for ordering diagnostic test and medication in a standardized and formalized way. Some systems also check for drug interaction and alert for patient allergy.

Booking service

Allow patients and clinicians to book appointments.

Communication/messaging

Facilitates communication between hospitals, General Practitioners, pharmacies, and laboratories.

Results management

Abnormal results warning, trending/graphing.

Charge capture/billing

Coding interventions (this module varies a great deal according to the organization and financing of the health sector).

Decision support

There are a lot of decision support modules in use, but they are rarely used outside the organization where they are developed.

Clinical practice guidelines

Module to manage and maintain clinical guidelines or national reference programs – sometimes categorized as decision support.

Disease management

Management of chronic diseases, like diabetes, etc.

Management of security issues

All EHR systems will have special facilities to manage authentication and authorization of user access according to national legislation

In practice there will be difference between the modules in which a specific functionality is positioned i.e. the above list can neither be regarded as exclusive nor exhaustive.

This is also true for the specific way a work task is executed in practice. E.g. one department can use the booking module to book an x-ray for a patient where another department will use facilities in the CPOE module. This makes it difficult to generate diffusion and adoption rates that can be generalized across different health care systems.

In this study the criteria for having an EHR implemented was a full functioning clinical documentation module, an order entry module able to handle medication functions.

Another limiting factor for the accuracy of surveying diffusion is the unit for measuring adoption. The best choice will depend on the organization and structure of the health sector. In systems where GP's and hospitals work independently, and where the physicians are employed by the hospital with rapidly changing workplaces physicians will not be an adequate unit to count. The number of hospitals will on the other hand be too coarse and the number of specific wards varies so much in size that the count will be inaccurate. In this study the number of beds covered by an EHR system has been chosen.

Questionnaire survey

In the present study the number of beds covered with an EHR system containing functionalities for at least clinical documentation and order entry for medication was surveyed. The ordering of lab tests (clinical chemistry) has already been implemented in all hospitals for a number of years, but is not integrated with medication. The ordering of other tests and X-rays etc. are only partly implemented, but are not included in this survey.

A questionnaire was sent to the county administrations (n=15) every year from 2001 to 2006. The county administrations are the ''owners'' of the public hospitals, and have the political and administrative responsibility for running the hospitals in Denmark. The questionnaire had a kernel of 12 questions that surveyed general aspects of strategy, diffusion, economy, benefits and barriers which remained the same throughout the 6 years and a number of detailed questions to a specific theme of investigation.

The diffusion questions asked the county administrations how many beds were in the hospitals and how many were covered by an EHR system (clinical documentation and medication). Furthermore they were asked how many beds

they were planning to cover the following three years (except in 2001 they were only asked for the current year).

The technology diffusion forecast model

Technology forecasting is a discipline focusing on estimating how technology adoption and diffusion will take place in the future. The goal is not to predict the future, but to create scenarios for planning and policy making. One of the sub-disciplines of technology forecasting is engaged in extrapolating diffusion trends – a quantitative approach to treating adoption data. One of the authors of a technology forecast textbook refers to it as "naive time series analysis" [19]. This implicates that presumptions are met with respect to structure and context of the technology in question. The basic foundation of the essential model however is well proven in countless empirical studies from many areas. The most frequent basic model rests upon the assumption that diffusion of technology follows a sigmoid curve [20]. The sigmoid curve is shown in figure 1 and can be expressed by the equations (1-3):

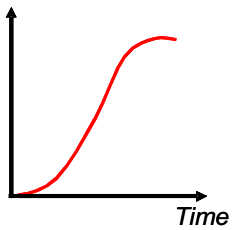


Figure 1 - Sigmoid curve

$$\frac{df}{dt} = b[f/(1-f)] \quad (1)$$

$$f = \frac{100}{1 + c^{*}e^{(-b^{*}t)}} \quad (2)$$

$$f = {}^{t}b[f/(1-f)]dt \quad (3)$$

Where f = the diffusion of the technology in % of the upper limit (here 100), c and b = constants determining the slope and placement of the curve, and t = time. Fisher and Pry have used this expression to study technology forecasting and social change (ref) and is referred to as "Fisher-Pry curve" [21]. If data is available for the first couple of years it is possible to make a best fit curve describing the most likely diffusion course. An optimistic scenario was added by increasing the growth by 20% for the future years, and a conservative scenario was added by reducing the growth in diffusion by 10% for the future years.

Results

The questionnaires were sent to the county administrations electronically each year in April and after one reminder the response was close to optimal as shown in table 1.

The number of beds actually covered and the number of beds the counties expected to cover in the following three years are shown in figure 2. In 2001 we did not ask for the expectations to the following years.

Table 1 - Response rate

Year	Responses	Response rate
2001	12	80,0%
2002	15	100,0%
2003	14	93,3%
2004	15	100,0%
2005	14	93,3%
2006	15	100,0%

The number of beds covered is steadily increasing over the years (the bars boxed with thick lines), but the expectations show a systematically overrating of what turns out to be possible.

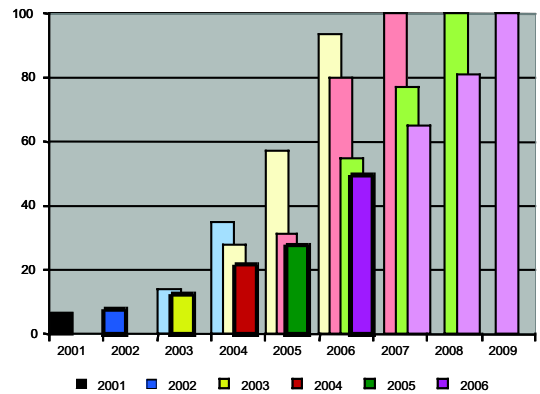


Figure 2 - Actual and expected national EHR diffusion

The data on the actual coverage of beds were fitted into the Fisher-Pry model and an optimistic as well as a conservative scenario was drawn. This is shown in figure 3.

In 2008 the optimistic scenario will have reached 77% adoption, the best estimate will have reached 69% and the conservative scenario 60% adoption.

The optimistic scenario will reach 90% adoption in 2009, the best estimate in 2010 and the conservative scenario in 2011.

Discussion

In figure 2 it is showed how the hospital owners overestimate the capacity to implement the EHR modules. The estimate for the year following the survey is reasonable accurate – only overrated with a few per cent, but the further they predict the implementation the more optimistic they get. This confirm to the optimistic policies on the national level, also seen internationally. In a public financed health care system - to a large extend controlled by politics – there will be a tendency to adhere to the policy goals, even if they seem unrealistic.

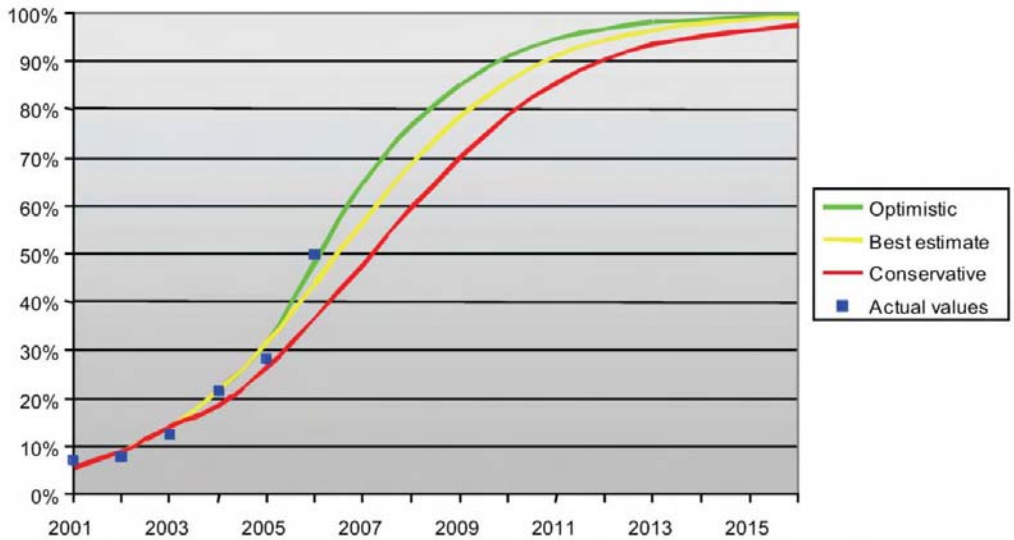


Figure 3 - Three models for diffusion of EHR system based on empirical data from the Danish Hospitals

Another aspect is also that the solutions are not known. There is not a single hospital in the world where a completely flawless system is in place, there are no known places that can be taken as a model. Essential parts have to be developed before usable and acceptable systems that are able to live up to the demands in the national plans exist. The NHS in the UK has a "MythBuster" homepage with the aim to dispel common misconceptions and enable a clearer understanding of the work being done. The answer to the "myth": "*The project is plagued with delays"* is that "*This is because the new software is challenging and there is a debate in the medical professions about what information should go in the records*". [22]

In the conclusion of his book about the history of medical informatics in the United States Morris Collen writes: "Developing a comprehensive medical information system was a more complex task than putting a man on the moon had been," [23]. From a local perspective the development and implementation of EHR systems is simply an unimaginable difficult task.

The scenarios for the diffusion of EHR systems shown in figure 3 are not prognosis. The scenarios are built by a rather static model applying naive time series analysis. Even though there has been numerous empirical studies showing that technology diffusion in general follow the trend in the sigmoid curve, it does not take into account that the investment in EHR systems can be increased by political initiative.

Increase in investment could promote the optimistic scenario, and there are aspects of the diffusion process that can be advanced by investment, such as educating the users to be ready when the systems are introduced, installing all the hardware and networking. For instance to day practically all the Danish hospitals are covered with wireless networks and secure infrastructure – ready for EHR systems, but as in the UK the new software is challenging and the clinical professions, the administrators and the policy makers are still debating the right structure and granularity of the data that must go in to the records.

On the other hand even the rigorous best estimate extrapolation of the diffusion trend might be too optimistic. Jos Aarts has studied implementation of CPOE systems in three Dutch hospitals and three hospitals in the United States. [24] The size of the hospitals he studied was comparable to hospitals in Denmark. The shortest implementation time was at the 315 beds VAPS hospital in Seattle where they replaced a system which had been used for many years. The implementation of the new system took three years. In the other end of the scale is the Radboud hospital in Nijmegen, where they implemented Eclipsys E7000 from 1988 to 2000 – twelve years. It has not been possible to find evidence for any remarkably faster implementations, which indicates that implementation of EHR systems, requires hard work and persistence.

The best estimate extrapolation, and even the optimistic scenario, clearly shows that full implementation by 2008 are unachievable. This initiated a debate about the EHR implementation policy in the beginning of 2006, and resulted in a revision of the policy and the establishment of a central EHR coordination office, which is under manning at the time of writing.

Conclusion

The EHR implementation in Denmark has been monitored by the EHR-Observatory since 2000 obtaining annual data on EHR diffusion. These data has been applied to a technology forecast model, and the results show that the policy for EHR implementation will not reach the goal of full implementation within the appointed time. This usage of technology forecast models can be useful for planning and evaluation of EHR implementation policies in hospitals.

References

[1] Canada Health Infoway. 2003-2004 Annual Report 2004-2005 Corporate Plan Summary. 31-3-2004.

[2] Commonwealth of Australia. HealthConnect Implementation Strategy. 3670 (JN9045). 6-7-2006. Canberra ACT, Commonwealth of Australia.

[3] Danish Ministry of the Interior and Health. National IT Strategy 2003-2007 for the Danish Health Care Service. 2003.

[4] Department of health. Delivering 21st Century IT Support for the NHS - National Strategic Programme. 2002. London, Department of health.

[5] The WAVE Advisory Board to the Director-General of Health. From Strategy to Reality The Working to Add Value Through e-Information (WAVE) Project: Health Information and Management and Technology Plan. 2001.

[6] Bush G W. Executive order: 13335: Incentives for the use of Health Information Technology and Establishing the Position of the National Health Information Technology Coordinator. 2004.

[7] Protti D, Johansen I. Further Lessons from Denmark About Computer Systems in Physician Offices. Electronichealthcare 2003; 2(2):16-21.

[8] MEDCOM. Medcom statistics in English. http://www.medcom.dk/wm109991 . 12-11-2006. 2-12-2006.

[9] Ford EW, Menachemi N, Phillips MT. Predicting the adoption of electronic health records by physicians: when will health care be paperless? J Am Med Inform Assoc 2006; 13(1):106-112.

[10] Jha AK, Ferris TG, Donelan K, DesRoches C, Shields A, Rosenbaum S et al. How common are electronic health records in the United States? A summary of the evidence. Health Aff (Millwood) 2006; 25(6):w496-w507.

[11] Andersen S.K., Nøhr C., Vingtoft S., Bernstein K., Bruun-Rasmussen M. EHR-Observatory Annual report 2002. 2002. Aalborg, EPJ-Observatoriet.

[12] Bruun-Rasmussen M., Bernstein K., Vingtoft S., Andersen S.K., Nøhr C. EHR-Observatory Annual report 2003. 2003. Aalborg, EPJ-Observatoriet.

[13] Nøhr C., Andersen S.K., Vingtoft S., Bruun-Rasmussen M., Bernstein K. EHR-Observatory Annual report 2004. 2004. Aalborg, EPJ-Observatoriet.

[14] Vingtoft S., Bruun-Rasmussen M., Bernstein K., Andersen S.K., Nøhr C. EHR-Observatory Annual report 2005. 2005. Aalborg, EPJ-Observatoriet.

[15] Vingtoft S, Lippert S, Bernstein K, Rasmussen MB, Nøhr C, Andersen SK et al. EHR Observatory Annual Report 2000. 2000. Odense, The County of Funen.

[16] Bernstein K, Rasmussen M.B., Nøhr C., Andersen S.K., Vingtoft S. EHR Observatory. Annual Report 2001. 2001. Odense, The County of Funen.

[17] Schloeffel P. Current EHR Developments: an Australian and International Perspective - Part 1. Health Care and Informatics Review Online™. 2004. 12-8-2005.

[18] Nohr C. Evaluation of electronic health record systems. Methods Inf Med 2006; 45 Suppl 1:107-13.:107-113.

[19] Porter A, Banks J, Mason T W, Rosinni F A, Roper T A. Forecasting and Management of Technology. Wiley-IEEE, 1991.

[20] Rogers E M. Diffusion of Innovations. 5th edition ed. New York: The Free Press, 1995.

[21] Fisher JC, Pry R H. A Simple Substitution Model of Technological Change. Technology Forecasting and Social Change 1971; 3:75-78.

[22] NHS connecting for health. Myth Buster. http://www.connectingforhealth.nhs.uk/faq/mythbusters . 3-12-2006. 3-12-2006.

[23] Collen M F. A History of Medical Informatics in the United States 1950 to 1990. American Medical Informatics Association, 1995.

[24] Aarts J. Understanding Implementation. A sociotechnical appraisal of the introduction of computerized physician order entry systems in Dutch and American Hospitals. Rotterdam: Erasmus University, 2005.

Address for correspondence

Christian Nøhr,
Virtual Centre for health Informatics,
Department of Development and Planning,
Aalborg University,
Fibigerstræde 13, 9220 Aalborg, Denmark
CN@v-chi.dk

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Text Categorization Models for Identifying Unproven Cancer Treatments on the Web

Yin Aphinyanaphongs[a], Constantin Aliferis[abc]

[a]*Department of Biomedical Informatics, Vanderbilt University, Nashville*
[b]*Department of Cancer Biology, Vanderbilt University School of Medicine, Nashville*
[c]*Vanderbilt Ingram Cancer Center, Vanderbilt University, Nashville.*

Abstract

The nature of the internet as a non-peer-reviewed (and largely unregulated) publication medium has allowed wide-spread promotion of inaccurate and unproven medical claims in unprecedented scale. Patients with conditions that are not currently fully treatable are particularly susceptible to unproven and dangerous promises about miracle treatments. In extreme cases, fatal adverse outcomes have been documented. Most commonly, the cost is financial, psychological, and delayed application of imperfect but proven scientific modalities. To help protect patients, who may be desperately ill and thus prone to exploitation, we explored the use of machine learning techniques to identify web pages that make unproven claims. This feasibility study shows that the resulting models can identify web pages that make unproven claims in a fully automatic manner, and substantially better than previous web tools and state-of-the-art search engine technology.

Keywords:

information storage and retrieval, medical informatics, internet, neoplasms, text categorization

Introduction

"The killing of all parasites and their larval stages together with removal of isopropyl alcohol and carcinogens from the patients' lifestyle results in remarkable recovery (from cancer), generally noticeable in less than one week [1]." This is one example of an unproven treatment claim made on the web. These unproven treatments are known as *quackery* with the *quacks* promoting them defined as "untrained people who pretend to be physicians and dispense medical advice and treatment [2]." The internet allows quacks to advocate inaccurate and unproven treatments with documented fatal, adverse outcomes in some situations [3-6].

In regards to cancer patients, Metz et al. reported that 65% of cancer patients searched unproven treatments and 12% purchased unconventional medical therapies online [7]. In another study, Richardson reported that 83% of cancer patients had used at least one unproven treatment [8]. Many patients are ill-equipped to evaluate treatment information [9]. The language and quality of web pages with unproven treatments is also highly variable [10]. The rapid growth of the internet, combined with the ease of publishing unproven claims leads susceptible and often desperately ill patients to further adverse outcomes, patient and family despair, and sunk costs. It is thus an important mandate of the medical profession to protect patients from inaccurate and poor medical information.

So far extensive research has developed several manual methods to combat the propagation of unproven claims on the web. The Health-on-the-Net Foundation advocates self-regulation of health related websites [11]. The foundation applies strict criteria to websites and grants them a seal of approval if they pass. However, most health care consumers ignore the seals [12]. In another approach, experts produced rating tools that consumers are supposed to apply to websites[13, 14]. Another method is manual review of individual websites that are published either in print or electronically.

Each method has limitations. Self-regulation relies on knowledge of the certification and a vigilant public to report failing web sites. Rating tools are dependent on a knowledgeable public to apply, they are difficult to validate, time consuming to produce, and do not always produce consistent ratings [15, 16]. Moreover, the rating tools are not appropriate for use on complementary/ alternative medicine sites [17]. Furthermore, manual review suffers from limits in reviewer time and the selection of web sites to review.

Ideally, we would like a solution that is validated, easy to apply by health care consumers, and works on any webpage. In this paper, we hypothesize that automated approaches to identifying web pages with unproven claims may provide a solution.

Previous work on automatic webpage identification

Previous research focused on automated or semi-automated approaches to identifying *high quality* medical web pages.

Price and Hersh [18] evaluated web page content by combining a score measuring quality proxies for each page. Quality proxies included relevance, credibility, bias, content, currency, and the value of its links. The authors evaluated the algorithm on a small test collection of 48 web pages covering nine medical topics labeled as desirable or undesirable by the investigator. In all cases, the

Selected for best paper award.

score assigned to the desirable pages was higher than the scores assigned to undesirable pages.

Even though the algorithm perfectly discriminated between desirable and undesirable webpages, several limitations exist. The test sample was small and not representative of the scale for a web classification task. The algorithm does not measure content quality directly, but used proxies for quality to compile a score for a web page. The usefulness of some of the explicit criteria may not correlate with content quality [19], and may not be valid or good features to include for scoring.

As a leading search engine, Google has become a de facto standard for identifying and ranking web pages. Pages that rank highly in Google are assumed to be of better quality than those at lower rank. Several researchers have explored this assumption for health pages. Fricke and Fallis [20] evaluated PageRank score as one indicator of quality for 116 web sites about carpal tunnel syndrome. Their results show that PageRank score is not inherently useful for discrimination or helping users to avoid inaccurate or poor information. Of the 70 web sites with high PageRank, 29 of them had inaccurate information.

Griffiths [21] evaluated PageRank scores for depression websites using evidence based quality scores. The authors obtained Google PageRank scores for 24 depression websites from the DMOZ Open Directory Project website. Two health professional raters assigned an evidence based quality score to each site. PageRank scores correlated weakly ($r = 0.61$, $P=0.002$) with the evidence based quality scores.

Tang, Craswell, and Hawking [22] compared Google results with a domain-specific search engine for depression. They found that of a 101 selected queries, Google returned more relevant results, but at the expense of quality. Of the 50 treatment related queries, Google returned 70 pages of which 19 strongly disagreed with the scientific evidence.

Hypothesis

Our fundamental hypothesis for this feasibility study is that we can model expert opinion and build machine learning models that identify web pages that make unproven claims for the treatment of cancer.

To the best of our knowledge, there is no research on validated automated techniques for identifying web pages that make unproven claims. In prior work, we showed that text categorization methods identified high quality content specific articles in internal medicine [23]. Extending this work into the web space, we reverse the hypothesis of the previous studies. Rather than identifying high quality pages, we explore automated identification of low quality pages, specifically pages that make unproven claims for cancer treatment.

Materials and methods

Definitions

Our gold standard relied on selected unproven cancer treatments identified by experts at http://www.quackwatch.org. The website is maintained by a 36 year old nonprofit organization whose mission is to "combat health related frauds, myths, fads, fallacies, and misconduct." The group employs a 152 person scientific and technical advisory board composed of academic and private physicians, dentists, mental health advisors, registered dietitians, podiatrists, veterinarians, and other experts whom review health related claims. By using unproven treatments identified by an oversight organization, we capitalized on an existing high quality review.

Corpus construction

For this feasibility study, we randomly chose 8 unproven treatments from 120 dubious cancer treatments listed by quackwatch.org [24]. The randomly selected treatments were "Cure for all Cancers", "Mistletoe", "Krebiozen", "Metabolic Therapy", "Cellular Health", "ICTH", "Macrobiotic Diet", and "Insulin Potentiation Therapy." We then identified web pages that have these treatments by appending the words "cancer" and "treatment" and querying Google. We retrieved the top 30 results for each unproven treatment. We used a python script to download and store each result as raw html for further labeling.

Corpus labels

We applied a set of criteria for identifying web pages with unproven treatment claims. First, of the initial 240 pages, we excluded not found (404 response code) error pages, no content pages, non-English pages, password-protected pages, pdf pages, redirect pages, and pages where the actual treatment text does not appear in the document[1]. Of the remaining 191 html pages, both authors independently asked the following question of each web page: does the web page make unproven claims about the proposed treatment and its efficacy. We labeled web pages with unproven claims as positive and the others as negative.

Web pages that are purely informational in nature but do not make any unproven claims about the cancer treatment and its efficacy were labeled as negative. Web pages selling a book with user comments that has unproven claims were labeled as positive. Portal pages that do not make any claim were labeled as negative. Web pages that attempted to present an objective viewpoint of the treatment were carefully reviewed for any unproven claims, and, if so, were labeled positive. Additionally web pages that sell unproven treatments but do not make claims were labeled negative.

Both authors applied the criteria independently. We calculated the inter-observer agreement (Cohen's Kappa [25]) at 0.76[2]. Of the 20 sites with discrepant labelings, the reviewers discussed the labels until consensus was reached. The final corpus was composed of 191 web pages with 93 labeled as positive and 98 as negative.

Webpage preparation

For this feasibility study, we chose the simplest web page representation. We converted web pages to a "bag of words"

1 The Google ranking algorithm relies on anchor text to identify web page content. Anchor text may point to a web page that does not use the anchor text in the web page itself.

2 We set a threshold of 0.70 for Cohen's Kappa. If kappa was below 0.70, we would refine the labeling criteria until the threshold was reached.

suitable for the machine learning algorithm[23]. First, for each web page, we removed all content between style and script tags. Second, all tags (including the style and script tags) were removed. Third, we replaced all punctuation with spaces. We split the remaining string on the spaces to obtain individual words. Finally, we stemmed each word [23], applied a stop word list [23], removed any words that appear in less than 3 web pages, and encoded as weighted features using a log frequency with redundancy scheme [23].

Learning model (support vector machines)

We employed Support Vector Machine (SVM) classification algorithms. The SVM's calculate maximal margin hyperplane(s) separating two or more classes of the data. SVMs have had superior text classification performance compared to other methods [23], and this motivated our use of them. We used an SVM classifier implemented in libSVM v2.8 [26] with a polynomial kernel. We optimized the SVM penalty parameter C over the range {0.1, 1, 2, 5, 10} with imbalanced costs applied to each class proportional to the priors in the data [23], and degree d of the polynomial kernel over the range {1, 2, 5}. The ranges of costs and degrees for optimization were chosen based on previous empirical studies [23].

Model selection and performance estimation

We used 10-fold cross-validation that provides unbiased performance estimates of the learning algorithms [23]. This choice for n provided sufficient high-quality positive samples for training in each category and provided sufficient article samples for the classifiers to learn the models. The cross-validation procedure first divided the data randomly into 10 non-overlapping subsets of documents where the proportion of positive and negative documents in the full dataset is preserved for each subset. Next, the following was repeated 10 times: we used one subset of documents for testing (the "original testing set") and the remaining nine subsets for training (the "original training set") of the classifier. The average performance over 10 original testing sets is reported.

To optimize parameters of the SVM algorithms, we used another "nested" loop of cross-validation [23] by splitting each of the 10 original training sets into smaller training sets and validation sets. For each combination of learner parameters, we obtained cross-validation performance and selected the best performing parameters inside this inner loop of cross-validation. We next built a model with the best parameters on the original training set and applied this model to the original testing set. Notice that the final performance estimate obtained by this procedure will be unbiased because each original testing set is used only once to estimate performance of a single model that was built by using training data exclusively.

Quackometer

We compared our algorithm to a heuristic, unvalidated, and unpublished quack detection tool available at http://www.quackometer.net. The exact details of the detection tool are proprietary. In general, the algorithm counts words in web pages that quacks use, and sorts the words into at least 5 dictionaries [27]. It looks for altmed terms such as "homeopathic" and "herbal", pseudoscientific words such as "toxins" and "superfoods", domain specific words such as "energy" and "vibration", skeptical words such as "placebo" and "flawed", and commerce terms such as "products" and "shipping". The algorithm counts the frequency of terms, applies a frequency threshold, and generates a corresponding score from 0 to 10. The tool is available at [28].

We compared our models to the Quackometer by calculating the corresponding area under the curve (AUC) for each 10 fold-split and reporting the mean and standard deviation.

Google PageRank

The PageRank algorithm [29] is used by Google to identify higher quality pages on the web. The basic tenet is that a web page will rank highly if the web page has more and higher quality links pointing to it. For example, if a web page has a link from Yahoo (a highly linked page), it would rank higher than a link from a less linked to web page. In detecting web pages with unproven claims, our assumption is that web pages with poor quality information should get fewer and lower quality links than web pages with better quality.

We used Google as a proxy for PageRank[3]. We made the comparison to our algorithms within each topic rather than within each 10 fold split. We compared within each topic to avoid bias in ranking situations where one topic has uniformly higher Google rank than another topic. We inverted the labels[4] in the 8 randomly selected topics, calculated the AUC, and reported the mean AUC and standard deviation.

Results

Table 1 shows the AUC performance between the machine learning filter models, Quackometer, and Google. The machine learning method identified web pages that make unproven claims with an AUC of 0.93 with a standard deviation of 0.05 across the 10 folds. Quackometer does worse with an AUC of 0.67 and a standard deviation of 0.10 across the same 10 folds. Finally Google performs least effectively in discriminating web pages with an AUC of 0.63 and a standard deviation of 0.17 across the 8 selected topics. Figure 1 shows the corresponding receiver operating curves for each method.

Table 1 – Area Under Curve for Each Discrimination Method

Model	Mean Area Under the Curve
Support Vector Machine	0.93 (std. 0.05)
Quackometer	0.67 (std. 0.10)
Google	0.63 (std. 0.17)*

* The mean and standard deviation are calculated across the 8 topics rather than across the test sets of the 10 folds.

3 Google uses a proprietary version of PageRank for ranking.

4 We test the assumption that PageRank will rank web pages with *proven* claims higher than web pages with *unproven* claims.

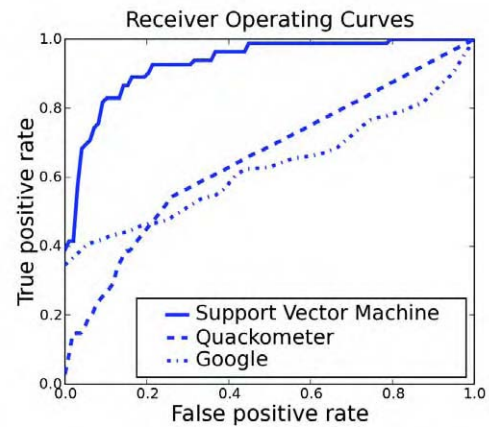


Figure 1: Receiver operating curves for each method.

Discussion

This feasibility study showed that machine learning filter models identify web pages that make unproven claims on a select, focused gold standard. The learning filters have superior performance over the Quackometer [27] and Google. We also note that the loose correlation between Google and high quality sites seems comparable to previous work [20-22].

This method has distinct advantages to rating instruments [13, 14] or manual review. First, there is no need to state explicit rating criteria. The model identified patterns in the data that label a page with unproven claims. Second, compared to the limited focus of manual review on select web pages, these models allow application to any web page.

We also highlight a subtle point in this work. We make a distinction between web pages that make unproven claims and web pages that promote the unproven treatment. Oftentimes, this distinction is blurry. For this work, we only want to identify pages that make unproven claims. Pages that promote a product but do not make unproven claims are not identified. In future studies, we are interested in evaluating models that identify web pages that promote treatments.

In Table 2, we present excerpts from pages where the previous models failed to identify pages with unproven claims. These pages should have been identified by the Quackometer [27] and should not have appeared in the top 30 Google results. Such failure to identify or mark these pages may result in patient's exposure to potentially harmful, unproven treatments.

Table 2 –Web page excerpts where previous tools fail to detect unproven claims. A page that makes unproven claims is identified as such if it has a small support vector machine rank, a large quackometer score, and a large Google rank, respectively. SVM rank is calculated over 10 fold cross validation test set composed of 9 positives and 9 negatives. Google rank is out of the top 30 results returned. Quackometer score provides ranks from 0 to 10. "S" denotes success of the corresponding filter, while "F" failure.

Failure Analysis Excerpts	Support Vector Machine Rank	Quackometer score	Google rank
I am convinced that our mind and emotions are the deciding factor in the cure of cancer.	1 (S)	1 (F)	16 (F)
The hundreds of clinical studies conducted by many competent physicians around the world, including those directed by Dr. Emesto Contreras Rodriguez at the Oasis of Hope Hospital hospital in Mexico, give us complete confidence that there is no danger.	3 (S)	0 (F)	9 (F)
The cure shows results almost immediately and lasts three weeks only. It is cheap and affordable for everybody and proved with 138 case studies.	3 (S)	8 (S)	3 (F)
Many advanced cancer patients are petrified of their tumor. This knee-jerk reaction is caused by orthodox medicine's focus on the highly profitable (and generally worthless) process of shrinking tumors.	1 (S)	1 (F)	18 (F)
IPT (Insulin Potentation Therapy) has an outstanding 135 doctor-year track record (115 years for cancer) over 72 years, and is ready for clinical trials and widespread use.	1 (S)	0 (F)	1 (F)
We are proud of these findings, which confirm that cellular medicine offers solutions for the most critical process in cancer development, the invasion of cancer cells to other organs in the body. Conventional medicine is powerless in this.	2 (S)	1 (F)	8 (F)

In practice, we envision implementing a system that works much like a spam filter works for e-mail. Spam filters identify illegitimate e-mails. In a similar fashion, we envision a system that runs on top of a search engine and flags any web pages that may have unproven health claims.

Limitations

We tested a small sample comprised of 8 unproven treatments in 240 web pages. We will explore how well the models generalize with an independently collected dataset, more unproven treatments, and more labeled web pages. Collecting an independent dataset would allow for validation of the labeling criteria and the model selection procedures.

For this feasibility study, we purposely limited the topic of this study to cancer treatment. In the future, we will build and evaluate other models identifying web pages that make unproven claims for other conditions such as arthritis, autism, and allergies.

Conclusions

We present a first of its kind feasibility study to build machine learning filter models that exhibit high discriminatory performance for identifying web pages with unproven cancer treatment claims. This work paves the way for building broadly applicable models involving more health conditions, more pages with unproven claims, and eventually applied systems to protect patients from quackery.

Acknowledgments

The first author acknowledges support from NLM grant LM007948-02. The second author acknowledges support from grant LM007948-01.

References

[1] Clark H. *The Cure for All Cancers*: New Century Press; 1993.
[2] *American Heritage Dictionary.*
[3] Hainer MI, Tsai N, Komura ST, Chiu CL. Fatal hepatorenal failure associated with hydrazine sulfate. *Ann Intern Med.* 2000 Dec 5;133(11):877-80.
[4] See KA, Lavercombe PS, Dillon J, Ginsberg R. Accidental death from acute selenium poisoning. *Med J Aust*. 2006 Oct 2;185(7):388-9.
[5] Bromley J, Hughes BG, Leong DC, Buckley NA. Life-threatening interaction between complementary medicines. *Ann Pharmacother.* 2005 Sep;39(9):1566-9.
[6] Mularski RA, Grazer RE, Santoni L, Strother JS, Bizovi KE. Treatment advice on the internet leads to a life-threatening adverse reaction: hypotension associated with Niacin overdose. *Clin Toxicol (Phila)*. 2006;44(1):81-4.
[7] Metz JM, Devine P, DeNittis A, Jones H, Hampshire M, Goldwein J, Whittington R. A multi-institutional study of Internet utilization by radiation oncology patients. *Int J Radiat Oncol Biol Phys*. 2003 Jul 15;56(4):1201-5.
[8] Richardson MA, Sanders T, Palmer JL, Greisinger A, Singletary SE. Complementary/alternative medicine use in a comprehensive cancer center and the implications for oncology. *J Clin Oncol*. 2000 Jul;18(13):2505-14.
[9] Sagaram S, Walji M, Bernstam E. Evaluating the prevalence, content and readability of complementary and alternative medicine (CAM) web pages on the internet. *Proc AMIA Symp*. 2002:672-6.
[10] Ernst E, Schmidt K. 'Alternative' cancer cures via the Internet? *Br J Cancer*. 2002 Aug 27;87(5):479-80.
[11] Health on the Net. [accessed 11-27-2006]; http://www.hon.ch/
[12] Eysenbach G, Kohler C. How Do Consumers Search For and Appraise Health Information on the WWW? *BMJ*. 2002 March 9;324.
[13] Bernstam EV, Shelton DM, Walji M, Meric-Bernstam F. Instruments to assess the quality of health information on the World Wide Web. *Int J Med Inform*. 2005 Jan;74(1):13-9.
[14] Kim P, Eng TR, Deering MJ, Maxfield A. Published criteria for evaluating health related web sites: review. *Bmj*. 1999 Mar 6;318(7184):647-9.
[15] Bernstam EV, Sagaram S, Walji M, Johnson CW, Meric-Bernstam F. Usability of quality measures for online health information. *Int J Med Inform*. 2005 Aug;74(7-8):675-83.
[16] Ademiluyi G, Rees CE, Sheard CE. Evaluating the reliability and validity of three tools to assess the quality of health information on the Internet. *Patient Educ Couns*. 2003 Jun;50(2):151-5.
[17] Walji M, Sagaram S, Sagaram D, Meric-Bernstam F, Johnson C, Mirza NQ, Bernstam EV. Efficacy of quality criteria to identify potentially harmful information. *J Med Internet Res*. 2004 Jun 29;6(2):e21.
[18] Price SL, Hersh WR. Filtering Web pages for quality indicators: an empirical approach to finding high quality consumer health information on the World Wide Web. *Proc AMIA Symp*. 1999:911-5.
[19] Fallis D, Fricke M. Indicators of accuracy of consumer health information on the Internet. *J Am Med Inform Assoc*. 2002 Jan-Feb;9(1):73-9.
[20] Fricke M, Fallis D, Jones M, Luszko GM. Consumer health information on the Internet about carpal tunnel syndrome. *Am J Med*. 2005 Feb;118(2):168-74.
[21] Griffiths KM, Tang TT, Hawking D, Christensen H. Automated assessment of the quality of depression websites. *J Med Internet Res*. 2005;7(5):e59.
[22] Tang TT, Craswell N, Hawking D, Griffiths KM, Christensen H. Quality and Relevance of Domain-specific Search: A Case Study in Mental Health. *Info Retr.* 2006;9(2):207-25.
[23] Aphinyanaphongs Y, Tsamardinos I, Statnikov A, Hardin D, Aliferis CF. Text Categorization Models for High Quality Article Retrieval in Internal Medicine. *J Amer Med Inform Assoc*. 2005;12(2):207-16.
[24] Cancer Patients Seeking Alternative Treatments. [accessed 11-26-2006]; http://www.quackwatch.org/00AboutQuackwatch/altseek.html
[25] Cohen J. A coefficient of agreement for nominal scales. *Education and Psych Measurement*. 1960;20(1):37-46.
[26] Chang C, C. L. LIBSVM. 3-13-2006 [accessed; http://www.csie.ntu.edu.tw/~cjlin/libsvm
[27] Science of Quackometrics. [accessed 11-26-2006; http://www.quackometer.net/blog/2006/04/science-of-quackometrics.html
[28] Quackometer. [accessed 11-26-2006]; http://www.quackometer.net/?page=quackometer
[29] Brin S, Page L. The anatomy of a large scale hypertextual Web search engine. *Computer Networks and ISDN Systems*. 1998;30:107-17.

Address for correspondence

Yin Aphinyanaphongs, ping.pong@vanderbilt.edu

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Currency of Online Breast Cancer Information

Funda Meric-Bernstam, M.D.,[1], Muhammad Walji Ph.D..,[2] Smitha Sagaram M.B.B.S., M.S.,[2] Deepak Sagaram M.B.B.S., M.S.,[2] Elmer Bernstam, M.D., M.S.E. [2]

[1]The University of Texas M. D. Anderson Cancer Center and
[2]School of Health Information Sciences, The University of Texas Health Science Center at Houston, Houston, TX

Abstract and Objective

Consumers are increasingly turning to the Web, expecting to find the latest health information. The purpose of this study was to assess the currency of online breast cancer information. We determined whether nine recent advances in breast cancer management were incorporated into 337 unique breast cancer Web pages. Two reviewers independently assessed content; if a Web page covered appropriate advances it was deemed to be "current." Of the 337 Web pages, 89 contained one or more advances. Of the 122 Web pages that had dates of update available, 49% had been updated within 6 months. Only 11%-37% of Web pages covered clinically accepted advances, even among Web pages that were updated after acceptance of the advance into clinical practice. We conclude that online health information is often not sufficiently current. Consumers searching for health information online should always consult an expert clinician before taking action.

Keywords:

breast neoplasms, internet, medical informatics, patient satisfaction

Introduction

The American Cancer Society estimates that over 210,000 women will be diagnosed with breast cancer in 2006. Over 40,000 will die of the disease in the same year (1). Nearly half of women newly diagnosed with breast cancer turn to the Internet for health information (2). The Pew Internet and American Life Project estimates that every day more patients seek health information online than consult physicians (3). The recent US National Cancer Institute-sponsored Health Information National Trends Survey (HINTS) found that health care providers remain a trusted source of health information. However, many respondents went online for health information before going to their physician (4). Consumers are satisfied with their online experience and are making treatment choices based on the information that they encounter (5, 6). Therefore, online information has the potential to significantly impact the health of breast cancer patients.

When breast cancer patients search for health information online, they expect to find the latest advances in medical science. Although there has been appropriate concern regarding the quality of online health information, there are few objective studies of online information currency that, in some cases, is equally important. In this study, we sought to determine whether Web pages that discussed specific breast cancer topics also presented information about important recent advances pertaining to those topics

Methods

Web site selection

We used five popular search engines—Google, Yahoo Directory, AltaVista, Overture, and AllTheWeb—to identify Web sites that consumers are likely to encounter. We used Yahoo Directory because at the time that the study was performed, Google provided the non-directory Yahoo Web search results. For each search engine, we performed 15 searches using the most frequently encountered topics from the first 322 breast cancer-related entries in the NetWellness database of user questions (7) . All searches were performed from June 1 to July 30, 2004. As most Web searchers review only the first page of results (8), we recorded all URLs found on the first search results page, including sponsored (advertisements) and unsponsored results (n = 1585). After we eliminated duplicate URLs, 870 remained.

The URLs were reviewed for relevance (i.e., presence of breast cancer content) by two independent reviewers. If a URL was not relevant, the links on the first page were reviewed and the first relevant link was included in the study. Reviewers were instructed to follow links sequentially, starting from the top of the page, left to right. The reviewers agreed that 344 URLs were relevant. However, each reviewer identified 25 URLs that were unique (e.g., followed different links from the original page). To avoid bias, we included all of these in the analysis, but excluded additional duplicates found at this stage (n=15) (final n = 379).

Web pages were downloaded to a local computer to create a static data set for analysis. Because online content changes continuously, the information stored in search engine databases is usually not an exact copy of the Web in real time. Thus, some sites and pages were unavailable (down) or required redirection and could not be downloaded. Pages were downloaded for two review efforts: the original URL was downloaded for technical quality assess-

ment and the original URL plus pages up to two links away were downloaded for assessment of topics and advances. Downloading the original URL generated four errors, leaving 375 pages. Downloading the page plus pages up to two links away from the original URL generated 30 errors, leaving 349 pages. Six more pages were eliminated at this stage. These included four additional duplicates and two URLs that could not be evaluated for technical quality criteria. One was a survey, not appropriate for analysis and another page was downloaded but without content (empty) and could not be located online. Therefore, a total of 343 pages were evaluated for technical quality criteria as well as breast cancer topics (Table 2, column 2) and advances (Table 2, column 1).

Table 1 - Advances in breast cancer, and number of times they were identified in the survey

Advance	**N**
Sentinel node biopsy*	10
Aromatase inhibitors for adjuvant therapy*	9
Taxanes*	9
Tamoxifen for chemoprevention*	6
Trastuzumab*	5
Minimally invasive breast biopsy*	2
Post-mastectomy radiation*	2
Preoperative chemotherapy for tumor downsizing	2
Partial breast irradiation*	2
Increased public awareness	2
Individualized therapy	2
Estrogen receptor testing for ductal carcinoma in situ	1
Tamoxifen for ductal carcinoma in situ	1
In situ ablation of breast tumors	1
Outpatient treatment	1
Support groups	1
Easier tolerated treatments	1
Demonstration of lack of efficacy of bone marrow transplantation	1
Focused physician training in breast disease	1

* Top nine chosen for further analysis

Identification of advances in breast cancer

A convenience sample of breast oncologists at the University of Texas M. D. Anderson Cancer Center (MDACC) was surveyed to identify advances in breast cancer screening, diagnosis and therapy within the past five years. Free text responses were requested in via email or via personal interviews. Twelve breast oncologists (four surgical oncologists, seven medical oncologists, one radiation oncologist) identified 19 advances (Table 1). Nine specific advances that were identified by more than one oncologist were chosen for further study (Table 2). One advance was related to surgical therapy, three were related to systemic therapy, two related to radiation therapy, one to cancer prevention, and one was related to timing of systemic therapy and surgery.

Results

To determine the currency of online breast cancer information, we identified advances associated with specific breast cancer topics. For example, if a Web page covered breast surgery, it should also discuss sentinel node biopsy, an important advance in surgical therapy. If a Web page covered the appropriate advance it was deemed to be "current." Of the 337 Web pages that covered one or more topics (98% of 343 pages analyzed), 89 (26%) contained one or more corresponding advance. The topics covered and the percentage of Web pages that were current for each advance are shown in Table 2. This was determined as the average of two independent clinically-trained reviewers. Reassuringly, the maximum deviation from the mean for any reviewer score was 2.1%.

Of the 337 Web pages in the study, 124 had listed their date of update. Of these, the date of update on two was not available, due to a technical problem with page fixation. The time between creation of static data set and the last date of update listed for the 122 remaining Web pages are shown in Figure 1. Of the Web pages that listed their date of update, nearly half (49%) had been updated in the past 6 months. The mean time from update was 20 months for commercial (.com) sites, 15 months for organizations (.org/.net) and 14 months for government (.gov) sites. For educational (.edu) sites, the mean time from update was 42 months, but this was based on only two sites.

Once an advance is shown to be effective and introduced as generally accepted or preferred therapy, it should be incorporated into information presented on breast cancer Web pages covering that topic. The date when the advance was considered to be generally accepted was determined as follows. When the advance was a medication, we used dates of approval by the U.S. Food and Drug Administration (FDA) for the corresponding indication. For other interventions, the date of publication of manuscripts describing acceptance of these advances based on consensus conferences or National Comprehensive Cancer Network (NCCN) clinical practice guidelines were used as the date of the advance. Of the advances selected, all but one, partial breast irradiation after breast-conserving therapy, had been determined to be either acceptable practice, or standard of care at the time of the study. Among the 122 Web pages with known dates of updates, we determined the percentage "current" among the Web pages updated

Table 2-Evaluation of content topic and currency

Advance	Topic	Current (%)	Date of Advance	Pages Updated after Date of Advance that are Current (%)
Non-surgical core biopsy	Diagnosis	37%	9/2001 (International Breast Cancer Consensus Conference) (9)	35%
Tamoxifen for chemoprevention	Cancer risk and risk reduction	23%	11/29/1998 (FDA)	27%
Sentinel node biopsy	Surgical therapy	26%	6/30/2002 (Consensus Conference on Sentinel Node Biopsy) (10)	35%
Aromatase inhibitors for adjuvant therapy	Systemic therapy	17%	9/5/2002 (FDA)	29%
Taxanes	Systemic therapy	11%	10/1999 (FDA)	12%
Trastuzumab	Systemic therapy	14%	11/25/1998 (FDA)	16%
Pre-operative chemotherapy for tumor downsizing	Systemic therapy Surgical therapy	13% 16%	11/2000 (NCCN guidelines) (11)	10% 14%
Post-mastectomy radiation	Radiation therapy	37%	9/15/1999 (American College of Therapeutic Radiology and Oncology Consensus statement) (12)	31%
Partial breast irradiation	Radiation therapy	2%	N/A	N/A

after each advance was introduced into standard clinical practice (Table 2).

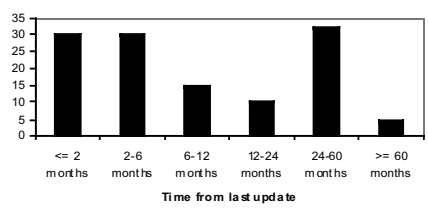


Figure 1 - Time from last date of update

Discussion

Consumers turn to the Web expecting to find state-of-the-art health information. We determined whether nine recent advances in breast cancer management were incorporated into 337 unique Web pages pertaining to those topics. Our data suggest that online breast cancer information is not adequately current, even when pages were recently updated.

Our study is limited by the fact that we focused on one disease state, thus the generalizability of our findings to other topics and search terms, is unknown. However, breast cancer is a common malignancy, and advances in breast cancer management often receive intensive media coverage. Thus, one would expect that online breast cancer information would be especially current.

Only 11-37% of Web pages covered the eight clinical advances that had been determined to be either acceptable practice, or standard of care at the time of the study. All eight advances had been "generally accepted" for at least 22 months prior to our study. It should be noted that the date we selected as "general acceptance" was based on FDA approval or consensus conference, which often lagged by months to years after proof of efficacy by clinical trials. In contrast, oral presentation of a single randomized clinical trial at a national conference in May 1998 was temporally associated with an increase in the use of taxanes for breast cancer in the community; even before study publication in a peer-reviewed journal or FDA

approval in October 1999 (13). This suggests that online information, at least in breast oncology, may lag behind clinical practice.

At the time of the study, the ninth advance, partial breast irradiation, had been shown to be effective in a study conducted at a single institution (14), but was still being tested against the standard of care, whole breast irradiation, in randomized multi-center trials. Only 2% of Web pages that discussed radiation therapy, also discussed partial breast irradiation.

Date of update is a widely recognized Web site quality criterion although it has not yet been shown to correlate with information accuracy (15). In this study, of the 122 Web pages that listed their date of update, nearly half had been updated in the past 6 months. However, the information displayed was "current" in only 10% - 35% of Web pages updated after the medical advances of interest. Thus "currency" based on date of update does not ensure the "currency" of the online information content.

Conclusion

Medicine is evolving rapidly, making it challenging for clinicians themselves to remain up to date. "Current" online health information could provide patients the opportunity to learn about the standard of care and all available options. This would empower the educated consumer to seek out additional options or second opinions in a timely manner. However, our data suggest that online information is not sufficiently current to enable this opportunity.

Acknowledgements

We thank Ms. Marlen Banda, The University of Texas M. D. Anderson Cancer Center, for assistance in data retrieval. We thank the breast oncology group at M. D. Anderson for their clinical input. This study was supported in part by a training fellowship from the Keck Center for Computational and Structural Biology of the Gulf Coast Consortia (NLM Grant No. 5T15LM07093) (M.W.), NLM Grant No. 5K22LM008306 (E.V.B.), a grant from the Robert Wood Johnson Foundation Health-e-Technologies Initiative (E.V.B., F.M-B), and NCRR Grant No. 1UL1RR024148 (E.V.B., F.M-B).

References

[1] Ries LAG HD, Krapcho M, Mariotto A, Miller BA, Feuer EJ, Clegg L, Eisner MP, Horner MJ, Howlader N, Hayat M, Hankey BF, Edwards BK (eds). SEER Cancer Statistics Review, 1975-2003. 2006 [cited 2006 December 2]; Available from: http://seer.cancer.gov/csr/1975_2003/

[2] Satterlund MJ, McCaul KD, Sandgren AK. Information gathering over time by breast cancer patients. J Med Internet Res. 2003 Jul-Sep;5(3):e15.

[3] Fox S, Rainie L. Vital decisions: how internet users decide what information to trust when they or their loved ones are sick. Washington, DC: Pew Internet & American Life Project; 2002 May 22, 2002.

[4] Hesse BW, Nelson DE, Kreps GL, Croyle RT, Arora NK, Rimer BK, et al. Trust and sources of health information: the impact of the Internet and its implications for health care providers: findings from the first Health Information National Trends Survey. Arch Intern Med. 2005 Dec 12-26;165(22):2618-24.

[5] Fox S, Rainie L. The online health care revolution: How the Web helps Americans take better care of themselves. Washington DC: Pew Internet and American Life Project: Online; 2000 Novermber 26.

[6] Helft PR, Hlubocky FJ, Gordon EJ, Ratain MJ, Daugherty C. Hope and the media in advanced cancer patients. American Society of Clinical Oncology 36th annual meeting; 2000 2000; New Orleans, LA; 2000.

[7] Ask an Expert. 2003.

[8] Eysenbach G, Kohler C. How do consumers search for and appraise health information on the world wide web? Qualitative study using focus groups, usability tests, and in-depth interviews. BMJ. 2002 Mar 9;324(7337):573-7.

[9] Image-detected breast cancer: state of the art diagnosis and treatment. International Breast Cancer Consensus Conference. J Am Coll Surg. 2001 Sep;193(3):297-302.

[10] Schwartz GF, Guiliano AE, Veronesi U. Proceeding of the consensus conference of the role of sentinel lymph node biopsy in carcinoma or the breast April 19-22, 2001, Philadelphia, PA, USA. Breast J. 2002 May-Jun;8(3): 124-38.

[11] Carlson RW, Anderson BO, Bensinger W, Cox CE, Davidson NE, Edge SB, et al. NCCN Practice Guidelines for Breast Cancer. Oncology (Williston Park). 2000 Nov;14(11A): 33-49.

[12] Harris JR, Halpin-Murphy P, McNeese M, Mendenhall NP, Morrow M, Robert NJ. Consensus Statement on postmastectomy radiation therapy. Int J Radiat Oncol Biol Phys. 1999 Jul 15;44(5):989-90.

[13] Giordano SH, Duan Z, Kuo YF, Hortobagyi GN, Freeman J, Goodwin JS. Impact of a scientific presentation on community treatment patterns for primary breast cancer. J Natl Cancer Inst. 2006 Mar 15;98(6):382-8.

[14] Vicini FA, Kestin L, Chen P, Benitez P, Goldstein NS, Martinez A. Limited-field radiation therapy in the management of early-stage breast cancer. J Natl Cancer Inst. 2003 Aug 20;95(16):1205-10.

[15] Walji M, Sagaram S, Meric-Bernstam F, Johnson CW, Bernstam EV. Searching for cancer-related information online: unintended retrieval of complementary and alternative medicine information. Int J Med Inform. 2005 Aug;74(7-8):685-93.

Address for correspondence

Elmer V. Bernstam, MD, MSE
Associate Professor
School of Health Information Sciences and Internal Medicine
The University of Texas Health Science Center at Houston
7000 Fannin Street, Suite 600
Houston, TX 77030
Email: Elmer.V.Bernstam@uth.tmc.edu

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Children's Contributions to Designing a Communication Tool for Children with Cancer

Cornelia M. Ruland, RN. Ph.D[ab], Laura Slaughter, Ph.D[a], Justin Starren, Ph.D[c], Torun M. Vatne, MS[a], Elisabeth Y. Moe, Bsc [a]

[a] *Center for Shared Decision Making & Nursing Research, Rikshospitalet Medical Center, Oslo, Norway,*
[b]*Department of Biomedical Informatics, Columbia University, NY*
[c]*Biomedical Informatics Research Center, Marshfield Clinic Research Foundation, Marshfield, WI, USA*

Abstract

In this paper, we describe the roles played as well as contributions made by child participants in the design of an innovative communication tool for children with cancer. SISOM is a handheld, portable computer application with a graphical user interface that is meant to: (1) help children with cancer communicate their symptoms / problems in a child-friendly, age-adjusted manner; and (2) assist clinicians in addressing children's experienced symptoms and problems in patient care. Unlike other applications for children, the purpose of SISOM is not to provide information to ill children but to elicit personal information from them. Thus the application has a unique set of design issues. Healthy and ill children played an important role in different stages in the design process. They made significant contributions to the graphical design of the system's interface; selection of understandable, child-friendly terms used by the system to describe symptoms; iconic and graphical representations; and its usability. We describe the participatory design methods we used that included children and share important insights from this collaborative design process.

Keywords:

participatory design, usability, children, cancer

Introduction

Children diagnosed with cancer are particularly vulnerable. There are many complex physical, functional, psychosocial and behavioral problems associated with having cancer. Children lack life experience and the maturity (both emotional and cognitive) that can equip them to cope with and make sense of their illness. To prevent unnecessary suffering, health care providers need to understand cancer symptoms and problems from a child's perspective. This can be difficult because children vary widely in their illness experiences. Furthermore, children may be prevented from conveying distressing symptoms due to their less developed verbal skills, as well as adults' communication styles and attitude that they should speak on child's behalf. Recent work has shown that clinicians are often unaware of symptoms and problems experienced by pediatric cancer patients [1; 2], which means that many symptoms remain under- diagnosed or untreated.

Children with cancer can benefit from communication support technology that provides them with a "voice"; such a tool can help health care providers understand and address symptoms from the child's perspective. A communication support system for adult cancer patients, called Choice, has been successfully used by patients to report their symptoms, problems and preferences for care. It significantly increased congruence between patient problems and patient care. [3;4] Children with cancer may have even more to gain from similar support, but so far no such systems have been developed for pediatric oncology. Therefore, we developed SISOM, a handheld, portable computer application to: (1) help children with cancer age 7-12 communicate their symptoms/problems in a child-friendly, age-adjusted manner; and (2) assist clinicians in better addressing children's experienced symptoms and problems in patient care.

Developing a tool like SISOM, a communication support system for seriously ill children, is a new area in medical informatics. When seriously ill children are the end users, the challenge is to adapt the application to children's cognitive and emotional developmental stages. SISOM is unique compared to other computer applications developed for children that are mainly computer games or have been developed for informational or educational purposes. These systems primarily deliver information or interactions *to* children. The purpose of most technology designed for ill children so far has been to help them manage and cope with their illness through educational material and play. Systems designed to support patient care have generally focused on obtaining information from parents rather than the sick children. For example, in a recent study Porter et al. [5] built an asthma kiosk within their hospital's emergency department so that parents can provide the critical information and symptoms experienced in order to drive guideline-based care for pediatric asthma. However, no applications have been developed so far that directly elicit and help children communicate their symptom experiences to their care providers as a means to improve patient-centered care.

Selected for best paper award.

Involving children in the design process of an application such as SISOM is critical. Children are experts on being children and can tell us what appeals to them and what they master and understand. When developing SISOM we actively included healthy and ill children in different stages of the design, and children made significant contributions. The design process involved several smaller studies with children as participants. In this paper, we describe the roles children played in the design of SISOM and their contributions to: (1) the graphical user interface (interaction and metaphors); (2) understandable, child-friendly terms that describe symptoms and problems; (3) icons and pictures within the system; and (4) usability. Working with children provided us with important insights that may be helpful for other system developers who wish to design support applications for ill children.

Materials and methods

Children's participation in designing the graphical user interface

Children can play an important role in creating new technologies for other children. Participatory Design (PD) is an essential aspect of good design practice for both adults and children [6]. In our work, we employed PD methods, outlined by Druin [5], that have been successfully applied in the design of story board software, digital libraries for children, computer games as well as educational software [6;7]

PD usually implies that the user participants are as similar as possible as the system's future users. However, when designing an application for seriously ill children, this is only partially possible. PD involves repeated meetings of 1,5-2 hours over an extended period of time [6]. This would pose an impossible burden on children undergoing cancer treatment. When designing SISOM, we therefore, worked together with healthy children, as they still have the aspect of being children in common with our target cancer population. However, ill children participated in other steps of the interface design described later.

Recruitment and sample

For the design sessions and all other methods described in this paper that involved children, IRB approval was obtained. To recruit children to participate in the design of the graphical interface, the Principal of a nearby elementary school in Oslo, Norway, was contacted, asking her permission to send letters to parents of 4th and 6th graders. The letter invited children to participate in the interface design of SISOM, explained the purpose of the system, and outlined the time requirements that PD sessions were 2 hours one afternoon a week, up to 6 weeks in total. Responses were overwhelming. Fifty children (17 boys and 33 girls) replied with interest in participating. As this was far more than needed, we selected final participants based on a phone conversation with their parents, where we screened children based on whether the child: 1) feels comfortable with and is usually active in group activities, 2) uses computers as an educational tool and to play games, 3) is creative, e.g. likes to draw or enjoys building things, and (4) if there was a particular reason why the child wanted to participate. The final group consisted of 12 children who worked in two separate design groups: one group of six 4th graders (9 year olds) and one group of six 6th graders (11 year olds). Other children who had volunteered did participate in other tasks described below.

Design sessions

Design sessions were held at the "Adolescent Club Room" within the pediatric department in Norway's National Hospital, Rikshospitalet, Oslo, Norway, in order to provide a child/adolescent-friendly atmosphere. Both groups attended four sessions after school over a period of two months. Each session lasted two hours. Two of the participatory design sessions were pilot tested with four children of hospital employees (2 boys 2 girls, 9 and 11 years old). Ideas that emerged from the pilot sessions were included in the analysis of design ideas.

Four adults who where trained as "participants" or "observers" were present during each session. Sessions proceeded with the idea that design is an iterative process and a mutual learning experience where both children and adults are experts. The children contributed design ideas, explained which aspects of computer interfaces they found appealing, and gave reasons why some interfaces are better than others. Adults showed the children interface and navigation examples and helped them understand what is technically feasible. Adults also ensured that the system's interface evolved according to a set of system specifications that were defined upfront. Our techniques included role play and scenarios, low-tech prototyping, observation, video-taping and note taking. All groups followed the same procedures.

The first session consisted of three tasks. After introducing each other, and informing the children about the project purpose and plans for the next couple of weeks, children were asked to work in pairs to test four computer programs or games each for about 30 minutes. Each pair had an adult facilitator and an observer who took notes. The purpose was to: (a) learn what interface features and interaction styles children like in computer applications; (b) show children what different features and interactions are possible, in order to stimulate their ideas for SISOM; and (c) as a warm-up. Children were asked to think aloud during this task and comment on what they liked/ disliked about each application, whether it was easy /difficult, engaging/boring and why. They were also asked to describe their favorite computer application and why it was their favorite.

In the next task, children were read a scenario about a child that was sick and had to go to the hospital. Children were asked to tell us and role play what these symptoms felt like. We choose a stomach flu scenario rather than a cancer story because we anticipated that children were more likely to have had a real life experience with flu rather than with cancer. We also felt that having a detailed scenario dealing with cancer might make the children worry or frighten them. However, children were asked to role play symptoms that are contained in SISOM, such as nausea, fatigue, and feeling sad. The story ended with a nurse com-

ing to the hospitalized child, handing her a tablet computer that the child could use to report its symptoms, problems and worries, so the doctor and nurse could help.

The children were then asked to start designing a system based on the scenario they had just heard about, and draw or build what it could look like and how it would behave. Children were provided with a large table equipped with low-tech prototype material, including paper in different colors, crayons, pencils, cardboard, cartoons, lego, glue etc. and asked to work on the task in groups of 2-3. Towards the end of each session, each sub-group was asked to summarize their ideas so that the rest of the group could provide feedback.

In sessions two-four, the team continued to draw and discuss different aspects of the interface, either in pairs or as a whole group. All sessions were videotaped, and notes were taken by adult observers.

Figure 1 - Drawings made during PD sessions

We made particular efforts to help our "healthy participants" understand the purpose and context of SISOM since they were not our real end-users. Therefore, all sessions started with a story of a sick child that the children could recognize or easily imagine. Additional symptoms and problems were introduced, again from the SISOM symptom list. In the third session, the children role-played an actual hospital scene where they took turns being dressed in a hospital gown, and put into bed with a pretend I.V. line. The other children played the role of doctor, nurse or worried parent. The "sick" and bandaged child in bed tested different computer types and input devices and was asked for feedback. They also tested the adult application for cancer symptom assessment that has been successfully used by cancer patients over the last years. Children found the adult version relatively easy-to-use, but naturally found its interface boring because symptoms are only presented as text in tables.

Between sessions, the adult team discussed ideas and observations from the previous session, monitored progress, summarized what was learned, and adjusted, when necessary, the agenda for the following session. The graphical designer on the team drew out children's ideas in Adobe Illustrator that were given back to the children for further evaluation and elaboration in the next session. Figure 1 shows an example drawn by two 11 year old girls that depicts their idea of how SISOM could help children report where it hurts. The left side displays the children's drawing; the right side how this idea was refined. In the final version the figure can be displayed as a boy or a girl, dependent on the gender of the child user, and it can also be turned to show the posterior side.

Table 1 – Children's Design Contributions

Idea Category	**Count**
Age /gender specific user choices	4
Animations, colors, graphics, media	19
Background scene	11
Interactions	30
Sound, text, voice	27
Main characters, avatar	16
Navigation	27
Help functions	11
Input	10
Likes, dislikes, experience	6
All	161

The final data from the design sessions consisted of the low tech prototypes that were developed, observation notes, and 22 hours of videotape. Videos and observation notes were annotated and organized according to topic, such as ideas for background, help, navigation, animation, graphics, age and gender specific content, as well as additional material. The children came up with a total of 161 design ideas altogether, but some of the ideas were the same or very similar; we counted a total of 109 (68% of 161) unique ideas. Children's idea types are summarized in table 1.

Children had many excellent and creative suggestions that the design team would not have thought of. For example, one design idea that emerged from the sessions and was endorsed by both boys and girls in both age groups was a "sailing from island-to-island" navigation theme. The children suggested that symptoms be placed on islands and that the sick children can sail from place to place in order to select symptoms. Also, one idea was that each child could create a main character and select its appearance (hair, cloth etc). However, not all ideas were workable. For example, when designing an interface for younger children who cannot read, one group suggested that a voice could be activated by clicking a button where: "Click here if you cannot read" is written on it. One group of boys suggested

shooting at the body picture with a gun to mark where it hurts.

Children's contributions to child-friendly terms

The participatory design sessions were preceded by the development of the list of symptoms to be contained in SISOM. This list was based on a critical review of 98 articles from the scientific literature that addressed symptoms and problems encountered by children with cancer along physical, functional, psychosocial and behavioral dimensions. In addition, we conducted focus groups with clinical specialists (physicians, nurses, psychologists, social workers) and parents who critically reviewed the symptoms abstracted from the literature and supplemented them with expert opinion. They were also asked what terms and expressions children used when communicating about symptoms with them. This resulted in a preliminary list of 78 child-friendly symptoms and problem expressions.

To ensure that children could understand the symptom and problem terms contained in SISOM, we interviewed 14 children (7 healthy volunteers, 5 with cancer, age 8-12). We presented them with one symptom at a time in random order and asked them about: (a) their understanding of the term: "what does it mean to be… (e.g. nauseated?)"; (b) alternative expressions: "how would *you* say it to let others know that you were…?"

Interviews were audio-taped, transcribed and analyzed. Symptom/problem terms were coded into: (a) clearly understandable terms; (b) ambiguous terms; (c) does not know what the term means; (d) familiar with the term but assigns a different meaning to it. Terms that were not clearly understandable were revised. Children provided several excellent suggestions for meaningful child-friendly terms that were used in these revisions.

Evaluations of graphical representations

PD sessions provided us with design ideas that were implemented in a prototype that subsequently had to be validated. Graphical elements function best when they are metaphorically meaningful and correctly represent the concept in question [8]. In SISOM each symptom is represented with a picture. Therefore, we conducted evaluation sessions with five children from our volunteer group age 9-11, to learn if they could correctly recognize the symptoms based on the graphical representations. Children were presented with one symptom picture at a time without providing them with any labels, or category headings under which they belonged. This approach evaluated the pictures only; in SISOM, symptoms are displayed in a graphical background that provides the context and text labels that can be heard by clicking on an icon if the child cannot read. Figure 2 displays a picture of the symptom "difficulty sleeping" that was immediately recognized by all children.

Figure 2 – Difficulty sleeping

A facilitator conducted the interviews and an observer took notes. Sessions lasted on average 45 minutes. For each picture, children were asked what symptom they thought it depicted and why. Their explanations helped us identify cues in the picture that made the symptom recognizable. If children did not recognize the symptom, we provided them with the category name under which it belonged. Children were provided with coloured pencils to revise the pictures or draw new ones.

Pictures were grouped into four categories: A: the symptom was correctly recognized immediately; B: was recognized after knowing the category name under which it belongs; C: knowing the symptom, the picture is a good representation; D: the picture does not represent the symptom well. 61% of the symptoms were immediately recognized by all children. 12% of the pictures were considered not a good symptom representation and were revised. Children provided a number of excellent ideas for revisions that we used.

Children as usability testers

We performed several usability tests with healthy and sick children at different stages of the development process. Usability testing encompasses a range of methods for identifying how users actually interact with a prototype or a complete system. It is an iterative process that involves testing the system and then using the test results to change it to better meet users' needs. The best process is to try out a prototype with a few users, fix it, and test it again. [9]

At the current stage of SISOM's development, four healthy school kids and two ill children at the hospital have participated in usability testing. They were asked to use the system and select a set of symptoms. They were prompted to think aloud during the task which was videotaped. Morae™ software was used for automatic recording and analysis of all events on the screen, such as when the user clicked on an object, opened a dialog box, or viewed /listened to specific text.

Usability testing showed that children in our target age group are very computer savvy, and have no fear of clicking on and finding things intuitively. Few initial instructions sufficed. All children liked the ability to select the main figure, and the metaphor of sailing through the island world provided them with a sense of discovery.

However, we made the interesting observation that a system can evoke very different associations in children with the experience of a life threatening illness than in healthy children. For example, in an early prototype, when the user was done with a symptom and zoomed out of the picture a spark appeared, the main character disappeared, and did not re-appear before the child had clicked on another picture. One of the sick children thought that the main figure exploded (= died). This made him very uncomfortable. Thus transitions between pictures were subsequently changed. The healthy kids had no such associations and thought that the spark feature was cool. It became clear from this experience that sick children who are the end-users are important usability testers, and healthy children cannot serve as their proxies in this task.

Discussion

We can summarize several important observations from working with children in the design of SISOM. In our project, children were involved in a number of tasks and made significant contributions to the system's graphical interface, child-friendly terms, iconic and graphical representations, and usability. Based on the tasks we gave them, children were able to contribute very useful ideas that the adult designers would not have thought of and considerably improved the software. It was also crucial that they participated in evaluations of the system.

However, children have also limitations. They are not professional designers and do not always have a good grasp of logical design. Good design for ill children requires insights and knowledge that children do not have (e.g. expert knowledge of children's motor and cognitive abilities, pedagogy, and child psychology). We had to make sure to meet the goals of SISOM and a set of pre-defined criteria. Specially, we had to ensure that the software allows health care providers to obtain valid and reliable data from children about their symptoms in a clinical context, without being too time-consuming and challenging. Our child participants often focused on designing fun and often time-consuming aspects, such as funny noises along with vivid animations for a symptom such as throwing up. In spite of reminders that we were not designing a game, the children had the tendency to slip back into a "game mode". Also, children could spend considerable time on fine details and lose sight of the overall purpose. For example, they could spend a whole session on drawing flowers in a background landscape or on ways to choose eye and hair color for a figure created to navigate the system. Thus the knowledge of professional designers is crucial when it comes to final decisions about which ideas to implement.

Comparing the contributions of 9-year olds to 11 year olds, we did not see any significant differences. Both did well in performing the tasks and suggested creative design ideas that could be used with children in their own age group. Working with two age groups was very valuable. Older children in our group were very sensitive to not creating a system that appeared childish. Smaller children had difficulty coming up with ideas suited for an older age group. Boys suggested a lot of action, guns, rockets, a race course, shooting etc. While not completely free of action, girls' ideas contained more flowers, animals, and "softer" effects. This observation suggests that it may be worthwhile to not only offer age-adjusted interface choices, but also gender-adjusted options.

A challenge in the design of SISOM was to balance participation of healthy and ill children. The burden and time required for some of the tasks prevented us from asking sick children to participate, e.g. in design sessions focusing on the graphical user interface. This however, raises the question whether healthy children can conceptualize what it is like to be suffering from a serious illness, and thus the degree to which they can serve as proxies in participatory design and evaluations. Our experience is that this is only partially possible. Role play and scenarios certainly helped to increase healthy children's understanding of what SISOM was for. Still, it was difficult for them to grasp the context in full. We made however an interesting observation. One of the 9-year olds had a cousin with lymphoma and thus had been exposed to cancer in her family. While we have no other empirical data to support this, this child appeared to have a more mature grasp of the purpose of the system. Her knowledge of some aspects of cancer, through her life experiences with her cousin, caused her to have more focused ideas based on a more sophisticated knowledge. Thus it seems that personal experience is an important factor for valuable design contributions. The limitation of healthy children may be partially compensated through extensive usability testing where participation of ill children as end-users is imperative. This will allow us to discover weaknesses in the design where sick children can contribute ideas for improvement.

Acknowledgements

This study was supported by grants from the Norwegian Research Council, The Norwegian Nurses' Association, Rikshospitalet's research fund, and Medinnova. For their valuable contributions we thank: Roar Andersen, Yizhak Itzchaki, Denis Pototylyuk, Birgitte Kolbeinsen, Simon Hjelle, and Glenys Hamilton.

References

[1] Patenaude AF, Last B: Cancer and children: where are we coming from? Where are we going? Psycho-Oncology 2001;*10:*281-283.

[2] Camp, LD: A comparison of nurses' recorded assessments of pain with perceptions of pain as described by cancer patients. Cancer Nursing 1988;*11:*237-243.

[3] Ruland CM. Handheld Technology to Improve Patient Care: valuating a Support System for Preference-based care Planning at the Bedside. *.Am.Med.Inform.Assoc.* 2002;9(2):192-201.

[4] Ruland CM, White T, Stevens M, Fanciullo G, Khilani SM. Effects of computerized system to support shared decision

making in symptom management of cancer patients: preliminary results. JAMIA 2003; 10(6):573-579.
[5] Druin A: The Design of Children's Technology. San Francisco CA, Morgan Kaufmann Publishers, 1999.
[6] Scaife M, Rogers Y, Aldrich F, Davies M: Designing for or designing with? Informant design for interactive learning environments; Proceedings of CHI '97, ACM Press, 1997, pp 343-350.
[7] Marcus A. Icons, Symbols and Signs: Visible languages to facilitate communication, 2003.
[8] National Cancer Institute, United States National Institutes of Health, Bethesda, Md Usability Website. http://www.usability.gov/methods/usabilitytesting.html
[9] Porter SC, Kohane IS, Goldmann DA: Parents as Partners in Obtaining the Medication History. JAMIA 2005;*12:* 299-305.

Address for correspondence

Cornelia Ruland, RN, Ph.D
Center for Shared Decision Making and Nursing Research
Rikshospitalet-Radiumhospitalet HF Forskningsvn. 2b
0027 Oslo, Norway.
E-mail: cornelia.ruland@rikshospitalet.no

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

"It's Your Game": An Innovative Multimedia Virtual World to Prevent HIV/STI and Pregnancy in Middle School Youth

Ross Shegog, Christine Markham, Melissa Peskin, Monica Dancel, Charlie Coton, Susan Tortolero

Center for Health promotion and Prevention Research, University of Texas School of Public Health, Houston, TX, USA

Abstract

Early sexual initiation is associated with increased risk of unintended pregnancy and sexually transmitted infection (STI). Effective HIV/STI/pregnancy prevention interventions for middle school youth are urgently needed. "It's Your Game, Keep It Real" (IYG) is a curriculum delivered in 7th and 8th grade that combines classroom activities with individualized, tailored computer-based activities embedded in a 'virtual world' environment. Interactive multimedia can offer a confidential, tailored, and motivational educational experience. Virtual world game interfaces offer further potential to immerse the learner. The purpose of this study was to evaluate the multimedia education program component of IYG on student attitudes of importance of the curriculum content, self-efficacy regarding refusal skills, and usability parameters of ease of use, credibility, understandability, acceptability, and motivation to determine that a broader efficacy field test would be indicated. Results of the study indicated acceptable usability criteria and impact on short-term psychosocial outcomes. IYG is currently being evaluated in a randomized controlled trial in ten Texas middle schools.

Keywords:

HIV/pregnancy prevention, multimedia, youth

Introduction

Although adolescent pregnancy and birth rates have declined steadily in the United States over the past decade, adolescent pregnancy remains a serious public health issue.[1] In Texas, 56,086 births reported in 2000 were to mothers aged 10-19; 8,465 of these were reported in Harris County.[2] In 2001, Texas also ranked joint first among states for rate of repeat births to teen mothers: 25% of births to mothers aged 15-19 were repeat births.[3] Teen pregnancy is particularly prevalent among minority populations, especially African American and Hispanics. [2,4] For the teen mother, consequences of pregnancy include decreased likelihood of completing high school and increased likelihood of relying on welfare. Infants born to teen mothers have lower birth weights, are more likely to perform poorly in school, and are at greater risk of abuse and neglect.[5] Overall, the U.S. government spends over $25 billion a year for social, health, and welfare services to families begun by teen mothers.[5]

Adolescent sexually transmitted infections (STI) including HIV also represent serious public health problems. In 2000, youth between the ages of 15 and 24 accounted for 9.1 million (48%) of all new STI cases.[6] The estimated medical cost of these cases was $6.5 billion.[7] At least half of all new HIV infections are estimated to be among those under the age of 25 and most young people are infected through sex.[8] Among youth, teen girls and minorities have been particularly affected.[9] Texas currently ranks fourth among states for the estimated number of persons living with HIV/AIDS.[10,11] In Houston, 319 (75%) of HIV cases reported between 1999-2003 among 13-19 year olds were among African American youth and 59 (13.9%) were among Hispanic youth.[11]

According to data from the 2003 Middle School Youth Risk Behavior Surveillance Survey (MSYRBS), 16% of 7th graders and 19% of 8th graders have engaged in sexual intercourse (implicitly, vaginal intercourse).[12] This is disturbing because early initiation of sexual intercourse has been associated with an increased risk of STIs and pregnancy.[13-16] Among older adolescents, condom rates for oral and anal intercourse are typically lower than for vaginal intercourse,[17-19] and a history of anal sex is predictive of non-use of condoms during vaginal sex.[20] This evidence points to the urgent need for effective HIV, STI and pregnancy prevention interventions at the middle school level to help delay or mitigate the consequences of early sexual activity.

A 2004 review of HIV, STI and pregnancy prevention programs for middle school populations identified 12 programs that have been rigorously evaluated.[21] Of these programs, seven showed positive behavior change. Two recently published studies also reported behavioral change for interventions at short-term (5 month)[22] and twelve-month follow-up.[23] These programs all incorporated small group activities to address peer pressure and decision-making, role-playing to practice refusal and communication skills, and other interactive, experiential learning techniques. Although these programs all demonstrated some positive behavior change, most showed

Selected for best paper award.

differential impact by gender or by sexual experience.[24-26]

The "It's Your Game" (IYG) curriculum offers an innovative application of computer-based gaming technology, interactive computer-based activities, and small group classroom interaction that does not currently exist in middle school HIV/STD prevention curricula. The purpose of this study was to evaluate a component of the curriculum, the IYG virtual world multimedia education program, to determine its impact on short-term outcomes and usability parameters to ensure feasibility of the program for field testing in a 10-site randomized controlled trial.

Methods

Study Design: The design was a single group pre-test post-test usability study conducted in facilities at the University of Texas, Houston Health Science Center.

Subjects: A convenience sample (n=14) of Houston middle school students was recruited who were representative of the Houston Independent School District demographic: primarily minority (50% African American) and of approximately equal gender (57% female). This sample tested the 7th grade lessons. Nine students from this sample returned 7 months later to test the 8th grade lessons. This 8th grade sample was 12-14 years of age, 55% African American, and 55% female. A smaller sample size, used in this study, was consistent with recommendations for usability testing given that statistical significance is not required to determine major usability problems and that the best cost-benefit ratio is achieved with 3-5 users in each representative group.[27] Participation was voluntary and written parental consent and child assent was obtained.

Intervention: The conceptual framework for the IYG curriculum is based on Social Cognitive Theory (SCT), social influence models and the Theory of Triadic Influence (TTI).[28-30] The curriculum consists of 12 lessons delivered in 7th grade and 12 lessons delivered in 8th grade. In each grade, the curriculum integrates group-based classroom activities (e.g., role plays, group discussion, and small group activities) with personalized journaling, and the IYG virtual world multimedia intervention delivered on laptop computers. The virtual world intervention comprises a total of 8 lessons spread through the 7th and 8th grade curriculum. These lessons comprise 3 functional elements: (1) A 3D virtual world interface featuring an entertainment complex motif (Figure 1); (2) tailored educational activities including interactive 2D exercises, quizzes, animations, peer video, and fact sheets that target determinants of sexual risk-taking (Figure 2); and (3) "real world"-style teen serials with on-line student feedback which allow for real time group discussion in the classroom. In addition, selected computer activities are tailored by gender or by sexual experience and intent so that students receive information and skills-training that is tailored to their needs.

Figures 1 & 2 - Screen captures of an aspect of the virtual world and an interactive FLASH activity

A life skills decision-making paradigm (*Select, Detect, Protect*) underlies the activities, teaching students to *select* personal limits or rules regarding risk behaviors, to *detect* signs or situations that might challenge these limits, and to use refusal skills and other tactics to *protect* these limits. Specific topics covered in the 7th grade include characteristics of healthy friendships, setting personal limits and practicing refusal skills in a general context (e.g., regarding alcohol and drug use, skipping school, cheating), information about puberty, reproduction and STIs, and setting personal limits and practicing refusal skills related to sexual behavior. The 8th grade curriculum reviews these topics and presents additional activities regarding the characteristics of healthy dating relationships including age compatibility, the importance of HIV, STI, and pregnancy testing if a person is sexually active, and skills training regarding condom and contraceptive use.

Study Protocol: The Virtual world intervention was tested separately from other classroom curriculum elements. The students accessed the program in a simulated class setting. Each student was provided with a laptop computer with headphones and asked to complete each of the four 35 minute 7th grade computer lessons individually. At the end of each lesson each student completed feedback questionnaires. At the end of the 4th (final) 7th grade lesson each student also completed overall assessment of the program. Sessions were observed by study personnel who logged problems (technical or content related) and provided assistance as required. This protocol was repeated 7 months later for the 8th grade lessons by the students, who were then in 8th grade.

Data Collection: Questionnaire, computer-based, and observational data was collected. Demographic data included age, race, gender, computer experience, type and frequency of computer use. Attitudes to the use of computer-assisted instruction were collected using a validated questionnaire.[31] Pre and post lesson ratings on the importance of content domains and self-efficacy (confidence) in performing skills was assessed using semantic differential scales developed for the intervention and embedded in the virtual world interface. Usability parameters including ease of use, credibility, understandability, acceptability, and motivation were assessed using Likert scale ratings adapted from usability assessment instruments reported by Sapperstein et al, 2004.[32] Open-ended

responses on recommendations for improvement of the program were collected via computer and paper and pencil questionnaire.

Data Analysis: Data was analyzed using descriptive and inferential statistics (paired t-tests) with SPSS analytic software.

Results

Demographics: The students were experienced computer users. Most accessed computers at their home (89%), friends (55%), and school (50%). Thirty three percent used computers for more than 1 hour on week days and 77% for more than one hour on weekends. Most frequent uses of computers were school work, visiting websites, and e-mail (all > 50%). Approximately 44% of the sample reported achieving A or B grades at school.

Short-term outcomes: Student attitudes toward the use of computers in education was enhanced in the 7th grade sample, and significantly enhanced in the 8th grade sample ($p<0.5$). Ratings of the importance of the program content in each lesson increased significantly for all content domains including the importance of keeping 'good' friendships, understanding how reproduction works and the possible consequence of sex (HIV/STDs/pregnancy), and the importance of enacting behaviors to limit sexual experience, described as selecting, detecting, and protecting personal rules about choosing not to have sex (all $p<0.05$). Ratings of self-efficacy for enacting behaviors in these domains also significantly improved (all $p<0.5$).

Usability: Usability parameters were highly rated across 7th and 8th grade lessons. *Ease of use*: Virtual world and educational activities were rated as easy to use by 78-100% of students. *Credibility:* A minimum of 92.9% of students perceived the content as correct and trustworthy across all lessons. *Understandability:* There was 100% agreement that most words in the program were understandable. *Acceptability:* Most students (71-100%) rated interface strategies and specific program activities within each lesson as fun and 92.9-100% rated it as helping them make healthy choices regarding sexuality. There was 64.3 – 92.9% agreement that each of the program lessons were as much or more fun than other lessons or favorite video games and 85.7-100% of students rated the lessons as "just right". Open ended responses suggested satisfaction with the lessons with just two participants suggesting a desire for more media elements in the form of characters and movies.

Discussion

The "It's Your Game" virtual world multimedia HIV/STD and pregnancy prevention lessons were found to significantly enhance attitudes toward the importance of sexual risk behavior and self-efficacy toward initiating such behavior. The lessons were reported to be easy to use, credible, understandable, acceptable, and of sufficient motivational appeal to students (12-14 years of age) to elicit confidence that the program was feasible for field testing in middle schools as a component of the IYG curriculum. An unexpected 'by-product' of this study was the immediate positive effects on attitudes toward the use of technology in education. Together, these results tender support for the emerging interest in gaming interfaces through burgeoning Serious Games and Games for Health research initiatives.[33,34] The capability of computer-based applications to provide interactive and individual tailored experiences that are also confidential is particularly salient when considering the sensitive nature of HIV/STD, pregnancy, and sexual health program content. Further, the technology can provide views and experiences beyond the limited classroom forum, allowing students to gain a realistic appraisal of what constitutes normative sexual behavior. A strength of the program is its ability to provide anonymous input that can be subsequently used within the small group setting of the classroom.

While the small sample size and limited scope is appropriate for the objectives of a usability study of this type [27], some results needs to be viewed with caution. A single group pre- post-test study design was the basis for the reported change in importance, self-efficacy, and attitude to computer-assisted instruction. While these results are encouraging, this design is open to internal threats to validity and does not imply causality.[35] Also, these effects were measured immediately prior to and following the intervention with no long term follow-up.

The combined IYG curriculum is currently being evaluated in a randomized controlled trial in 10 Texas middle schools. Baseline data have been collected from a largely minority sample of 1,321 7th graders comprising 57.1% female, 43.5% black, and 41.9% Hispanic, with a mean age 12.5 (SD=0.69) years. Fourteen percent had engaged in any type of intercourse (12% vaginal, 7.9% oral, 6.5% anal). Impact and outcome data are being collected 5, 14 and 24 months post-baseline. Measures include sexual behavior and intentions (lifetime/current vaginal/oral/anal sex, condom use), beliefs, perceived norms, knowledge, self-efficacy, reasons to/not to have sex, exposure to risky situations. This study will provide behavioral outcome data for the curriculum. Initial 5 month post-test results indicate students receiving the IYG curriculum report program lower prevalence of any type of intercourse during the past 3 months, and positive change in abstinence beliefs, perceived friends' beliefs about sex, exposure to risky situations, and reasons not to have sex (all $p<0.05$).

If the IYG curriculum proves effective, dissemination strategies will include a web-based application, a Spanish language version, and "institutionalization of the curriculum within the Houston Independent School District. Future development plans for the IYG virtual world multimedia program are to evaluate its impact when used as a stand alone intervention, separated from the IYG curriculum. The potential for the program to provide multi-user education for naturally occurring peer groups is also being investigated.

Conclusion

These results indicate that the “It’s Your Game” virtual world computer-based program is a feasible modality for HIV/STD, and pregnancy prevention for middle school students. The program meets usability criteria and shows potential in impacting attitudes and self-efficacy regarding refusal skills. The program also shows promise in translating positive attitudes to the use of technology in education. This usability study indicated that field testing on behavioral effects was indicated. The IYG curriculum, containing the IYG virtual world multimedia program, is currently being evaluated in an NIMH NIH funded randomized controlled trial in 10 Texas middle schools to determine its impact on sexual behaviors and their antecedents.

Acknowledgement

This study was funded by the national Institutes of Mental Health, NIH. Award # NIMH R01 MH66640.

References

[1] Klein JD, Barratt MS, Blythe MJ, Diaz A, Rosen DS, Wibbelsman CJ, Hertweck SP, Kaufman M, Shain B. Adolescent pregnancy: Current trends and issues. Pediatrics. 2005;116(1):281-286.

[2] Texas Department of Health Bureau of Vital Statistics. 2000 Birth Report: Births by Mother's Age and Ethnicity. Data geocoded by the Center for Health Policy Studies, UT-Houston, School of Public Health, 2000.

[3] Texas Department of Health Bureau of Vital Statistics. 2000 Birth Report: Births by Mother's Age and Ethnicity. Data geocoded by the Center for Health Policy Studies, UT-Houston, School of Public Health, 2000.

[4] Alan Guttmacher Institute. U.S. Teenage Pregnancy Statistics Overall Trends, Trends by Race and Ethnicity and State-by-State Information. Alan Guttmacher Institute. Feb 19, 2004. Available at: http://www.agi-usa.org/sections/youth.html ; accessed: May 12, 2004.

[5] Card JJ. Teen pregnancy prevention: Do any programs work? Annual Review of Public Health. 1999;20:257-285.

[6] Weinstock H, Berman S, Cates W. Sexually transmitted diseases among American youth: Incidence and prevalence estimates, 2000. Perspectives on Sexual and Reproductive Health. 2004;36(1):6-10.

[7] Chesson HW, Blandford JM, Gift TL, Tao G, Irwin KL. The estimated direct medical cost of sexually transmitted diseases among American youth, 2000. Perspectives on Sexual and Reproductive Health. 2004;36(1):11-19.

[8] Centers for Disease Control and Prevention. HIV Prevention Strategic Plan through 2005. 2001. Available at: http://www.cdc.gov/nchstp/od/hiv_plan/default.htm ; accessed: May 23, 2005.

[9] Advocates for Youth. Serving youth of color. Transitions. Jan 1, 2004; 15(3). Available at: http://www.advocates foryouth.org/ publications/ transitions/transitions1503.htm ; accessed: Sep 8, 2005.

[10] Centers for Disease Control and Prevention. Cases of HIV Infection and AIDS in the United States, 2002. HIV/AIDS Surveillance Report. 2003; 14(2002):1-40. Available at: http://www.cdc.gov/hiv/stats/hasrlink.htm ; accessed: Feb 26, 2004.

[11] Houston Department of Health and Human Services. AIDS: Summary of Houston/Harris County Cases, Cumulative - Reported 1981 through 6/30/2005. Jan 9, 2004. Available at: http://www.houstontx.gov/health/ HIV-STD/ 2ndQRT2005.pdf ; accessed: Sep 8, 2005.

[12] Whalen, L. G., Grunbaum, J. A., Kinchen, S., McManus, T., Shanklin, S. L., and Kann, L. Middle School Youth Risk Behavior Survey 2003. Web. 2005; 1-292. Available at: http://www.cdc.gov/healthyyouth/yrbs/middleschool 2003/ pdf/fullreport.pdf; accessed: Nov 8, 2006.

[13] Kaestle CE, Halpern CT, Miller WC, Ford CA. Young age at first sexual intercourse and sexually transmitted infections in adolescents and young adults. American Journal of Epidemiology. 2005;161(8):774-780.

[14] O'Donnell L, O'Donnell CR, Stueve A. Early sexual initiation and subsequent sex-related risks among urban minority youth: The Reach for Health study. Family Planning Perspectives. 2001;33(6):268-275.

[15] Coker AL, Richter DL, Valois RF, McKeown RE, Garrison CZ, Vincent ML. Correlates and consequences of early initiation of sexual intercourse. Journal of School Health. 1994;64(9):372-377.

[16] Flanigan CM. Sexual activity among girls under age 15: Findings from the National Survey of Family Growth. In: Albert B, Brown S, Flanigan CM, eds. 14 and Younger: The Sexual Behavior of Young Adolescents. Washington, DC: National Campaign to Prevent Teen Pregnancy, 2003:57-64.

[17] Boekeloo BO, Howard DE. Oral sexual experience among young adolescents receiving general health examinations. American Journal of Health Behavior. 2002;26(4):306-314.

[18] Stone N, Hatherall B, Ingham R, McEachran J. Oral sex and condom use among young people in the United Kingdom. Perspectives on Sexual and Reproductive Health. 2006;38(1):6-12.

[19] Halperin DT. Heterosexual anal intercourse: prevalence, cultural factors, and HIV infection and other health risks, Part I. AIDS Patient Care and STDs. 1999;13(12):717-730.

[20] Koniak-Griffin D, Lesser J, Uman G, Nyamathi A. Teen pregnancy, motherhood, and unprotected sexual activity. Research in Nursing and Health. 2003;26(1):4-19.

[21] Manlove J, Franzetta K, McKinney K, Papillo AR, Terry-Humen E. No Time to Waste: Programs to Reduce Teen Pregnancy among Middle School-Aged Youth. Washington, DC: National Campaign to Prevent Teen Pregnancy, 2004.

[22] Borawski EA, Trapl ES, Lovegreen LD, Colabianchi N, Block T. Effectiveness of abstinence-until-marriage sex education among middle school adolescents. American Journal of Health Behavior. 2005;29(5):423-34.

[23] Dilorio C, Resnicow K, McCarty F, De AK, Dudley WN, Wang DT, Denzmore P. Keepin' it REAL! Results of a mother-adolescent HIV prevention program. Nursing Research. 2006;55(1):43-51.22. Thomas B, Mitchell A, Devlin M, Goldsmith C, Singer J, Watters D. Small group sex education at school: The McMaster Teen Program. In: Miller BC, Card JJ, Paikoff RL, Peterson JL, eds. Preventing Adolescent Pregnancy. Newbury Park, CA: Sage Publications, 1992:28-52.

[24] Howard M, McCabe JB. Helping teenagers postpone sexual involvement. Family Planning Perspectives. 1990;22(1):21-26.

[25] Aarons SJ, Jenkins RR, Raine TR, El Khorazaty MN, Woodward KM, Williams RL, Clark MC, Wingrove BK. Postponing sexual intercourse among urban junior high school students-a randomized controlled evaluation. Journal of Adolescent Health. 2000;27(4):236-247.

[26] Coyle KK, Kirby DB, Main BV, Gomez CA, Gregorich SE. Draw the Line/Respect the Line: A randomized trial of a

middle school intervention to reduce sexual risk behaviors. American Journal of Public Health. 2004;94(5):843-851.
[27] Nielsen J. Usability Engineering. San Diego:Morgan Kaufmann, 1993.
[28] Bandura A. Social Foundations of Thought and Action: A Social Cognitive Theory. Englewood Cliffs, NJ: Prentice Hall, Inc., 1986.
[29] McGuire W. Social psychology. In: Dodwell PC, ed. New Horizons in Psychology. Middlesex, England: Penguin Books, 1972:219-242
[30] Flay BR, Petratis J. The theory of triadic influence: A new theory of health behavior with implications for preventive interventions. Advances in Medical Sociology [4: A reconsideration of models of health behavior change], 19-44. 1994. Greenwich, Conn., JAI Press.
[31] Asker P, Yavuz H, Koksal M. Student's perceptions of computer-assisted instruction environment and their attitudes towards computer-assisted learning. Educ Res. 1992;34:133-9.
[32] Sapperstein MS, Gold RS, Atkinson N, Peterson KJ. BodyFun Online: How usability studies can improve internet-based health education. Am. Pub. Hlth Assoc. 132nd Annual Meeting, Washington DC, Nov. 2004.
[33] Serious Games Initiative. Available at: http://www.seriousgames.org/index2.html. Accessed Dec 3rd, 2006.
[34] Games for Health. Available at: http://www.gamesforhealth.org/index3.html. Accessed Dec 3rd, 2006.
[35] Cook TD and Campbell DT. Quasi-Experimentation. Design & Analysis Issues for Field Settings. 1979. Houghton Mifflin Co. Boston, 1979.

Address for correspondence

Ross Shegog, PhD.
Assistant Professor
Center for Health Promotion & Prevention Research
UT-School of Public Health
7000 Fannin, Suite 2668
Houston, TX 77030
Phone: 713-500-9677
Fax: 713-500-9602
e-mail: Ross.Shegog@uth.tmc.edu

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

HeartCareII: Home Care Support for Patients with Chronic Cardiac Disease

Patricia Flatley Brennan [a], Gail Casper [a], Susan Kossman [a,b], Laura Burke [c]

[a] *University of Wisconsin-Madison, Madison, WI*
[b] *Illinois State University, Normal, IL*
[c] *Aurora HealthCare, Milwaukee, WI*

Abstract

Systematic engagement of patients in disease management requires design and deployment of innovative technologies that complement and extend professional nursing services. We describe here a model of nursing practice that capitalizes on a web-based resource (HeartCareII) to support patient self-management, symptom interpretation, and self-monitoring. Research staff provided computers and technical assistance; visit nurses trained patients in the components of the HeartCareII Website most relevant to their care needs. This paper describes the nursing practice model and the web resource, and reports the experience of patients recruited in the early phase of the study.

Keywords:

heart disease, consumer health informatics, home care services, nursing models

Introduction

Advances in computerized home monitoring could lead to a disease management approach in which the patient is a passive recipient in the assessment and management of key health parameters. While the benefits of passive home monitoring strategies are recognized, the philosophy underlying our approach is that technology should enhance and facilitate patients' *active* engagement in self-monitoring and self-management. Research demonstrates that Web-based information resources that not only alert patients about what symptoms to expect but also provide coaching about how to manage them improve self-care and health outcomes.[1, 2] We propose to create a new model of home care nursing, Technology Enhanced Practice (TEP), and support it with a website, HeartCareII, accessible by *both patients and nurses*. HeartCareII provides tailored coaching information, communication with peers and professionals, and personal monitoring tools designed to make patients more active co-creators of their health and to facilitate effective nursing care. In this study, we integrate the technology within the formal care delivery system, building upon the parallel approach used in our prior work. This paper describes the core of the intervention, and reports early evidence from the field experiment of the ways in which 24 patients who have completed the study used the HeartCareII resources.

Patients with complex cardiac disease, including Congestive Heart Failure (CHF) Acute Myocardial Infarction (AMI), Coronary Artery Bypass Graft (CABG) and Valve Surgery and various combinations of these conditions are the targets of this HeartCareII project. Clinical management of these chronic heart conditions involves a partnership between patients, physicians and nurses. Therapeutic interventions may include pharmacotherapy, diet, and activity consultation[3, 4] and may rely on care provided by home care nurses. Positive effects often result from nursing interventions and home-care approaches, but these interventions are expensive, may over-utilize scarce professional resources, and have effects that cease once the home care nursing interventions are stopped. Even under optimal management, patients with chronic heart disease experience frequent, serious exacerbations. The challenge is to institutionalize an intervention strategy that provides patients with timely access to relevant health information, self-management guidance and self-monitoring resources, while facilitating contact with clinicians for rapid intervention. Consumer health informatics (CHI) solutions, appropriately fit to the care situation and nursing practice model, may improve self-management, reduce demand on hospital and home care services and ensure timely, appropriate treatment.

The quality chasm and nursing shortage require sophisticated information solutions to replace the naïve view that simply providing WWW-based health information will lead to improved disease management and adherence to healthy behaviors. Safran's work demonstrated that the integration of professional presence within an electronic outreach intervention contributed strongly to desirable clinical outcomes.[1] Studies of the equivalence of technology-mediated care with traditional approaches [5] are no longer adequate. They must now be superseded by complex large-scale studies that examine the integration of CHI within existing care approaches, rather than supplanting existing care delivery models with electronic substitutes. Recent evidence suggests that tailoring the information provided in the technology intervention to the patient's health status and learning needs improves outcomes. [6, 7] In this study we draw information regarding the needs for tailored health information from the patient *in addition to* considering how information technology resources complement or expand resources available in the nurse-patient encounter. In creating TEP we provide a

Selected for best paper award.

means to close a gap among consumer health informatics tools, evidence-based nursing practice, and home health nursing practice by integrating the informatics innovation into the routine process of home care nursing. [8] Practice innovations such as TEP may provide sufficient support to ensure that all nurses have the competence to address the needs of the person with heart disease.

Materials and methods

HeartCareII project

The goal of the HeartCareII project, a collaboration between an academic nursing informatics research team, a health system clinical research group and a practice setting, is to develop and test a sustainable model of home care nursing supported by a technology core and home based technology. The practice setting is a large metropolitan visiting nurse association (VNA). We believe home based technology will help the nurse provide better care during home visits and also support extended asynchronous care.

The innovative home care nursing practice model labeled, *Technology Enhanced Practice*, provides an approach that allows home care nurses to balance interpersonal and technologically-mediated interventions in the care of patients with heart disease, thus maximizing their time and effectiveness while supporting patient's self-management efforts.[8] The major aim of the HeartCareII project is to determine whether TEP leads to improved outcomes in patients with heart disease. This multiyear project, conducted within a large integrated health care delivery system in Wisconsin, will involve 300 patients and approximately 80 home care nurses in the study's five VNA offices. Most nurses are female; the average age is 43 years old (range 24-60). Racial composition is predominantly Caucasian. Although all nurses use computers for planning and documenting the care they provide, they exhibit considerable variation in their level of comfort and skill in using them.

The first phase of the study, the design phase, applied human factors techniques to analyze the work of home care nurses. It resulted in development of the technology core for TEP, the HeartCareII website, which is housed within the clinical partner's patient portal. The second phase of the study, initiated in August 2005, implements the TEP and its technical core, the HeartCareII website, in a randomized field experiment that will examine its impact on nurses' workflow and patient outcomes including clinical status, quality of life, self-monitoring, satisfaction with care, and health care service utilization.[9]

Nursing practice models

We envision TEP as a new nursing practice model for home care nurses. Nursing practice models (NPMs) can be broadly defined as the "manner in which nurses assemble to accomplish clinical goals".[10] They represent structural and contextual features of care settings, including organizational, nursing resource and support dimensions[10,11] No universal, ideal NPM exists and there are many variations across care settings. Additionally, actual practice may deviate from anticipated practice specified in organizational NPM plans. NPMs share common elements that provide a framework to understand factors that may support, or interfere with, nursing practice. Basic elements common to NPMs and some associated attributes include 1) nurses' role (autonomy, accountability, philosophy of care); 2) professional relationships (teamwork, collaboration, communication, coordination); 3) care delivery systems (primary, case management); 4) continuity of care; 5) management systems (shared governance, decentralized decision making); and 6) compensation.[10-13] The degree to which the attributes listed in each category are present influences nurses' ability to administer effective interventions. Ideally, nurses have the autonomy and accountability to make appropriate care decisions, effectively communicate and collaborate with healthcare team members and patients, ensure continuity of patient care through adequate staffing with skilled and knowledgeable nurses, and be active participants in decision making.[13] In the HeartCareII study, the homecare nurses' existing NPM served as the framework on which to build the technology nurses could use in providing TEP. We designated this existing NPM as the "usual care" practice for the purposes of the experiment.

Usual care

The usual care for patients with heart disease is provided by Registered Nurses accountable for delivering evidence-based nursing care to homebound patients based on their assessment of the patient's needs, physician's orders and clinical guidelines developed by advanced practice nurses. The patient's diagnosis, acuity, physician's orders, insurance coverage and nursing care needs determine the number of home visits to be made, which typically ranges from 1-9. Continuity of care is provided by having a single nurse provide as many of the home visits as possible.

Nurses interact with patients through home visits and telephone for the specified number of visits typically spanning several weeks. The usual nursing care for patients with any diagnosis of chronic heart disease includes initial and ongoing comprehensive health assessments, patient and family education, medication management and administration, skilled interventions, and coordination of other necessary health services. Patient and family education are integral components of care. Nurses use an institutional set of care management initiative education materials to assist patients to adhere to medications, modify their life style, understand the disease, and recognize early manifestations of disease progression or complications. In addition to education, the practices include surveillance of cardiopulmonary symptoms, and prompt response to a change in status. Nurses coordinate services which include dietary, intravenous therapy, laboratory, physical and occupational therapy, behavioral health, referral to the VNA heart failure program, cardiac rehabilitation services, and telemanagement.

Nurses at this VNA use a set of guidelines as the benchmark for usual care. These guidelines are based on Barrella and Monica's Congestive Heart Failure at home clinical pathway [14] and are extensible to patients with many

types of chronic cardiac disease. The clinical pathway aids the nurse in developing a plan of care to facilitate patient integration of, and compliance with, the medical plan of care. Practice innovations in combination with such guidelines may provide sufficient support to enable nurses to have the competence required to address a full range of patient needs.

Nurses are organized in care teams of one to three nurses and a care coordinator; five VNA offices located in southeastern Wisconsin participate in the study. A single nurse takes primary responsibility for a given patient's care plan and makes as many of the visits to that patient as possible. Communication and collaboration occur during weekly nursing care team meetings and monthly staff meetings. Nurses also collaborate with physicians and other healthcare team members as needed. The VNA follows a shared governance structure, and nurses are paid per visit incorporating visit intensity.

Technology enhanced practice

Technology Enhanced Practice is a nursing practice model in which nurses selectively and deliberately employ computer and information technologies in a manner designed to meet individual patient care goals. We built the HeartCareII website to serve as the technology core for TEP based on the existing nursing practice model and added resources to expand, support and supplement nursing work, to enhance nursing care and to facilitate patient self-monitoring and self-management. Existing paper based tools nurses use to provide usual care were modified to an electronic format and supplemented with additional tools to address care needs (e.g. a drug checker, several health trackers).

The technology core (HeartCareII website) provides a standard suite of technology services which nurses can select to use based on patient needs and illness trajectory.[9] Additionally, we anticipate that the technology features that support TEP may affect NPM elements such as nurses' role, communication, collaboration and continuity of care. The HeartCareII website is housed within the patient portal section of the clinical partner's public website and may be accessed by nurses and patients through an internet connection from a variety of geographic settings (point of care, physician or VNA office, remote locations such as cardiac rehabilitation centers or family members' homes). Interactive tools on the HeartCareII website address symptom monitoring (symptom checklist, weight tracker, blood pressure tracker, heart rate tracker), education (patient education papers on a variety of topics, and trackers such as the food and fluid trackers), and communication (my goals, my journal, email, bulletin boards). A sample display from the HeartCareII website is provided in Figure 1.

Nurses encourage patients to use pertinent tools daily and to explore the website on their own. The nurse can monitor the patient remotely by accessing their account (with permission) to check on symptom monitoring, and communicate in person, on the telephone, or by computer. Additionally, patients may communicate with other study participants through a private bulletin board.

Nurses' actual use of the technology resources in the delivery of patient care (e.g. type of service, frequency, and duration) varies both within an individual nurse's caseload and among different nurses based on patients' needs and abilities as well as nurses' levels of clinical expertise, preferences, and experience with technology. Nurses use professional judgment to individualize TEP for each patient based on their skills, preferences and assessment of patient needs. The following description illustrates how a nurse might use the TEP intervention during the home visit phase:

Figure 1 – The HeartCareII Home Page

Upon assignment of an eligible patient to his or her caseload, the nurse will complete the standard VNA intake protocol and conduct an information needs assessment. Nurses will use this information to plan a TEP intervention for this patient, identifying practice interventions and selecting features of the computer support for the patient. If the nurse appraises this patient information needs as high, he or she may use the HeartCareII website to provide ongoing, between-visit education for the patient, directing the patient to relevant readings. If he or she also determines this person has strong needs for coaching and support, and he or she may use the HeartCareII web site to message the patient daily, alternating home visits with daily electronic monitoring and communication, thus ensuring that the patient has adequate support. Thus, TEP supports creative use of human resources and information technology to meet patient needs.

Conduct of the HeartCareII experiment

Through a field experiment we will compare selected outcomes of TEP to those of usual home nursing care. Three of the VNA offices are designated as "experimental sites" and deliver care to patients with chronic cardiac disease using TEP; the remaining two VNA offices provide usual care. Patients in the TEP arm receive computers and related software or have their own computer reconfigured for the study; patients in the usual care condition receive a booklet including clinical care guidelines, standard patient teaching materials and paper forms for documenting health parameters. The study period for each patient lasts six months, with a home care phase during which the nurse delivers TEP (about 2 to 8 weeks) and the post-discharge phase in which the patient assumes responsibility for self-

monitoring and self-management that lasts the remainder of the six month study period.

Results

Sample

In the first 15 months of the experimental phase of the project, we enrolled 152 patients and 47 have completed the 6-month study period (24 in the experimental group and 23 in the comparison group). We report here on the 24 patients in the experimental group who have completed the study to date. Patients included in this analysis range in age from 43 to 88 years old; mean age is 69 (S.D. 11.6). There are 10 women and 14 men. Fifteen patients have a primary diagnosis of CHF, with the other nine patients recovering from coronary artery bypass graft or valve replacement surgery.

We used the Specific Activity Scale (SAS) [15] to measure functional capacity. Fifteen respondents reported being able to walk down a flight of steps without stopping; 17 were able to dress themselves without stopping because of symptoms. Only three of the patients were able to strip and make a bed. The remaining items (ability to walk up steps carrying something or carrying 25 pounds) of the SAS require careful interpretation because inability to do these activities may be related medical restrictions rather than functional capacity.

Patient diagnoses included congestive heart failure (CHF), coronary artery bypass graft surgery (CABG), aortic valve replacement and heart transplant. Their experience with computers prior to participation in the study ranged from "none" to "competent". Length of time in the study while receiving home nursing care ranged from one week to greater than two months. Approximately one-third of the patients used the Internet access device and dial-up connection provided by the study (Wyse thin client with a CRT); the remaining two-thirds used DSL connections with the study device or with their own computers.

All patients in the experimental arm received TEP; however, the extent to which they used the HeartCareII website varied, and examining this use in depth will provide a foundation to later characterize TEP. Thus, we turn to an in depth characterization of use of the HeartCareII website. Because the study is on-going, we will not be providing any outcome data in this report.

Patient use of the HeartCareII website

Patient use data are collected through passive capture of logins and webpage tracking. Login counts are reported weekly, while webpage tracking is provided "on-demand". Accuracy is validated by comparing the two sources of data.

Use of the HeartCareII resources is highly variable. Logins to the system (defined as a successful connection to the site after submitting user ID and password) range from two to 144 with a mean of 36. Length of use of the system (days between first and last login) varies from zero to 194 days, with a mean of 86 days. Two-thirds of the patients used the HeartCareII website for 50 days or less and tended to have 20 or fewer logins. Approximately one-third of the patients used the system for greater than 50 days and tended to have more logins. Once connected to the HeartCareII website, the most commonly used resources included the weight tracker, food tracker and an action plan that assisted in symptom interpretation. Figure 2 below plots the number of logins against the number of days the patient used the system.

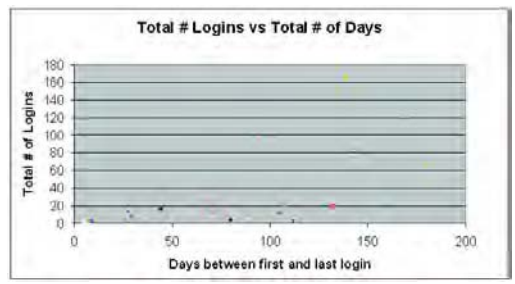


Figure 2 – Plot of Use Interval with Login Count

Illustrative case studies

The following case studies summarize two patients' actual use of the HeartCareII website resources and illustrate nurses' and patients' use of the website during the home visit and self-management phases. These cases provide beginning insights into the nature of TEP.

Case 1: Mr. A. was admitted to home care services nine days after receiving a heart transplant. He consented to participate in the HeartCareII study and had the study-provided internet device installed 11 days later. He had used a computer for personal and professional purposes and described himself as "familiar" with using computer technology.

Although there was no nursing documentation of his instruction in the use of the HeartCareII website or evidence of collaborative use by any of his three nurses, Mr. A. used many of the self-monitoring and information tools extensively, logging in almost daily for periods of 35 to 110 minutes. He even reported returning the study-provided computer and dial-up ISP after purchasing a PC with cable Internet to make his access easier and faster. He primarily used the health trackers - recording his weight, fluid and sodium intake, and blood pressure daily. He also reported using the HeartCareII website to learn about his diet and to seek information about heart disease. He continued to access the HeartCareII website for two months after his discharge from home care.

Case 2: Following his hospitalization for heart surgery, Mr. F. was referred to the VNA for post-operative care and instruction. He was enrolled in the study four days later. A PC owner with cable Internet connection, Mr. F. described himself as "competent" in the use of a computer. He opted to use his own computer for access to the HeartCareII website.

Due to limited insurance authorization, Mr. F.'s nurse was able to make only one home visit. Following that visit, she documented that he was able to access the website, locate information, and that he "found the information valuable". Mr. F. logged on to the HeartCareII website only two times prior to his discharge from the VNA.

Discussion

We envisioned a typical course of TEP for an experimental patient in the HeartCareII study to last for six months, encompassing about one week to two months of intensive nurse-patient interaction (laying the groundwork for subsequent self-care management by the patient and his or her family caregiver) and approximately four months of patient self-management aided by the HeartCareII website. We envisioned that use of the resources would vary over time and by individual. Our observations to date confirm that duration of VNA service and use of the HeartCareII resources does indeed vary across patients and nurses. Patients' use of the HeartCareII tools reflect that about two-thirds were engaged with the technology for 50 days or less, while one-third continued to use the resources for the entire 6 month period that it was available to them. About 25 percent of these opted to keep the internet device provided by the study, suggesting potential continued use of the electronic resources beyond the study period.

The variability in use of the HeartCareII resource follows patterns of use that parallel those found in earlier studies.[16,17] They provide evidence supporting patients' active engagement in web-based information and use beyond the initial home visit phase, suggesting that this informatics tool extends nursing care impact.

The variability in use of the HeartCareII website may stem from many factors. Patients may have found the technology off-putting and difficult to manage, although this has not been our experience in earlier studies with similar-aged participants.[2] Nurses may vary in their ability to understand and encourage patients to use the HeartCareII website. Or, most interestingly, we may be seeing the beginnings of differentiated practice under a TEP model, in which the use of the technology varies with the nursing goals.

- Nurses' role in encouraging and assisting patients to use the HeartCareII resources remains an important, yet unexplored explanatory aspect of use. Follow-up work system analyses are planned to further evaluate delivery of TEP.
- Evidence presented here demonstrates that home care patients with chronic cardiac disease can and will use a web-based resource for self-management, self-monitoring and clinical communication. Explication of the full nature of technology enhanced practice and determining its impact on patient outcomes awaits completion of this study.

Acknowledgments

Support for this project was provided by the National Library of Medicine LM 6247 and Aurora Health Care.

References

[1] Gray JE, Safran C, Davis RB, Pompilio-Weitzner G, Stewart JE, Zaccagnini L, et al. Baby CareLink: using the internet and telemedicine to improve care for high-risk infants. *Pediatrics*. 2000 Dec;106(6):1318-24.

[2] Brennan PF, Moore SM, Bjornsdottir G, Jones J, Visovsky C, Rogers M. HeartCare: an Internet-based information and support system for patient home recovery after coronary artery bypass graft (CABG) surgery. *J Adv Nurs*. 2001 Sep;35(5):699-708.

[3] Rich MW. Non-pharmacological treatment of heart failure: Just as important? In: McMurray JJV, Cleland JEF, editors. Heart failure in clinical practice. London Martin Dunitz; 1999.

[4] Gheorghiade M, Cody RJ, Francis GS, McKenna WJ, Young JB, Bonow RO. Current medical therapy for advanced heart failure. *Heart Lung*. 2000 Jan-Feb;29(1): 16-32.

[5] Jenkins RL, McSweeney M. Assessing elderly patients with congestive heart failure via in-home interactive telecommunication. *J Gerontol Nurs.* 2001 Jan;27(1):21-7.

[6] Strecher VJ. Computer-tailored smoking cessation materials: a review and discussion. *Patient Educ Couns*. 1999 Feb;36(2):107-17.

[7] Revere D, Dunbar PJ. Review of computer-generated outpatient health behavior interventions: clinical encounters "in absentia". *J Am Med Inform* Assoc. 2001 Jan-Feb;8(1):62-79.

[8] Brennan P, Burke L, Casper G, Sebern M, Krause C, Kossman S, et al. Creating technology-enhanced practice: A university-home care-corporate alliance. In: Park H-A, Murray P, Delaney C, editors. Consumer-centered computer supported care for healthy people Proceedings of NI 2006; 2006; Seoul. S. Korea: IOS Press: Amsterdam. ; 2006. p. 664-48.

[9] Casper GR, Karsh BT, Or CK, Carayon P, Grenier AS, Brennan PF. Designing a technology enhanced practice for home nursing care of patients with congestive heart failure. AMIA Annu Symp Proc; 2005; 2005. p. 116-20.

[10] Brennan PF, Anthony MK. Measuring Nursing Practice Models using Multi-Attribute Utility theory. *Res Nurs Health*. 2000 Oct;23(5):372-82.

[11] Mark BA. Characteristics of nursing practice models. *J Nurs Adm*. 1992 Nov;22(11):57-63.

[12] Arford PH, Zone-Smith L. Organizational commitment to professional practice models. *J Nurs Adm*. 2005 Oct;35(10):467-72.

[13] Hoffart N, Woods CQ. Elements of a nursing professional practice model. *J Prof Nurs*. 1996 Nov-Dec;12(6):354-64.

[14] Barrella P, Della Monica E. Managing congestive heart failure at home. *AACN Clin Issues*. 1998 Aug;9(3):377-88.

[15] Goldman L, Hashimoto B, Cook EF, Loscalzo A. Comparative reproducibility and validity of systems for assessing cardiovascular functional class: advantages of a new specific activity scale. *Circulation*. 1981 Dec;64(6):1227-34.

[16] Brennan PF, Moore SM, Smyth KA. The effects of a special computer network on caregivers of persons with Alzheimer's disease. *Nurs Res*. 1995 May-Jun;44(3): 166-72.

[17] Shaw BR, Han JY, Baker T, Witherly J, Hawkins RP, McTavish F, et al. How women with breast cancer learn using interactive cancer communication systems. *Health Educ Res*. 2006 Jul 7.

Address for correspondence

Patricia Flatley Brennan, RN, PhD, FAAN
University of Wisconsin-Madison
School of Nursing H6/241 CSC
600 Highland Ave
Madison, WI 53792

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

A Web-Based Communities of Practice Support System for Caregivers

Shirley L. Fenton[a,b], H. Dominic Covvey[a], Douglas W. Mulholland[b], Donald D. Cowan[a,b], Judith Shamian [c], Bonnie Schroeder[c]

[a] *Waterloo Institute for Health Informatics Research, University of Waterloo, Canada*
[b] *Computer Systems Group, University of Waterloo, Canada*
[c] *Victorian Order of Nurses (VON), Canada*

Abstract

CareNet is an interactive Web-based system intended to support informal caregivers (ordinary citizens who are engaged in providing residential healthcare to their families and friends). The design of CareNet uses concepts from a number of areas including: Communities of Practice, software engineering, content authoring, and knowledge management to create a supportive environment for the caregiver. The specific objectives of the CareNet project are: (1) To create a highly effective interactive environment that addresses the needs of caregivers to: (1a) obtain information and guidance, (1b) achieve efficient communication with professional healthcare workers with whom they collaborate and with other caregivers from whom they can gain advice, emotional support and solace, (1c) access other information and physical resources that are needed for proper care, and (1d) document their observations, interventions and insights that can in future become a knowledge resource for other caregivers; and (2) To demonstrate the beneficial impact of CareNet on caregivers who collaborate with professional care providers and on their interactions with professional care providers.

Keywords:

Communities of practice (CoP), caregivers, web-based system, consumer e-health, educational systems, Internet, Web, knowledge management.

Introduction

Healthcare faces many challenges if it is to remain affordable and accessible. Specifically, we require innovative approaches and alternatives to institutional solutions if we are to keep costs under control. One such solution is to discharge patients with short-term illnesses as well as individuals with chronic conditions earlier than in the past with the home, family and friends then becoming the locus for recovery. Using the home as the centre of care is definitely a trend as patients are discharged from institutional care at a higher level of acuity than the levels at which they were admitted just a few decades ago. This approach imposes many challenges on these "informal caregivers," who generally lack adequate care-related knowledge and skills, and are already living a full and often independent existence.

How can we provide adequate support to these informal caregivers? Can modern information and communications technology, such as the web, provide useful tools and valuable experiences? In this paper, we present a description of such an interactive web-based information systems called "CareNet." In addition to describing the technology, we will explain the motivation for and the process of arriving at the CareNet solution and the preliminary results that have already been observed. This project has motivated us to look more deeply into web-based information systems that can support the informal caregiver, as well as other communities of practice, that exist in healthcare and elsewhere. Our current work continues to examine the information technologies that would be useful in supporting communities of practice and how members of such communities interact with these technologies.

The Waterloo Institute for Health Informatics Research (WIHIR) and the Computer Systems Group, both at the University of Waterloo (UW), worked closely with the Victorian Order of Nurses (VON) to develop a web-based system to support informal caregivers in:

- providing effective and sustainable care in the home setting and
- achieving effective and efficient communications with professional healthcare personnel when needed.

CareNet is intended to provide informal caregivers with access to "support channels," such as care-related information resources, and the means to communicate with professional care providers and other informal caregivers in similar situations. CareNet is intended to provide experience-based insights of caregivers and providers, directories of physical resources (e.g., pharmacies), event co-ordination via booking and scheduling services, and other capabilities. All of these aim to reduce the burden of care giving and increase the interaction with other caregivers who can provide assistance and emotional support.

Our approach allowed for a team of professional caregivers from the Victorian Order of Nurses (VON) to be completely involved in the requirements, design and development of the system using a rapid prototyping tool developed at the University of Waterloo called the Web-

based Informatics Development Environment (WIDE)[4]. The resulting system (CareNet) is an example of a Community of Practice support system[7],[8],[9],[10],[11],[12] which facilitates interaction among caregivers, and provides information and other resources. According to Wenger et al., "Communities of practice are groups of people who share a concern or a passion for something they do and learn how to do it better as they interact regularly."

Our research objectives were to:

a) Determine and document caregivers' detailed needs;
b) Define the capabilities required of CareNet to address the needs of caregivers and VON Canada;
c) Design and implement a proof of concept version of CareNet;
d) Apply CareNet to address key needs of patients and providers; and
e) Evaluate the impact of CareNet on caregivers and care professionals.

Methods

Our approach to developing CareNet included the following steps:

1. Form a small joint UW-VON Project Team to develop the prototype. The VON members of the team were individuals who were knowledgeable about "informal caregivers" and the type of issues they face.
2. Identify the goals and objectives that VON staff believes address the needs of informal caregivers who currently seek advice from and collaborate with VON professionals in care provision.
3. Analyze these goals and objectives and use the analysis to define the needs of caregivers that can be supported by a web-based interactive information system.
4. Design the system and review this design with the VON. This step could go through several iterations as both the VON staff and the systems designers develop a more comprehensive understanding of the issues related to the problem.
5. Create a pilot version of the CareNet portal suitable for demonstration to caregivers and VON professionals using a set of rapid development frameworks developed at the University of Waterloo called the Web-based Informatics Development Environment (WIDE). Web-based information systems created using WIDE can be built and modified quickly. Thus the system itself can be and was used as a requirement elicitation tool to refine the design further with the feedback received from the iterations. Shared elicitation and design sessions with VON staff and development staff were held where the two groups were remote from each other. The nature of WIDE also allowed system changes on the fly thus supporting "what-if" scenarios.
6. Review the pilot version by VON staff supporting caregivers and informal caregivers. This review would help determine whether the functionality of the system meets user needs.

Once the goals and objectives were identified and analyzed, a number of applications for the prototype were discussed and prioritized. From this analysis and prioritization, it was determined that three primary applications would be created for the prototype:

- A directory of services offered by VON branches across Canada.
- Information resources for the caregiver community.
- A tool to allow both caregivers and supporting professionals to augment the online content based on personal experiences and to make a vetted version of this content available to both parties.

A description of each of these applications will be presented later in this paper.

As mentioned earlier in this paper the system design and implementation was developed iteratively and interactively by the UW and VON Project Team members. Each step of the implementation was reviewed with VON staff responsible for the project and with some staff members who are care professionals. One of the design criteria specifically focused on maintenance of the information handled by the CareNet portal. It was intended that all information available through the portal could be maintained and manipulated by VON staff involved in healthcare and that the training required would be no more complex than learning to use the basic functions of a word processor. This included vetting and manipulating all the online content that was supplied by the users including both informal caregivers and supportive VON staff.

The project was not field tested with informal caregivers as there was not enough information available in the prototype system to support an adequate test. However, it was felt that the variety of experiences that the VON staff had with informal caregivers would provide a sufficient level of validation at this moment in time, and with the rapid development nature of the WIDE platform, essential changes to CareNet could be identified later and easily implemented. Based on the results of this research project, the VON management has elected to contract the development and operation of a complete system to a Canadian company.

Results

The CareNet portal is more than just an information source and chat room. It provided a test bed for a number of concepts related to user support and also to methods of software development in domains where the experts themselves are still trying to understand the problem at hand. From the user perspective, the CareNet portal is a proof of concept of:

1. a Community of Practice support system. The CareNet portal supports two identifiable groups of people who share a concern or need for the delivery of informal care namely the informal caregiver and a specific distributed group of VON staff members. They come from different perspectives, bring different expertise to the community, and can be mutually supportive of each

other. The system itself allows interaction on an as needed basis, providing reinforcement, assistance and interaction on demand.

2. an experimental platform for supporting caregivers. The platform allowed the members of the community of practice both to share existing information and also to add new sources and experiences. It is recognized that new information would need to be vetted and automatically indexed to ensure that inappropriate information is not inserted into the system and that it can be easily found at a later date. However, it should be possible through appropriate controls to ensure that vetting of information would not be an onerous task.
3. a web-based system which can be maintained and enhanced by VON staff with minimal technical expertise support. This property is key to the successful operation of any web-based information system, particularly the interactive one that we have proposed. The information experts should have the ability to maintain the content themselves with little or no help from technical support staff. This property is essential to the ongoing maintainability of the system.
4. a rapid development environment for web-based applications. With the advent of more public systems such as the web, we are exploring many web applications with wider audiences. In many cases, these applications are not well understood and need to be modified frequently based on the experience and feedback from both the users and the service providers. Thus a rapid development environment is an essential tool where the information system can be modified quickly and reliably, preferably while it is in operation.

Application details

Three applications were developed for the CareNet portal prototype which is now operational. The applications are quite advanced and include the following capabilities:

VON directory

This application created a searchable directory of VON branches and their services. The information on each branch includes a list of its services as well as a map showing its street location. The directory can be viewed in English and in French. The directory database is maintained online. Maintenance of the directory information can be done easily by authorized central or branch staff using simple web-based forms. This directory would not only be useful for the general public seeking the closest VON branch or the branches which offer a particular service, but it is also an online resource for VON staff in answering client questions and a tool to analyze VON resources. For example, “How many branches offer certain services?;” “Where are possible gaps in services?;” “Should names of services be consistent?;” and if so “What should they be called?”

Caregiver support system

To create this application, three types of information a caregiver needs were considered:

a) Information about a particular disease or health-related problem (Constraints).
b) What organizations can provide help (Resources).
c) How to do something (Protocols).

Collectively, the Constraints, Resources and Protocols were called CPR. A system was designed to include these three types of information resources. This application was based on a section of the "Resource Guide for Family Caregivers", a manual [5],[6] produced by individuals associated with the Family Caregivers Network Society in Victoria BC. A prototype interface was designed to determine the caregiver’s request and display the appropriate information resource or set of resources. The system can provide information that provides support for both the caregiver and the recipient of the care. For example, it handles questions such as: “How do I get support so I can take a break from my care giving role?,” (Resource) or “What kind of help can I get for my parent with Parkinsons disease? ” (Constraint, Protocol)

Caregiver information and knowledge sharing tool

There is a wealth of useful information that professionals have not documented or that caregivers have learned from personal experiences. This application allows both caregivers and professionals to document this knowledge and information. The tool was designed to allow the collection of this information and to make it available so that others could share and gain value from this information in a simple way. The initial tool is based on a discussion forum. The information is categorized (indexed) and can be searched. Such a tool normally would also contain a vetting mechanism to assist in determining appropriateness with perhaps some editing capability.

The WIDE toolkit

The CareNet system was developed using the WIDE Toolkit, a set of technologies that have been used to construct frameworks[3] for over 30 portals. The WIDE Toolkit contains the following functionality:

- The ability to connect to multiple distributed, disparate databases. This function allows each VON site to produce and maintain databases for local services while supporting a global view of the information.
- An indexing and search engine This function supports indexing and subsequent searching of all databases and documents placed in the portal.
- Structured database content presentation (listings, reports). Reports such as a listing of services and where they are available can be easily introduced and modified.
- Database content administration (remote updates). Data can be changed by any authorized person from any location.
- Structure administration (maintenance of content presentation). The entire WIDE toolkit is built on the premise that we may want to modify both the presentation and the structure of the application.

- Online distance education tool. Succinct online documentation is a key to portal usability and maintenance. The system incorporates an easy-to-use online educational tool that provides the user with the ability to read and annotate documentation.
- Geographic mapping of data. The mapping tools are simple to implement and use but are quite powerful. They allow viewing of data by geographic location, searching of an area for geographic data with specific properties and posting and recording of geographic data. The mapping tools are not based on geographic information systems (GIS) but can communicate with them in both directions (read, write) through accepted standards.
- Interactive visual display (display charts and diagrams constructed from database content). The display tools are derived form the mapping tools and have many of the same properties.
- Access control framework. Access to the data and various reports and other displays is role based. Thus a person can easily be included and excluded from changing data.
- Notifications of events. Based on criteria, users can request notification by e-mail of additions, changes or deletions to content of interest.
- Agents. Agents typically play housekeeping roles and can perform functions such as reporting and/or repairing discrepancies between databases/web sites automatically.
- A self-assessment framework. The self-assessment framework which is an application of WIDE can be populated with data appropriate to a particular field of study and the user can determine their knowledge in that field. The framework also connects to supplementary materials for self-study.

Discussion

CareNet provides caregiver support and demonstrates what can be gained by using tools, such as the WIDE Toolkit, that reduce the barriers to end-user development and sustainability[1],[2]. This is particularly important when new types of requirements and little user knowledge of potential solutions exist. CareNet allows significant systems support in the home setting using browser accessible technologies.

This research project is a stepping stone to further investigation. Many questions remain unanswered and new issues became evident that should be explored. For example, what are the best methods for the capture of undocumented organization information? How can we easily classify this type of information to enable the rich retrieval of this information? What are effective mechanisms to capture the needs of the caregiver? How do we enable the professional caregiver organizations to use this system as a powerful analytical tool?

Conclusion

Home and community care organizations can benefit from applications systems like CareNet and toolkits that allow affordable and rapid system construction and revision. If components of the EHR are to be sourced from the community setting, solutions like these are essential.

Finally tools such as the web provide new methods of supporting communities of practice. What are the most effective way of using these tools? Many unanswered questions still remain to be explored in this area as well.

Acknowledgments

This work was supported by the VON Canada.

References

[1] Cowan DD, Fenton SL, Mulholland DW. Lowering technological barriers to developing, deploying and maintaining web-based interactive information systems. International Journal of Technology, Knowledge and Society 2005; 1(4):175-82.

[2] Cowan DD, Covvey HD, Fenton S, Miller J, Mulholland D, de Groot M, Matyas C, McGarry F. Sustainability models for web-based information systems supporting communities of practice. International Journal of Technology, Knowledge and Society, to appear.

[3] Cowan DD, Alencar PSC, Luo M. A framework for community information systems. Annals of Software Engineering (ASE), Special Issue on Object-oriented Web-based Software Engineering 2002; 13:381-411.

[4] Cowan DD, Fenton S, Mulholland D. The web-based informatics development environment (WIDE), http://csg.uwaterloo.ca/wide.htm, 2006.

[5] Family Caregivers Network Society, Resource Guide for Family Caregivers, Victoria, BC, 2006.

[6] Family Caregivers Network Society, Educational Activities to Support Family Caregivers: Facilitator's Manual, Victoria, BC, 2006.

[7] Hildreth P, Kimble C. Knowledge Networks: Innovation through Communities of Practice, Idea Group Publishing.

[8] Kimble C, Hildreth P, Wright P. Communities of Practice: Going Virtual, Chapter 13 in Y. Malhotra (Ed.) Knowledge Management and Business Model Innovation, Idea Group Publishing, Hershey (USA)/London (UK), 2001, pp 220-34.

[9] Kimble C., Li F, Barlow A. Effective Virtual Teams through Communities of Practice, University of Strathclyde Management Science Research Paper No. 00/9, 2000.

[10] Pan Shan L, Leidner D. Bridging communities of practice with information technology in pursuit of global knowledge sharing. The Journal of Strategic Information Systems 2003;12(1): 71-88.

[11] Wenger E, McDermott R, Snyder W. Cultivating communities of practice: A guide to managing knowledge, Harvard Business School Press, 2002.

[12] Wenger E. Communities of practice: Learning, meaning, and identity. Cambridge University Press, New York, 1998.

Address for correspondence

Shirley L. Fenton
Managing Director
Waterloo Institute for Health Informatics Research
University of Waterloo
Waterloo, Ontario, CANADA
N2L 3G1

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Core Features of a Parent-controlled Pediatric Medical Home Record

Roberto A. Rocha[a,b], Alfred N. Romeo[c,d], Chuck Norlin[c,e]

[a]RemedyMD, Inc., Utah, USA; [b]Department of Biomedical Informatics, University of Utah School of Medicine, USA; [c]Utah Collaborative Medical Home Project, Utah, USA; [d]Utah Department of Health, USA; [e] Department of Pediatrics, University of Utah School of Medicine, USA

Abstract

We describe a coordinated effort to identify the core features of a parent-controlled personal health record for children with special health care needs, involving parents, care givers, and healthcare providers. A summary of the core features is presented, emphasizing needs that are not commonly recognized as functions of a generic personal health record. Our goal was to identify requirements for personal records that empower parents to effectively obtain, organize, understand, and communicate the information necessary to help their children receive the best possible care.

Keywords:

Personal Health Record, pediatrics, children with special health care needs, consumer informatics

Introduction

Children with developmental disabilities and other complex medical problems (collectively, "*children with special health care needs*" or CSHCN) comprise about 12.8% [1] of children in the US and are defined as "those who have or are at increased risk for a chronic physical, developmental, behavioral, or emotional condition and who also require health and related services of a type or amount beyond that required by children generally" [2].

Some CSHCN have relatively common disorders, such as asthma and attention deficit hyperactivity disorder, but the majority have uncommon diagnoses and many have associated disabilities. Roughly 6.5% of US children experience some disability as a result of a chronic condition, the most common ones being respiratory and mental impairments [3]. Rates of disability have increased over the past two decades, as have racial disparities in disability prevalence. The rate of activity-limiting disability for white children increased from 4.07% to 5.97% between 1979 and 2000, while the rate for black children increased from 3.79% to 6.71% [4].

The value of comprehensive and coordinated medical care for children with chronic conditions and disabilities has been demonstrated by several studies [5-8]. The importance of such care for CSHCN has led the American Academy of Pediatrics and the federal Maternal and Child Health Bureau to promote the "Medical Home" concept [9]. Ideally, care provided in the Medical Home is "accessible, family-centered, continuous, comprehensive, coordinated, compassionate, and culturally-effective" [10]. The Medical Home was conceptualized around one of the key elements of comprehensive care: ready access to all relevant information about a patient with a chronic or complex medical condition [11].

Typically, care for CSHCN is provided by numerous individuals (physicians, therapists, dentists, counselors, etc.) and institutions (clinics, outpatient centers, hospitals, etc.), and in multiple environments (home, day care, school, etc.). It is rare that these providers and settings share information systems or communicate information in ways that result in comprehensive, coordinated care. The primary care provider ("Medical Home") is envisioned as the "central repository" for collecting, managing, and appropriately sharing this information, but this goal has yet to be realized.

Personal child health records have been promoted by national and international institutions and have been widely used and appreciated by parents for many years [12]. A national survey published in 2006 found that only 21.3% of primary care pediatricians used an "Electronic Health Record" (EHR) [13]. Until there is widespread adoption and use of interoperable electronic records by primary care providers and other providers, schools, etc., we probably should not expect physicians to offer a viable EHR solution to CSHCN.

Parents of CSHCN, particularly those whose children have developmental, physical, or mental disabilities, spend much of their time caring for their child, managing their multiple health care providers and the information generated by them, and interfacing between those providers, insurance companies, day care and schools, interested family members, and many others. The amount and complexity of information that these parents, and the children themselves as they near adulthood, need to understand, remember, and communicate is daunting. Accurate and complete information is critical to assuring the quality and safety of the health care such children receive, optimizing the coordination of care among providers, and minimizing the duplication of diagnostic and therapeutic interventions. Ready access to records of care, instructions, and reliable information about their child's condition and recommended therapies may encourage parental compliance, as well as educate family members or other caregivers. The ability to access, record, and share information with providers can save time, improve care, and prevent errors. In essence, a parent-controlled health record may offer a bet-

ter and more feasible solution that could take advantage of and enhance collaboration with the Medical Home.

Materials and methods

The overarching needs that guided the identification of core features for a parent-controlled "Pediatric Medical Home Record" (*PedMHR*) were derived from the "Personal Health Record" (PHR) principles recently published [14], including: a) parents are ultimately responsible for decisions about their children's health; b) parents should have access to a reliable and complete record of their children's health information; c) parents should have control and accountability over how these records are used and shared; and d) information in the PHR should be understandable to parents and other caregivers.

In order to identify the core features of an electronic record that would be perceived as helpful for families of CSHCN, a better understanding of the daily life of these children and their families was considered a critical step. Our intent was to identify a small group of parents with extensive experience in advocating for improved care for their children, and who were passionate about improving systems of care for other families.

Five parents were ultimately identified. The chronic clinical conditions found in their children included: Down Syndrome (3 children), Celiac Disease (1 child), Cerebral Palsy (1 child), Cystic Fibrosis (1 child), and developmental delays not yet associated with a diagnosis (1 child). Within these families, the mothers were the primary caregivers. Local groups involved with families of CSHCN, such as Utah State University's Center for Persons with Disabilities[1], the Utah Down Syndrome Foundation[2], and the Utah Family Voices[3], provided great assistance in the family selection process.

Each family received a survey designed to help identify their computer literacy and the level of complexity of the care required by their children. The survey was preceded by a brief description of the project, along with phone and email contacts that could be used to obtain further explanations. Table 1 presents the survey questions.

The five families were also invited to participate in a focus group discussion. The focus group discussion was organized as a mediated phone conference call where each family described a typical day in their lives related to their disabled children. Each family was asked to elaborate on the potential value and desired features of an electronic parent-controlled health record. The session lasted almost 2 hours and was recorded after consent was obtained from each participant. Table 2 presents a summary of the topics discussed.

The information obtained from the families was reviewed and annotated by the authors, leading to a comprehensive set of needs and features. A review of the literature pertinent to PHRs, with a special emphasis on systems proposed for children was also performed.

1 http://www.cpd.usu.edu/
2 http://www.udsf.org/
3 http://www.familyvoices.org/

Table 1 – Survey questions distributed to the families

Questions	Answer
How many medical providers is your child involved with?	[Numeric]
How many other types of professionals is your child involved with?	[Text]
Do you have a computer at home?	Yes; No
Do you have Internet access?	Yes-Dial-up; Yes-DSL; No
How often do you use the Internet (World Wide Web)?	Daily; Weekly; Monthly; Rarely; Other (specify)
How comfortable are you with using a computer?	Very Comfortable; Somewhat Comfortable; Not Comfortable; Other (specify)
Do you use a Personal Digital Assistant (PDA) or hand-held organizer?	Yes; No
What type of medical equipment do you use in your home?	Blood pressure device; Oximeter; Nebulizer; Glucometer; CPAP; Other (specify)

Results

A long list of desiderata was generated from the focus group discussions and subsequent interactions with the families and the local groups involved with CSHCN. Many identified features are recognized as traditional PHR features, including the desire to track provider visits, ask questions prior to the next visit, review treatments, labs and tests ordered and medications prescribed, award password protection and proxy rights to designated others, edit the record but with a provider or care-coordinator validating its accuracy, link to patient-friendly medication information, and "make sure" that providers read what was entered in the record.

The sections below highlight features considered important by families of CSHCN, with special emphasis on items that are not described as standard PHR functions [15-17].

Core features

A *PedMHR* should expand on the important universal elements of a personal child health record (e.g., growth charts, developmental monitoring, immunization records, and advice/information [18]) to develop a comprehensive record that will meet the more complex requirements of CSHCN, their families, and their providers of health and other services.

The key features of this record should include: ease of accessing, recording, and organizing information about care; secure and customizable sharing of information with others in electronic or printed formats; focused links to relevant and reliable information and decision support; links to services and other resources, both locally and nationally; integration of voice recording, transcription, and indexing of visits with providers; integrating financial management tools; and tools for creating and maintaining care plans, with reminders for parents and providers.

Table 2 – Topics selected for the focus group discussion

Topic
Please describe a "day in your life", focusing on activities related to your child, either directly or indirectly.
If you could design the perfect system of storage and access to your child's health records, what would you want in that system?
How would you want to receive new information and learn new activities that may be helpful to your child?
What would you want to help you objectively record and analyze the progress your child is making based on the care plan from the doctor?
What would you want to help you and your doctor detect problems that may affect the treatment and therapies that your child is receiving?
Would you be opposed to having your child's medical record on the Internet at a password-protected site?
Would you want a record of your child's appointments, diagnoses, history, medications, providers, etc, as a reminder for you and to hand to new providers so you don't have to repeat your story?
Should it include a way to send messages to your doctor or other providers, or schedule appointments, refill prescriptions, etc?
How about the ability to write down things to remember, notes, questions to ask, progress your child made, etc., and to link that to your provider(s) or to guidelines to help identify problems?
What other things have you thought of today that you want to share with us?

Information acquisition and storage

Conversations during physician visits are highly valued by patients/parents and generally assumed by physicians to effectively transmit their opinions, explanations, and recommendations. However, much of what is said is missed or forgotten by parents after the visit [19], impairing compliance with treatment and depriving family members and other caregivers of the opportunity to understand the "what, why, and how" of the physician's suggestions. The written record, transmitted electronically or in printed format (and later scanned into the electronic record), may be sufficient to communicate the "what", but is not likely to adequately communicate the "why and how." Some patients have found audio recordings of physician visits to be helpful both to review and remember details and to allow others (e.g., family members) to better understand the communications [20].

A *PedMHR* should provide multiple mechanisms to acquire information, including electronic forms for direct data entry, paper scanning into digital documents, and uploading of digital audio and video. However, information captured using these mechanisms has to be properly structured and encoded to enable computerized repurposing and decision support.

The integration of pre-visit questionnaires, screening instruments (e.g., for development, behavior, depression), medication response assessment tools, and treatment diaries into a *PedMHR* should assure access to them, facilitate their completion (via reminder notices), transmission, and evaluation or validation (prompted by an email notice to the provider), and their integration into the ongoing record.

Perhaps the most useful feature of a *PedMHR* for a child with a chronic, complex condition would be a documented care plan. Such plans allow details of care needed by CSHCN to be recorded and shared with parents, providers, and designated others (e.g., family members, educators). Care plans may also include goals against which progress can be measured, timelines able to trigger reminders of needed labs, appointments, or phone calls, and links to information, instructions, and data entry forms.

Information access and reporting (repurposing)

Parents often lament the need, particularly when their child is being admitted to a hospital, to repeat their "story" over and over to interns, residents, attendings, specialists, etc. A *PedMHR* should dynamically provide preformatted and customizable reports to alleviate much of this frustration and assure the accurate transmission of appropriate and current information. Ideally, such reports should be provided electronically and include a rich set of links and annotations that would offer additional details about specific problems, prior reactions to medications, parent-recorded symptom diaries, and other features. These reports should also be configured with detailed utilization monitoring, enabling parents to ascertain who accessed the information and when.

Similarly, templates for "standard reports" such as letters of necessity, plans of care, school forms, insurance reports, etc., should direct the extraction of information from a *PedMHR* and its subsequent customization and formatting. Condition-specific growth charts (e.g. Down syndrome) should plot growth from measurements entered by providers or caregivers.

Links to pertinent information and knowledge resources

A key need identified by parents of CSHCN is access to information about their child's condition, available services and other resources, educational interventions, growth and development, support groups, financial management, and other aspects of caring for their child and family [21-22]. A number of web sites have been developed over the past several years aimed at providing such information for parents, physicians, and others. The Med-

Home Portal[4] has been serving many of these information needs in Utah since 2001. A *PedMHR* should offer configurable context-aware and data-driven "infobuttons" [23] to sources of specialized information, enabling parents to continuously improve their knowledge and understanding of their child's condition. The utilization of these links should also trigger simulated scenarios followed by focused tests, helping parents to practice what they have learned.

Knowledge sharing and collaboration

A *PedMHR* should enhance the ability of parents and physicians to share knowledge, information, and ideas. Parents of children with complex conditions often become experts in those conditions, particularly in their child's manifestations and responses to treatments. A *PedMHR* should enable parents and providers to share valued resources (e.g., web sites, articles, experiences, observations, best practices) using a variety of online tools (e.g., blogs, message boards, podcasts, wikis, vlogs). A *PedMHR* should make it possible for parents and providers to critique existing resources and suggest new ones, and also to allow parents and providers to develop, individually or collaboratively, a wide variety of brand new educational resources. Given the significant effort required to create such educational resources, a *PedMHR* community should let parents and providers receive monetary rewards for their authoring and editorial efforts, with the perceived usefulness of a resource defining its commercial value.

Communication with providers and care givers

A *PedMHR* should provide for asynchronous messaging and consultations among parents, providers, and caregivers. Asynchronous communications between physicians and patients via email and web-based secure messaging is gradually increasing and some insurers have piloted compensating physicians for such services. Though some resist providing care without face-to-face contact, there are many situations in which electronic consultations may be preferable. "Curbside consultations" by primary care physicians with subspecialists provide potential value in saved time and expense for patients, expediting obtaining an opinion or beginning an intervention, and enhancing the knowledge of the primary care provider. Accomplishing such consultation electronically, linked to authorized access to a *PedMHR*, could potentially be more efficient and useful. With integration of demographic records, these messages could also serve as documentation for billing patients or third parties. Compensation would provide added incentive to physicians and acknowledge the value of the service to all parties.

Financial management

For parents of CSHCN, managing family finances and negotiating the maze of insurance benefits can be a frustrating, nearly full-time job. A *PedMHR* should help families by integrating financial management features, including tools for tracking medical bills and payments, links to insurance companies, suppliers, and providers, and providing reports for use with tax preparation software. Similarly, the clinical activities tracked by a *PedMHR* should be used to document and trigger the appropriate charges and reimbursement. Parents should retain complete control over the content and routing of electronic transactions.

4 http://medhomeportal.org/

Integration with local and regional repositories

As communications among health data repositories become more standardized and patient privacy issues are better integrated into information systems, linking a *PedMHR* to those systems and to service sites should serve to further enhance its utility. In Utah, the "Child Health Advanced Records Management" (CHARM)[5] system is compiling personal health information from several state databases, serving as an electronic broker to improve access. Similar efforts should enable a *PedMHR* to share data with vital records, the state's immunization registry, newborn screening and birth defects registries, and other programs' databases. "Utah Clicks"[6] is a web-based "universal application system" enabling parents to apply for multiple services for children, including: Medicaid, "Baby Your Baby" (prenatal support services), "Head Start and Early Head Start", "Baby Watch" (early intervention), and the CSHCN program (evaluation and treatment for developmental and medical problems). A *PedMHR* should seek integration with such systems and repositories, thereby not only providing and retrieving relevant data in real-time, but also using "intelligent" (context-aware) interfaces to find and access the appropriate services.

Discussion

Despite the significant achievements of the informatics community in defining standardized and interoperable systems, substantial efforts are still required to properly structure and encode the information associated with complex clinical conditions, such as those presented by CSHCN. Similarly, automated methods for consistently extracting precise information from a wide variety of textual and non-textual documents also remains an important challenge.

Without the availability of "computable" data most of the advanced information retrieval and repurposing features identified will not be feasible, dramatically reducing the usefulness of a *PedMHR*. The consistent extraction and repurposing of information represented in clinical documents is one of the core features of a *PedMHR*. However, the structure and content of these documents, as well as their metadata and provenance, must be standardized and predictable, otherwise such features become restricted to documents authored locally by *PedMHR* editing tools.

Particularly in the case of care plans, a number of templates have been published and, in working with our participating parents and providers, it is clear that several new "standard" templates that can be customized to fit various clinical diagnoses and life situations will have to be created. Multiple aspects of the challenges just mentioned regarding the structure and content of the information apply directly to care plans, potentially restricting their usefulness and functionality.

Significant advances in how web-based information resources are structured, indexed, and encoded need to be completed before context-aware and on-demand learning can become a reality. Most of these features are encompassed by what is described as the "Semantic Web"[7]. Our

5 http://charm.health.utah.gov/
6 https://utahclicks.org/index.cfm?
7 http://www.w3.org/2001/sw/

group was recently awarded three years of funding from the National Library of Medicine to substantially expand the MedHome Portal offerings and features, making it compatible with these innovative learning efforts. In terms of web-based communication and collaboration, while most of the technologies mentioned (e.g., blogs, podcasts, wikis) are currently available, user literacy and access must be addressed if all members of the *PedMHR* communities are to become active and knowledgeable collaborators.

A successful implementation of a *PedMHR* clearly has to rely not only on ubiquitous secure communications, but also on advances related to privacy and data security, and technology access. Mechanisms for recognizing and compensating providers for electronic messaging and teleconsultations have to be widely implemented. Potential liability issues must also be addressed. Similarly, ongoing "e-commerce" standardization efforts will greatly benefit the integration of a *PedMHR* with other non-clinical systems. Substantial progress on the specification and standardization of real-time interoperable web services are required before seamless integration can be obtained.

Conclusion

We have attempted to identify the core features of a special kind of PHR, characterized as a *PedMHR*. In essence, a *PedMHR* would empower parents of CSHCN to effectively obtain, organize, understand, and communicate the information necessary to help their children receive the best possible care. This vision was promptly endorsed by the families of CSHCN we had privilege to interact with and also by their providers. Our intent is to eventually create a *PedMHR*, where the parents/family have complete control over the clinical data and can use these data not only to ensure optimal care planning and coordination, but also to acquire and disseminate knowledge about the conditions and needs of their children. We are certain that CSHCN, along with their families and providers would greatly benefit from a *PedMHR*.

Acknowledgments

We gratefully acknowledge Gina Pola-Money and Jefferson C. Sheen, MSW for their help in conceptualizing a *PedMHR* and in organizing the parent focus group.

References

[1] van Dyck, P. C., Kogan, M. D., McPherson, M. G., Weissman, G. R., & Newacheck, P. W. (2004). Prevalence and characteristics of children with special health care needs. *Arch Pediatr Adolesc Med*, 158(9), 884-890.

[2] McPherson, M., Arango, P., Fox, H., Lauver, C., McManus, M., Newacheck, P. W., et al. (1998). A new definition of children with special health care needs. *Pediatrics*, 102(1 Pt 1), 137-140.

[3] Newacheck, P. W., & Halfon, N. (1998). Prevalence and impact of disabling chronic conditions in childhood. *Am J Public Health*, 88(4), 610-617.

[4] Newacheck, P. W., Stein, R. E., Bauman, L., & Hung, Y. Y. (2003). Disparities in the prevalence of disability between black and white children. *Arch Pediatr Adolesc Med*, 157(3), 244-8.

[5] Chernoff, R. G., Ireys, H. T., DeVet, K. A., & Kim, Y. J. (2002). A randomized, controlled trial of a community-based support program for families of children with chronic illness: pediatric outcomes. *Arch Pediatr Adolesc Med*, 156(6), 533-539.

[6] Farmer, J. E., Clark, M. J., Sherman, A., Marien, W. E., & Selva, T. J. (2005). Comprehensive primary care for children with special health care needs in rural areas. *Pediatrics*, 116(3), 649-656.

[7] Liptak, G. S., Burns, C. M., Davidson, P. W., & McAnarney, E. R. (1998). Effects of providing comprehensive ambulatory services to children with chronic conditions. *Arch Pediatr Adolesc Med*, 152(10), 1003-1008.

[8] Palfrey, J. S., Sofis, L. A., Davidson, E. J., Liu, J., Freeman, L., & Ganz, M. L. (2004). The Pediatric Alliance for Coordinated Care: evaluation of a medical home model. *Pediatrics*, 113(5 Suppl), 1507-1516.

[9] American Academy of Pediatrics (2002). The medical home. *Pediatrics*, 111, 184-186.

[10] Ziring, P., Brazdziunas, D., Cooley, W., Kastner, T., Kummer, M., Gonzalez de Pijem, L., et al. (1999). Care coordination: integrating health and related systems of care for children with special health care needs. *Pediatrics*, 104(4 Pt 1), 978-981.

[11] Sia, C., Tonniges, T. F., Osterhus, E., & Taba, S. (2004). History of the medical home concept. *Pediatrics*, 113(5 Suppl), 1473-1478.

[12] Walton, S., Bedford, H., & Dezateux, C. (2006). Use of personal child health records in the UK: findings from the millennium cohort study. *BMJ*, 332(7536), 269-270.

[13] Menachemi, N., Ettel, D. L., Brooks, R. G., & Simpson, L. (2006). Charting the use of electronic health records and other information technologies among child health providers. *BMC Pediatr*, 6, 21.

[14] *The Value of Personal Health Records - A Joint Position Statement for Consumers of Health Care*. (2006). Retrieved Dec 1, 2006, from http://www.amia.org/inside/releases/2006/ahima-amiaphrstatement.pdf

[15] Kim, M., & Johnson, K. (2002). Personal health records: evaluation of functionality and utility. *J Am Med Inform Assoc*, 9(2), 171-180.

[16] Simons, W., Mandl, K., & Kohane, I. (2005). The PING personally controlled electronic medical record system. *J Am Med Inform Assoc*, 12(1), 47-54.

[17] Tang, P., Ash, J., Bates, D., Overhage, J., & Sands, D. (2006). Personal health records: definitions, benefits, and strategies for overcoming barriers to adoption. *J Am Med Inform Assoc*, 13(2), 121-126.

[18] Hampshire, A. J., Blair, M. E., Crown, N. S., Avery, A. J., & Williams, E. I. (2004). Variation in how mothers, health visitors and general practitioners use the personal child health record. *Child Care Health Dev*, 30(4), 307-316.

[19] Kessels, R. P. (2003). Patients' memory for medical information. *J R Soc Med*, 96(5), 219-222.

[20] Liddell, C., Rae, G., Brown, T. R., Johnston, D., Coates, V., & Mallett, J. (2004). Giving patients an audiotape of their GP consultation: a randomised controlled trial. *Br J Gen Pract*, 54(506), 667-672.

[21] Farmer, J. E., Marien, W. E., Clark, M. J., Sherman, A., & Selva, T. J. (2004). Primary care supports for children with chronic health conditions: identifying and predicting unmet family needs. *J Pediatr Psychol*, 29(5), 355-367.

[22] Perrin, E. C., Lewkowicz, C., & Young, M. H. (2000). Shared vision: concordance among fathers, mothers, and pediatricians about unmet needs of children with chronic health conditions. *Pediatrics*, 105(1 Pt 3), 277-285.

[23] Fiol, G. D., Rocha, R., & Cimino, J. J. (2006). *Infobutton Standard API Proposal*. Retrieved Dec 1, 2006, from http://www.hl7.org/Library/Committees/dss/infobutton%2Dproposal%2D2006%2D01%2D04%2Ezip

Address for correspondence

Roberto A. Rocha, MD, PhD
RemedyMD, Inc;
9350 South 150 East - Suite 950;
Sandy, UT 84070; USA;
Email: rrocha@remedymd.com

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Empowering Patients to Improve the Quality of Their Care: Design and Implementation of a Shared Health Maintenance Module in a US Integrated Healthcare Delivery Network

Eric G Poon, MD MPH[1,2,4]; Jonathan Wald, MD MPH[1]; Jeffrey L Schnipper, MD MPH[2,4]; Richard Grant, MD MPH[3,4]; Tejal K Gandhi, MD MPH[2,4]; Lynn A Volk, MHS[1]; Amy Bloom, MPH[1,2]; Deborah H Williams, MHA[2]; Kate Gardner[1]; Marianna Epstein[1]; Lisa Nelson[1]; Alex Businger[1,2]; Qi Li, MD MBA[1]; David W Bates, MD MSc[1,2,4]; Blackford Middleton, MD MPH MSc [1,2,4]

[1]Partners Information Systems, Partners HealthCare, Boston, MA, USA [2]Brigham and Womens Hospital, Boston, MA, USA [3]Massachusetts General Hospital, Boston, MA, USA [4]Harvard Medical School, Boston, MA, USA

Abstract

We describe a health maintenance module within a personal health record designed to improve the quality of routine preventive care for patients in a large integrated healthcare delivery network. This module allows patients and their providers to share an online medical record and decision support tools. Our preliminary results indicate that this approach is well-accepted by patients and their providers and has significant potential to facilitate patient-provider communication and improve the quality of routine health maintenance care. Further research will determine the long term impact and sustainability of this approach.

Keywords:

personal health records, quality of care, preventive care, patient-centered care.

Introduction

In spite of spending more per capita on healthcare than any other country in the world, the United States (US) faces significant gaps in the quality of care patients receive(1). Recent research suggests that the US population only receives about half of the recommended evidence-based care, including routine health maintenance (HM) care such as cancer screening tests and adult vaccinations(2). Not only do these gaps in care endanger patients well-being, they represent lost opportunities to deliver cost-effective care to the population at large.

Several reasons likely account for these gaps in quality. Medical care in the US historically has focused on acute care, and not chronic or preventive care. Providers typically are financially rewarded for the quantity of care they deliver, not the quality of care(1). In the ambulatory setting, these factors pressure providers into scheduling more patients per hour rather than spending more time with each patient to address key care issues. In addition, the typical ambulatory practice relies on each individual clinician to detect areas where a patients routine HM care is out-of-date, an unreliable strategy for improving care given the limitations of human vigilance(3).

More recently, healthcare information technology (HIT) with decision support has been touted as promising means to improve quality in the ambulatory setting. However, many systems in the ambulatory setting do not provide adequate decision support, and even when decision support is available, not all providers use it in their clinical workflow(4). In addition, while research in academic settings has shown that decision support aimed at physicians does lead to increased adherence to preventive care guidelines, the improvements have been variable and at best modest(5-7). Given the challenges busy clinicians face in the ambulatory setting, improvement efforts that are directed solely at them may continue to fall short.

By empowering patients to become active participants in their care, an interactive personal health record (PHR) has the significant potential to overcome these quality improvement challenges(8). If patients are given the opportunity to anticipate the discussions that may occur during the clinical encounter about HM issues, they may be more likely to bring those issues to the attention of busy clinicians. Furthermore, if patients can be more informed about the nature of HM tasks, the patients and their clinicians will be able to discuss these issues more efficiently, thus mitigating time pressure as a barrier to good quality care(9).

At Partners HealthCare (Massachusetts, USA), we have developed and deployed a solution to leverage the power of a tethered PHR(8) to address quality and communication gaps in the deliverance of routine health maintenance care. This manuscript describes the rationale and design of a *health maintenance (HM) module* embedded within our personal health record, and will also present data on how this module is being used and patients attitudes toward it.

Methods

Study setting

Partners HealthCare was formed in 1994 through the financial merger between the Massachusetts General Hospital and Brigham and Womens Hospital. Since its inception, this integrated delivery network has grown to include five community hospitals, four rehabilitation and long term care facilities, and a large network of primary care and specialty physicians. Results of clinical tests performed throughout the network, including those that are involved in routine HM care, are stored in a common clinical data repository (CDR). Clinical activities in the ambulatory setting are supported by the enterprise electronic medical record the Longitudinal Medical Record (LMR) which can access data stored in the CDR in addition to other data types such as vaccinations and clinical notes. There are currently over 7,000 clinical users of the LMR.

Personal health record platform patient gateway

As part of an enterprise-wide strategy to facilitate communication between patients and their physicians, Partners HealthCare began in 2001 to develop a secure patient portal to serve as a communication gateway for participating primary care practices. The product, Patient Gateway (PG), allows patients to renew their medications, review their medication lists, request appointment and referrals, communicate with their practice via secure email, and access a licensed health information library(10). As of November 2006, PG supports 20558 patients across 23 primary care and specialty practices, with more than 4200 unique patients using it in any given month. The functionality described in the remainder of this manuscript was developed as additional features to the base PG product.

Design principles

The design of the HM module within PG was guided by several key informatics and patient-centered care principles:

- Data in the medical record ultimately belong to the patient, and the patients privacy must be protected at all times.
- When patients are given online access to their medical record, they should receive guidance on how to interpret the information stored in their record.
- Successful PHR solutions must facilitate workflow and communication for both patients and clinicians.
- Usability is key to patients and providers acceptance of any informatics solutions.

Functional goals for the HM module

We assembled a team of physicians, informaticians, health services researchers to define how to design the HM module to maximize its likelihood of acceptance by clinicians and patients and its potential to improve health care quality. This multi-disciplinary team set forth the following functional goals:

- To inform patients prior to visiting their primary care physicians about health maintenance care that is appropriate given their age, gender, family history and other co-morbidities.
- To remind patients about HM care items for which they may be overdue, giving them an opportunity to learn why they need them and why they are overdue.
- To inform patients about HM care items for which they are up-to-date and congratulate them.
- To allow patients to initiate the process of updating their HM record if they have had a care item performed at an outside facility.
- To encourage patients to state before the visit how they want to take care of HM care items for which they are overdue, passing that information to the clinician who can use it as a starting point for further discussion and planning.
- To facilitate documentation within the LMR by allowing clinicians to review data submitted by patients and to save them (with modifications if necessary) into the patients medical record.

Defining the knowledge base

In deciding on the types of *patient-centric* reminders the PG HM module would provide, we first looked to prior work that had reviewed well-accepted quality indicators in the US such as HEDIS measures, and trusted sources such as the US Preventive Services Task Force recommendations and other locally accepted clinical guidelines(6). Using the results of that review, a previous project had implemented a set of *clinician-centric* reminders within our electronic medical record LMR to improve clinicians adherence to established HM care guidelines(6). For the current project, we critically reviewed the knowledge base (KB) supporting the LMR *clinician-centric* reminders and, where necessary, updated it. The updated KB addresses the following 3 areas of routine HM care:

- *Womens health*: breast cancer screening, cervical cancer screening, osteoporosis screening
- *Adult Vaccinations*: influenza vaccine, pneumococcal vaccine, tetanus/diphtheria booster vaccine
- *Lipid Assessment*: for patients with average cardiac risk factors, those with diabetes and those with documented coronary artery disease

Colon cancer screening was not addressed in the KB because our clinical data repository did not contain results of endoscopy reports. The absence of these data might cause *clinician-centric* reminders within LMR and *patient-centric* reminders in the HM module to fire inappropriately. We therefore made the decision not to implement colon cancer screening logic at this point.

Because of a separate PG module that also allowed patients and physicians to collaborate on entering coded family history into the electronic medical record, we were able to adjust the recommendations to patients based on their family history. This benefited the logic for breast cancer screening, osteoporosis screening, and lipid management in that high-risk patients would be prompted

in the PG HM module to start screening at an earlier age and at closer time intervals.

Decision support architecture

Beyond allowing patients and clinicians to share information in the medical record, our project allows both parties to share decision support tools to improve the quality of routine HM care. To ensure consistency for the logic behind *clinician-centric* reminders in the LMR and the *patient-centric* reminders in the PG HM module, our team elected to manage the knowledge base (KB) for both types of decision support centrally. As demonstrated in Figure 1, a *reminder engine*, implemented as a series of services that can be used by a variety of applications, determines whether a patient qualifies for a particular HM care item using the *risk group* definitions within the KB. If a patient does not qualify for the HM item (e.g. a male patient in the case of mammography), neither the patient nor the clinician will see a reminder. If the patient does fall into the risk group (e.g. female patient over the age of 50 in the case of mammograms), the reminder engine further determines whether the *overdue condition is* met. If a patient falls under a *risk group* and the *overdue condition* is met (e.g. female patient over the age of 50 who has not had a mammogram within the past year), the PG HM module will inform the patient that the item is overdue and the LMR will inform the provider that the patient needs a mammogram. If, however, the patient falls under the *risk group* but the *overdue condition* is not met (e.g. a female patient over the age of 50 who had a mammogram 6 months ago), the PG HM module will inform the patient that the HM item is up-to-date, and the LMR will not remind the clinician to order a mammogram.

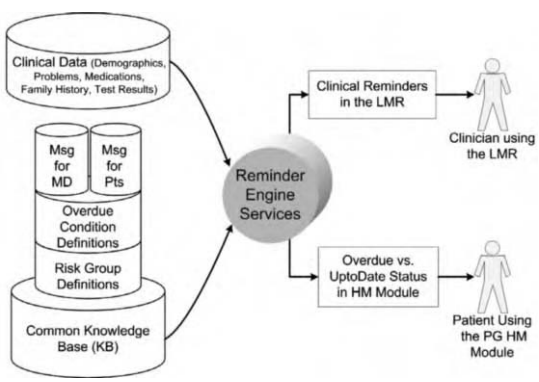


Figure 1 - Clinical Decision Support Architecture

While the patient and clinician may share the same decision support logic, the messages displayed to the two groups are necessarily different: the physician may get a very terse prompt about overdue items while the patient receives more extensive information to help him or her interpret the decision support prompts provided in the HM module. The content for these two groups is maintained separately within the KB.

User interface design process

As the multi-disciplinary team gathered the functional requirements, we iteratively developed a series of low and then high fidelity prototypes to explore different design approaches and refine the details of the design. We also solicited input from an advisory council comprised of primary care physicians from practices that were already using the base PG product. Finally, we conducted a series of usability tests with volunteer patients going through mock scenarios to enhance the ease of navigation and clarity of the user interface.

Functionality review patients experience

Patients with access to PG were invited to participate in this study to evaluate the HM module. Once they completed the consenting process, they were given access to their HM records within the HM module. Figure 2 illustrates the current functionality of this module. HM items that are overdue are grouped together and highlighted. Patients may review, for each overdue HM item, i) a short description of the HM item (e.g. what is a mammogram and why might women need one?); ii) a link to a more detailed description of the HM item, and iii) an explanation of why the patient may be overdue for this HM item (e.g. because the patients age and risk factor call for an annual mammogram and her last one was more than 2 years ago).

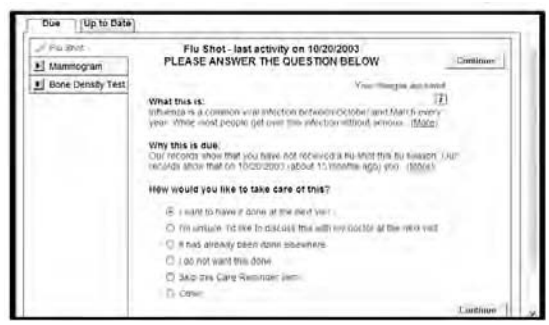


Figure 2 - Patients Experience in PG HM Module

Three weeks before a scheduled visit, patients are sent a secure message in which they are invited to review their HM record online. At this point, patients are prompted to indicate how they may want to address each overdue item in their HM record during the upcoming visit. Patients may indicate one of the following: i) they want to take care of this item at the next visit; ii) they are unsure what to do and would like to discuss it at the next visit; iii) they state that they have had this item taken care of elsewhere, after which they will be prompted to enter the specifics so that their clinician can update the medical record; iv) they do not want to address this item, or v) they want to defer the decision. After reviewing one or more items in their HM record, the patient can view a summary of their desired actions and submit the data for review by their clinician at the upcoming visit.

Functionality review clinicians experience

After a patient has submitted information through the HM module before the visit, the physician is prompted visually during the visit that data have been submitted by the patient. Upon reviewing the health maintenance screen within the electronic medical record (a commonly-used screen regardless of whether patient has used the HM module), data submitted by the patient through the PG HM module, if present, become visible (Figure 3). This information serves as a starting point for further discussion regarding each care item that may be out-of-date, and the information entered by the patient can facilitate documentation. For example, if the patient indicates that a vaccine was given outside the system and enters the date the vaccine was given, that information can be easily saved into the patients medical record once the clinician has reviewed it.

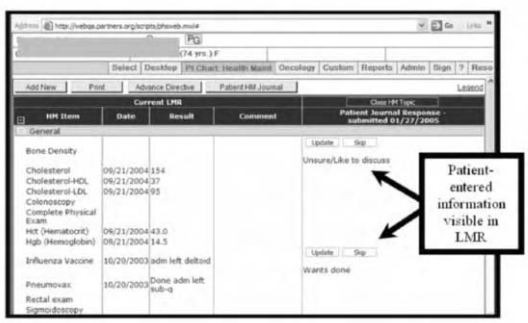


Figure 3 - Clinicians Experience in the LMR

Deployment strategy

We relied on several approaches to deploy the HM module within the PG product platform. First, we identified physician champions in each practice and solicited their input for the optimal ways to deploy the HM module within their practice. Second, we presented the HM module to physicians at practice meetings to introduce its functionality to the clinicians. Third, we provided on-site and online support, and returned to practices to hear feedback about the HM module after rollout. Fourth, we provided marketing materials to practices so that patients became aware of this new feature. Fifth, we sent reminders through the base PG product to all patients who had consented to the study to prompt them to review and update information in the HM module prior to the visit.

Short-term evaluation of the HM module

The HM Module was rolled out over the course of 14 months between October, 2005 and November, 2006 to 7 primary care practices within our integrated delivery network in conjunction with a family history module[1]. To evaluate the usage of the HM by patients and clinicians, we recorded the following: i) the number of patients who accessed the HM module to review their HM records; ii) the number of patients who were invited to update their record prior to their scheduled visit; iii) the number of patients who submitted their information to the clinicians, and iv) the number of times clinicians electronically reviewed data submitted by the patients through the HM module.

To evaluate patients attitudes towards the use of the HM and family history modules, patients who were invited to review their data prior to a scheduled visit were invited to complete a short online survey 3 days after the scheduled visit. On this survey, patients were asked to rate on Likert-scales their experience in reviewing and updating their medical records online, including whether they found the HM and family history modules easy to use, whether they felt more prepared for the visit, and whether the use of the modules led their providers to have more accurate clinical information. Patients were also invited to express other comments in free-text.

Long-term evaluation plan for the HM module

To evaluate the impact of the HM module on quality of care, we are conducting a cluster randomized controlled trial (RCT) in which the 7 practices whose patients are granted access to the HM and family history modules (intervention group) will be compared with 7 practices whose patients are granted access to other modules related to medication management and diabetes care (control group). Clinical outcomes to be assessed in the RCT include adherence to established HM guidelines. Patients knowledge about HM guidelines will also be assessed with surveys administered before and after the intervention period in both the intervention and control arms. We will also assess physicians and their staffs attitudes towards the HM module through separate clinician surveys. Results of these evaluation efforts should be available by late 2007.

Results

Between July 2005 and November 2006 patients who sought primary care at one of the 7 study practices were invited to participate in the study. Of them, 2,779 completed the consenting process, which included a baseline attitude survey. Of the 2,779 patients who consented, 63% were female. Their mean age was 47.4 at the time of consent.

Of these 2,779 patients who completed the consent process, 2,361 (85%) reviewed their HM records. In addition, 970 of 2,779 (35%) patients had a routine visit scheduled at least three weeks in advance and were therefore invited to update their HM record and state their preference for taking care of overdue HM care items. Of these 970 patients, 696 (72%) completed the review and updating process and submitted the information for their clinicians to review. Clinicians reviewed the data electronically within the electronic medical record for 460 (66%) of these patients.

Between July and November 2006, 437 patients who opened their invitation to update their HM record prior to a visit were further invited to respond to an online survey to assess their experience with the HM module. Overall, 179

1 The details of the family history module are being described in a separate manuscript.

patients (response rate = 41%) responded, and 81% of the respondents found the journal very easy or easy to complete. 51% of respondents either strongly agreed or agreed that the use of the journal led their providers to have more accurate information, with another 37% feeling neutral about whether the use of the journal had an impact in this area; 48% of respondents agreed that they felt more prepared for the visit, with another 41% feeling neutral.

While patients response to the HM module was largely positive, review of the qualitative comments from the survey revealed that not all clinicians were aware of their ability to review HM data submitted by the patient. Some patients felt that their clinician did not review data submitted by them. Others reported that their physicians asked them to fill out another paper survey in the waiting room that largely duplicated their interaction with the HM module. Certain patients desired the ability to enter information in free-form to the physician, such as reason for the visit or other active concerns.

Discussion

Our preliminary findings demonstrate that sharing the medical record and decision support tools between patients and their providers is a promising approach for improving quality of care. While not all eligible patients and clinicians used this new set of tools when offered the opportunity, those who did generally found the toolset easy to use and many thought that it made them feel more prepared for the visit and allowed their clinicians to have more up-to-date information.

There are several potential limitations to this approach to improve the quality of routine HM care. First, this approach does not benefit patients who do not have internet access or who do not have sufficient computer literacy to use our system. However, this concern is partially mitigated by the fact that internet access in the US is rapidly increasing (11). In addition, we have taken every effort through usability testing to ensure that our system is easy to use. Second, in our current implementation, only patients who are scheduled to have a visit are invited to update their HM records, and those without upcoming visits do not benefit from this aspect of the HM module. Also, patients are only invited to update their records in focused and structured ways, and cannot express their thoughts and concerns in free-form. We would have liked to provide these features, but physician practices were concerned early on during the project that allowing patients to submit updates to their HM record without an upcoming scheduled visit or providing for free-form entry might impose undue workflow and liability burden on the practices. Further research is therefore needed to examine these issues that arise from the deployment of tethered PHRs. In addition, if our long-term evaluation demonstrates a clinical benefit to the use of the HM module, then it may be possible to articulate the business case for the broader deployment of this approach. Finally, successful deployment of this approach requires the full support of local clinical leaders and extensive training. While we invested significant resources in these areas, our efforts did not reach all clinicians, as evidenced by the fact that some clinicians were not aware that they could review HM data submitted by patients. Optimal ways for deploying this technology at various types of institutions deserve further investigation.

Conclusions

We have implemented a novel approach to improve the quality of routine HM care by allowing patients and their providers to share the medical record and decision support tools. Our preliminary results indicate that this approach is accepted by patients and their providers. Further research will determine the long-term impact and sustainability of this approach.

Acknowledgements

The authors would like to thank the US Agency for Healthcare Research and Quality (Rockville, MD) for supporting this work via grant R01 HS 13326-03.

References

[1] Institute of Medicine. Crossing the Quality Chasm: A New Health System for the 21st Century. Washington, DC: National Academy Press; 2001.

[2] MCGlynn EA, Asch SM, Adams J, et al. The Quality of Health Care Delivered to Adults in the United States. NEJM 348(2635), 2645. 2003.

[3] Reason J. Organizational accidents: The management of human and organizational factors in hazardous technologies. Cambridge, England: Cambridge University Press; 1997.

[4] Simon SR, Kaushal R, Cleary PD, et al. Correlates of Electronic Health Record Adoption in Office Practices: A Statewide Survey. JAMIA. 2006;in press.

[5] Tierney WM, Overhage JM, Murray MD, et al. Effects of Computerized Guidelines for Managing Heart Disease in Primary Care: A Randomized, Controlled Trial. JGIM. 2003;18:967-76.

[6] Sequist TD, Gandhi TK, Karson AS, et al. A randomized trial of electronic clinical reminders to improve quality of care for diabetes and coronary artery disease. JAMIA 12(4), 431-437. 2005.

[7] Shea S, DuMouchel W, Bahamonde L. A meta-analysis of 16 randomized controlled trials to evaluate computer-based clinical reminder systems for preventive care in the ambulatory setting. JAMIA 1996;3:399-409.

[8] Tang PC, Ash JS, Bates DW, et al. Personal Health Records: Definitions, Benefits, and Strategies for Overcoming Barriers to Adoption. JAMIA 13(2), 121-126. 2006.

[9] Markle Foundation and Robert Wood Johnson Foundation. Connecting Americans to Their Healthcare. 2004.

[10] Poon EG, Wald J, Bates DW, et al. Supporting patient care beyond the clinical encounter: three informatics innovations from Partners Healthcare. Proceedings, Ann AMIA Symp, 1072. 2003.

[11] U.S.Census Bureau. Computer and Internet Use in the United States: 2003. U.S.Census Bureau . 2005.

Address for correspondence:

Eric Poon, MD MPH;
Email: epoon@partners.org

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Clinical Communication Ontology for Medical Errors

Yang Gong[a], Min Zhu[b], Jun Li[b], James P. Turley[b], Jiajie Zhang[b]

[a]*School of Medicine, University of Missouri-Columbia, 65211, USA*
[b]*School of Health Information Science, University of Texas Health Science Center at Houston, 77030, USA*

Abstract

Clinical communication failures caused 60% of sentinel events reported by the Joint Commission on Accreditation of Healthcare Organizations. The difficulties of communication have been the primary cause of errors leading to patients' death. For analyzing medical error events, uncovering the patterns of clinical communication, this paper reports the design and development of clinical communication ontology. The ontology contains eight axes and was validated using ten medical error cases, where communication was the main factor. The coding process demonstrates that the ontology can be used as a guideline for future medical error reporting system, through which the root cause of medical error due to communication will be revealed in a clear pattern. This ontology contributes to the generation of proper interventions and effective strategies for reducing medical errors.

Keyword:

medical error, communication, ontology, cognitive factors

Introduction

According to the Institute of Medicine, medical errors killed 44,000 to 98,000 people each year in the United States, which is more than breast cancer or highway accidents[1, 2]. Improper clinical communication is nearly twice as common as improper skills in causing preventive injury or death[3].Widespread attention to medical error and its prevention is becoming part of the culture in most health care organizations.

Researchers have found that the causes of errors are complex, involving human slips and mistakes and that systems weaknesses that predisposed to human error are important and recommended checks[4]. It is not surprising that big mistakes can lead to bad outcomes. However, unnoticed and un-rectified small mistakes, when they are accumulated, can cause bad outcomes as well. In this situation how errors occur is similar to a game, called *whispering game* or *telephone*, where a line of people are given a message at one end which is whispered to the next person down the line until it gets to the other. The final recipient then announces the phrase that he or she has heard, and it is compared to the original. Most of the time, the original message is distorted due to the cognitive distance, such as the difference of memory, language skill, and attention, etc. among the people in the communication line.

The *whispering game/telephone* phenomenon is a basic face-to-face communication, falls into the core level of the system hierarchy illustrating cognitive factors in medical errors [4], where the individuals trigger errors, and no technology usage is involved. Beyond this core level, individuals and technologies in clinical communication such as telephone, patient chart, whiteboard, post-it sticker, etc interact with each other; therefore, can cause more complex errors. At the level of distributed systems, communication is more complex among groups of people and technologies used by the groups. For example, it is prone to errors due to interaction when moving a patient to another unit or to a new team during a shift change. A flubbed handoff in 1995 caused one man in Florida having the wrong leg amputated[5].

Clinical communication is the main contributing factor to medical errors. Problems related to communication have taken 60% of sentinel events reported to the Joint Commission on Accreditation of Healthcare Organizations[6]. The information exchanged through such communication negatively affects the quality of decision making as well as healthcare quality. Earlier in 2006, in response to this report, the Joint Commission on Accreditation of Healthcare Organizations released a draft requiring U.S. hospitals to standardize their approach for handoff communications or risk losing their accreditations.

Clinical communication is carried out in various patterns, which can be involved with or without technologies. Interactions during communication can occur between person and person; person and technology; technology and technology. Observational studies have revealed that clinicians prefer face-to-face communication which is supposedly efficient in most occasions but frequently causes interruption and high communication loads[7-10]. The high workload with an interruptive nature in multitask setting may negatively affect clinicians' memory, attention, and work performance and causes medical errors. The difficulties of communication have been the primary cause of errors leading to the death[11]. Researchers have identified that clinicians spend 80% of their times in communication, with 30% of all communication events classified as interruptions[9]. Therefore, improving clinical communication efficiency and quality is helpful and significant in reducing medical errors.

Although researchers [12, 13] in different domains have classified various types of communication errors that occur between humans, and between humans and comput-

ers, little effort has been made to identify the cognitive factors which cause miscommunication clinical culture rather than device malfunction, channel failure, and low quality of signal, etc. Medical errors are products of cognitive activities when situated in a multitask, time critical and high workload domain[14]. Since medical errors are often triggered by human errors that occur during teamwork[4], the quality of communication among the team members has become our research interest.

In the context of communication, individuals, technologies, cognitive factors, and other factors can be used as key concepts to describe how communication is carried out. The communication ontology provides formal definitions and coverage of various concepts relevant to the domain of communication. It can be used to develop a knowledge-based reporting system through unification of information from data across various organizations, practice domains, and applications, and have sufficient detail to be of practical use. This ontology is an important component of any medical error ontology.

The purpose of this article is to describe the development of ontology of clinical communication for medical errors and explain how we use the ontology to support collection, storage and interpretation of the communication activities in clinical settings.

Materials and methods

Communication error ontology

An ontology can be defined as a specification of a conceptualization[15], it defines a common and controlled vocabulary for the purpose of enabling knowledge sharing and reuse of information[16]. Ontology has a broader application and richer functionality than a taxonomy which is a collection of controlled vocabulary terms organized into a hierarchical structure. A formal ontology is a controlled vocabulary expressed in an ontology representation language, for example OWL. This language has a grammar for using vocabulary terms to express something meaningful within a specified domain of interest. The grammar contains formal constraints on how terms in the ontology's controlled vocabulary can be used together.

To the best of our knowledge, there is no ontology on medical errors due to clinical communication. We have successfully retrieved some observational studies in the medical field on the medical errors due to communication [4-6] and a prototype of communication error taxonomy in the technology field [9]. These are of great help in collecting key concepts used in the ontology.

Taking grounding as a theoretical framework, we propose the communication ontology covering three levels of cognitive factors which include Level I, individuals; Level II, Individual-Technology Interaction within a team; Level III, Individual-Technology Interaction among teams. The collection of concepts included in the ontology was based on observational results of clinical activities and existing communication error taxonomy.

Although there are no methodologies of building ontology mature enough and widely accepted. The main methodology we employed to build ontology is ENTERPRISE methodology[17], which proposes the following stages:

1. identify the purpose and scope of the ontology;
2. build the ontology by capturing knowledge, coding knowledge and reusing appropriate knowledge from existing ontology;
3. evaluate the ontology;
4. document the ontology.

This methodology contains a set of techniques, methods and guidelines for each stage. Examples are, during capturing knowledge stage, identifying key concepts and relations, producing unambiguous text definitions for such concepts and relationships, identifying terms to refer to such concepts and relationships via brainstorming and meetings with domain experts.

Ontology scope and requirements

The concepts we collected for developing the ontology have been designed to satisfy Cimino's desiderata for controlled medical vocabularies[18], which are the core concepts in our ontology. Cimino has articulated that medical controlled vocabularies should include the following: Comprehensive content, concept-based, formal definitions, concept permanence, Multiple hierarchies, meaningless concept identifiers, do not use "not elsewhere classified", multiple granularities, multiple consistent views, context specific information, graceful evolution, and composition-decomposition.

An ideal ontology can introduce a host of structural and conceptual relationships including superclass/ subclass/ instance relationships, property values, time relationships, and others depending on the representation language used. As a collection of communication concepts in clinical settings, first step was to build up a thesaurus, and define each concept and the relationships. The collection of concepts is expected to be exhaustive and mutually exclusive. To be exhaustive is to reach the full coverage of the concepts needed for the description in clinical communication. To be mutually exclusive is to maximally reduce the conceptual overlapping between concepts used in the ontology.

Development approach

Knowledge acquisition was conducted through brainstorming and meetings with domain expert on the basis of literature review which provides multiple observational results on clinical communication[8, 9, 19] and some communication taxonomies for other domains.

Conceptual model of the ontology was developed by employing a middle-out approach. Instead of a bottom-up or top-down approaches, the middle-out approach begins by conceptualizing and defining the concepts that are more highly connected to other concepts. The definitions of core concepts, therefore, can be used as references to build up the definitions of simpler concepts. By doing so, it allows the use of an ontology reasoner to check whether or not all of the statements and definitions in the ontology are mutually exclusive and consistent, and to maintain a correct hierarchical structure of the conceptual model.

Protégé-OWL editor was used to build the ontology. RacerPro was used as a reasoner to check the definitions and

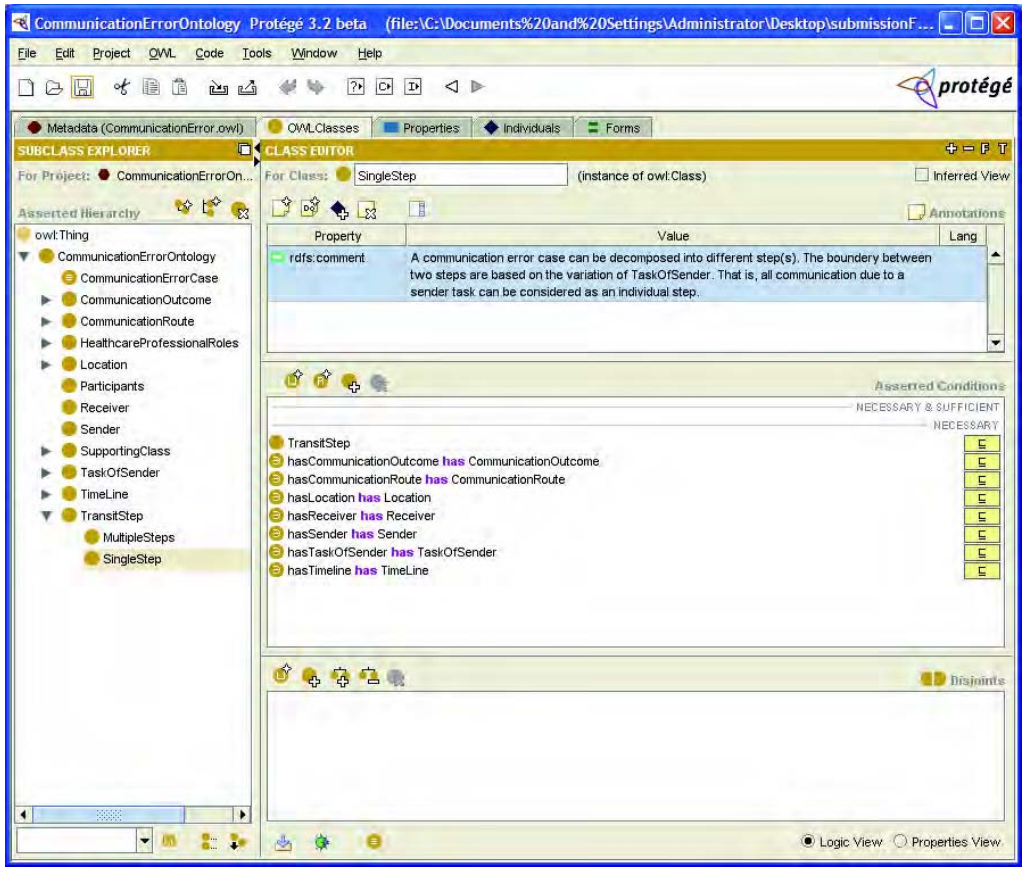


Figure 1 - CommunicationErrorCase Class presented in Protégé 3.2 Beta

logics in the ontology. The core part is the CommunicationErrorCase class which models a medication error due to communication. All the other classes are either the properties of the CommunicationErrorCase class or supporting class for these properties. In the ontology, a communication case typically contains steps by which information is communicated among senders and receivers and patterns associated with the steps. Two levels of classes were used in the ontology to describe the information relay in specific steps and the overall case pattern in general.

At the step level, each step is attached a time when the step happens, a location where the error occurs, the sender(s) or receiver(s) who involved in the activity. It also includes the description of messages which is communicated, the media used to carry the content of communication, and an outcome for each communication step. The CommunicationOutcome class is further defined with four levels of coordination for grounding mutual understanding. The levels are Conversation, Intention, Signal, and Channel. Each level is built on top of its lower level. The communication error could occur in any of these levels. In the situation where a communication error occurs at a level, none of its higher levels can be complete.

At the case pattern level, overall communication patterns are built in such as single step, linear step, circuit step, hybrid of linear and circuit step and so on. Communication participants are identified according to their roles. For example, clinicians usually play roles of sender in one step and may become receivers in the other. Patient outcomes are defined and incorporated with the standard medical nomenclature or medical error ontology[20].

As a result of such consideration, totally eight axes have been created for the ontology, which are TimeLine, Location, Participants, MessageCharacteristics, Media, TaskOfSender, TransitStep, and CommunicationOutcome classes. *Figure 1* shows the eight classes presented in Protégé 3.2 beta version. In each class related to the communication error case, we defined each class by adding subclasses, properties, and restrictions. The example of how to code communication error cases by our ontology is demonstrated in *Figure 2*.

Case coding as a process of validation

As a result of online search and literature, we collected 250 medical error cases for examining the usability of the communication ontology. Authors scrutinized all the cases and identified the cases reported communication as the cause of medical errors.

In general, the authors coded cases using the ontology in the following steps:

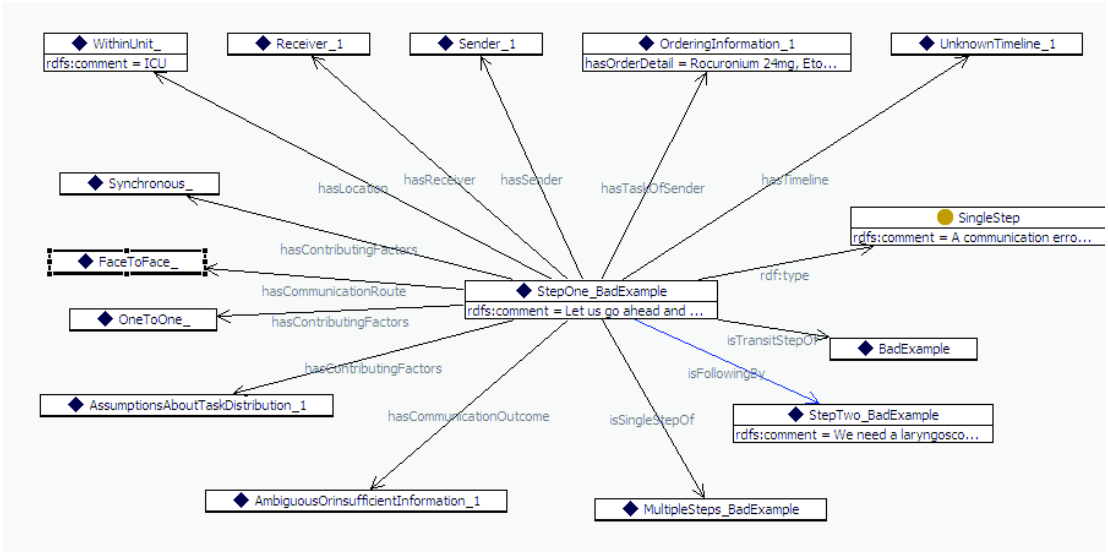


Figure 2 - Case Coding Results

1. Selecting communication from the medical error cases we had collected.
2. Selected communication cases were analyzed and coded individually by authors based on previous identified classes in protégé.
3. Three authors then discussed the case content, granularity and possible assumptions.
4. One of the authors then coded the case into protégé as final version. Meantime, we upgraded the ontology towards a more exclusive and exhaustive version as well.

Figure 2 shows the results of coding a case occurred in ICU where a group of clinicians worked collaboratively on a patient of respiratory failure. An attending physician tried to place an order with Rocuronium 24mg, Etomidate 10mg and atropine 0.4mg. Unfortunately, he did not specify the time point when these medications should be admitted. Shortly after his verbal order, an inexperienced nurse admitted the medication directly before a respiratory therapist had prepared the patient's respiratory track. As a result, the patient lost his spontaneous breath due to the medication admitted in an incorrect time point. In response to this emergency, the attending physician and respiratory therapist placed an artificial airway immediately. The entire case was decomposed into 16 steps based on the communication ontology classes. The 15 steps are coded as success and the very first step was defined as a communication error due to unspecified medication admission.

Results

Using the ontology, we are able to code all of the selected medical error cases due to communication. However, the information described in the reports has various levels of granularity. We identified two patterns of case reports while analyzing 10 cases. In one pattern, cases were coded successfully by making minimum assumptions. Under this pattern there were two subcategories. For the case with a detailed description, there was no need to get hospital-specific information. For some cases, coders had to be familiar with the hospital culture and procedures without which coders could not complete the coding tasks. In the other pattern, cases were coded with an amount of assumptions based on coder's understanding of healthcare domain. There were some cases that could not be coded due to the lack of detailed information. This detailed information is expected during the reporting of medical errors. The reporter of the medical error event could be aware of this missing information, if the ontology had guided the reporter to report the incident in a complete and useful manner. An ontology driven reporting system has the advantage in collecting the expressive and mandatory information. In return, the collection of information guided by ontology will play an important role in analyzing error cases, revealing communication patterns and reducing reoccurrence of the same kinds. The cases cover a variety of methods of communication used in the clinical workflow. Analyses of the cases indicate that clinicians favor the synchronous form of communication even though this kind of communication is highly interruptive to their workflow [8]. This result is consistent with other researcher's observational studies.

Discussion and future work

This paper describes our experience of developing clinical communication ontology to serve as a means of describing clinical communication error, a subset of medical errors. Concepts, relationships and properties and so on in the ontology are subject to modification during analyzing more communication cases.

The ontology primarily provides eight axes for description of the clinical communication. It allows the users to categorize the communication activities and ultimately figure out the patterns for effective communication among clinicians. The development is an iterative process, the

modification and reconstruction will be continuous during the case coding process. We aim at an exhaustive and exclusive ontology in clinical communication which not only describes the activities but also be able to analyze communication pattern, explain communication error, and eventually suggest proper solutions.

We noticed that there is a tradeoff between the ontology and the case description. We tried to balance these two components to ensure that the ontology was exclusive and exhaustive. Based on the case description we did not make any major changes to the ontology and did not make too many assumptions to make the case adequate for coding. This implies that the ontology can be tailored into different granularity fitting the case reporting requirement.

One limitation of this study was that the medical error cases collected were only a small sample of all events and hence they do not sufficiently represent the population. We would require a larger sample from a healthcare quality reporting system in practical use to ensure that the cases adequately represent the population.

Conclusion

The communication ontology for medical error has its practical implication in improving patient safety and reducing medical error. One way to learn from errors is to establish a medical error reporting system, where medical error data are collected in a structured format and can be useful for the detection of patterns, discovery of underlying factors, and generation of solutions. Such ontology serves as a guideline for hospital reporting medical errors. It can be customized into different scales so as to meet hospital needs and minimize the burden of data entry in health care organizations.

This ontology will be served as a foundation for analyzing the communication activities within medical errors, revealing the communication patterns and providing information to possible intervention strategies. The ultimate goal is to improve the communication quality and efficiency and thus reduce medical errors. Validating the ontology through medical error cases gives us an insight into the usability and validity of such ontology. Possible modifications to the ontology can also be made based on the findings of such a validation. It would also help us to identify critical missing components of medical errors that are expected to be reported towards a more useful reporting system.

Acknowledgments

This project was supported by Grant R01 LM07894 from the National Library of Medicine. The authors express heartfelt thanks to Julianna W. Brixey, Anwar M. Sirajuddin for their contribution in the design and development of the ontology.

References

[1] O'Connor E. Medical errors kill tens of thousands annually, panel says. CNNcom. 1999 1999-11-29.
[2] Kohn LT, Corrigan JM, Donaldson MS. To err is human. Washington, DC: National Academy Press 1999.
[3] Zinn C. 14,000 preventable deaths in Australian hospitals. BMJ. 1995;310(1487).
[4] Zhang J, Patel VL, Johnson TR, Shortliffe EH. A cognitive taxonomy of medical errors. Journal of Biomedical Informatics. 2004;37(3):193-204.
[5] Naik G. A hospital races to learn lessons of ferrari pit stop. The wall street journal. 2006 2006-11-14.
[6] Unknown. Sentinel Event Alert, June 23, 2003. *Sentinel event statistics*: Joint Commission on Accreditation of Healthcare Organizations 2003.
[7] Bailey LL, Willems BF, Peterson LM. The effects of workload and decision support automation on enroute r-side and d-side communication exchanges. Oklahoma City, OK 73125: Civil Aerospace Medical Institute Federal Avaition Administration; 2001 Dec 2001.
[8] Coiera E, Tombs V. Communication behaviours in a hospital setting: an observational study. bmj. 2005;1998(316):673-6.
[9] Coiera EW, Jayasuriya RA, Hardy J, Bannan A, Thorpe MEC. Communication loads on clinical staff in the emergency department. MJA. 2002;176:415-8.
[10] Parker J, Coiera E. Improving clinical communication: a view from psychology. Journal of the American Medical Informatics Association. 2000;7(5):453-61.
[11] Lingard L, Espin S, Whyte S, Regehr G, Baker G, Reznick R, et al. Communication failures in the operating room: an observational classification of recurrent types and effects. Qual Saf Health Care. 2005(12):330-4.
[12] Paek T. Toward a taxonomy of communication errrors. ISCA Workshop on Error Handling in Spoken Dialogue Systems; 2003; 2003. p. 53-8.
[13] Te'eni D, Sagie A, Schwartz DG, Zaidman N, Amichai-Hamburger Y. The process of organizational communication: a model and field study. IEEE transactions on professional communication. 2001;44(1):6-20.
[14] Zhang J, Patel, VL., Johnson, TR., Turley, JP. . Health informatics and medical error. . Business Briefing: US Healthcare Strategies. 2005:34-5.
[15] Gruber T. A translational approach to portable ontologies. Knowledge Acquisition. 1993;5(2):199-220.
[16] Turley JP, Brixey JJ, Johnson TR, Mokkarala P, Zhang J. Comprehensive medical error ontology for the codification of published literature. Bulletin of the Japanese Cognitive Science society. 2006 March;13(1):1-11.
[17] Uschold M. Building Ontologies: Towards A Unified Methodology. *Proc Expert Systems 96*. Cambridge 1996.
[18] Cimino JJ. Desiderata for ControlledMedical vocabularies in the Twenty-FirstCentury. *IMIA WG6 Conference*. Jacksonville,Florida 1997.
[19] Coiera E. When conversation is better than computation. Journal of the American Medical Informatics Association. 2000;7(3):277-86.
[20] Mokkarala P, Brixey, JJ., Johnson, TR., Patel, VL., Zhang, J., & Turley, JP. Development of comprehensive medical error ontology. Manuscript submitted for publication. in review.

Address for correspondence

Mailing Address: 404 Clark Hall,
Department of Health Management and Informatics,
School of Medicine, University of Missouri-Columbia,
Columbia, Missouri, 65211, USA
Email: gongyang@missouri.edu

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

u-SHARE: Web-based Decision Support / Risk Communication Tool for Healthcare Consumers with Unruptured Intracranial Aneurysms

Aoki N [a, b, c], Sakai M [d], Nakayama T [e], Fukuhara S [d], Ohta S [a, b, f], Kikuchi N [a, b, g], Oishi M [a, h], Kiuchi T [c], Nozaki K [i], Hashimoto N [i]

[a] *Center for Health Service, Outcomes Research and Development – Japan (CHORD-J), Tokyo, Japan*
[b] *School of Health Information Sciences, University of Texas Health Science Center at Houston, Houston, TX, USA*
[c] *University Hospital Medical Information Network (UMIN) Center, The University of Tokyo Hospital, Tokyo, Japan*
[d] *Department of Epidemiology and Healthcare Research, Kyoto University School of Public Health, Kyoto, Japan*
[e] *Department of Health Informatics, Kyoto University School of Public Health, Kyoto, Japan*
[f] *Okayama Central Hospital, Okayama, Japan*
[g] *Palliative Care Service, Tohoku University Hospital, Sendai, Japan*
[h] *Oishi Clinic, Kyoto, Japan*[i] *Department of Neurosurgery, Kyoto University Faculty of Medicine, Kyoto, Japan*

Abstract

Purpose: Clinical management for unruptured intracranial aneurysms (UIA) is controversial and requires professional knowledge which is the main reason that patients have difficulty in making decisions. The purpose of this study is to develop a tool that aids healthcare consumers in making optimal shared decisions with decision analysis.

Methods: The decision model and relevant data were derived from published literature. A web-based decision analytic tool was designed to provide a systematic guide for patients to understand favorable treatment options, intrinsic uncertainty, and critical factors for decision making. Twenty-nine testers evaluated content appropriateness, usability and clinical usefulness of the tool.

Results: The decision analytic tool has been successfully implemented and evaluated. Testers generally judged the web-based decision analytic tool as functional and useful. Acceptance rate for decision analysis was higher in non-healthcare professionals than health care professionals.

Conclusions: Our decision analytic tool was well accepted especially by healthcare consumers. The tool enables UIA patients to enhance their knowledge and understanding toward optimal shared decision making and can be an alternative "structured informed consent tool".

Keywords:

decision analysis, unruptured intracranial aneurysms, decision support tool, health communication, risk communication

Introduction

In spite of the recent prospective multi-center cohort study conducted by the International Study of Unruptured Intracranial Aneurysms Investigators (ISUIA) [1], optimal management for unruptured intracranial aneurysm (UIA) remains controversial. In Japan, medical-related disputes between patients and healthcare providers regarding decision making on UIA management has been gradually increasing with cases brought to court.

Due to the controversy, several decision analyses have been conducted and published [2-4]. A decision analytic framework could provide comprehensive information regarding what is known and unknown about the disease, options for treatment, possible outcomes, and encourage elicitation of preference for outcomes [5]. In addition, results of the analysis could provide recommendations, important factors for decision making, and a framework for discussion with a physician.

Despite the above scientific merits for physicians and researchers, however, healthcare consumers may not be able to fully utilize the evidence-based information. Difficulties include perception of uncertainty, preference measurement, and interpretation of the result in application to "my case".

In this study, we developed, implemented, and performed a preliminary evaluation of a web-based decision analytic tool for healthcare consumers called u-SHARE (ubiquitously-Support and Heal patients with intracranial Aneurysms with Risk communication and Empowerment).

Materials and methods

Conceptual framework

Figure 1 demonstrates the conceptual framework for decision making by UIA patients using a decision analytic approach.

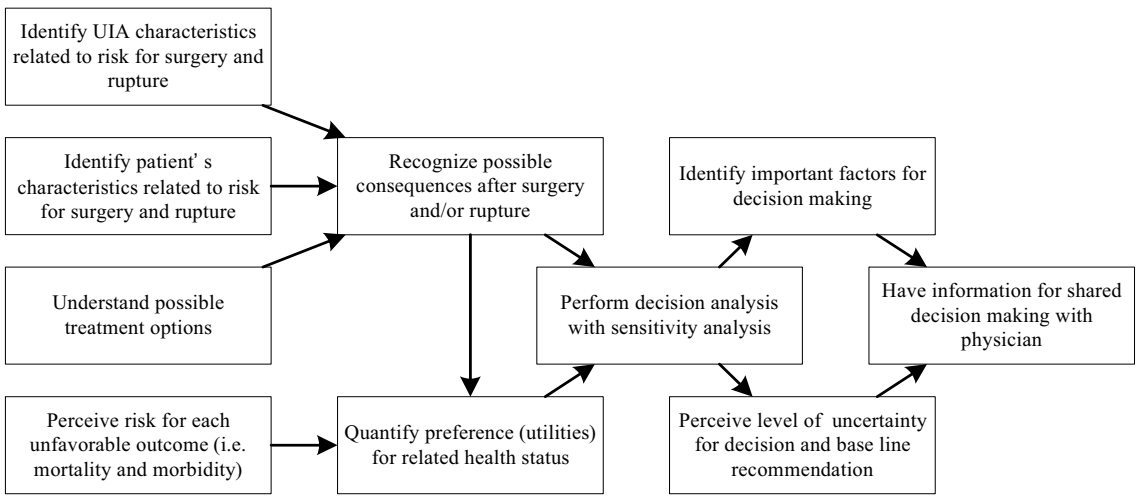


Figure 1 - Conceptual model for a decision analysis support system

Our ultimate goal is to equip patients with enough information and knowledge for appropriate shared decision making with their physicians. To achieve this goal, u-SHARE is designed to provide systematic steps for patients to understand (1) their UIA and physical conditions, which influence the risk of rupture and surgery, (2) treatment options, (3) possible consequences after surgery or rupture, and (4) their preference in unfavorable outcomes.

Decision analytic model

The published Markov decision analytic model was used in this study [2]. The decision analysis compared preventive surgery for a detected UIA and "watchful waiting", with quality adjusted life years (QALYs) as an outcome. The decision analytic model includes four health states: well, living with UIA, severe neurological deficit defined as a Rankin Scale of 3-5, and death. Quality of life with UIA and severe neurological deficits are designed to be collected from each user.

Although the original paper utilized probability data from a previous ISUIAI report [1], we updated the data with the current ISUIAI report published in 2003 [6]. The ranges for on-way sensitivity analysis for probabilities were defined with 95% confidence intervals. The ranges for utilities were set as 0 – 1.0. The decision model was developed with TreeAge Pro 2006 software (TreeAge Software, Inc., Williamstown, MA. 2006).

Table 1 reveals key probability data for the decision analysis.

Table 1 – Baseline data for decision analysis system design

Items	Baseline
Preventive Surgery	
Mortality	4.5%
Physical complication	68.6%
Relative Risk	
Age 50	2.4
Past history of strokes	1.9
Aneurysmal symptoms	1.6
Annual Risk of Rupture	
Anterior UIA: < 7 mm	0.5%
Anterior UIA: 7 - 12 mm	3.1%
Anterior UIA: 13 - 24 mm	10.2%
Aneurysmal rupture	
Mortality	43.1%
Physical complication	24.3%

Only part of the data is shown due to space limitation.

The u-SHARE application is developed as a web-based application run by Active Server Pages (ASP). Default probability data reported in the ISUIAI report [6], such as surgical mortality and morbidity for preventive surgery and annual rupture rate for various type of UIA, is stored in a relational database management system (RDBMS) developed in MySQL. The system schema of u-SHARE is illustrated in the Figure 2.

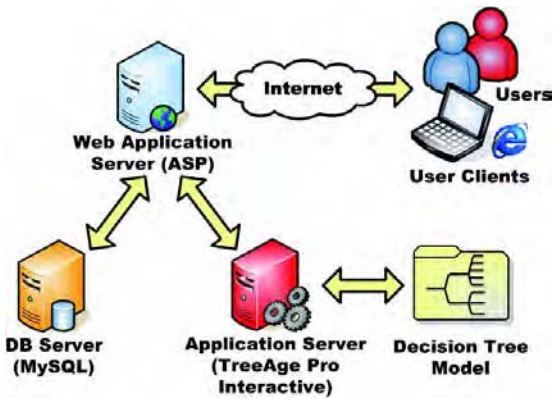


Figure 2 - Web-application overview

The web server, MySQL RDBMS, and the decision analytic model created with TreeAge Pro are actively connected to one another by a special ASP program called TreeAge Pro Interactive (TAPI) (TreeAge Software, Inc., Williamstown, MA. 2006).

All data entered through the web site is passed from the ASP server and stored in the MySQL RDBMS via ODBC (Open DataBase Connectivity). TAPI is an ActiveX interface application, which connects data from the MySQL database and the web browser to a decision model developed by the TreeAge Pro software. Active X control enables users to employ any command in the TreeAge Pro application through the web site. In addition, the results can be transferred to the TAPI which can dynamically generate an HTML file.

Preliminary evaluation

A total of 29 testers (14 healthcare professionals and 15 non-medical individuals) were asked to evaluate the u-SHARE. We modified a questionnaire, which was reported in a previous MedInfo paper, for IT decision support system evaluations [7]. The questionnaire consists of 13 questions corresponding to the following three assessment categories

- Content appropriateness - corresponding to the conceptual model described in Figure 1
- Usability
- Clinical usefulness for shared decision making

Table 2 shows all questions in the questionnaire.

Questions are asked with a scaling of 1 to 7 (1= strongly disagree, 2= disagree, 3= somewhat disagree, 4= somewhat, 5= somewhat agree, 6= agree, 7= strongly agree).

Participants were also interviewed to gain further detailed comments regarding the general concept of a web-based decision analytic tool.

Table 2 – Questions for preliminary evaluation

No	Questions
Content appropriateness	
1	Can you confirm the type of UIA you have?
2	Do you perceive the risk of unfavorable outcomes?
3	Do you understand the available treatment options?
4	Do you understand the health conditions of severe neurological deficit?
5	Was your preference for the severe neurological deficit successfully elicited?
Usability	
6	Is the text readable?
7	Are the images easy to recognize?
8	Did you have difficulty proceeding with the analysis?
9	Is appropriate guidance provided?
Clinical usefulness for shared decision making	
10	Do you understand the recommendation?
11	Do you recognize the uncertainty of decision making?
12	Do you perceive important factors for your decision?
13	Would you want to use this tool for decision making?

Results

The u-SHARE is currently available on the web site (http://www.u-share.org). The following is a description of u-SHARE.

Description of the web-based decision analytic tool

Visitors of u-SHARE will obtain general background information, current medical evidence, basic steps for decision analysis and understand the limitations of the decision analytic tool. Each UIA patient is required to register a username and password in the system so they can return to the system for further decision analysis.

The patient is asked to input factors related to risks, reported in an ISUIAI paper [6], such as age, gender, presence / absence of aneurysmal symptoms other than rupture and previous ischemic cerebrovascular disease.

In the next screen, patients are required to enter UIA characteristic data (i.e. size, location) which are closely related to risks for surgery and rupture.

Once, the UIA data is entered, graphical information about the annual rupture rate, mortality and morbidity after UIA rupture, and surgical mortality and morbidity are summarized as pie charts. (Figure 3) .

Figure 3 – Risk of Surgery and Rupture

After the patient understands the risks and types of unfavorable outcomes, u-SHARE explains the average health condition in regards to the morbidity (neurological deficit defined by a modified Rankin Scale of 3-5) [8]. The explanation is made with written sentences and a voice and video clip using Windows Media Player. Figure 4 demonstrates examples of the video clip demonstrating severe neurological deficit.

Figure 4 – Examples of the video clip

After understanding the health states, u-SHARE will elicit utility values from patients for a QALYs calculation. Figure 5 demonstrates a tool for utility elicitation using a time trade-off (TTO) method [9].

After examining the necessary information for decision analysis (i.e., probability data for each UIA and utility for health states), the patient can perform a decision analysis. In order to visualize and emphasize uncertainty of the results, we employed a Monte Carlo simulation to display the recommendation. (Figure 6).

Figure 5 – Graphical results of the Monte Carlo simulation

The Monte Carlo simulation in Figure 5 reveals the proportion of patients who benefit from each strategy. For instance, Figure 6 shows 56% of UIA patients benefit from preventive surgery, and 44% benefit from watchful waiting. This format was selected as we intend to provide information for decision making rather than a number / recommendation.

u-SHARE then performs a one-way sensitivity analysis for annual UIA rupture rate, morality and morbidity after preventive surgery, mortality and morbidity due to rupture, quality of life with UIA, and assigns a Rankin Scale between grades 3-5. The factors within the predefined threshold range, appear as "important factors for decision making" in the following format. (Figure 6). The graph reveals current value, threshold, and range of probability or utility for further consideration and discussion.

Figure 6 – Graphical demonstration of threshold analysis

Finally, all of the information used in the analysis and results by u-SHARE is summarized in a printable format. We expect that each patient would use the summary for further discussion with their family, friends, or healthcare professionals.

Preliminary evaluation

Table 3 shows results of evaluation. Decimal number indicate average score (7= strongly agree to 1= strongly disagree), and percentage shows proportion of testers who put score greater than or equal to five. .

Content appropriateness

Generally, the average score was greater than 4.0, indicating that most testers felt that u-SHARE had appropriate content. Healthcare professionals scored higher than non-healthcare ones, which indicated that more detail guidance may be needed for this latter group although the contents do include appropriate information of risk and outcomes.

Ease of use

Although characters and images in the u-SHARE appear easy to understand (Qs 6 and 7), more than half non-healthcare professionals felt comfortable with the decision analysis, but more than half of the physicians felt uncomfortable about it (Qs 8 and 9).

Clinical usefulness for shared decision making

More than 85% of non-healthcare professionals successfully recognized the recommendation and its uncertainty, which was a higher proportion than for healthcare professionals. As well, more than 70% of non-healthcare professionals understood "important factors for decision making", which was also a higher score than for the healthcare professionals'.

Table 3 - Result of evaluation

No	Healthcare Professionals		Non-Healthcare individuals	
1	5.2	(71.4%)	4.9	(66.7%)
2	5.7	(85.7%)	5.3	(73.3%)
3	5.9	(85.7%)	4.4	(46.7%)
4	5.8	(92.9%)	5.2	(73.3%)
5	5.6	(78.6%)	5.7	(86.7%)
6	5.1	(71.4%)	5.5	(66.7%)
7	5.1	(64.3%)	5.2	(73.3%)
8	4.1	(35.7%)	4.6	(53.3%)
9	4.3	(42.9%)	4.9	(66.7%)
10	5.1	(71.4%)	5.3	(86.7%)
11	5.1	(64.3%)	5.3	(86.7%)
12	4.6	(50.0%)	5.0	(73.3%)
13	4.1	(42.9%)	4.4	(46.7%)

Discussion

This study demonstrated that the integration of current information technology could overcome the difficulty of deploying decision analysis systems. Since a large barrier to utilizing decision analysis is requirement of special knowledge, even for clinicians, this web-based tool could be an alternative format of publication for practical usage of decision analysis in clinical settings in addition to scientific reports in medical journals.

In addition, the study showed decision analysis systems were accepted as an evidence-based second opinion acquisition source for autonomic decision making by patients. Our preliminary evaluation implied that decision analytic tool might be more acceptable for healthcare consumers than for healthcare professionals. Such a tool may help healthcare consumers to make appropriate decisions, and as such could be considered as a form of structured informed consent

Finally, during development of u-SHARE the authors recognized the difficulty in developing common understanding among professionals in healthcare, information technology specialists, and decision analysts. It is important to encourage collaboration among people who can produce interdisciplinary research and development which combine various disciplines to generate marketable, medically significant, and sustainable services, especially in the health informatics area.

Conclusion

Our study showed that a web-based decision analytic tool could have significant potential in assisting healthcare consumers achieve appropriate shared decision making and which can also be used as "structured informed consent". A clinical trial is planned, as the logical next step for the evaluation, with actual UIA patients.

Acknowledgments

The authors thank the CAC, Inc. (Tokyo, Japan) for providing programming skills.

References

[1] International Study of Unruptured Intracranial Aneurysms Investigators. Unruptured intracranial aneurysms--risk of rupture and risks of surgical intervention. N Engl J Med 1998: 339(24): 1725-1733.

[2] Aoki N, Beck JR, Kitahara T, et al. Reanalysis of unruptured intracranial aneurysm management: Med Decis Making 2001: 21(2): 87-96.

[3] Aoki N, Kitahara T, Fukui T, et al. Management of unruptured intracranial aneurysm in Japan Med Decis Making 1998: 18(4): 357-364.

[4] Dippel DW, Vermeulen M, Braakman R, et al. Transient ischemic attacks, carotid stenosis, and an incidental intracranial aneurysm. A decision analysis. Neurosurgery 1994: 34(3): 449-457; discussion 457-448.

[5] Sanders GD, Hagerty CG, Sonnenberg FA, et al. Distributed decision support using a web-based interface: prevention of sudden cardiac death Med Decis Making 1999: 19(2): 157-166.

[6] Wiebers DO, Whisnant JP, Huston J, 3rd, et al. Unruptured intracranial aneurysms: natural history, clinical outcome, and risks of surgical and endovascular treatment Lancet 2003: 362(9378): 103-110.

[7] Aoki N, Ohta S, Masuda H, et al. Edutainment tools for initial education of type-1 diabetes mellitus - initial diabetes education with fun Medinfo 2004: 2004: 855-859.

[8] Bonita R, and Beaglehole R. Recovery of motor function after stroke Stroke 1988: 19(12): 1497-1500.

[9] Torrance GW. Utility approach to measuring health-related quality of life J Chronic Dis 1987: 40: 593–600.

Address for correspondence

Noriaki Aoki, MD, PhD, MS, FJSIM, CPE, Assistant Professor, School of Health Information Sciences, University of Texas Health Science Center at Houston, 7000 Fannin, UCT-600, Houston, TX 77030, USA;
TEL: (713) 500-3938
Email: Noriaki.Aoki@uth.tmc.edu

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Health On the Net Foundation: Assessing the Quality of Health Web Pages All Over the World

Célia Boyer[a], Arnaud Gaudinat[a], Vincent Baujard[a], Antoine Geissbühler[a,b]

[a]*Health on the Net Foundation, Geneva 14, Switzerland*
[b]*Service of Medical Informatics, Geneva University Hospitals, Switzerland*

Abstract

The Internet provides a great amount of information and has become one of the communication media which is most widely used [1]. However, the problem is no longer finding information but assessing the credibility of the publishers as well as the relevance and accuracy of the documents retrieved from the web. This problem is particularly relevant in the medical area which has a direct impact on the well-being of citizens. In this paper, we assume that the quality of web pages can be controlled, even when a huge amount of documents has to be reviewed. But this must be supported by both specific automatic tools and human expertise. In this context, we present various initiatives of the Health on the Net Foundation informing the citizens about the reliability of the medical content on the web.

Keywords:

patient safety, natural language processing, internet, information retrieval, information quality

Introduction

The Internet has become one of the communication media that is most widelyused. With the availability of Web server software, anyone can set up a Web site and publish any kind of data which is then accessible to everyone. The problem is therefore no longer finding information but assessing the credibility of the publishers, as well as the relevance and accuracy of the documents retrieved from the web. This problem is particularly relevant in the medical area: the information found (explanations of diseases, recommended treatments, available medicine ...) has a direct impact on the well-being of the citizens. Indeed, in many cases, web sites provide no appropriate information regarding the scientific design of a medical study, nor are studies made available that support given claims. Various initiatives have been proposed for managing the quality of web pages:

- *Selection or referencing* of pages was first implemented by the *Yahoo!*[1] directory. In the medical area, the American *MedlinePlus* [2] and the French *CISMeF* [3] portals are referred to. The principle of referencing consists of publishing a selection of web pages arranged according to domains which are organised more or less hierarchically.
- *The accreditation* of web pages, such as performed by HONcode [4] and URAC [5], specifically aims at promoting the quality of information. According to this intiative, sites must respect a set of criteria to be accredited.
- *The popularity* of web pages is detected by the use of PageRank [6] or similar models. The popularity is correlated with the judgement of webmasters when they decide whether links to other web pages should be estalblished.
- *The social or collaborative initiative* consists of using the judgment of all users in order to characterise web content and thus, to create networks of reliable users in each addressed area. *Lijit* [7] and *Google co-op* [8] implement this initiative.
- *The self-regulation* of web sites consists of their auto-accreditation by respecting given rules [9,10].
- *The education of patients* consists of helping them to find reliable web sites on the basis of a set of criteria [11,12].

The objective of the Health on the Net Foundation (HON[2]) is to provide citizens with information on the reliability of the medical content of web documents. HON assumes that the quality of web sites should be certified by a neutral third party organisation such as HON, according to quality criteria such as the HONcode. Since 1995, the HONcode has been accepted by the health web community. In 2001, HON Foundation was recognized as a Non Governmental Organisation and currently provides consulting services to ECOSOC. Moreover, HON has participated in the workgroup of the European Community being in charge of the elaboration of eEurope 2002: Quality Criteria for Health related Websites. And HON is a partner of WHO for the development of quality health information in African countries. In addition to the experience acquired by the HON Foundation, we believe that the difficulty of promoting high quality health information on the web, can be reached by combining the development of suitable automatic tools with human expertise. In the rest of this paper, we introduce the HON Foundation by presenting three

1 www.yahoo.com

2 Health On Net: *http://www.healthonnet.org*

major aspects of its activity: (1) the HONcode of Ethical Conduct, (2) the database of accredited medical web sites and (3) the automatic tools created in order to help users and human reviewers when assessing the quality of medical sites. We then draw a conclusion and provide some perspectives to the HON Foundation activity.

HONcode principles and accreditation protocol

Ethical HONcode of conduct

The HONcode [4] is a set of high ethical principles defined by the HON Foundation in order to assess the quality of on-line health information on the web. These principles are applied during an evaluation of the web sites which have demanded to be accredited by HON. Each principle has both, short patient-oriented (here above) and long expert-oriented formulations:

1. *Authoritative*: Indicate the qualifications of the authors
2. *Complementarity*: Information should support, not replace, the doctor-patient relationship
3. *Privacy*: Respect the privacy and confidentiality of personal data submitted to the site by the visitor
4. *Attribution*: Cite the source(s) of published information, date and medical and health pages
5. *Justifiability*: Site must support claims relating to benefits and performance
6. *Transparency*: Accessible presentation, identities of editor and webmaster, accurate email contact
7. *Financial disclosure*: Identify funding sources
8. *Sponsorship*: Clearly distinguish advertising from editorial content

The long description of principles is more explicit and should remove any remaining uncertainties about their meaning. For instance, the complete statement of the first principle *Authoritative* is: *Any medical or health advice provided and hosted on this site will only be given by medically trained and qualified professionals unless a clear statement is made that a piece of advice offered is from a non-medically qualified individual or organisation.* It indicates particularly the interest for medical information being provided by medically trained people. The HONcode principles can be read online in 32 languages. Additionally, The HON Foundation provides the guidelines of the HONcode accreditation process [13].

Accreditation process and active seal

During the accreditation process, human experts evaluate web sites, which have demanded for the accreditation, in order verify if these sites respect the ethical HONcode. If web sites respect the principles, the dated accreditation seal (Figure 1) and the personal identification number (PIN) corresponding to the accreditation certificate are provided to these sites. The provided seal is *active*, and when one clicks on it in order to verify its validity, the generated screen contains the personalised HON page presenting the Web site's accreditation status. This technique allows us to reflect in the status of an accredited web site in real time and, if needed, to change it from "accredited" to "under reviewing process" or "not compliant" (see Figure 1).

Figure 1 - Active Seals, "accredited", "under review" and "not compliant" from the certificate

During the accreditation review, the HON experts identify modifications that are required for the site in order to be accredited. Each missing ethical information is indicated and must be added to the content of web site pages. Once this has been done, the accreditation seal and PIN are provided. All the accredited web sites are reviewed each year. If a web site does not anymore respect the HONcode principles, the webmaster receives warnings and, if required modifications are not made, the site loses its accreditation. The HONcode ethical principles and accreditation are being adopted by health web publishers all over the world. The accreditation process, as it is implemented, allows to conduct an educational work with webmasters and to naturally improve the quality of on-line health information. The accreditation request can be made on-line. A pre-evaluation is proposed to the webmaster in order to highlight the missing principles.

Accredited HONcode database of websites

Currently, the database is composed of over 5,500 accredited web sites in 72 countries which represents over 1,200,000 web pages indexed in Google. 52% of the accredited sites are in English and about 11% of the sites in French, followed by a considerable amount in Spanish, Italian, Portuguese and Dutch. For each evaluated evaluated site the database provides the following information, which assists us in developing automatic tools:

- respect of ethical HONcode labels,
- excerpts corresponding to these eight principles,
- indexing with MeSH headings (i.e., *diabetes*, *asthma*),
- general content type labels (i.e., *Woman health*, *Information for Patients*).

Automatic tools

The automatic tools developed by the research and development team of the HON Foundation address two main objectives:

- help citizens to manage the ever increasing amount of information on the web and provide them with reliable medical web resources,
- help human reviewers during the web sites' evaluation and reviewing processes.

According to these objectives, we addressed the following tasks when conceiving and developing tools: crawling task with MARVIN tool, indexing of medical pages with the MeSH terms, selection of medical content, development of

the specific HON toolbar, and automatic detection of statements on HONcode principles. The majority of these tools can be tested on-line through the HON Foundation web site (http://www.healthonnet.org). They can be used by medical experts (doctors, nurses, students) and citizens.

MARVIN the crawler

The objective of the MARVIN (Multi-Agent Retrieval Vagabond on Information Networks) crawler [14] is not only to find web pages which respond to the query, but to locate the right piece of information among them. Indeed, general web search engines, indexing most of the web, return a long list of documents, often to the detriment of precision. The search result is then barely usable because of the large number of answers from different domains and topics. Only complex queries may, in a given situation, produce a limited number of potentially relevant documents. In order to make searches more efficient and useful to ordinary users, intelligent and specialised search engines are needed on the web. The HON crawler MARVIN was first applied to the medical domain. Armed with a dictionary of medical terms and the MeSH terminology, MARVIN skims the Web for new sources of medical information.

Automatic indexing of web pages with the MeSH keywords

MeSH [15], Medical Subject Heading is a terminology developed by the National Library of Medicine. It contains currently over 33,000 terms. MeSH is translated into various languages. In our applications, we use its versions in English, French, German, Spanish, Portuguese, Danish, Dutch and Italian. MeSH has been conceived for information retrieval purposes. Its hierarchical organization thus offers important search opportunities by using the "concept exploding". For instance, a search with *diabetes mellitus* will also bring documents indexed with the following terms: *diabetes mellitus, experimental, diabetes mellitus, type 1, diabetes mellitus, type 2, diabetes, gestational, diabetic ketoacidosis,* and *prediabetic state*. Indeed, all these terms are hierarchical children of the term *diabetes mellitus*, and the concept exploding function will "explode" *diabetes mellitus* to all its children and, possibly, to other related terms. Through the interfaces of our search engines, the user may select the terms of interest and refine his or her search. The multilingual capability of MeSH transforms the MeSH thesaurus into a search tool which is even more powerful. Several of our tools exploit the multilinguality of the MeSH.

In our approach the indexing of web pages with MeSH terms is performed in two steps there are: the extraction of these terms and their weighting. The extraction of the MeSH terms relies on lexical normalisations and the detection of synonymous relations between terms. The weighting of terms has first been performed by learning algorithms, which showed interesting results. Furthermore, in order to improve these results, we exploited semantic relations between terms, as recorded in the table of co-occurring terms of the UMLS [16]. The UMLS (Unified Medical Language System) is a knowledge source of the biomedical area in which various terminological resources have been merged. The UMLS provides, among other things, the file of co-occurrences that is computed on all the MEDLINE databases. In order to compute the weighting of terms on the basis of these co-occurrences, we assume that the more co-occurring a concept is with other concepts, the more important is this concept. During the reranking process of extracted terms according to this methodology, we proceed in two steps [17]:

- for each pair of terms in the list of extracted terms, we calculate the cumulative weight of their relations;
- we express each term according to its related pairs to obtain a unique score by the term.

Evaluation of this indexing system in comparison with similar French systems (CISMeF-NLP [18] and NomIndex [19]), showed that the HON MeSH-indexer presents the best F-measure [20] as compared with manually built gold standard.

Hunt and HONselect services

Hunt services are based on a flexible search engine especially tailored to specific domains. Each Hunt database can further be organised according to a specific domain such as the medical one or the domain medical biology. For instance, the MedHunt [21] database is dedicated to the health domain and is composed of documents which are automatically retrieved from the web and selected by the HON team. In addition MedHunt is capable of narrowing down a search to Web sites in either French or English, and it has currently also been updated to function in German.

HONselect [22] is a multilingual search tool integrating heterogeneous web resources. It covers the languages English, Dutch, French, German, Italian, Spanish, and Portuguese. HONselect offers a comprehensive collection of medical terms and corresponding pictures, bibliographic references, news and Web sites. To our knowledge, no other search engine offers such easiness in searching the medical web. HONselect integrates several separate databases (Medline, HONmedia, NewsPage, MedHunt, HONcode, clinical trials as well as scientific articles). Moreover, the HONselect proposes more options: (1) its *search* function checks automatically for spelling errors in seven languages; (2) its *translate* function allows to easily switch from one language to another so one can obtain additional and complementary information in other languages. Search results are displayed in the language of the formulated query.

Currently, the crawling task is also supported by the Google Co-op Health service. Indeed, it becomes possible to define one's own search engine on the basis of the ommon Google crawler. According to the HON Foundation policy, the tailoring of the Google crawler is performed by providing a list of trustworthy web sites from the HONcode accredited database. This helps to increase the precision of obtained results. In the mean time, the performance and coverage of the Google crawler and its storage capacity helps to increase the recall. Com-

bining expertise of both the HON Foundation and the Google incorporation, we can contribute to satisfy the citizens when they look for reliable health information on the web.

Verification of web pages' reliability with WRAPIN

The WRAPIN engine [17] (Worldwide online Reliable Advice to Patients and Individuals) has the objective to enhance the capabilities of HON's retriever and indexer MARVIN. The innovations gained by the WRAPIN engine, in addition to the already integrated functions are:

- processing of health and medical documents in any format (HTML, PDF, etc) or length,
- querying of more medical and health databases: i.e., Bookshelf, ClinicalTrials, PubmedCentral, PubMed, FDA, OESO and Urofrance,
- determining information quality by comparing the documents with the interconnected knowledge base,
- providing a summary of the ideas contained in the processed documents,
- using of the entire document and of the URL as the search query, in addition to a normal series of search terms,
- highlighting of MeSH terms in the summary in order to better define the correctness of results,
- identifying results according to the user's profile (specialist or new to the field of medicine).

Automatic identification of HONcode accredited sites

For the automatic identification of HONcode accredited web sites, The HON Foundation proposes a toolbar which detects the HONcode status of a web site. This toolbar has been accomplished within the framework of the European project ActiveHealth (Active Environment for Health Promotion and Disease Prevention). The aim of this project was to transfer information to individuals. This toolbar is downloadable from the HON web site [23]. It has two functions:

- Automatic checking of the accreditation of the web site being read by the user;
- Searching in the HONcode accredited web sites: (1) by entering a word or phrase into the search box or (2) by selecting text using the mouse, and right-clicking to begin the search.

Automatic quality criteria extractor

The proposed methodology for the automatic extraction of quality criteria focuses on processing vast amounts of health data on the Internet. An automatic tool [24] is currently under development in order to help the daily work of human reviewers when they evaluate the transparency of health websites and annotate them by using the HONcode principles. This tool relies on the application of machine-learning algorithms such as Naive Bayes, Support Vector Machine, k-Nearest Neighbours and Decision Tree [25,26]. This tool is trained on the HON database of accredited web sites. It is currently applied to medical web sites and shows up to 78% of precision and 73% of recall. The contingency computed between precision and recall indicates that some of the principles of the HONcode are rather easy to detect. For instance, the *Privacy* principle shows one of the best contingency rates, 92% / 90%. While other principles remain problematic: the *Justifiability* shows the precision/recall contingency of only 45% / 33%. Moreover, it appears to be ambiguous with the *Complementarity* principle.

Conclusion and perspectives

During the past ten years, the HON Foundation has undertaken important efforts in order to promote the quality of health information on the web worldwide. To address the size of the web, the necessary human expertise is supported by various automatic tools (crawler, indexer, search engine, HONcode principle extractor) in order to make the evaluation of web sites more systematic and rapid. From the user's point of view, easy mechanisms are provided such as the specific toolbar and the dynamic seals, and particularly, a database of high quality medical web sites.

All the methodologies and tools developed by the HON Foundation can also be applied to other areas, where trust and transparency are important issues.

Additional solutions can be found in order to assist the citizens and medical professionals even more in searching the medical web. Some of these conceptions can be realised with partners from the addressed areas, such as the National Library of Medicine, Bethesda, USA or Google Inc.

By proposing the HONcode, HON educates information providers to be more transparent. However, users are not aware enough regarding the coexistence of reliable and unreliable information on the web. HON should conduct awareness campaigns in order to educate the citizens to efficiently use the medical and health information on the Web. Collaboration and integration with search engine mainly used is also crucial to directly reach the user in his/her daily life. So, HON has worked closely with Google to highlight the trustworthy health information to the citizen by using a set of health labels they jointly developed.

The approach so far conducted by HON is a global one covering 32 languages worldwide. However, HON should respond to the local needs and the various languages spoken. The creation of chapters located in different regions of the world might help us to locally and thus globally improve the quality of medical and health information.

Acknowledgments

The foundation activities are supported by the State of Geneva and more particularly by the Geneva Ministry of Health (DES) and the European Union through the WRAPIN, ACTIVEHealth and PIPS projects.

References

[1] Fox S. Online Health Search 2006, Online Health Search 2006. Washington, DC: Pew Internet & American Life Project, October 29, 2006.

[2] Miller N, Lacroix EM and Backus JE. MEDLINEplus: building and maintaining the National Library of Medicine's

consumer health Web service. *Bull Med Libr Assoc* 2000;88(1): 11-7.
[3] Darmoni SJ, Leroy JP, Baudic F, Douyere M, Piot J and Thirion B. CISMeF: cataloque and index of French speaking health resources. *Stud Health Technol Inform* 1999:493-6.
[4] Boyer C, Baujard V and Scherrer J. HONcode: a standard to improve the quality of medical/health information on the internet and HON's 5th survey on the use of internet for medical and health purposes. In *6th Internet World Congress for Biomedical Sciences (INABIS 2000)*, 1999.
[5] *www.urac.org/MMandQualityChasm.asp*
[6] Page L, Brin S, Motwani R and Winograd T. The pagerank citation ranking: Bringing order to the web. *Technical report*, Stanford Digital Library Technologies Project, 1998.
[7] *www.lijit.com*
[8] *www.google.com/coop*
[9] Risk A. Dzenowagis J. Review of Internet information quality initiatives. *Journal of Medical Internet Research.* 2001; 3(4):e28
[10] Sheldon T. Trust mark launched as a guarantee of safety in the Netherlands. *BMJ* 2002;324(7337):567
[11] Charnock D. The DISCERN Handbook. 1998. Radcliffe Medical Press
[12] *omni.ac.uk*
[13] *www.hon.ch/HONcode/Guidelines/guidelines.html*
[14] Baujard O, Baujard V, Aurel S, Boyer C and Appel R. A multi-agent softbot to retrieve medical information on internet. *Medinfo* 1998; 150–4.
[15] National Library of Medicine, Bethesda, Maryland. Medical Subject Headings, 2001. http://www.nlm.nih.gov/mesh/meshhome.html.
[16] NLM. UMLS Knowledge Sources Manual. National Library of Medicine, Bethesda, Maryland, 2005. www.nlm.nih.gov/research/umls/.
[17] Gaudinat A, Joubert M, Aymard S, Falco L, Boyer C, and Fieschi M. WRAPIN: new generation health search engine using UMLS knowledge sources for MeSH term extraction from health documentation. In M. F. at al., editor, *Medinfo*, p. 356–360, 2004.
[18] Névéol A, Rogozan A and Darmoni S. Automatic indexing of online health resources for a french quality controlled gateway. *Information Processing & Management*, 42(3):695–709, 2005.
[19] Pouliquen B, Delamarre D and Le Beux P. Indexation de textes médicaux par extraction de concepts, et ses utilisations. In *JADT*, pages 617–628, 2002.
[20] Névéol A, Mary V, Gaudinat A, Boyer C, Rogozan A and Darmoni SJ. A benchmark evaluation of the french mesh indexing systems. In S. Miksh, J. Hunter, and E. Keravnou, editors, *Artificial Intelligence in Medicine*, Springer's Lecture Notes in Computer Science, pages 251–255, 2005.
[21] Boyer C, Baujard O, Baujard V and Appel R. New developments of MedHunt medical search engine. In *MedNet 98*, Brighton, U.K., 1998.
[22] Boyer C, Baujard V, Nater T, Appel R and Scherrer J. Honselect: A new assisted-search facility for health information and resources integrating heterogeneous databases. In *MedNet 99*, Heidelberg, Germany, 1999.
[23] *www.hon.ch/HONcode/Plugin/Plugins.html*
[24] Gaudinat A, Grabar N and Boyer C. Utilisation de méthodes d'apprentissage pour la détection des critères de qualité des documents en ligne du web médical. In O. Ly and C. O. Bayaoko, editors, *JFIM 2007*, Bamako, Mali, 2007. To appear.
[25] Grobelnik M and Mladenic D. Efficient text categorization. In Mining workshop on the *10th European Conference on Machine Learning ECML98*, 1998. citeseer.ist.psu.edu/grobelnik98efficient.html.
[26] Manevitz LM and Yousef M. One-class SVMs for document classification. *Journal of Machine Learning Research*, 2:139–154, 2001.

Address for correspondence

Célia Boyer, Health on the Net Foundation, HUG/SIM
24, rue Micheli-du-Crest, 1211 Geneva 14 Switzerland
tel: +41 22 372 62 50 fax: +41 22 372 88 85
email: celia.boyer@healthonnet.org

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

The Use of Electronic Medication Reconciliation to Establish the Predictors of Validity of Computerized Medication Records

Alexander Turchin, MD, MS[a,b,d], Tejal K. Gandhi, MD, MPH[b,d], Christopher M. Coley, MD[c,d], Maria Shubina, DSc[b], Carol Broverman, PhD[a]

[a] *Clinical Informatics Research and Development, Partners HealthCare, Boston, MA*
[b] *Brigham and Women's Hospital, Boston, MA*
[c] *Massachusetts General Hospital, Boston, MA*
[d] *Harvard Medical School, Boston, MA*

Abstract

Medication records in clinical information systems (CIS) are frequently inaccurate, leading to potentially incorrect clinical decisions and preventing valid decision support interventions. It is not known what characteristics of electronic medication records are predictive of their validity.

We studied a dataset of 136,351 electronic medication records of patients admitted to two academic hospitals that were individually validated by admitting providers using novel medication reconciliation software. We analyzed the relationship between characteristics of individual medication records and the probability of record validation using a multivariable linear regression model.

Electronic medication records were less likely to be validated if more time had passed since their last update (14.6% for every 6 months), if they represented an antiinfective (61.6%) or a prn (50.9%) medication, or if they were in an outpatient CIS rather than on an inpatient discharge medication list (18.1%); $p < 0.0001$ for all.

Several characteristics of electronic medication records are strongly associated with their validity. These findings could be incorporated in the design of CIS software to alert providers to medication records less likely to be accurate.

Keywords:

clinical information systems, electronic medical records, medications

Introduction

A growing body of literature supports the overall superior safety of clinical information systems that capture medication orders [1]. Accurate medication information in clinical information systems is therefore crucial for patient safety and quality of care [2]. Nevertheless, many investigators report that electronic medication data is frequently incomplete and / or outdated [2-4]. Other sources of medication information (e.g. patients, insurance claims or pharmacies) can be helpful but may also be inadequate [5, 6]. Consequently providers frequently face the task of identifying inaccurate medication information in the clinical information systems in absence of other clues. However, it is not known what characteristics of medication records in clinical information systems are predictive of the validity of the records.

Since January 2006, the Joint Commission on Accreditation of Healthcare Organizations (JCAHO) has mandated that medication information be reconciled across the continuum of care [7], with a particular emphasis on transitions of care, such as hospitalizations. At our integrated healthcare delivery network we have developed a software application, Pre-Admission Medication List (PAML) Builder, that is used to document the complete list of medications the patient was taking prior to the admission to the hospital in a structured electronic format using a combination of information transfer from existing clinical information systems, electronic lists of discharge medications from previous hospital admissions and new data entry (described in more detail below) [8]. During the process of medication reconciliation using the PAML Builder, providers admitting the patient to the hospital validate or discard each of the existing electronic medication records based on the information obtained from the patient / caregiver interview and other relevant sources, and the results of these activities are permanently stored in our data warehouse. We have therefore analyzed the data generated during the medication reconciliation process to identify the characteristics of medication records that are predictive of their validity.

Materials and methods

PAML builder and medication reconciliation

The PAML Builder is a standalone software application designed to assist the process of medication reconciliation for hospitalized patients (Figure 1). The left side of the application presents a list of the patient's medications collected from four clinical information systems in use at Partners HealthCare: the Longitudinal Medical Record (LMR; outpatient), OnCall (outpatient), Brigham Integrated Computer System (BICS; inpatient) and Massachusetts

General Hospital Order Entry (MGH OE; inpatient). Medication records imported from the outpatient clinical information systems are all of the patient's active medications in that system, and medication records imported from inpatient systems are the last set of discharge medications from the corresponding hospital (from hospital admissions within the last 18 months). For each medication record, the PAML Builder displays the source clinical information system and the date of the last update of this medication record in the system. In the outpatient clinical information systems the date of the last update of a medication record is set to the current date when either a prescription is printed from the record or any of the data elements in the record are changed. In the inpatient systems the date of the last update is the date of the patient's discharge from the hospital. The right side of the application contains the list of the medications the patient was taking at the time of admission (PAML), as collected and validated by the admitting provider. The application allows providers to select all accurate electronic medication records from the left side for inclusion in the PAML. Patient's medications not found in the clinical information systems are then entered manually into the PAML. Creation of an accurate PAML is mandatory for all patients who are admitted to one of the two hospitals where the application has been rolled out.

Study patients and settings

We conducted a retrospective cohort study of all patients who were admitted to the Brigham and Women's Hospital and Massachusetts General Hospital between Aug ust 1, 2006 and October 30, 2006 and had a PAML created. For each patient, we obtained the PAML, the list of active medications in four clinical information systems available at the time of the creation of the PAML (as displayed on the left side of the PAML Builder application), and identified all medication records that were copied over to the PAML (i.e. validated to be accurate). The institutional review board at Partners HealthCare System approved the study, and the need for written informed consent was waived.

Measurements

A unique electronic medication record in one of the clinical information systems was the unit of analysis in this study. Only medication records active at the time of the hospital admission were analyzed. We also defined a unique medication entry as a combination of medication name, route and a particular admission; there could be more than one unique medication record corresponding to a unique medication entry (e.g. records documenting the same medication in different clinical information systems). Medication records active at the time of more than one hospital admission over the study period were analyzed separately for each admission. For every electronic medication record we identified the following characteristics:

1. *Inpatient vs. Outpatient*. Medication records that originated in the inpatient clinical information systems (BICS, MGH OE) were labeled "Inpatient".
2. *Prn (as needed) medication* (Yes or No). This category included medications prescribed to be taken as needed rather than regularly.
3. *Antiinfective* (Yes or No). This category included antibacterial, antiviral and antifungal agents administered systemically.
4. *Medication record age* was calculated as the number of months between the date of the last update of the medication record and the date of the patient's admission to the hospital.
5. *Validated*. Medication records that were copied by providers from the left side of the PAML Builder to the PAML were considered validated. If the medication's dose or frequency were changed after it was copied, it was still considered validated. If the medication route was changed, it was not considered validated because the same chemical substance administered by different

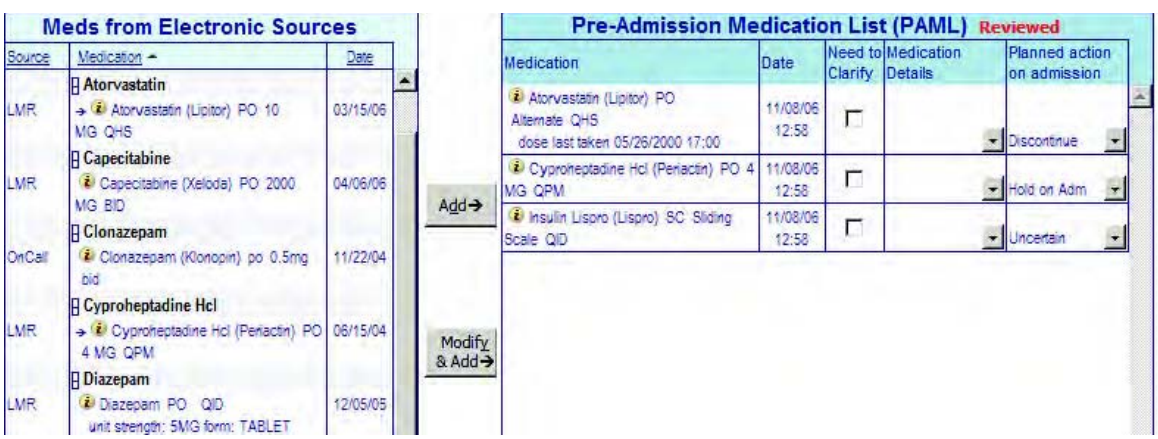


Figure 1 - Pre-Admission Medication List (PAML) Builder

The left side (Meds from Electronic Sources) displays the list of currently active medications imported from four EMRs. The "Date" column lists the date of the most recent update of the medication record in the source EMR. The right side (PAML) contains the list of medications the patient was taking prior to admission as collected and verified by the admitting provider. Users can transfer valid EMR medication records to the PAML by clicking on "Add" or "Modify and Add" buttons.

routes (e.g. oral vs. topical) typically has different indications and therefore would more likely represent a new prescription than a validation of the old one. If the medication was removed from the PAML after it was originally copied to it, it was not considered validated.

6. *Duplicate*. If there was more than one unique medication record for a given medication entry, all of the medication records corresponding to that medication entry were considered duplicates.

All data was obtained from the electronic medical records systems at Partners Healthcare.

Statistical analysis

Summary statistics were constructed by using frequencies and proportions for categorical data and by using means, standard deviations, medians, and ranges for continuous variables. We constructed a multivariable logistic linear regression model to evaluate the association between characteristics of the medication record in a clinical information system and the probability of the record being copied to the PAML (validated). All analyses were performed with SAS statistical software, version 9.1 (SAS Institute, Cary, North Carolina).

Results

Electronic medication records

We identified 17,335 hospital admissions over the study period for which a PAML was created. 10,785 (62.2%) of these admissions had at least one medication record in any of the clinical information systems and were included in the study. There were 136,351 unique medication records in all four clinical information systems linked to these admissions. Among these, there were 111,231 unique medication entries. 92,403 (83.1%) of medication entries had only one medication record linked to each of them, while 18,828 (16.9%) medication entries were linked to multiple medication records. 52.2% of study admissions had medication records only in one clinical information system, 41.6% in two systems, and 6.2% in more than two systems. On average each study admission had 12.7 medication records in the clinical information systems.

Table 1 - Characteristics of electronic medication records

Variable	Value
All records	136,351
Inpatient discharge medications	53,689 (39.4%)
PRN	21,481 (15.8%)
Antiinfectives	8,889 (6.5%)
Medication record age, months	9.8 (± 21.7)
Validated	48,286 (35.4%)

Data are means (± SD) or n (%).

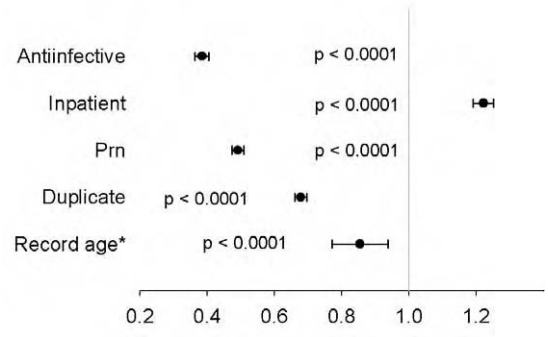


Figure 2 - Predictors of medication record validation

Circles indicate the estimated adjusted odds ratios for the medication record to be copied to the PAML (validated). Wisps indicate the 95% Wald confidence limits for the odds ratios.

* for every 6 months prior to the date of hospital admission

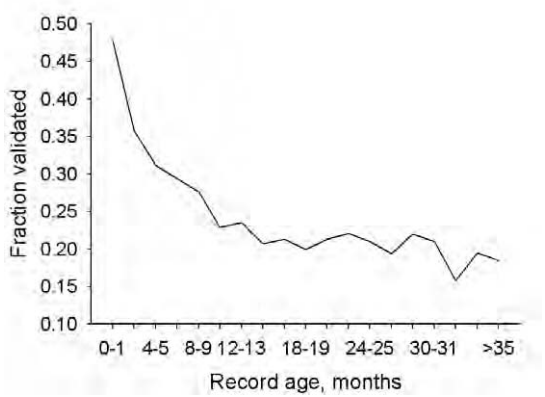


Figure 3 - Medication record age and validation

Characteristics of the medication records are shown in Table 1. Median age of the medication record was 3.5 months. 4,695 (43.5%) of study admissions had both inpatient (hospital discharge medications) and outpatient medication records, while 5,124 (47.5%) had only outpatient and 966 (9.0%) had only inpatient medication records. Over a third of medication records from clinical information systems were copied to the PAML.

Predictors of medication record validation

The effects of the characteristics of a medication record on the probability that it would be validated are illustrated in Figure 2. Inpatient discharge medications had 22.1% higher odds of being validated than the ones from outpatient EMRs. Medications ordered *prn* (as needed) and antiinfective medications had 50.9% and 61.6% smaller odds of being validated, respectively. Medications for which there were duplicate electronic records had 32.2% smaller odds of being validated. Finally, in the linear

model the odds of the medication record being validated decreased by 14.6% for every six months between the last date when the medication record was updated or the date of admission (all relationships significant at $p < 0.0001$). When the medication record age was plotted directly against the fraction of the records validated (Figure 3), it appeared that the continuous decrease in the probability of validation described by the linear model only occurred when the interval between the last update of the record and date of admission was less than one year. After one year the probability of validation stabilized at about 20%.

Discussion

In this retrospective analysis of over 130,000 medication records we have identified several characteristics of the medication records that were associated with a higher probability of record validation. As in several previously published reports [2-4], a large fraction of the medications from clinical information systems was not validated. While some investigators have reported lower prevalence of outdated medication information in clinical information systems than what we observed [9], accuracy likely varies with local practice, and based on the literature our findings appear typical.

As expected, the medication record age was an important factor and older medication records were less likely to be validated. However, further univariate analysis identified a threshold (approximately one year) after which medication record age no longer appeared to play a significant role. Only one in five medication records over one year old were validated compared to nearly half of the medication records that were updated or had a prescription printed from them within the last month. This finding confirms previous reports that a large fraction of medication records in clinical information systems may be outdated [2], and demonstrates that the date when the record was last updated or had the last prescription printed from it is an important indicator of accuracy. It corroborates previously published smaller investigations limited to a single pharmacological domain that reported that a record of the prescription being filled within the last six months had a high positive predictive value for identification of active medications [10]. As reported in other studies, medications that the patient no longer takes but that are still listed as active in the clinical information system put the patient at particularly high risk for adverse drug events [2]. It may therefore be helpful if the date of the last update or prescription for each medication record is prominently displayed in the clinical information system in order to alert providers to possible inaccuracy of an old record.

Medication records representing antiinfective agents were significantly less likely to be validated. Previous reports that focused on medications missing from clinical information systems did not find that antibiotics were missing more frequently than average [3] while other studies, similarly to ours, found that outdated antibiotic records were common [2]. It therefore appears that there is an asymmetric lack of quality of antiinfective medication data in clinical information systems. This phenomenon could likely be explained by the fact that antiinfective agents are more commonly than other medications taken only for a limited period of time. If the medication records are not set to be automatically marked as inactive after the patient has completed the course of treatment, they will remain on the patient's record until manually removed, making them susceptible to errors of omission.

Medications prescribed on as-needed basis were also less likely to be validated. Similarly to the antiinfective agents, this discrepancy was asymmetric and other studies did not report that as-needed medications were missing from the medical record more frequently than medications taken on a regular basis [11]. It is therefore likely that the explanation for the high rate of outdated medication records in this group also lies in the increased prevalence of transient prescriptions that are not subsequently manually deactivated. Based on these findings, it may be helpful to incorporate special handling of antiinfective and *prn* medications in clinical information systems design to include additional alerts for providers to inactivate these medications when the original prescription has expired.

Medication records representing discharge medications from previous admissions were slightly more likely to be validated that the ones from outpatient clinical information systems. One possible reason for this is that patients admitted to the hospital usually have their medications reconciled [7]. While this process may not always be complete [12], it could nevertheless be more comprehensive than outpatient medication reconciliation due to more extensive resources available in the hospital.

Our study has multiple strengths. It was a large-scale investigation that included over 130,000 medication records from more than 10,000 admissions at two hospitals. To our knowledge, it is the first investigation of the characteristics associated with medication record validation. We took advantage of a novel medication reconciliation software – PAML Builder – to assemble a unparalleled database of medication records that were individually validated by admitting providers. These records included patients being admitted to all hospital services, creating a clinically diverse study population. Finally, most previously published studies focused on the validity of medication records in outpatient clinical information systems but our analysis also included hospital discharge medication lists from inpatient clinical information systems and provided a direct comparison between the two.

The study has several limitations. We assumed that all electronic medication records copied to the PAML were validated, and that the patients were not actually taking medications that were not copied to the PAML. It is possible, however, that PAMLs could also be inaccurate. If identical medication entries were present in more than one clinical information system, only one would be copied to the PAML, even if all were valid. However, the majority of medication entries analyzed in the study did not have duplicates, and in multivariable analysis of medication

record characteristics predictive of validation all of our findings were independent of duplicate records. Our study was limited to two academic medical centers in Boston; it is therefore possible that the findings may not apply elsewhere. We only studied patients who were admitted to the hospital, and the characteristics of the electronic medication records of the patients who were not admitted could be different. Finally, as with any retrospective analysis, the data were not collected specifically for the study which could potentially introduce bias in our findings.

In summary, this large retrospective study of validation of medication records in clinical information systems has confirmed and quantified findings previously reported in smaller investigations and demonstrated new relationships between the characteristics of electronic medication records and their validity. These findings could be incorporated in the design of the future clinical information systems to ensure that information predictive of non-validity of medication records is prominently displayed and that special procedures are devised for handling classes of medications (for example, antibiotics or *prn* medications) that are particularly likely to be outdated.

Acknowledgments

We would like to thank Partners Medication Reconciliation team for their hard work that made this project possible. We would like to express our gratitude to James Shalaby and Jaime Chang for assistance with medication class mapping.

References

[1] Bates DW, Gawande AA: Improving safety with information technology. *N Engl J Med* 2003, 348(25):2526-2534.

[2] Manley HJ, Drayer DK, McClaran M, Bender W, Muther RS: Drug record discrepancies in an outpatient electronic medical record: frequency, type, and potential impact on patient care at a hemodialysis center. *Pharmacotherapy* 2003, 23(2):231-239.

[3] Ernst ME, Brown GL, Klepser TB, Kelly MW: Medication discrepancies in an outpatient electronic medical record. *Am J Health Syst Pharm* 2001, 58(21):2072-2075.

[4] Grant RW, Devita NG, Singer DE, Meigs JB: Improving adherence and reducing medication discrepancies in patients with diabetes. *Ann Pharmacother* 2003, 37 (7-8):962-969.

[5] Vilke GM, Marino A, Iskander J, Chan TC: Emergency department patient knowledge of medications. *J Emerg Med* 2000, 19(4):327-330.

[6] Vermeire E, Hearnshaw H, Van Royen P, Denekens J: Patient adherence to treatment: three decades of research. A comprehensive review. *J Clin Pharm Ther* 2001, 26(5): 331-342.

[7] Joint Commission 2006 National Patient Safety Goals Implementation Expectations http://www.jointcommission.org/NR/rdonlyres/DDE15942-8A19-4674-9F3B-C6AE2477072A/0/06_NPSG_IE.pdf

[8] Poon EG, Blumenfeld B, Hamann C, Turchin A, Graydon-Baker E, Baker G, McCarthy PC, Poikonen J, Mar P, Schnipper JL *et al*: Design and Implementation of an Application and Associated Services to Support Interdisciplinary Medication Reconciliation Efforts at an Integrated Healthcare Delivery Network. *J Am Med Inform Assoc* 2006, in press.

[9] Wagner MM, Hogan WR: The accuracy of medication data in an outpatient electronic medical record. *J Am Med Inform Assoc* 1996, 3(3):234-244.

[10] Curtis JR, Westfall AO, Allison J, Freeman A, Kovac SH, Saag KG: Agreement and validity of pharmacy data versus self-report for use of osteoporosis medications among chronic glucocorticoid users. *Pharmacoepidemiol Drug Saf* 2006, 15(10):710-718.

[11] Yang JC, Tomlinson G, Naglie G: Medication lists for elderly patients: clinic-derived versus in-home inspection and interview. *J Gen Intern Med* 2001, 16(2):112-115.

[12] Schnipper JL, Kirwin JL, Cotugno MC, Wahlstrom SA, Brown BA, Tarvin E, Kachalia A, Horng M, Roy CL, McKean SC *et al*: Role of pharmacist counseling in preventing adverse drug events after hospitalization. *Arch Intern Med* 2006, 166(5):565-571.

Address for correspondence

Alexander Turchin, MD, MS
Clinical Informatics Research and Development
Partners HealthCare System
93 Worcester Road, Suite 201
Wellesley, MA 02481
aturchin@partners.org

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Evaluation of an Electronic Medication Reconciliation System in Inpatient Setting in an Acute Care Hospital

Abha Agrawal[a], MD, FACP, Winfred Wu[b], MD, MPH, Israel Khachewatsky[a]

[a] *Kings County Hospital, Brooklyn, NY, USA*
[b] *Suffolk County Department of Health Services, Hauppauge, NY, USA*

Abstract

Background: Medication reconciliation (MedRecon) is being implemented in many healthcare facilities as a means to reduce medication errors. However, there is scant literature on the evaluation of electronic MedRecon systems.

Objective: To evaluate the rate and type of discrepancies between a patient's home medication history and admission orders and to analyze factors affecting their occurrence using an electronic MedRecon system.

Design / Methods: We analyzed 3,426 consecutive inpatient admission MedRecon events from August to October 2006 in an acute care hospital using a recently implemented electronic MedRecon system.

Results: Overall, discrepancy rate was 3.12% (n=107) with omission of a home medication being the most common type (56.52%, n=65) of discrepancy. Admission time (8 PM to 8 AM), and total home medications >4 were found to have a significant positive correlation with discrepancy rate.

Conclusion: Using multidisciplinary MedRecon process based on an electronic system, we found a low discrepancy rate between patient's home medication history and admission orders compared with the rate in the literature, implying that an electronic MedRecon system is an important tool for improving patient safety.

Keywords:

medication reconciliation, medication errors, adverse drug events, patient safety, JCAHO, computerized medical record, quality of health care

Introduction

Occurrence of medication errors in acute inpatient care facilities in United States is a widely published issue and prevention of medication errors has become a critically important national priority (1). Many of the medication errors occur during care transition points such as hospital admission, transfer and discharge due to multiple changes in medication regimens (2).

In response to these concerns, the Joint Commission on Accreditation of Healthcare Organizations (JCAHO), in 2006, mandated that all JCAHO-accredited facilities must "accurately and completely reconcile medications across the continuum of care."

The Medication Reconciliation (MedRecon) process on admission involves compiling a complete and accurate list of a patient's home medications and comparing that list to a provider's admission orders. Many health care facilities have implemented paper processes to meet the challenging task of performing effective MedRecon (3-5). Recent publications have reported some experience with electronic MedRecon systems (6-7); however the use of such systems is a relatively new phenomenon and evaluation of their effectiveness in preventing potential medication errors is of much importance as many hospitals are expected to implement such systems in the coming years.

The objective of this study is to evaluate the rate and type of discrepancies between a patient's home medication history and inpatient admission orders and to analyze the factors affecting their occurrence from the recently implemented electronic MedRecon system at our facility. Published research strongly suggests that a low rate of such discrepancies would imply a low risk of potential adverse drug events and a safer hospitalization (8).

Materials and methods

Setting and participants

The study was conducted at Kings County Hospital Center (KCHC), a 630-bed acute tertiary care teaching hospital with approximately 24,000 discharges per year. The inpatient services included in the study were medical, surgical, behavioral health, obstetrics / gynecology and other specialty services. Table 1 describes the patient, clinician, and environment of care characteristics for the 3,426 MedRecon events in the study.

Electronic Medication Reconciliation System at KCHC

In 2006, an electronic MedRecon system was implemented facility-wide as a module within the existing commercial electronic medical record (EMR) system, MISYS. The EMR is

Selected for best paper award.

Table 1 - Patient, clincian, and environment of care characteristics

Total n=3,426 unique MedRecon events	
Patient Characteristics	
Age, mean± SD, y	38.4 yr +/- 23.2
Sex	
Male	1418 (41%)
Female	2008 (59%)
Insurance status	
Insured	2686 (78%)
Uninsured	740 (22%)
Clinician Characteristics	
Resident	2405 (70%)
Attending physician	804 (23%)
Physician Assistant/Nurse Practitioner	217 (7%)
Environment of care characteristics	
Admission day	
Weekday	2631 (77%)
Weekend (Saturday or Sunday)	795 (23%)
Admission time	
Daytime, 8 AM or later	1584 (46%)
Night time, 8 PM or later	1842 (54%)
Service type	
Non-pediatrics	589 (17%)
Pediatrics	2837 (83%)

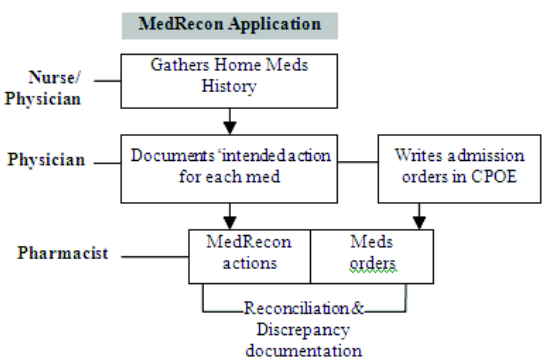


Figure 1 - Schemata of Admission Medication Reconciliation

extensively used across KCHC with all outpatients and inpatients orders being placed using the computerized physician order entry (CPOE) application in the EMR.

To comply with the mandatory JCAHO requirement of MedRecon, a dedicated multidisciplinary team consisting of physicians, nurses, pharmacists, information technology personnel, and clinical educators was convened. The team was responsible for the design, development, implementation, and evaluation of the electronic MedRecon module. The outpatient MedRecon module was implemented in all 100+ clinics from December 2005 to February 2006. The inpatient MedRecon system

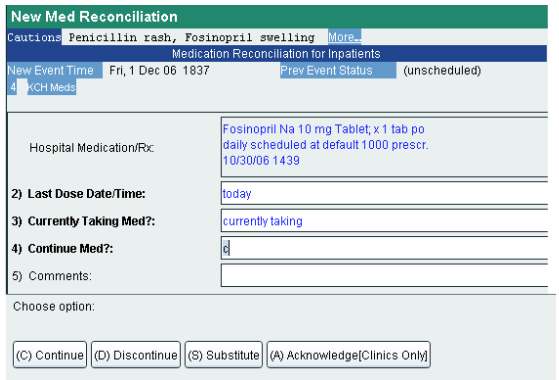


Figure 2 – Screen Shot Displaying Physician Medication Reconciliation Documentation Screen

was implemented for all services in May 2006. The inpatient system incorporates MedRecon processes for all three stages: admission, transfer and discharge. The current study evaluated the admission MedRecon process only, described below, as a majority of medication errors have been found to occur on admission (9).

The admission MedRecon process involves three steps (Figure 1). The first step consists of documenting a comprehensive medication history of a patient from all sources and can be performed by a nurse or a physician. The clinician invokes the MedRecon application on every admission by clicking on a MedRecon menu button in the EMR. The system automatically compiles a list of medications preexisting in that patient's profile (including outpatient or discharge prescriptions written using CPOE) in the EMR and presents them on the screen. These medications are labeled as "KCHC meds". For each KCHC med, the clinician documents the "currently taking" status (yes or no), last dose date and time, and optionally, any relevant comments. To ensure a complete home medication history, the next screen asks the clinician to add any medications being prescribed outside KCHC and non-prescription substances, such as Over-the-counter medications, nutraceuticals, herbals etc., labeled "Non-KCHC meds". The clinician documents "currently taking" status, last dose date and time, and an optional comment for "Non-KCHC meds" as well.

In the second step, a physician documents the 'intended action' for each medication in the MedRecon application by selecting one of the options: "continue", "discontinue", "substitute", or "unable to verify" (Figure 2). The MedRecon documentation is then automatically routed to an electronic work queue for pharmacy. The physician also places admission medication orders using CPOE. At the time of order writing, patient's home medication history, derived from the MedRecon application is made available in real-time to the provider, on the CPOE screen, to minimize discrepancies between home medication history and admission orders.

In the final step, a pharmacist performs 'reconciliation' by comparing the physician's 'intended action' for home medications with the admission orders. The pharmacist screen is configured to display both MedRecon documentation and CPOE admission orders side-by-side using the split-screen technique. The pharmacist records whether any discrepancies were found (yes or no) and if yes, categorizes them into one or more of the following: "continued" medication(s) not ordered, "discontinued" medication(s) ordered, dose discrepancy, frequency discrepancy, therapeutic duplication, and other (show screen shot). The taxonomy of discrepancies was derived from the literature (10-11). In addition, the pharmacist also communicates with the provider to resolve any discrepancies found.

Study design

The study was conducted as an observational study based on retrospective data analysis, performed as a quality improvement project. For data collection, an automated query was performed in the EMR to retrieve medication reconciliation data on all inpatient admissions from all services from August 1 2006 to October 30 2006. The data were analyzed for overall discrepancy rate, type of discrepancies, and the association of various patient, provider and environment of care characteristics with discrepancy rates.

Statistical analysis

Statistical analysis was performed using Microsoft Excel and SPSS for Windows, version 13.0. Chi-square analysis was used for comparisons of proportions. Patient, clinician, and environment of care characteristics were also analyzed using logistic regression. These characteristics were entered into the model and adjusted odds ratios with 95% confidence intervals were calculated. All P-values were 2-sided and a significance level of 0.05 was used.

Results

A total of 3,426 consecutive unique MedRecon events were found in the system during the study period. Of these, 107 events had at least one discrepancy recorded, yielding an overall rate of 3.12%. The categorization of discrepancies is described in Table 2. A total of 115 types were recorded as one event may have more than one type of discrepancy recorded. Unintentional omission of a home medication from admission orders was the most frequent type of discrepancy.

Table 3 summarizes various characteristics according to the presence of discrepancy based on bivariate analysis. Table 4 describes the association between selected variables and discrepancies based on multivariate analysis. Total medications >4 and night time admission were significant positive predictors of discrepancy and weekend admission was a significant negative predictor.

Table 2 - Categorization of discrepancies found

Type of discrepancy	No.	%
"Continued" medication not ordered	65	56.52%
"Discontinued" medication ordered	12	10.43%
Dose discrepancy	11	9.57%
Frequency discrepancy	1	.087%
Therapeutic discrepancy	4	3.48%
Other	22	19.13%
Total	**115**	

Table 3 – Patient, clinician and environmental characteristics according to the presence of discrepancy

	Discrep. present	Discrep. absent	p value
Patient Characteristics			
Age >65	28%	14%	<0.01
Uninsured	20%	22%	0.61
Total meds >4	31%	13%	<0.01
Clinician Characteristics			
Resident	80%	70%	0.02
Nurse didn't perform meds history	59%	66%	0.11
Care Environment Characteristics			
Weekend admission	14%	24%	0.02
Nighttime admission	69%	53%	<0.01
Pediatrics service	13%	17%	0.25

Table 4 - Association between selected variables and discrepancies

	Adjusted OR*	95% CI*
Patient Characteristics		
Age >65	1.59	0.99-2.55
Uninsured	1.08	0.66-1.76
Total meds >4	2.25	1.41- 3.57
Clinician Characteristics		
Resident	1.57	0.94 - 2.62
Nurse didn't perform meds history	1.23	0.81-1.88
Care Environment Characteristics		
Weekend admission	0.50	.284-.863
Nighttime admission	1.84	1.20-2.81
Pediatrics service	1.03	0.55-1.95

* OR - Odds Ratio; CI- Confidence Interval

Discussion

Current study reports a relatively low rate (3.12%) of discrepancies between a patient's home medication history and admission orders. Other studies have reported that 54-67% of all admitted patients have at least one discrepancy

between home medications and the actual admission orders (8, 12-14).

Similar to reported literature, we found that error of omission is the most common prescribing error on admission, followed by error of commission. We also observed that about 20% of the discrepancies were recorded under the free text category of 'other'. The contents of this field are being analyzed to understand if another category of discrepancy should be included in the system.

A earlier preliminary 2-week (June 3-June 16, 2006) investigation of MedRecon events at our facility using the aforementioned system, shortly after implementation, had found a 20% rate of discrepancies (15). This significant reduction in the discrepancy rate over the next 4 months implies that the system is remarkably effective in preventing discrepancies as its implementation has matured. Reduction of discrepancies is an important factor in improving patient safety as a recent analysis reported that discrepancies have the potential to cause moderate to severe discomfort or clinical deterioration in 38.6% cases (8).

Our study is novel in reporting not only the discrepancy rate using an electronic MedRecon system but it also reports various factors influencing the discrepancy rate. On adjusted analysis, two factors were found to have a significantly positive correlation with the occurrence of a discrepancy – night time admission and total home medications >4. This finding underscores the risk of polypharmacy and reiterates the need to exercise extra caution in writing admission orders during night hours. Weekend admission was found to have a significant negative correlation with the occurrence of discrepancies. This may be secondary to reduced workload on weekends, but further investigation is required to ascertain the implications of this finding.

This study demonstrates that although CPOE decision support tools are an important component of medication error prevention strategies, they alone are not sufficient to prevent errors of prescribing. An effective MedRecon process should complement CPOE systems in preventing medication errors for the following reasons. First, similar to prior research, we found that the most common error is omitting a medication from admission orders that is taken at home (10-11). CPOE systems with decision support tools such as drug-drug interactions, dose-range checking, therapeutic duplication etc. have proven quite effective in addressing the issues of lack of prescriber's knowledge about drugs; however, a CPOE system would not be capable of detecting unintentional omission of medications taken before admission. Second, CPOE, unless linked to a community pharmacy database, only 'knows' the medications prescribed at the primary institution and is 'unaware' of medications prescribed elsewhere. An effective MedRecon process, on the other hand, ensures a comprehensive home medication history. Third, CPOE profile only indicates that a medication was prescribed and not whether the patient is actually taking the medication – another issue addressed effectively with MedRecon solution.

The discrepancy rate in our study essentially highlights the 'safety gap' between gathering the home medication history list and (via MedRecon) the admission ordering (via CPOE), even when both processes are computerized. To eliminate medication errors, this gap - though small as reflected in our low discrepancy rates – needs to be bridged. One potential solution to bridge this gap is to allow clinicians to convert entries in home medication history into orders easily 'with a click'. Clinicians at our institutions expressed an interest in this solution as it would save them time. We deferred linking MedRecon directly to CPOE, at least in the first phase, for several reasons. First, several reports on unintended consequences of information technology caution that over-automation of workflow may introduce new errors (16-17). Second, a standardized normalization and data mapping approach would be required before converting outside home meds into CPOE entries. Therefore, currently we decided to keep home medication history detached from the CPOE, although it is always available on the order entry screen as a reminder to the physician.

The current study also highlights the important role of clinical pharmacists in preventing potential adverse events at admission. A recently published systematic review reported that the addition of clinical pharmacists to care teams resulted in improved care (18). Most studies in this review focused on their role in medication history taking or as a liaison service. Our study is unique in comprehensively evaluating their active role in the actual reconciliation process.

A useful 'side effect' of the medication reconciliation implementation at our facility has been a heightened awareness among prescribing physicians of the importance of a comprehensive home medication history. In addition, the system provides a standardized method of documenting outside medications and sharing the information among various providers across the continuum of care.

A limitation of our study is that no chart review was done to ascertain the impact of the discrepancies on clinical outcomes or length of stay. The study would also be strengthened by categorizing each instance of discrepancy into a severity score according to the potential to cause harm (12).

Conclusion

Using a multidisciplinary MedRecon process based on an electronic system, we found a low discrepancy rate between patient's home medication history and admission orders. Further research is needed to evaluate the impact of these discrepancies on clinical outcome. As informatics infrastructure for data mapping among disparate systems advances, integrating MedRecon system with CPOE should be explored.

Acknowledgements

The authors wish to thank other key members of the MedRecon project team at Kings County Hospital – Munish Khaneja, MD, MPH, Eugene Edynak, MD, Christopher Russo, RPh.

References

[1] Aspden P, Institute of Medicine (U.S.). Committee on Identifying and Preventing Medication Errors. Preventing medication errors. Washington, DC: National Academies Press; 2006.

[2] Bates DW, Spell N, Cullen DJ, Burdick E, Laird N, Petersen LA, et al. The costs of adverse drug events in hospitalized patients. Adverse Drug Events Prevention Study Group. JAMA 1997;277(4):307-11.

[3] Bartick M, Baron D. Medication reconciliation at Cambridge Health Alliance: experiences of a 3-campus health system in Massachusetts. Am J Med Qual 2006;21(5):304-6.

[4] Haig K. Medication reconciliation. Am J Med Qual 2006;21(5):299-303.

[5] Pronovost P, Weast B, Schwarz M, Wyskiel RM, Prow D, Milanovich SN, et al. Medication reconciliation: a practical tool to reduce the risk of medication errors. J Crit Care 2003;18(4):201-5.

[6] Hamann C, Poon E, Smith S, Coley C, Graydon-Baker E, Gandhi T, et al. Designing an electronic medication reconciliation system. AMIA Annu Symp Proc 2005:976.

[7] Poon EG, Blumenfeld B, Hamann C, Turchin A, Graydon-Baker E, McCarthy PC, et al. Design and Implementation of an Application and Associated Services to Support Interdisciplinary Medication Reconciliation Efforts at an Integrated Healthcare Delivery Network. J Am Med Inform Assoc 2006;13(6):581-592.

[8] Cornish PL, Knowles SR, Marchesano R, Tam V, Shadowitz S, Juurlink DN, et al. Unintended medication discrepancies at the time of hospital admission. Arch Intern Med 2005;165(4):424-9.

[9] Tam VC, Knowles SR, Cornish PL, Fine N, Marchesano R, Etchells EE. Frequency, type and clinical importance of medication history errors at admission to hospital: a systematic review. Cmaj 2005;173(5):510-5.

[10] Lau HS, Florax C, Porsius AJ, De Boer A. The completeness of medication histories in hospital medical records of patients admitted to general internal medicine wards. Br J Clin Pharmacol 2000;49(6):597-603.

[11] Beers MH, Munekata M, Storrie M. The accuracy of medication histories in the hospital medical records of elderly persons. J Am Geriatr Soc 1990;38(11):1183-7.

[12] Gleason KM, Groszek JM, Sullivan C, Rooney D, Barnard C, Noskin GA. Reconciliation of discrepancies in medication histories and admission orders of newly hospitalized patients. Am J Health Syst Pharm 2004;61(16):1689-95.

[13] Akwagyriam I, Goodyer LI, Harding L, Khakoo S, Millington H. Drug history taking and the identification of drug related problems in an accident and emergency department. J Accid Emerg Med 1996;13(3):166-8.

[14] Bikowski RM, Ripsin CM, Lorraine VL. Physician-patient congruence regarding medication regimens. J Am Geriatr Soc 2001;49(10):1353-7.

[15] Wu W, Khaneja M, Agrawal A. Evaluation of Electronic Inpatient Medication Reconciliation Implementation at an Acute Care Hospital [Poster]. In: 17th Annual Symposium on Health Care Services in New York: Research and Practice; 2006; New York, NY; 2006.

[16] Ash JS, Berg M, Coiera E. Some unintended consequences of information technology in health care: the nature of patient care information system-related errors. J Am Med Inform Assoc 2004;11(2):104-12.

[17] Koppel R, Metlay JP, Cohen A, Abaluck B, Localio AR, Kimmel SE, et al. Role of computerized physician order entry systems in facilitating medication errors. Jama 2005;293(10):1197-203.

[18] Kaboli PJ, Hoth AB, McClimon BJ, Schnipper JL. Clinical pharmacists and inpatient medical care: a systematic review. Arch Intern Med 2006;166(9):955-64.

Address for correspondence

Abha Agrawal, MD, FACP
Kings County Hospital Center
451 Clarkson Avenue, Brooklyn, NY 11203
E-mail: abha.agrawal@nychhc.org

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Computerized Management of Chronic Anticoagulation: Three Years of Experience

Beatriz H. Rocha,[a] Laura H. Langford,[a] Steven Towner[b]

[a] *Medical Informatics, Intermountain Healthcare, United States*
[b] *Department of Internal Medicine, Intermountain Healthcare, United States*

Abstract

Chronically anticoagulated patients taking the drug Warfarin require time intensive management and follow-up processes to avoid complications. The "Chronic Anticoagulation Clinic" (CAC) protocol is a set of production rules that help manage, treat, and follow-up such patients. The CAC protocol has been in regular use at Intermountain Healthcare (Salt Lake City, UT, USA) for over three years. The results demonstrate an improvement on the number of patients with anticoagulation levels within the desired target range. The protocol alerts have a high acceptance rate (83.4%) and were able to help patients remember to collect their next coagulation test. The CAC protocol results show that production rules can improve the management of chronically anticoagulated patients. Additional studies are required to verify if this experience can be transferred to other institutions.

Keywords:

clinical decision support systems, computerized protocols, anticoagulation, Warfarin

Introduction

Chronically anticoagulated patients are those that receive oral anticoagulation treatment with Warfarin for a long period of time. Medical reasons for chronic anticoagulation therapy include diseases such as Stroke, Atrial Fibrillation, and Deep Vein Thrombosis (DVT). Many published studies highlight the importance of properly managing these patients in order to avoid complications of excessive or sub-therapeutic Warfarin dosage [1, 2]. Improper anticoagulation management results in medical complications such as embolism or bleeding, along with increased health care costs.

The proper dosage of Warfarin is monitored using the International Normalized Ratio (INR) coagulation test. Studies have shown that patients that are managed through specialized anticoagulation services have better control of INR and fewer complications [1, 2]. The objective is to maintain the patient's INR within a "target range" that is specific for the patient.

A specialized anticoagulation clinic was created at Intermountain in 2001. From its inception, this clinic has been supported by a paper protocol (set of production rules) that helps with the management, treatment, and follow-up of chronically anticoagulated patients. The intent of this paper is to report on the experience obtained by using a computerized version of this protocol for a period of over three years.

Materials and methods

Patients who are receiving Warfarin are referred by their primary care physicians to the Salt Lake Chronic Anticoagulation Clinic (CAC). CAC is a nurse practitioner-led telephone-based clinic responsible for educating patients about Warfarin usage, and also for managing their Warfarin dosage through periodic INR tests [3]. Patients come to the clinic just once a year for a scheduled educational session. INRs can be collected in any Intermountain laboratory, taking into account patient preferences and time availability. Providers are alerted electronically about test results which are then communicated to patients by phone.

The initial set of production rules was created in "paper format" by one of the authors (ST), taking into account his personal experience and pertinent literature. These rules were put in routine use at the clinic in 2001. The "paper format" rules were later refined through knowledge engineering sessions and implemented as computerized production rules by another author (BHR). The computerized rules are commonly known as the "CAC Protocol". Since its initial deployment, the CAC Protocol has had a few minor revisions to better integrate the rules with the workflow of the clinicians.

The CAC Protocol was implemented using "Foresight" [4] as its platform. Foresight is a flexible decision logic execution engine coupled with a sophisticated clinical data monitor, both developed at Intermountain Healthcare. Foresight is integrated with Intermountain's outpatient "Electronic Medical Record" (EMR) known as "HELP2" [5]. Patients referred to the clinic are enrolled in the CAC Protocol through HELP2. After enrollment, the clinician also enters in HELP2 the INR target range for the patient (e.g., "2.0 to 3.0" or "2.5 to 3.5").

Selected for best paper award.

Table 1 - Examples of CAC Protocol rules for the target INR ranges

Target INR	Current INR	Last INR (past 25 days)	Message	Severity	Method of Delivery
2.0 to 3.0	< 1.6	If no previous INR or previous INR was >= 1.8.	Action Point Low. Message if dosage of Warfarin was changed in the past 25 days. Inquire about signs and symptoms of clotting, and if necessary, refer to an appropriate facility for care. Consider extra dose of Warfarin. Increase weekly dose by 5-15%. Retest in 7-14 days. Previous INR zone: XXXXX (if available)	Medium	Message Log
	2.0-3.0	If previous INR was between 1.8 and 3.3 (and was a "Repeated Yellow Zone")	Possible sliding green zone. Message if dosage of Warfarin was changed in the past 25 days. Retest in 14-21 days. Previous INR zone: XXXX (if available)	Low	Message Log
2.5 to 3.5	5.0-8.9	If no previous INR or previous INR < 4.0	Action Point High. Message if dosage of Warfarin was changed in the past 25 days. Omit 1-2 doses. Inquire about bleeding and refer to appropriate facility for care if needed. MD to review. Retest in 24-48 hours. Previous INR zone: XXXXX (if available)	Critical	Message Log and Pager
	>= 9.0	N/A	Critical INR. Message if dosage of Warfarin was changed in the past 25 days. Hold Warfarin. Consider Vitamin K. Consult the medical director or nurse practitioner for advice. Retest in 24-48 hours. Previous INR zone: XXXXX (if available)	Critical	Message Log and Pager

The clinical data monitor activates Foresight every time a patient event (e.g., laboratory result, pharmacy order,

allergy documentation, etc.) is stored in the HELP2 database. The CAC Protocol rules are triggered every time a new INR result is stored into the HELP2 database. POC (point-of-care) INR

results can be entered directly into HELP2, activating the CAC Protocol in real-time.

Once triggered, the protocol rules start by verifying if the INR result is from an enrolled patient. If the patient is enrolled, the rules verify if the result is valid and if it has not been interpreted before, avoiding duplicated alerts. The rules also verify if an INR target range has been selected for the patient. If no target range can be found, an alert is generated requesting the clinician to select a target range.

Once these verifications have been completed, the rules interpret the new result by comparing it to both the last INR in the past 25 days and the last alert generated. Examples of the rules are shown in Table 1. During the interpretation phase, the new INR result is classified into a "zone" (Table 2). For instance, a "green zone" means that the new INR result is within the patient's selected target range while a "yellow zone" means that the new INR is slightly outside the target range (see Table 2 for possible zones).

Table 2 - Possible INR interpretation "zones"

Zones	Severity	Late alert
Critical	Critical	73 hours after 1st alert
Action Point High	Critical	97 hours after 1st alert
Action Point Low	High or Medium	21 days after 1st alert
Red (High or Low)	High or Medium	21 days after 1st alert
Possible Sliding Yellow (High or Low)	Medium	21 days after 1st alert
Repeated Yellow (High or Low)	Medium	28 days after 1st alert
First Yellow (High or Low)	Low or Medium	31 days after 1st alert
Possible Sliding Green (within target range)	Low or Medium	31 days after 1st alert
Green (within target range)	Low	50 days after 1st alert

At the end of the interpretation phase an alert message is generated classifying the current INR into a zone. The alert message also contains recommendations about dosage management, information if the Warfarin dosage has changed in the past 25 days, when the patient should collect the next INR, and the previous INR zone (if available). The alerts are also classified by severity (Table 1). A dif-

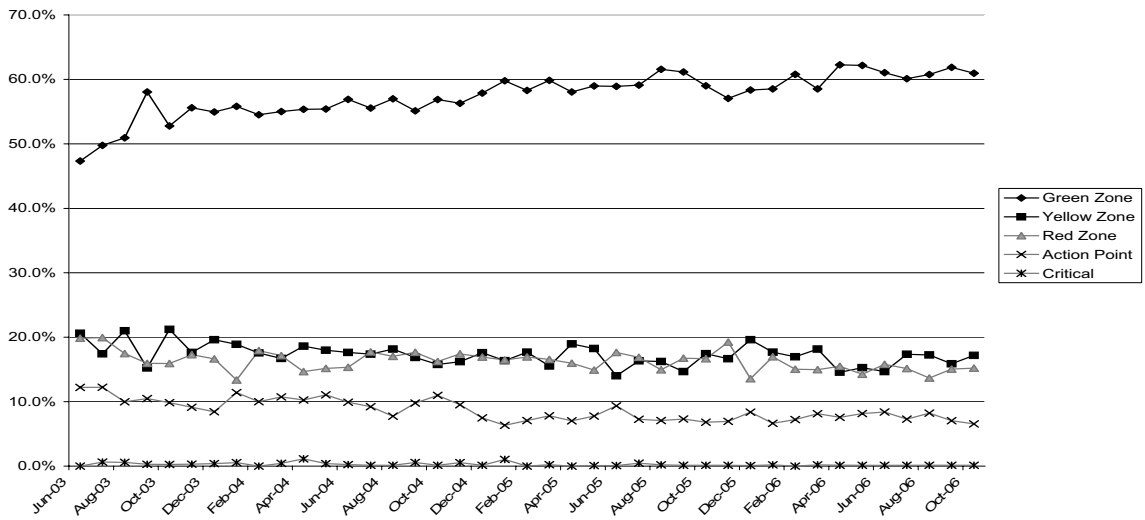


Figure 1 - Alert distribution by interpretation zone

ferent alert message is generated when an INR result is originated from an inpatient setting, since CAC clinicians do not manage Warfarin usage while patients are hospitalized.

The alert message is stored in the HELP2 database and also routed to the HELP2 "Message Log". Message Log is where the HELP2 users receive internal emails, notifications, and alerts. The alert message is sent to all clinicians that work in the CAC and that have an association with patient. Besides sending the alert to the Message Log, "Critical" and "Action High" alerts are sent to the pager of the clinician responsible for the clinic. If the alerts are not acknowledged within one hour, the alert message is resent to the same pager.

Clinicians use Message Log to review the CAC Protocol alerts that were issued since their last access. The effort to review the alert messages is divided among the clinicians, taking into account the severity of the alerts. The nurse practitioners (NPs) review "critical" and "high severity" alerts, while the licensed practical nurses (LPNs) review "medium" and "low severity" alerts. Besides the alert messages, the clinicians can also access from the same screen the INR result that triggered the alert. After reviewing the alerts, clinicians call each patient to communicate the INR result and, if necessary, the appropriate Warfarin dosage changes.

After contacting a patient, the clinician acknowledges the alert by accepting or rejecting it. If the clinician decides to accept the alert, she is presented with a set of possible actions taken, such as "Left Message with Household Contact" or "Spoke with Patient". The clinician can also enter other actions, but only as free-text. If the clinician decides to reject the alert, a set of rejection reasons is presented, such as "Perioperative management by another provider" or "Surgery/Procedure (Warfarin held)". Once an alert is acknowledged, it is automatically removed from the Message Log of all clinicians. If a clinician does not want to acknowledge an alert, she can add a comment to the alert such as "Called patient twice; not able to reach anyone at home". All alert acknowledgement details, including the clinician identity, the transaction time stamp, and the action taken are stored with the alert in the HELP2 database. Previously stored alerts can be reviewed at any time by the clinicians through the HELP2 Alert Review module.

During the INR interpretation phase, a "future event" is also created by Foresight depending on the resulting INR zone. A future event reactivates the CAC Protocol after a predetermined amount of time has elapsed. Each INR interpretation zone has a predefined "late" retest time period (Table 2). Whenever a new INR result becomes available before the expiration of the retest period, the future event is canceled and a new time event is created for the new INR result. If a new INR result is not stored before the predefined retest period expires, a "late alert" is generated reminding the clinician that the patient is due for a new INR test. Once the first late alert is generated, it is reissued once a week until a new INR result becomes available, or until the patient is un-enrolled from the CAC Protocol. Also, if the patient is enrolled in the CAC Protocol but an INR result is not made available, an alert is generated after 50 days reminding the clinician that the patient needs to have an INR test.

Results

The CAC Protocol was first deployed in June of 2003 and has been in production use ever since. The CAC Protocol has 46 rules total. The rules have been changed 11 times (9 times during the first 9 months of production use). One example of the changes made were the "late" retest time periods.

The Salt Lake CAC currently has 734 patients enrolled in the computerized protocol. A total of 48,552 alerts have

been generated between June of 2003 and October of 2006. An average of 52 alerts per day are generated during weekdays and 9 alerts per day during weekends. Figure 1 shows the distribution of alerts by interpretation zone. Alerts of the same type were aggregated to facilitate visualization, i.e., "green zone" includes "Green Zone" and "Possible Green Zone" alerts and "red zone" includes "Red Zone High" and "Red Zone Low" alerts.

Table 3 shows the number of alerts distributed by type. Of all patients ever enrolled in the protocol, 81.8% have received at least one "Patient late for follow-up INR" alert. The average number of late alerts is 8.1 alerts/patient (median is 5.0 late alerts/patient). All Salt Lake CAC alerts have been acknowledged: 40,492 (83.4%) have been "accepted" and 8,060 (16.6%) have been "rejected". The average time for acknowledging an alert after it has been generated is 13.6 hours (median is 3.0 hours). Only 3 patients, out of the 734 currently enrolled in the CAC Protocol, have not received an alert in the past 50 days. All 3 patients were recently enrolled in the protocol and are still in the initial 50 days waiting period.

Table 3 – Number of alerts by type

Type of alert	# of alerts	Percentage of total
Green Zone	15,966	32.9%
Patient late for follow-up INR	9,408	19.4%
Possible sliding Green Zone	4,328	8.9%
INR drawn for hospitalized protocol patient	3,896	8.0%
Red Zone High	3,369	6.9%
First Yellow Zone Low	2,968	6.1%
Action Point Low	2,475	5.1%
First Yellow Zone High	2,302	4.7%
Red Zone Low	2,293	4.7%
Action Point High	506	1.0%
Repeated Yellow Zone Low	447	0.9%
Repeated Yellow Zone High	252	0.5%
Patient INR goal has not been set. Please chart the goal.	200	0.4%
Critical INR	89	0.2%
Possible Sliding Yellow Zone High	27	0.1%
Possible Sliding Yellow Zone Low	26	0.1%
Total	48,552	100%

The most common reasons for accepting alerts were "Notified by phone" (23,975), "Spoke with Patient" (6,912), "Send this alert through voicemail to the patient" (4,386), and "Notified in person" (694). The most common reasons for rejecting an alert were "Hospitalization" (2,741), "Perioperative management by another provider" (578), "Deviation from the protocol" (437), "Noncompliance with Medication" (291), and "Incorrect Zone" (289). Out of the 289 alerts rejected as "Incorrect Zone" (0.6% of all alerts), 200 were "Patient late for a follow-up INR" alerts and 46 "INR drawn for hospitalized patient" alerts. The rest of the alerts rejected as "Incorrect Zone" were "Green Zone" (12), "Yellow Zone" (23), "Red Zone" (4), and "Action Point Zone" (4).

The Salt Lake CAC is run by two LPNs (one full-time and one part-time) and two NPs (both part-time), or 2.25 full-time equivalents (FTEs). Since 2005, 13 other CACs have been created within Intermountain. The 14 CACs are now responsible for the chronic anticoagulation management of 1,760 patients distributed within the state of Utah. All 14 clinics use HELP2 and the same computerized CAC Protocol.

Discussion

There are several publications describing the use of computerized decision support systems to help manage chronically anticoagulated patients [6-10]. These studies have shorter follow-up time and fewer patients when compared to the experience here reported. Most studies are related to Warfarin dose management [6,7,9,10]. However, all studies describe stand-alone systems that require users to enter pertinent patient data and also the recurring INR results. Some systems also suggest when the next INR should be collected [8-10]. No other study has reported the results of using a computerized oral anticoagulation decision support system fully integrated with an EMR, or where clinicians actively use the system to manage a large group of patients for a period of more than three years.

The results observed at the Salt Lake CAC confirm that a computerized protocol can help with the management, treatment, and follow-up of chronically anticoagulated patients. The CAC Protocol can be considered a relatively simple computerized protocol, given the small number of rules and the intent of handling only two INR target ranges. Despite the frequent changes to the rules during the first 9 months, which were primarily required for handling events that were not conceived during the initial design of the protocol, the CAC Protocol has been rather stable. Despite the simplicity of the protocol, its reliance on features of Foresight that are not commonly available in other EMRs makes its transferability to other healthcare institutions potentially difficult.

Figure 1 shows a steady improvement on the number of INRs maintained within the target range since the CAC Protocol was first implemented (from 47.3% to 60.9%). The proportion of patients with INRs within the target range is comparable or better than similar results observed in the literature [2,6,7]. It is conceivable that similar improvements could be obtained with a non-computerized version of the same protocol, but most likely not with the same staffing level given the relatively large number of enrolled patients. Furthermore, the clinic had been using the "paper format" protocol for two years and the observed improvements were after the implementation of the computerized version. The other 13 anticoagulation clinics that are now using the CAC Protocol will hopefully confirm the same improvements observed at the initial implementation site (Salt Lake CAC).

One of the main advantages of the CAC Protocol and its implementation using Foresight is that clinicians do not

need to know when and where the INR test was performed. Similarly, patients do not need to come for an in-person visit to the clinic and are able to go to the laboratory near their residence at their convenience. Clinicians also do not need to re-enter information that is already available in the EMR, including the recurring INR results. Once the INR result is stored in the EMR, clinicians can use the HELP2 Message Log to view the alert message and the triggering INR. Late INR collection alerts are also considered very useful, enabling proper management of all enrolled patients.

The alerts generated by the CAC Protocol had a high acceptance rate. The main reason for rejecting an alert was the fact that the patient had been hospitalized. These hospitalization events occurred outside the Intermountain network and were not recorded in the HELP2 database, i.e., the CAC Protocol interpreted the INR results as if the patient had not been admitted to a hospital. Since CAC clinicians do not manage the anticoagulation during hospitalization periods, they appropriately rejected the alerts. Conversely, the "Incorrect Zone" rejection reason had many different causes. The most common cause was lack of knowledge about the protocol rules by recently employed clinicians. The CAC Protocol was not designed to explain the reason why an interpretation zone was selected and new employees sometimes did not agree with the selection. These misinterpretations were normally solved with education, where a "paper" version of the CAC Protocol was made available to each clinician. The other two causes for "Incorrect Zone" alert rejections were patients that received late alerts but the clinician did not consider them late (e.g., INR test while patient was hospitalized, INR test done outside the Intermountain network and result not entered in HELP2), or INR results that were later corrected.

Due to the good experience with the current rules, a new set of rules was created for a new INR target range ("1.5 to 2.5"). These rules are being validated and should be implemented in January of 2007.

Acknowledgments

We would like to express our gratitude to Ilene Tippets, RN, MS for championing the implementation of the computerized protocol at the various clinics, and to Noman Rahman, David Steiner, and Randy Secrist for programming and maintaining Foresight.

References

[1] Witt SM, Sadler MA, Shanahan RL, Mazzoli G, and Tillman DJ. Effect of a Centralized Clinical Pharmacy Anticoagulation Service on the Outcomes of Anticoagulation Therapy. Chest 2005: 127 (5): 1515-22.

[2] Chiquette E, Amato MG, and Bussey HI. Comparison of an Anticoagulation Clinic With Usual Medical Care: Anticoagulation Control, Patient Outcomes, and Health Care Costs. Arch Intern Med 1998: 158: 1641-1647.

[3] Waterman AD, Milligan PE, Banet GA, Gatchel SK, and Gage BF. Establishing and running an effective telephone-based anticoagulation service. J Vasc Nurs 2001: 19(4): 126-32.

[4] Rocha RA, Bradshaw RL, Hulse NC, and Rocha BHSC. The Clinical Knowledge Management Infrastructure of Intermountain Healthcare. In: Greenes RA, editor. Clinical Decision Support: The Road Ahead. Boston: Academic Press, 2006.

[5] Clayton PD, Narus SP, Bowes WA 3rd, Madsen TS, Wilcox AB, Orsmond G, Rocha B, Thornton SN, Jones S, Jacobsen CA, Udall MR, Rhodes ML, Wallace BE, Cannon W, Gardner J, Huff SM, and Leckman L. Physician use of electronic medical records: issues and successes with direct data entry and physician productivity. AMIA Annu Symp Proc. 2005: 141-5

[6] Wurster M and Doran T. Anticoagulation Management: A New Approach. Dis Manage 2006: 9(4): 201-9.

[7] Fitzmaurice DA, Hobbs FDR, Murray ET, Holder RL, Allan TF, and Rose PE. Oral Anticoagulation Management in Primary Care With the Use of Computerized Decision Support and Near-Patient Testing. Arch Intern Med 2000: 160: 2343-48.

[8] Fihn SD, McDonell MB, Vermes D, Henikoff JG, Martin DC, Callahan CM, Kent DL, and White RH. A computerized intervention to improve timing of outpatient follow-up: a multicenter randomized trial in patients treated with warfarin. National Consortium of Anticoagulation Clinics. J Gen Intern Med. 1994: 9(3): 131-9.

[9] Margolis A, Flores F, Kierszenbaum M, Cavallo Z, Botti B, D'Ottone E, Tavella N, and Torres J. Warfarin 2.0--a computer program for warfarin management. Design and clinical use. Proc Annu Symp Comput Appl Med Care. 1994: 846-50.

[10] Ryan PJ, Gilbert M, and Rose PE. Computer control of anticoagulation dose for therapeutic management. BMJ 1989: 299: 1207-1209.

Address for correspondence

Beatriz H. Rocha, MD, PhD - Medical Informatics; Intermountain Healthcare; 4646 Lake Park Blvd; Salt Lake City, UT 84120; USA; email: Beatriz.Rocha@intermountainmail.org

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Physicians' Response to Guided Geriatric Dosing: Initial Results from a Randomized Trial

Josh F. Peterson, Benjamin P. Rosenbaum, Lemuel R. Waitman, Ralf Habermann, James Powers, Debbie Harrell, Randolph A. Miller[a]

[a]*Vanderbilt University School of Medicine, Department of Biomedical Informatics, Nashville, USA*

Abstract

Guided dosing within a computerized provider order entry (CPOE) system is an effective method of individualizing therapy for patients. Physicians' responses to guided dosing decision support have not been extensively studied. As part of a randomized trial evaluating efficacy of dosing advice on reducing falls in the elderly, CPOE prompts to physicians for 88 drugs included tailored messages and guided dose lists with recommended initial doses and frequencies. The study captured all prescribing activity electronically. The primary outcome was the ratio between prescribed dose and recommended dose. Over 9 months, 778 providers entered 9111 study-related medication orders on 2981 patients. Physicians using guided orders chose recommended doses more often than controls (28.6% vs. 24.1%, p<0.001). Selected doses were significantly lower in the intervention group (median ratio of actual to recommended 2.5, interquartile range [1.0,4.0]) than the control group (median 3.0 interquartile range [1.5,5.0], p<0.001). While physicians selected the recommended dose less than a third of the time, guided geriatric dosing modestly improved compliance with guidelines.

Keywords:

computerized physician order entry system; decision support systems, clinical; geriatrics.

Introduction

Dosing advice in computerized provider order entry (CPOE) systems frequently helps to individualize prescriptions [1-3] Systems have been developed to provide guided drug therapy based on age [2,4], renal function [5], and microbiological target [6]. Evaluations often show modest response rates to such offered recommendations. Iterative refinement of decision support systems may improve acceptance [7]. Yet, the evidence about what choices physicians will make when interacting with a guided dosing application remains empirical and less than definitive.

Previous work by author JP showed CPOE-based geriatric dosing guidelines can successfully improve inpatient fall rates [2]. The present study performed a randomized trial of guided dosing for patients 65 and older, to determine physicians' response rates.

Methods

Design of intervention: During care of elderly patients, the guided dosing system delivered advice to physicians about appropriate initial dosing for sedatives, neuroleptics, anti-emetics and skeletal muscle relexants for the most common indications. The geriatric dosing advisor also discouraged prescription of contraindicated drugs as defined by Fick and Beers (i.e. the "Beers Criteria") [8]. Because appropriate sedative and neuroleptic dosing ranges are broad for high-acuity (e.g., intensive care unit) patients, no barriers prevented selecting higher doses than recommended. The system utilized the same computational infrastructure that project members had previously implemented for specialized age- and weight-based pediatric dosing [4]. Figure 1 shows the editor screen that enables customization of dose lists, setting minimum and maximum single doses, and designating default doses for each study medication. Authors derived study-related dosing information for FDA-approved indications from "package insert" monographs, and for "off-label" indications from the medical literature and published textbooks. The CPOE-based text messages displayed along with study dosing information communicated titration strategies, possible adverse effects, and key monitoring parameters (Figure 2). An advisory group of 2 geriatricians [JP, RH], a geriatrics pharmacist [DH], and an internist [JFP] reviewed the doses and associated messages for accuracy. The knowledge base developed for this project will be submitted to a publicly accessible repository, the POGOe Geriatrics Web site sponsored by the D.W. Reynolds Foundation.

Figure 1 – Web editor for geriatric dosing knowledge base

Setting: The study took place in a tertiary care academic health center with 10 years experience with a self-developed CPOE system. Clinicians enter all inpatient medications orders into the CPOE system except for rare periods of downtime. At the study site, the majority of medication orders are entered directly by resident physicians (>70%), with the remaining 30% entered by attending MDs, nurses (verbal orders), pharmacists (transcribed written orders) and nurse practitioners.

Population: Patients 65 years and older receiving care on one of the order entry wards of the hospital including the emergency room, intensive care units, and a subacute unit were enrolled if admitted between December 8th and August 31st 2006. Patients who received no orders for an intervention were not analyzed. Likewise, only physicians caring for a study patient were analyzed. The CPOE system recorded all study orders including dosing parameters, ordering provider, and location of patient. Additionally, the system logged whether the physicians viewed the advice for intervention patients. For the analysis, study medication orders entered via order sets or specialized sedation protocols were excluded**.**

Analysis: The primary outcome was the ratio of prescribed dose over recommended dose. Medication orders with continuous frequencies (e.g., "q6h") were represented as a projected 24 hour dose and compared to the projected 24 hour dose of the recommended prescription. Single doses (e.g. "once only") were compared to the initial dose of the recommended prescription. Descriptive statistics were performed with medians and inter-quartile ranges (IQR); all dosing distributions were skewed. Significance tests for comparing intervention and control arms were performed with the Mann Whitney rank sum test. To assess for a crossover effect, where dosing advice on intervention patients influences decisions on control patients, we compared "control-only" physicians and "intervention-only" physicians and a pre-trial period to the trial period. All analyses were performed with R statistical package (http://www.r-project.org/).

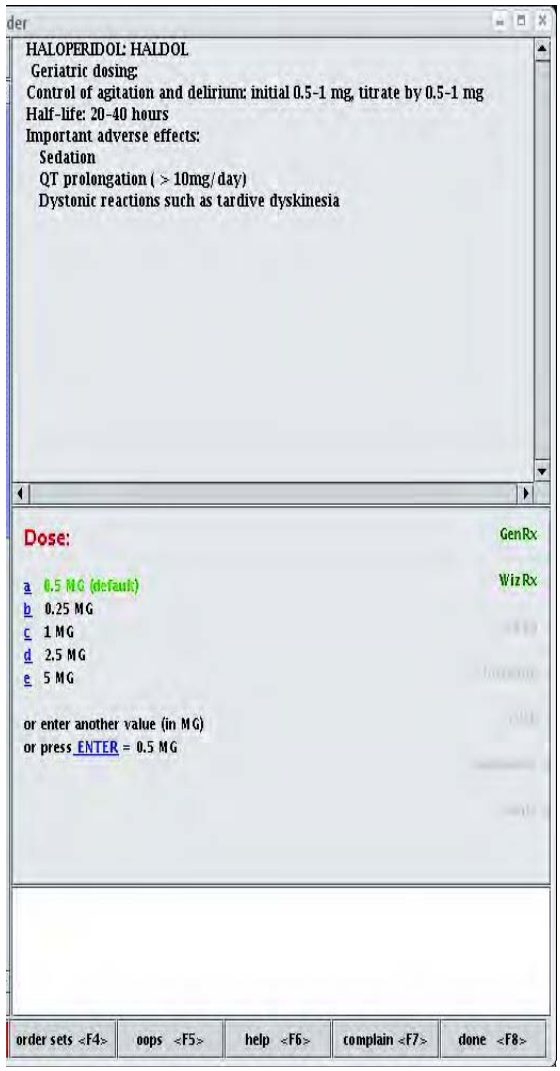


Figure 2 – Intervention integrated into institutional CPOE system

The institutional review board (IRB) approved the trial, including a waiver of consent for both patients and physicians.

Results

Over 9 months, 9111 study-related orders by 778 providers were entered for 2981 patients. Among the 88 study medications in the knowledge base, 23 were never ordered; many of these drugs were members of the Beers criteria list of potentially inappropriate medications and were not available from the hospital pharmacy. The overall acceptance rate of recommended doses was 28.6% in the intervention group vs. 24.1% in the control group, $p<0.001$). *Table 1* lists the median doses for all orders and predefined subgroups expressed as a ratio between the pre-

scribed dose and recommended dose. Overall, the intervention group of patients received lower doses than the control group (median 2.5, IQR [1.0,4.0] vs median 3.0, IQR [1.5,5.0], p<0.001). The difference comprised predominantly reduced doses of single-dose prescriptions, muscle relaxants, and slightly reduced benzodiazepine and anti-emetic prescription. The calculated 5th and 95th percentile dose ratios were 0.7 and 15 in the control group and 0.5 and 15 in the intervention group suggesting few differences at the extremes of the distribution.

In order to assess the effect of physician crossover, where physicians cared for both control and intervention patients, we compared the prescribed dosing from physicians who only cared for control patients (n=117) to physicians who only cared for intervention patients (n=103). Intervention-only physicians prescribed a significantly lower dose than control-only physicians (median 2.0 [1.0,4.0] vs median 4.0 [2.0,6.0], p<0001). The potential for crossover was also assessed by comparing a pre-trial period of 2 months (2315 medication orders) to the trial period. Pre-trial dosing was significantly higher with median 3.0 (2.0, 6.0) vs. median 3.0 (1.0, 5.0) for the trial period, p<0.001.

Beers criteria medications were prescribed in 34% of intervention prescriptions and 33% of control prescriptions. Dosing in the intervention group for medications in the Beers criteria list was not different than in the control group (Table 1).

*Table 1 – Median ratio of prescribed to recommended doses for control and intervention patients** IQR = inter-quartile range

Category	**Ratio of prescribed dose to recommended dose (median [IQR*])**		
	N	**Intervention**	**Control**
All Orders	9111	2.5 [1.0 , 4.0]	3.0 [1.5 , 5.0]
Drug Class			
Antihistamine/anti-emetic	2311	4.0 [2.0 , 4.0]	4.0 [2.0 , 6.0]
Benzodiazepines	2645	2.0 [1.0 , 4.0]	2.5 [1.2 , 4.2]
Neuroleptics	1473	4.0 [1.0 , 10]	4.0 [1.0 , 10]
Antihypertensives	1050	2.0 [1.0 , 4.0]	2.0 [1.0 , 4.0]
NSAIDs	442	4.0 [1.5 , 4.0]	4.0 [2.0 , 4.0]
Antispasmodics	292	2.0 [1.0 , 4.0]	3.0 [1.1 , 6.0]
Opiates	297	1.0 [0.5 , 1.5]	1.0 [0.4 , 1.5]
Sulfonylureas	305	4.0 [2.0 , 6.5]	4.0 [2.0 , 8.0]
Other anticholinergic	198	2.5 [2.0 , 5.0]	2.5 [1.0 , 5.0]
Other	98	1.0 [1.0 , 1.6]	1.3 [1.0 , 2.0]
Beers criteria medications	3051	2.0 [1.0 , 4.0]	2.0 [1.0 , 4.0]
Order Types			
Scheduled	5619	2.0 [1.0 , 4.0]	2.0 [1.0 , 4.0]
PRN	3492	4.0 [3.0 , 6.0]	4.0 [3.0 , 7.5]
Single dose	2580	1.0 [1.0 , 2.0]	1.25 [1.0 , 2.0]
Multiple dose	6531	4.0 [2.0 , 6.0]	4.0 [2.0 , 6.0]
Patient location			
Non-critical care unit	5028	2.5 [1.0 , 4.0]	3.0 [1.3 , 5.0]
Critical care unit and Procedure Suites	2463	3.0 [1.5 , 6.0]	3.0 [2.0 , 6.0]
Emergency Room	1338	2.0 [1.0 , 4.0]	2.0 [1.0 , 4.0]
Subacute unit	279	3.0 [1.5 , 6.0]	4.0 [2.0 , 4.0]

* IQR = inter-quartile range
Beers criteria medications are potentially inappropriate or with significant dose limitations as proposed by a consensus panel of geriatric experts

Discussion

This randomized trial demonstrated that physicians receiving advice from a guided geriatric dose advisor selected recommended doses for a minority of drug orders. Nevertheless, the guided dose advisor had a modest positive overall effect in decreasing the variation from recommended dosages for elderly patients. A low acceptance rate may be explained, in part, by the diverse indications for sedatives and neuroleptics in hospitalized patients. For example, benzodiazepines are used for both low acuity problems such as sleep as well as urgent indications such as combative behavior, pre-procedure sedation and acute alcohol withdrawal. Dosing may legitimately be higher for patients who have demonstrated tolerance or have chronically been taking a higher dose prior to admission. To improve physician acceptance and adherence to dosing recommendation, dosing guidance will have to cover a broader range of indications and design methods of capturing indication from physician users in order to route the user to a relevant dosing pathway. Single-doses were lower in the intervention group which may reflect a greater use of recommendations for "test-doses" which assess a geriatric patient's tolerance to a drug.

The differences between control and intervention groups were significant but of low magnitude. One explanation may be a "crossover" effect where physicians' learned of recommended dosing from treating intervention patients and made similar choices with control patients. Crossover was suspected because pre-trial dosing was higher than the control arm during the study, and "control- only" physicians prescribed doses significantly higher than "intervention-only" physicians. A crossover effect would bias the results towards the null. The trial was designed to randomize patients instead of providers because patient factors are most important when considering the primary patient outcome of the trial, falls. A crossover effect was considered unlikely during development because of the unobtrusive nature of the intervention. Potentially, the crossover effect represents evidence of physician learning. While this result will need more rigorous examination, if it is confirmed, clinical decision support systems may be serving an important teaching function in hospitals that implement CPOE.

The messages discouraging use of Beers criteria medications were generally ineffective with similar prescription rates in both control and intervention arms. Reasons for the lack of efficacy may include the lack of offered alternatives within the application or the inability for hospitalized patients to switch to preferred alternatives (e.g. due to allergies or intolerance). This outcome may also have been affected by the crossover effect mentioned above.

The interpretation of this trial was limited by several factors. Little information about the clinical context of the medication order was collected, and the overall appropriateness of the selected doses cannot be determined. There was no method to determine whether the physician making dosing decisions interacted directly with the guided dosing system. Some orders may have been dictated by proxy decision makers (such as senior residents or attendings) who were not directly exposed to the intervention. Finally, the influence of the specific CPOE system that hosted this project could not be determined. Compared to the CPOE system that hosted the previous implementation [2], the current system, has greater ease and fewer barriers to directly enter a higher dose which may encouraged physicians to bypass the dosing advice.

Conclusion

Guided medication decision support can affect geriatric dosing by modestly increasing agreement with guidelines. Higher levels of physician acceptance of recommendations will likely require greater use of indication-specific dosing advice.

Acknowledgments

Authors thank Osman Jalloh for his assistance with data collection. Dr. Peterson received support from the American Federation of Aging Research via the Merck/AFAR Junior Investigator Award in Geriatric Clinical Pharmacology. Drs. Miller, Peterson and Powers receive support from a Geriatrics Education grant from the D.W. Reynolds Foundation.

References

[1] Peterson JF, Bates DW. Automated selection of drugs and drug dose in patients with renal insufficiency. Medscape medication error and adverse drug reaction column 2002 March.

[2] Peterson JF, Kuperman GJ, Shek C, Patel M, Avorn J, Bates DW. Guided prescription of psychotropic medications for geriatric inpatients. Arch Intern Med 2005 April 11;165(7):802-7.

[3] Kuperman G, Anne Bobb, Bates DW. Medication-related clinical decision support in computerized provider order entry systems: a review. Journal of the American Geriatrics Society 2006.

[4] Waitman LR, Pearson D, Hargrove FR et al. Enhancing Computerized Provider Order Entry (CPOE) for neonatal intensive care. AMIA Annu Symp Proc 2003;1078.

[5] Chertow GM, Lee J, Kuperman GJ et al. Guided medication dosing for inpatients with renal insufficiency. JAMA 2001 December 12;286(22):2839-44.

[6] Evans RS, Pestotnik SL, Classen DC et al. A computer-assisted management program for antibiotics and other antiinfective agents. N Engl J Med 1998 January 22;338(4):232-8.

[7] Bates DW, Kuperman GJ, Wang S et al. Ten commandments for effective clinical decision support: making the practice of evidence-based medicine a reality. Journal of the American Medical Informatics Association 2003 November;10(6):523-30.

[8] Fick DM, Cooper JW, Wade WE, Walker JL, Maclean JR, Beers MH. Updating the beers criteria for potentially inappropriate medication use in older adults. Archives of Internal Medicine 2003;163:2716-24.

Address for correspondence

Josh F Peterson, MD
Center for Health Services Research and
Department of Biomedical Informatics
Vanderbilt University School of Medicine
Ph: (615) 936-1465
josh.peterson@vanderbilt.edu

Chapter 7.

Usability

This page intentionally left blank

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Graphical Overview and Navigation of Electronic Health Records in a Prototyping Environment Using Google Earth and openEHR Archetypes

Erik Sundvall[a], Mikael Nyström[a], Mattias Forss[a], Rong Chen[a], Håkan Petersson[a], Hans Åhlfeldt[a]

[a] *Department of Biomedical Engineering, Linköpings universitet, Sweden*

Abstract

This paper describes selected earlier approaches to graphically relating events to each other and to time; some new combinations are also suggested. These are then combined into a unified prototyping environment for visualization and navigation of electronic health records. Google Earth (GE) is used for handling display and interaction of clinical information stored using openEHR data structures and 'archetypes'. The strength of the approach comes from GE's sophisticated handling of detail levels, from coarse overviews to fine-grained details that has been combined with linear, polar and region-based views of clinical events related to time. The system should be easy to learn since all the visualization styles can use the same navigation.

The structured and multifaceted approach to handling time that is possible with archetyped openEHR data lends itself well to visualizing and integration with openEHR components is provided in the environment.

Keywords:

information visualization; overviews; timelines; openEHR archetypes; medical informatics; medical records systems, computerized.

Introduction

Relating events in time plays an important role in getting an overview of a record. This paper first briefly provides an overview to and exemplifies some previous approaches to represent relationships between events and time. It then describes how Google Earth (GE) by simple means can be extended to become a prototyping environment capable of combining and extending the previous visualizations in a unified environment. New combinations and the use of time-lapse animation are also presented.

Currently some elementary script programming and XML authoring skills are needed to create the described visualizations, but we are extending the GE based environment with more tools and examples that will make it even easier to use by non-programmers. We have created Java based components that work 'behind the scenes' to provide easier translation from events in electronic health records (EHRs) to 2- or 3-dimensional space with optional time-based animation capabilities in the GE-based environment. Aggregation and summary functions are also handled by the components. The toolkit has a built in integration with openEHR[1] based data using the archetype approach to modeling, but could be connected to other data sources as well.

Towards end users this approach tries to exploit the ease of use in GE so that the user only needs to learn one way of navigating to use several different kinds of visualizations. Some functions are also provided to ease usability testing of developed visualizations. We are publishing our approach at an early stage, hoping to broaden the research community using it for future stages of further visualization development and usability testing. Color images and demonstrations illustrating the previous approaches and our prototyping environment can be found by looking up references, footnotes and the corresponding author's webpage.

Background

In the research field of Information Visualization, time series data have been of interest a long time. A number of approaches are exemplified below.

Linear time views

Improvements of overview and navigation of EHRs have been reported, e.g. by using graphical timelines, in the often cited LifeLines [1] project. The events visible on the timeline had short labels, and longer labels were revealed by pointing the mouse cursor at the event. The timeline view allowed visual correlation between events and also acted as a 'giant menu' that improved navigation since events could be 'clicked' to get direct access to detailed information in the EHR. Information categories, such as notes, medications, lab tests, were called *facets*[2] and were displayed as horizontal ribbons containing associated events. The user could control which facets should be open (showing events), or closed (showing a compressed 'silhouette' of the contained events without labels). Bade et al [2] explored visual timelines further and describe methods to show alerting qualitative levels in streams of quantitative data in compact ways. They also present methods for showing uncertainty of timepoint and value, trustability of data and periods of missing data. Further they suggest a method for interactive timeline distortion in order to decrease the screen space used by less interesting periods of time. Data from an intensive care unit was used to exemplify their approach and they included methods to compact high frequency data into efficient visual forms.

1 http://www.openehr.org/

2 The term facet will be used in the rest of the paper for information categories etc.

Polar time views

Time can also be represented in polar coordinates. One way is to do as in most 'analog' clocks where time runs clockwise around the circle. This has been explored in tulip plots[3], where different facets are represented as concentric arcs along the time circle. In order to investigate periodic patterns in time-series data, Carlis [3] and others have used spiral timelines where periodical data patterns surface visually if the period for a turn around the spiral corresponds to the periodicity in the data (e.g. yearly cycles of allergy symptoms).

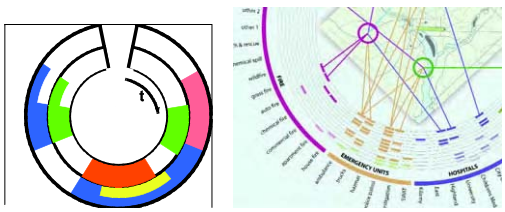


*Figure 1- **a** (left) A tulip plot[3] shows time clockwise and shows facets as concentric arcs. In Figure **b** (right) from [4] Livnat et al. put time radially outward and facets around the circle.*

Livnat et al. [4] use a polar coordinate system in an alternative way by putting the time axis radially outwards and distributing facets around the circle. Many polar representations leave space in the middle in order to avoid clutter. That space can be used to convey additional information. Livnat et al. put maps or network charts in the center and then relate the events around the circle to them using lines.

Time overlaid on maps and images

Kapler & Wright [5] describe a combined temporal and geospatial display, GeoTime, by putting a time (z-)axis perpendicularly upwards from a more or less flat (xy-)map. Events are shown further away from the map surface the further away they are in time. Kapler et al. also describe methods for aggregation of items based on proximity or defined regions.

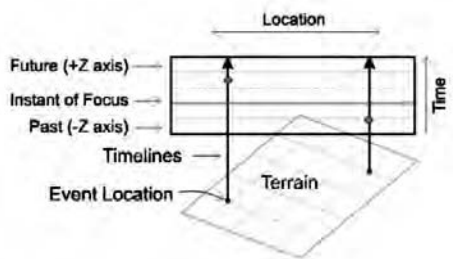


Figure 2 – GeoTime from [5]

Region based data entry

In Clinergy/Pen&Pad[4] [6] maps or charts of the body and organ systems are used as one way to specify parameters for data entry into an EHR. The user can point to coarse or detailed regions on the body charts and zoom in further to more detailed areas. The symptoms etc. available change depending on active region.

Visualization dimensions and variables

In [7] (section 4.3) parameters to use for visualization are summarized by Andrienko as being *dimensional* (referring to position in space and time plus various arrangements of display space) or *retinal* (size, shape, color, texture, orientation etc.) A recommendation is given that referrers of the dataset should be represented by dimensions and that attributes of the referrers should be represented by retinal variables.

Materials and methods

Archetypes

Archetypes are promoted by e.g. the openEHR foundation and the standards body CEN[5] as formalism to support modeling of clinical information structures and improved semantic interoperability between EHR systems. A detailed description of archetypes is outside the scope of this paper, see [8] for details. For the purpose of discussion they can be thought of as design descriptions of limited parts of the information structure in EHRs. Many such parts are then combined into hierarchies that constitute an EHR.

There are different aspects of time in a healthcare setting and for the purpose of recording they need to be treated carefully. The openEHR foundation has developed specifications for EHR systems and the topic of 'time' is extensively discussed in [9].

The information generated (e.g. using several *observation* archetypes) for a care event is collected into a *composition* that can be signed by clinical staff and put under version control. They can be flagged as being either *event compositions* that are intended for recording care events, or *persistent compositions* used for items of long term interest such as medication list, vaccination history, allergies etc. Compositions can contain entries of different kinds referring to events in the past (symptoms started a year ago) present or future (planned actions). In [10] *history* structures for *point events* and *interval events*, (possibly periodic) are defined. Methods for structuring e.g. long timeseries of measurements into compact summarized interval forms are described; this interval approach has some interesting similarities with summary approaches in [2].

3 Inspired by http://www.cas.lancs.ac.uk/alcd/visual/tulip_plots.html that refers to Barry, J.T. et al (1990) Graphical exploration of work history data. *Quad. Statist. Mat. Appl. Sci. Econ Sociali*., **12**, 65-74.

4 A Clinergy demo can be downloaded at http://www.opengalen.org/sources/software.html

5 EN13606, 'EHRcom', see http://www.centc251.org/

Google Earth

Google Earth[6] (GE) is an application that enables navigation of a digital globe with satellite imagery and map data. Developers can add new content in the form of XML files and images that can either be locally loaded into GE as files or published on a web server. If a server is used, then dynamic features such as reporting the current view of the user or updating previously loaded information are available.

Placemarks are icons with an optional text label that can be placed anywhere in a GE map. They can be clamped to the ground or put at an arbitrary height above the ground. When single clicking a placemark, a detailed description in the form of a 'text balloon' containing e.g. HTML-formatted text, possibly including images and hyperlinks, shows up. Double-clicking a placemark can change the 'camera' view of GE to a suitable position stored in the placemark by the placemark author. Hyperlinks in placemark description can be followed to related information that can either be a web page that opens in a built in or external web browser, or another set of GE objects (more images, placemarks etc.) that opens within the main GE navigation view. In this environment, these features can be used to go deeper and deeper into the EHR content. Transitions to the stored camera view in a placemark can be used as an efficient way to zoom in and position the view to a suitable angle that fits the next level of detail used in the EHR visualization.

Results

By placing an arbitrary image to be used as background over a piece of the globe we can hide the map and use that area as a 'desktop' for visualizations unrelated to the original map content. Initial tests show that most of the system features described in the introduction chapter can be reproduced in GE based visualizations. Linear and polar timelines can be added as images or drawn using built in GE geometric objects on top of the desktop.

We have also experimented with placement of notes that have relation to body parts (finding site of a tumor etc.) on organ system charts inspired by Clinergy. The notes are placed as placemarks above the relevant anatomical part in the chart and time is (as in [5]) represented by distance from the map. Just like real objects in a pile, old notes are at the bottom close to the map, and newer information is stacked above. By using the GE feature of region based loading and display of details, controlled by 'level of detail' (LOD) settings, different versions of a visualization piece can be shown depending on how close the user zooms in. If there are many notes in one region, they can be summarized into a single node indicating the number of notes when viewed at a distance, but shown as individual notes when zooming in. This has an effect similar to the region based aggregation discussed by Kapler et al. [5].

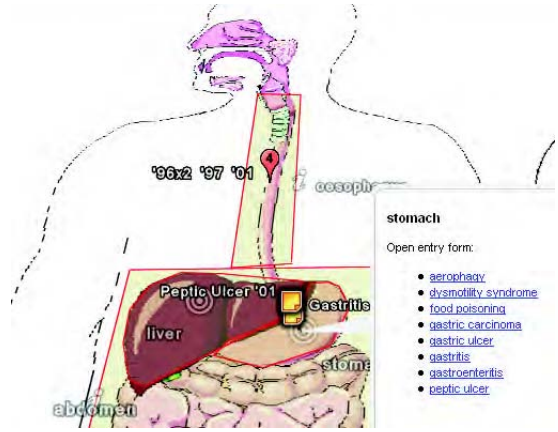


Figure 3 - A screenshot from Google Earth showing a tilted region based visualization in the prototyping environment. The abdomen region has been zoomed in so that subregions are visible. Two previous notes relating to stomach are seen, with the most recent one on top. The oesophagus region is too far away to show subregions and indicates that it contains four previous notes. A stomach related entry form selector has been opened by clicking the stomach 'target'.

Since it is possible to open web pages within GE, placemarks or information about current view can be used to select and open web based entry forms. This allows us to facilitate display and 'Clinergy-style input' in the same view. The usability of this combination remains to be tested. Maybe it is better to switch between dedicated viewing mode and a dedicated entry mode. Two-dimensional data like X-ray images or ECG plots are easy to show directly in GE. By using LOD, images with higher resolution can be fetched incrementally and shown when zooming closer to a low resolution thumbnail image.

If recent time is of most interest, then logarithmic time scaling can be used both for linear and polar diagrams. Related effects can also be achieved by simply tilting the view in GE when looking from recent time backwards. We have not yet attempted to develop ways to distort several different regions of a timeline simultaneously as in [2]. It would be interesting to compare this to the usability and understandability of regular GE-based navigation. Separate parallel timelines can be used for documentation events (versions of compositions) and clinical event timing (e.g. observations) if visual separation is desired. Important information (Patient ID, allergy warnings etc.) can be shown e.g. in a corner by using 'screen overlays' that are fixed to the screen and not to the map surface.

Facets and aggregation

How to section categories of information into facets like LifeLines [1] did with medications, lab tests etc. is an interesting research problem. In a (potentially) highly structured EHR e.g. based on archetypes we see these fairly simple computable divisions, based on:

6 http://earth.google.com/

- archetypes used, based on groups manually listed when creating the visualization, or by archetype class (observation, evaluation, instruction etc.) or external ontologies and systems [11] for classifying archetypes.
- openEHRs 'folder' structure that e.g. can be used for linking entries to health problems, care episodes etc.
- terminology system entities used in EHR entries e.g. ICD diagnosis codes or codes from SNOMED CT's main categories and/or grouping according to the target of relationships like 'finding site'.
- provider of the entry (organization, role or profession, person).

Aggregation that is needed e.g. in zoomed out views can be based on facets and/or time. In some cases (e.g. in the openEHR class 'History') some events may already have been manually summarized at the time of data entry and this could of course be used. Another obvious aggregation is over 'natural' time chunks like 'the number of entries per year'. In region based visualizations (body part maps etc.) geo-located aggregation based of the containment hierarchy of regions can be used. Better suited facets and aggregation can potentially be generated by decision rules and automated reasoning taking the above mentioned aspects and the current role and task of the user into account.

Time as a fourth dimension

In GE (from version 4) time is available as a fourth dimension. Objects (placemarks etc.) can be given time spans for validity. A time axis with sliders allows the user to select time span restrictions of what should be visible. By pressing the 'play button' the sliders can also be moved automatically which results in a time-lapse animation of changes over time. One way this can be used is to easily view previous states of the EHR in order to see how the medical picture grows over time. We want to explore if and how this can be used efficiently. We believe that a time-based view e.g. for the body/organ-system map where you can see information about problems etc. appearing (and possibly disappearing) over time may serve as a complement to using the z-axis to represent time. For medico-legal purposes it may also be useful to easily browse earlier states of the record. Clinicians will hopefully use the toolkit to find many other creative views and usages for time animations of EHR data – perhaps for studying the progress rate of diabetes complications etc. Users' changes of 'camera view' can be time-stamped and recorded. This information can be used e.g. in usability studies to record where most time has been spent etc. Time-lapse replay of frames showing the view trail in different ways is possible. Another possible use of recording can be to highlight for users which views in an EHR that have been most visited by other users—analogous to which pages of large paper based record that carry signs of frequent reading. Recording of view trails can also be used for medico-legal logging purposes.

Intended use

We do not expect that clinicians will use the toolkit in every day use; rather it can be used by them and others as a prototyping tool to invent and explore designs that can be used as parts of descriptions when ordering more polished and specialized systems from system providers.

Existing clinical images, either literally used (e.g. growth curves and partographs/partograms) or conceptually used (e.g. care-flows) can be used as a basis for fast prototyping of overviews provided that they can be captured as images.

We are aiming at decreasing the programming knowledge required to create visualizations. A user will be able to use GUI features to select the kind of visualization they want, e.g. linear, polar, region based (Clinergy inspired) and which axis to use for what (time, facets etc.), start and end points, width/radius etc. Connecting to openEHR data sources and selecting what data to fetch currently requires some scripting[7], but we are investigating the possibilities to select nodes using our Archetype Editor [12] or archetype visualization and browsing tools. A use case scenario could be as follows.

An overview based on a sketch of a tree of significant blood vessels and a linear timeline diagram for lab values is sketched on a whiteboard during a meeting. It is then captured by camera and the pictures are transferred into the GE based environment and anchored to the desktop on the map next to some previously developed overviews. Hierarchical regions containing each other are drawn and named on the 'vessel tree' and then mapped to entities in the archetype or terminology used in the EHR. For the lab result timeline the start- and endpoints on the time (x-)axis are marked in the image. Then the different lab value fields are mapped to positions for intervals and marked on the facet (y-)axis forming (possibly overlapping) ribbons along the diagram. Finally, color, icons, aggregation and summary strategies for the lab value plots are chosen.

Discussion

The technical solutions behind these extensions for GE are not very complex[8], instead it is more the possibilities for rapid development and evaluation that are of interest. Will the solutions be easy and efficient enough to be used by clinicians? If so, how will that affect the future development of interaction with EHR data? Will more efficient overviews and interaction possibilities be put into real clinical every-day use and would that have an impact on medical practice? Early feedback from clinicians and medical informaticians that have seen the current prototyping system has been positive and has resulted in comments like: "It opens up some new ways of thinking", "This would be nice when seeing new patients in primary care, but less useful for patents I already know well.", "[It] can change your cognitive level of interaction with the information" and "Information that is subtle or hidden beneath the surface might be found easier".

Future work

The prototyping environment presented in this paper is a prerequisite for planned user studies. Usability aspects of the visualization creation process will be one of the first

7 http://freemarker.sourceforge.net/

8 The approach can easily be reproduced by others. We intend to release a version of our solution as Open Source.

aspects to evaluate. Another aspect to study is the usability for end users of different created visualizations.

This paper focused on the 'dimensional' aspects of visualization – where to place information entities. How to create concise and efficient text labels has not been focused, neither have the 'retinal' parameters [7] (color, shape etc.). Systematic selection and use of icons and pictograms has not been the focus of our studies yet. Adding a structured approach to icon use along the lines of VCM by Lamy et al. [13] looks promising.

Our focus has been on viewing EHRs of individual patients one at a time, but the visualization principles would be applicable to studies of groups of patients as well.

The normal use of GE is to relate information to geographic positions, here we have not discussed this, but future work could include relating entry and retrieval of EHR content related to positions. In, e.g., ambulatory or distributed care EHR content and maps could be accessed in the same environment. We have started a location related experiment by using a map of a hospital ward as background image in GE and then placed the EHR visualizations for patients in their respective rooms.

If visualizations in GE are flat and the tilt function is not used, then the environment essentially becomes a zoomable 2D interface. Even though 3D visualizations often are more appealing Chen [14] (section 6.6) summarizes several 2D vs. 3D studies and concludes that increasing an interface from 2D to 3D is unlikely to improve the users task performance 'unless additional functions are provided so that users can have greater controls of objects in 3D interfaces'. Hence comparative studies of visualizations dependent on 2D and 3D respectively should be preformed. We believe that access to 'multi touch'[9] interfaces or hardware[10] dedicated to 3D interaction can affect user performance in 3D.

Conclusion

The capability and usability of geographical information systems of today like GE combined with the push for more structured and semantically well defined EHRs can in combination be used to create a powerful environment for prototyping overviews and interaction style for EHR systems.

We have summarized and unified approaches that may be used to create visualizations of temporal, casual and possibly anatomical relationship between events.

Acknowledgments

This work was performed in the framework of the EU-funded Network of Excellence entitled Semantic Interoperability and Data Mining in Medicine (SemanticMining). We want to thank SemanticMining network members, the OpenEHR community and Isabella Scandurra (Centre for eHealth, Uppsala University) for feedback during our work, Ocean Informatics for providing test data from their EHR Bank and the authors of [4] and [5] for permission to reuse illustrations.

9 E.g. http://cs.nyu.edu/~jhan/ftirtouch/

10 E.g. SpaceNavigator by http://www.3dconnexion.com/

References[11]

[1] Plaisant C, Mushlin R, Snyder A, Li J, Heller D, Shneiderman B. LifeLines: using visualization to enhance navigation and analysis of patient records. Proc AMIA Symp. 1998:76-80.

[2] Bade R, Schlechtweg S, Miksch S. Connecting time-oriented data and information to a coherent interactive visualization. CHI '04: Proceedings of the SIGCHI conference on Human factors in computing systems; 2004: ACM; 2004. p. 105-12.

[3] Carlis J, Konstan J. Interactive visualization of serial periodic data. UIST '98: Proceedings of the 11th annual ACM symposium on User interface software and technology; 1998: ACM; 1998. p. 29-38.

[4] Livnat Y, Agutter J, Moon S, Foresti S. Visual Correlation for Situational Awareness. INFOVIS '05: Proceedings of the 2005 IEEE Symposium on information Visualization; 2005: IEEE Computer Society; 2005. p. 13.

[5] Kapler T, Wright W. GeoTime information visualization. Information Visualization. 2005;4(2):136-46.

[6] Kirby J, Rector AL. The PEN&PAD data entry system: from prototype to practical system. Proc AMIA Annu Fall Symp 1996:709-13.

[7] Andrienko N, Andrienko G. Exploratory Analysis of Spatial and Temporal Data: A Systematic Approach: Springer 2005.

[8] Beale T. Archetypes: Constraint-based Domain Models for Futureproof Information Systems. OOPSLA 2002, workshop on behavioural semantics.

[9] Beale T, Heard S, Kalra D, Lloyd D. EHR Information Model; 2006.

[10] Beale T, Heard S, Kalra D, Lloyd D. Data Structures Information Model; 2006.

[11] Garde S, Knaup P, Hovenga EJS, Heard S (2007): Towards Semantic Interoperability for Electronic Health Records: Domain Knowledge Governance for openEHR Archetypes. Accepted for publication in Methods of Information in Medicine. http://healthinformatics.cqu.edu.au/downloads/Garde_openEHR_Methods_2007.pdf Accessed 2006-12-04.

[12] Sundvall E, Qamar R, Nyström M, Forss M, Petersson H, Åhlfeldt H, Rector A. Integration of Tools for Bind-ing Archetypes to SNOMED CT. In: Schulz S, Klein G, editors. Semantic Mining Conference on SNOMED CT; 2006 October 1 - 3; Copenhagen, Denmark; 2006.

[13] Lamy JB, Duclos C, Rialle V, Venot A. Which Graphical Approaches should be Used to Represent Medical Knowledge? Stud Health Technol Inform. 2005;116:719-24

[14] Chen C. Information Visualization: Springer 2004.

Address for correspondence

Erik Sundvall, Department of Biomedical Engineering, Linköping University, SE-581 85 Linköping, Sweden.
erisu@imt.liu.se,
http://www.imt.liu.se/~erisu/ (demo & images available)

11 Web links to references available at http://www.imt.liu.se/~erisu/

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Which Parts of a Clinical Process EPR Needs Special Configuration

Anders Barlach, Jesper Simonsen

Department of Communication, Business and Information Technologies, Roskilde University, Denmark

Abstract

Subject: *Which parts of an electronic patient record (EPR) can initially form a stable standard solution to be used by all clinicians? And which parts of an EPR can we predict needs initial as well as on-going re-configuration to meet the needs from diverse medical specialties.*

Purpose: *To analyze which screen types in a clinical process that can be standard configured and which are subject to initial as well as on-going re-configuration.*

Methods and results*: A pilot-project implementing a fully functional clinical process EPR was configured and used at a neurological ward, replacing all paper records 24/7. The analysis characterizes the different types of screens, a total of 243 included in the EPR solution. All screens have been extracted from the application and analyzed for changes in total 222 changes.*

Discussion and conclusion: *Most screens (87%) are very stable. Few (13%) are subjected to several re-configurations and they stabilize after an average of six iterations: Some may further stabilize over time since they address new but also general ways of working. Other screens relate to the specific medical specialty and cannot be part of a standard solution.*

Keywords:

integrated advanced information management systems, software design, interprofessional relations, user-computer interface, problem-oriented medical records

Introduction

Clinical Process form a core module of an Electronic Patient Record (EPR) that supports clinical documentation and decision making and comprises the on-going documentation of medical patient information made by the clinical staff (physicians, nurses, therapists, medical secretaries, etc.). It is hard to imagine that one single standard configuration would be optimal for all clinicians throughout the hospital: The clinicians at the neurological ward needs patient overviews focusing at parameters addressing cerebral haemorrhage and cerebral stroke while the clinicians from for example psychiatry needs completely different data and information in their patient overviews. Much basic patient data however (e.g. address information, family relations, drug profile, previous diagnosis, etc.) as well as functionality for common data entries (e.g. temperature, blood pressure, drug prescriptions, etc.) might be presented in a uniform way through standard screens used by all clinicians regardless of their medical specialties.

Modern EPR platforms today support international health standards (e.g. HL7), terminologies (e.g. SNOMED CT), and database platforms (e.g. Oracle HTB) while at the same time offering a high degree of configurability by means of e.g. XML-based templates. EPR technologies has reached a level where EPRs can be developed as a standard tool for all clinicians while the EPR at the same time can be configured to serve as a tool customized for specific needs supporting the high level of specialization that each clinical specialty represents. A main question that hospital managers and EPR developers face is thus the research question of this article: *Which Parts of a Clinical Process EPR Needs Special Configuration?*

The article is based on a project where a fully functional clinical process EPR was configured and used at a neurological ward, replacing all paper records 24 hours a day throughout a pilot lasting one week. We have analyzed how each of the EPRs 243 screens was configured and reconfigured to meet the needs of the clinicians. Based on this study we are able to indicate: 1) Which parts (screens) of an EPR that initially might be developed as a stable standard solution and used by all clinicians throughout many hospitals (no or few further re-configurations are to be expected when the EPR is rolled out to other wards); and 2) which parts of an EPR we can predict needs initial as well as on-going re-configuration to meet the needs from clinicians representing diverse medical specialties (e.g. neurology, surgery, gynaecology, psychiatry, etc.)?

In the following we introduce the study and the method used for the analysis. Then we present the results in terms of our categorization of screen types and the completed as well as requested re-configurations for each type. Finally we discuss and conclude our findings, the limitations of the study and the implications for practice and research.

The clinical process EPR project

The project was part of a research project on effects-driven IT development [1] (http://Effects-DrivenIT.dk) and was formed by 3 partners: Clinicians from the neurological stroke unit and project managers from the EPR unit at Roskilde County Hospital, researchers from the Department of Communication, Business and Information Technologies at Roskilde University and business architects from the vendor, CSC Scandihealth A/S. One main aim of the project was to experience how to configure a clinical process EPR module in participation with clinicians and to test how a configured solution would work in a real clinical process.

The project involved a neurological stroke unit treating patients with acute apoplexy where all paper-based patient

records were replaced with a configured and fully functional prototype EPR system. The aim was to evaluate an EPR with complete patient records tested on-line on real clinical processes [2, 3]. The project thus required thorough planning involving development of new EPR-supported patient trajectories, configuration and implementation of all screens needed in the EPR system, real-time integration with other systems, migration of patient data, and training of the clinical staff in using the system and working according to the revised patient trajectories.

The content of the EPR was identified during three workshops, i.e. the structure, content and placement of clinical notes and result templates, standard plans, concept lists etc. At the final workshop the complete specification was presented and reviewed before the actual configuration of the XML-based templates and load of the templates to the EPR. During this process the content of the EPR was elaborated in up to three iterative events. First, mock-ups were drawn on flip-over paper. Secondly, a preliminary non-interactive prototype was discussed. Finally, a running prototype was demonstrated, discussed, and evaluated. The vendor undertook the technical development of the prototype, along with interfaces to various legacy systems currently used at the hospital (ADT system, laboratory system, and medication module). A number of tests and re-configurations of the system were made in parallel with training the clinical staff in using the prototype. A final rehearsal was performed by testing the system under laboratory conditions using real patient-cases in a scenario setup on the solution that was due for release in the pilot. This was the final reassurance within the project team that the EPR was ready.

The most complicated part of the screens concerned those that should provide the clinical staff the ability to efficiently obtain overview and assessment of patients as well as on more efficient coordination in three specific and highly cooperative situations:

Nursing handover, which happens three times a day at the beginning of each nursing shift (7am, 3pm, and 11pm) and last about an hour. There is no time for the nurses that leave the ward to discuss patients with the nurses on the next shift. During the nursing handover, one nurse is designated as the team leader and provides an overview of the patients at the ward and manages the necessary coordination and exchange of information. This nurse reviews the patient records and orally informs the others about status and plans for the shift.

Team conference, which takes place once every weekday, lasts approximately 15 minutes, and includes all clinical staff members (physicians, nurses, and therapists). An inter-disciplinary assessment of each patient is carried out and plans are revised. The current status of each patient is given orally by a nurse and an overview of current plans is available by means of a table on a large whiteboard or, in the prototype EPR system, a full screen projected on the wall.

Medical ward round, which happens once every weekday and lasts for three to six hours. It includes evaluation, reviewing, and discharging of patients. The chief physician visits all patients and reviews the plans for their treatment. Usually there is no time for nurses to follow the physician during the ward round. Information exchange and coordination is obtained through the patient record and by ad hoc communication with the nurses on shift.

The required content was configured as XML-based templates that were loaded into the clinical framework tool, CSC Clinical Suite, based on the Oracle Healthcare Transaction Base (HTB). CSC Clinical Suite is not an EPR per se, but a clinical framework tool that can contain and present the clinical content as specified by the clinicians by use of XML-based tem-plates for overviews, clinical notes, results, standard plans, work situations and structure of the patients medical record. This makes it possible to configure a complete medical record in accordance with the clinicians requirements and is able to evolve dynamically as new requirements emerge.

In the final part of the project (the pilot), the configured EPR system was online 24 hours a day and replaced the paper-based records for all patients during one week in December 2005. Five years of patient data (in total more than 26 million data records from more than 300.000 patients) had been migrated to the EPR system and interfaces were established to the legacy systems in order to receive updated data during the project. The EPR system included screens projected on the wall during nursing handovers and team conferences, stationary and portable PCs, and PDAs used for obtaining measurements at the patients bedside (temperature, blood pressure, etc.). All clinicians used the EPR system during the pilot. Management oversaw to the project ensuring both legal requirements and patient-ethics were respected.

Data analysis method

All screens in the EPR solution 243 in total have been extracted from the application and analyzed for changes 222 in total made in the project period. The analysis is based on the vendors systematic documentation of all the changes made to each screen, from an initial first version of a screen and throughout the project period including the pilot where the system was used 24 hours a day.

In order to analyse the screens they have been categorized as follows. The screens have been divided into general and specific screens. A general screen can serve the same purpose on any medical ward: E.g. screens for recording basic vital values such as blood pressure, pulse and temperature might be the same on a medical and a surgical ward. Specific screens serve a special purpose within the given clinical speciality: E.g. screens for recording and monitoring a SIP score (Stroke In Progress) are specific to the Neurology speciality.

All screens in the EPR system (as well as in information systems in general) can be divided into two different categories, as either a form or a view:

- Form, resembling a paper form for *recording* (registration and submitting) data. This can be free-text or structured information in various degrees. Typical forms could be observations, notes, and basic vital values.
- View, is the *presentation* of data either recorded in the EPR system or received from external systems. A view retrieves data from one or several sources and presents it as information to the user or as an indexing service.

Typical information views could be: Graph presentation of basic vital signs, or overviews' creating information bulletin boards with focused information for a specific clinical situation (including e.g. nursing handover, team conference, and medical ward round). Views also include Journal structure views that present the user with all available data in a structure for navigation. This navigation hierarchy was designed to resemble that of the paper record.

In order to determine who made decisions regarding changes we assigned each view or form a primary user in terms of professional discipline (doctor, nurse, therapist or shared by doctor & nurse). This is to indicate the coordination involved among professional groups in the design and implementation process.

All changes analyzed were changes that were actually implemented. Several change request where also collected but not implemented because they were considered non-essential (nice-to-have as opposed to need-to-have) to the continued use of the system during the pilot. The changes made to the screens were analyzed with regard to when they occurred in the project (before, during or after completion of pilot). Types of changes include content (new fields in forms, new selections in views, labels changed); rules (business logic, validations); computations (adding or changing calculation functionality); and cancelled (retirement of screens due to time pressure or obsolescence due to other screens delivering similar services). The changes are summarized into 3 major groupings:

- None (0): No changes were necessary.
- Few and initial changes (1-2): One or two changes were made initially in the project during the prototyping process. These types of changes reflect a low complexity or uncertainties in design.
- Several and sustained changes (>2): More than two changes occurring including changes beyond the initial prototyping process. These changes reflect either uncertainty among clinicians or complexity in the implementations. It also reflect screens that needs to be configured by an *experimental* approach which entail several successive changes throughout the project, in some cases including changes made within the pilot period.

Based on the categories listed above all screens and changes were analysed and the resulting patterns are presented below.

Results

We have identified a number of interesting patterns with regard to the changes made to the screens representing the overall configuration (and re-configuration) of the clinical process EPR system. The implementation resulted in an EPR with a 4:1 ratio between forms and views. Less than 10% of the total 243 screens were specifically configured to the neurological specialty (16 out of 183 forms and 7 out of 60 views).

The majority of screens (87%) were not changed at all or only subject to few initial changes (table 1). Thus the major part of the total system may be considered as being quite stable. These stable screens were both medical specific forms (45%, 7 out of 16 totals) and general forms (90%, 152 out of 167 totals). Most of the stable forms were quite simple, in terms of containing only one or two data fields (e.g. registration of simple results like blood glucose) and often they were serving as a sub-template in larger and more complex forms. Views that present data from other systems were also very stable, e.g. views presenting X-ray results. Another characteristic for stable forms and views was that only one professional discipline was involved as main user, or the design was known from other systems as e.g. views presenting aggregated laboratory-results.

Distribution among screens changed

Total screens	None (0)	Few and initial (1-2)	Several and sustained (>2)
243	184	27	32
100%	76%	11%	13%

Table 1 - Changes made to the screens during the entire project.

The total number of changes accounts to 222. Out of these 83% (184 changes) were made to the 32 screens that received more than 2 changes each. This verifies that the configuration of 13% of the screens (32 out of 243) reflects a need for experimentation. These screens where subjected to a more thorough analysis and present interesting change patterns as seen below in table 2 and 3.

Screen change pattern specific vs. general

	Specific		General	
	Screens	Changes	Screens	Changes
Form	7	39	15	79
View	5	38	5	28
32	12		20	
184		77		107

Table 2 - Analysis of the 32 screens subjected to several and sustained changes (from table 1) distributed among screen requirements attributed to the specific neurology speciality or of a general clinical nature.

Screen change pattern among professional disciplines

	doctor	nurse	multi
Form	5	14	3
View	0	3	7
32	5	17	10

Table 3 - Analysis of the 32 screens to support either a professional discipline (doctor or nurse) or information collaboration among disciplines (multi).

Content changes where dominant in these patterns (82%, 184 out of 222 total changes), including adjusting labels on fields, adding new fields, removing obsolete fields, in some cases later to be added again. The need for experimentation grew according to the complexity of views, for example as more than one professional discipline was identified or data had to be drawn from several forms.

Motivations for changes in forms where often driven by their dependency to deliver data in the views. If for example a view is changed to include additional data this often entail that a form needs change in order to capture this data. Changing one view sometimes indirectly contributed to changes made in index-views that display a structure for navigating the various documentation models in the EPR. The scenario would typically be that each time a new view was available, it also had to be accessible without the search functionality, and this sometimes entailed that a logic entry or indexing in the Journal structure had to be assigned adding to changes accumulated by this index-view.

If we focus on the 32 screens from table 1 that were subject to several and sustained changes and display the results in table 2 and 3, we observe that the general forms are in majority to the general views (3:1). They are primarily owned by only one group of professionals (19 out of 22 have only one profession as primary owner). The nurses account for 11 out of 15 general forms shown in table 2 which also sparks attention to why they are represented with so relative many forms?

The views are equally distributed among specific and general, but are characterized by having more than one owner (7 out of 10, see table 3). Specific forms include forms for clinical plans with regard to stroke, which in the project constituted an entirely new way of applying their knowledge. These plans account for 4 out of 7 Specific Form screens shown in table 2, and 25 of 39 changes made to these screens.

Common factors contributing to changes among all the 32 screens has been identified as complex computations required on the client side, specific forms (e.g. Scandinavian Stroke Scale), or views (both specific and general) involving multiple professional disciplines where forms and views should support coordination of data or tasks. Especially when supporting an inter-disciplinary approach to EPR: E.g. complex views supporting the ward round or the team conference draw on information from radiology systems and clinical laboratory systems, and in addition including observations and notes made by doctors, nurses, and therapists.

The fact that the systems delivered 24/7 service during the pilot could entailed that only needed changes were implemented during the pilot (need-to-have changes as opposed to nice-to-have changes). During the pilot a few changes were deemed necessary in order to continue efficient operations. These changes occurred only to 3 views while the remaining 240 screens (99%) remained unchanged. Nevertheless the pilot and the use of the EPR in general were evaluated as being successful and measurements on clinical practice using the EPR has documented several significant improvements [1, 2, 3].

Discussion

The results indicate patterns of changes displaying themes that are predominant in the process of designing and implementing the clinical process EPR. Our study indicate that the majority of a clinical process EPR does not require special configuration with, regard to the different clinical specialties, as 87% of all screens in the EPR remained stable by requiring no or only few changes. The stable screens include simple forms, views presenting data from other known systems as well as forms and views addressing only one professional discipline. Content changes where dominant representing 82% of all changes. A substantial part of the changes is a result of chain reactions, typically where a change to a view subsequently trigger other changes in related forms or views for navigating the EPR. Screens with more than 2 content changes account for 143 out of 222 in total or 64% in only 29 screens. They were often related, e.g. FORM; Stroke Scale, Apoplexy Observations relate to VIEW; Apoplexy Overview AND FORM; Apoplexy Plan. Approximately 3 times as many forms were needed as views giving an idea of how many forms are required to sustain views.

A number of screens was subjected to several and sustained changes reflecting a need for an *experimental* approach to the process of configuring the EPR. The configuration of these parts of the system addresses application areas where the EPR introduce new ways of working. Potentially this might result in far-reaching improvements by ways of efficient support of inter-disciplinary coordination among multiple professional disciplines. We can predict that some parts of this configuration will stabilize over time since they address new but also *general* ways of working with EPR. Other parts of this configuration addresses themes related to the *specific* clinical specialty which indicate parts of EPR that can hardly be standard configured to serve clinicians throughout the hospital.

Forms supporting new ways of structuring documentation and views presenting the journal structure are examples of general parts of the EPR that faced several and sustained changes.

In our study the doctors applied their existing documentation model to the EPR and they retained dictating as usual with the medical secretary entering the dictate into the EPR. The nurses on the other hand, had to invent and specify their documentation model and integrate it with the doctors model in the journal structure (it was a deliberate part of the project to experiment with adding structure to the nurses documentation). This resulted in a higher activity regarding the design of forms with nurses as professional discipline (as indicated in table 3). Throughout the project this sparked several general discussions among the nurses about how they where using the paper based journal structure and how to use EPR. They could see a new perspective with the EPR, and the need to evolve their documentation models to include how they decode clinical data into nursing information. In general it was a challenged to figure out how to merge multiple documentation models serving their interdisciplinary needs without compromising their professional knowledge to accommodate other professionals. E.g. the doctors did not have to give up describing the paμtients anamnesis from the diag-

nostic perspective, just because the nurses would insist on describing the anamnesis from a holistic perspective.

Views presenting the journal structure support navigating the EPR and provide an alternative to the pre-defined information clustering implemented in the overviews. This also provides the users with the possibility to verify, in case of uncertainty, if they had missed some information in the overviews. They came to rely on the patient-journal structure for completeness. The upper levels in the journal structure must be general throughout the hospital and this we can be predict to become relatively stable over time (though this was not the case in our study where introducing clinical process EPR). However the lower parts patient-journal structure hierarchy might become more specialized and susceptible to changes thereby requiring occasionally experimentation and dynamic technological solutions for the medical specialties.

The parts of the clinical process EPR where an ongoing and experimental configuration can be identified as addressing the *specific* clinical specialty comprises support for highly cooperative activities such as planning the patient treatment and activities such as team conferences and nursing handovers.

Planning was the primary contributor to many changes in the doctors group. Planning account for 4 out of 7 specific forms having sustained changes and 25 of 39 changes made to the specific forms included among the top 32 screens listed in table 2. This was due to the fact that many of the planning and coordination tasks traditionally are handled by other professionals (nurses or secretaries). The story repeats itself as with the nurses lacking a documentation model, since the doctors had no prior system to rely on. Plans can overall be divided into 2 categories: The initiating or basic plan and the follow-up or supplementing plan. It was relatively easy to design the initiating plans as they to a high degree resemblance with the department guidelines. However the follow-up proved more difficult as they where often conditional (e.g. if X-ray result is positive order antibiotics) or involved coordination of tasks between professional disciplines or other medical specialities. This complexity is contributed to the innovation requirements of the professionals as they become aware of one-anothers areas of responsibilities and explore the possibilities of coordinating and sharing information in new ways.

The views supporting the coordinating activities during team conferences and nursing handovers were also subject to sustained changes. Although not many in numbers, they account to a significant number of changes: 33% of all changes listed in table 2 and 3. The changes were primarily content changes to views supporting interdisciplinary cooperation (team conference) or single disciplines being derived from an entirely new documentation model nursing observations.

Conclusion

The majority of screens (87%), were stable and include simple forms, views presenting data from other known systems as well as forms and views addressing only one professional discipline.

Relatively few screens (13%, or 32 out of 243) were subjected to several re-configurations and a part of these may further stabilize in the future since they address new but also general ways of working. Another part are screens specific to the clinical specialty. There are indications that only few specific screens are necessary per medical speciality.

The screens with sustained change requirements include both general and specific screens and comprise different types of views displaying the potential of an EPR: They present new ways of decoding and sharing information and supporting highly cooperative activities. These screens are characterized by the clinicians having no previous experience from a mainly paperbased everyday work environment, or clinicians involved in multi-disciplinary content and cooperative activity. Our project documents that such screens can be efficiently configured through an experimental and participative approach [4]. It is also clear that it requires continuing the experimental approach to include using the EPR in a real clinical everyday work environment. From the technological point of view it sets the standards for how the EPR vendors must be ready to meet the dynamic requirements and where to expect more confidence in the stability of the EPR.

The perspective of our study gives an indication as to what to expect when engaging in the implementation of a dynamic EPR. This paper present the result of just one pilot-test, and more tests are necessary to investigate the issues of the dynamic versus stable parts of a clinical process EPR. We are now applying our experience from the pilot to new projects where several medical specialities are involved; neurology, cardiology and paediatrics across three different hospitals.

References

[1] Hertzum M and Simonsen J. Effects-Driven IT Development: Specifying and Measuring Usage Effects during Systems Development (forthcoming).

[2] Mller-Jensen J, Pedersen IL and Simonsen J. Measurement of the Clinical Usability of a Configurable EHR. In Hasman A et al. eds. Ubiquity: Technologies for Better Health in Aging Societies, Proceedings of the 20th International Congress of the European Federation for Medical Informatics (MIE 2006), Maastricht, the Netherlands, August 27-30. Amsterdam: IOS Press, 2006; pp. 356-361.

[3] Mller-Jensen J, Simonsen J, and Iversen RK. Measuring Effects on the Clinical Practice from a Configured EHR. In: Hejlesen O et al., eds. Proceedings of the 4th Scandinavian conference on Health Informatics (SHI 2006), Aalborg University, Aalborg, August 24-25. Aalborg: Virtual Centre for Health Informatics, 2006; pp. 58-62.

[4] Simonsen J and Hertzum M. A Regional PD Strategy for EPR Systems: Evidence-Based IT Development. In Jacucci G et al. eds. Expanding Boundaries in Design, Proceedings of the ninth biannual Participatory Design Conference, Vol. II (PDC 2006), Trento, Italy, August 1-5. Palo Alto: CPSR, 2006; pp. 125-128.

Address for correspondence

Anders Barlach
Email: barlach@ruc.dk

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

User Driven, Evidence Based Experimental Design; a New Method for Interface Design Used to Develop an Interface for Clinical Overview of Patient Records

Niels Boye[a,f], Frey Eberholst[b,f], Richard Farlie[c,f], Lene B. Sørensen[d,f], Karen Marie Lyng[e,f]

[a] *Regional Hospital of Randers, University of Aarhus, Denmark*
[B] *Acure – an IBM Division, Aarhus, Denmark*
[c] *Regional Hospital of Herning, Denmark*
[d] *Regional Psychiatric Services, North Jutland , Denmark*
[e] *Tieto-Enator, Denmark*
[f] *Master Education Programme in Health Informatics, University of Aalborg, Denmark*

Abstract

A novel method of interface design - user driven, evidence based experimental design - was developed which approximates the usual clinical way of maturing science and technology in the healthcare domain.

The method is user-driven and the clinician remains in control of gathering and evaluating evidence of relevance to the project - as well as specifying the details of the user interface.

Information not obtainable from the literature was gained experimentally and used to achieve a deeper understanding of the problem before the design phase. The design was subsequently validated experimentally by ordinary users with no connection to the software or design team.

After applying this method to the problem of gaining a satisfactory clinical overview of a single patient's record, we recommend that clinical IT interfaces have clinical logic, sufficient complexity, and are well structured. Developers should use computer power to support "building blocks" such as anatomical problem lists and summaries of history, status and treatment, personal notes, and should support clinical browsing using text and graphics.

Keywords:

medical informatics; computing methodologies; software design; user-computer interface; medical informatics applications; medical record systems, computerized; decisions support systems, clinical; physician practice patterns; clinical reasoning

Introduction

The scientific basis of healthcare is in natural sciences, but the clinical execution also draws on methods and concepts from social sciences. This makes development, formalization, computation and presentation of health related information in electronic systems for clinical use a complex task (1). Attempts to develop a comprehensive and satisfactory overview of a patients status and history to the clinical user in the form of an Electronic Health Record (EHR) has been regarded as inferior to the paper based record.

Various design methods, such as socio-technical design (2-4) and user-centred design or participatory design (www.cpsr.org) have been developed with the aim of promoting a common understanding between IT-professionals and professionals in social sciences such and health care.

In the user-centred design method, the initiative and control in establishing the necessary common understanding between user and IT-professional are in the hands of the latter.

This paper describes the initial efforts to develop a general design method for interfaces, where the common understanding between healthcare- and IT-professionals is established but where control remains with the end users as opposed to the user-centred design method. Furthermore this novel method is more in line with the usual way healthcare technology is matured and developed. We have named it "user driven, evidence based experimental design" to describe the principal components of the method.

The method was developed during work to attempt to solve a general problem in IT systems for health care design; how to obtain a clinical overview of a single patients record. Evidence for the design was drawn from the literature and from our own experiments. The first iteration design was subsequently validated through experiments with end-users (doctors).

Materials and methods

A loosely coupled research group for this specific project was formed at the Master Education programme in Health Informatics at the University of Aalborg, Denmark (http://www.v-chi.dk/english/index.htm). A subgroup of two (RF, KML) investigated "genuine features of overview in paper based records", with a combination of a literature survey and experiments. Another subgroup of two (FE, LBS) investigated "specific features in clinical IT systems conveying the overview", also with a combination of literature survey and experiments. One member (NB) coordinated and supervised the two sub projects. All project members had a common basis of extensive clinical experience, three specialist doctors (RF in gynaecology and obstetrics, KML

in surgery, NB in internal medicine), one nurse (LBS) and one radiographer (FE). No trained IT professionals participated. All participants had years of experience as users and clinical participants in user-design efforts and/or clinical specialists in software companies and all participated in the Master Program in Health Informatics at Alborg University as teacher or students.

Literature overview

Literature surveys and compilation into a common "concept of clinical overview" for the planning and execution of the experimental part was done by Medline searches and subsequent "drilling down" from reviews and papers of cognitive tasks research ((5-9)), decision-theory ((10)), graphical perception ((11-13)) and interface design (e.g. www.useit.com). Literature information was discussed and shared in the total project group.

Empirical information collection

Overview in paper based records

This subgroup employed semi-quantitative interviews with eight doctors from five different medical specialities (four of whom used an EHR in their daily clinical work) to elucidate "the nature of the clinical information overview". This information was analysed using a grounded theory method and led to an observational study with five junior-doctors each performing three constructed scenarios. Data was acquired from the scenarios by the "think aloud method" and various parameters were analysed afterwards from video recordings of each scenario.

Overview in EHRs

A complex patient history based on true data was constructed and presented in two electronic prototypes. The overall construction of the prototypes was similar, but the design in one was mainly text-based and very like EHRs modelled over the paper record, with text and tables.

The other prototype developed was based on the evidence acquired from literature about cognitive mechanisms, clinical and human reasoning, graphics, interface-design and experiments with the overview strategies in paper-based records and it had additional visualisation of clinical information in graphs and displays, constructed in accordance with knowledge of graphic displays and interface design. Both prototypes were programmed "by hand" using Microsoft PowerPoint.

Twenty-three doctors were tested about different problems using either the text-based prototype or the graphics-based prototype of the patient record using the same structured questionnaire. The doctors were randomly allocated to the two groups. User-prototype interaction was observed, timed and further quantified using Camtasia Studio from Techsmith.

Results

Literature review

The literature review showed that doctors use a variety of different strategies to gain an overview of a patient's situation ((14;15)) from the textual information on paper or on screen. The overall strategy employed depends on, whether the information is obtained for the first time or the patient is known to the clinician - as well as time constraints. Reading of textual information can be categorized as: 1) Reading, where the whole text is read, 2) Skimming, where some words in each sentence are read or 3) Skipping, where only few words across a page are read. Hornbæk and Frøkjær (16) have constructed three ways of formatting text: *normal text, fisheyes view* – where less important text-parts were shown with a smaller font, or *overview + detail*, where keywords were extracted and shown in the margin. The *overview + detail view* was found to be the most effective and satisfying for the viewer. A rigorous structure in text and layout and consistent use of design elements supports focused information retrieval (medication, adverse reactions, description of previous results from a procedure) and the general sense of overview by the reader (14). Personal annotations should be supported. Graphical methods should be prioritized and follow the heuristic research based rules for perception of graphical information as set up by Cleveland (11).

Condensation of interview and observational study result for establishing overview of clinical problems using paper based information

Although the paper record employed was unknown to the participants in the observational part, they quickly established an overview of the record structure and formed a preliminary hypothesis for verification or falsification against the specific information in the record by pattern recognition. Data such as lab. results are in tables with a certain structure, daily notes are in ordinary text separated by dates. There was no general agreement on what kind of information it was important to start with.

Observation showed that even junior doctors have developed individual styles in handling the task of establishing an overview of paper based patient record information. Some start from the last entry and read backwards, some start from the beginning and work chronologically towards the most recent entries. The reading style can be combined with ether a tendency to separate the record into several heaps or to keep papers in chronological order, except when comparing individual documents in the record. All alternated between tables, descriptions of results and the entries in the main section of the record. Summaries and compilations of results formulated by colleagues, such as indications for a treatment plan, are easily and quickly identified and used as the "building blocks" of an overview.

Comparison of design elements in the two prototypes

The text based prototype was not just "paper on screen" but contained features that took advantage of computer power to compile chronological overviews, problem lists and lists of active medications and investigations, but maintaining layout and textual features similar to a paper-based record (figure 1). No graphical elements in pathology lab reports, medication lists or diagnostic imaging were present in this prototype.

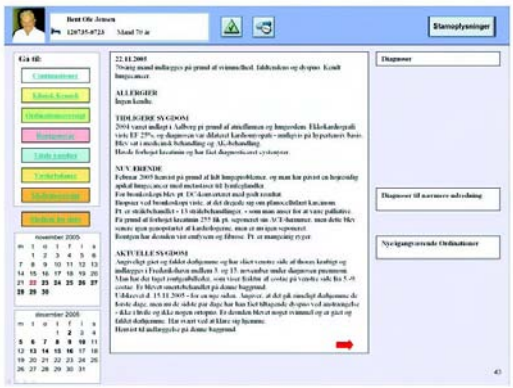


Figure 1 - The text-based prototype with view-menu at the left, a calendar and plain text.

The graphic prototype had the same features as the text based system, with the addition of keyword highlighting to enable text skimming or text scanning, this feature was linked to a graphic anatomical representation of the patients problems. Extensive use of graphical views was used for pathology lab results, vital signs, fluid balance and diagnostic imaging (not suitable for diagnosis) (figure 2).

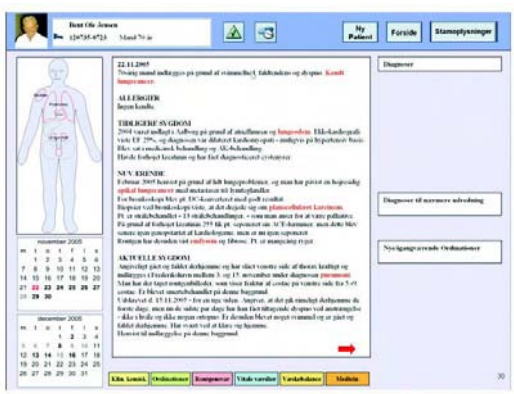


Figure 2 - Same "view" as in figure 1 but in the graphic prototype. The view menu is at the bottom of the screen.

Performance in relation to establishing clinical overview in a text- or a graphic design prototype.

The clinical overview questionnaire of the patient case had 30 questions (each giving 1 point). Ten doctors were allocated to the text based prototype and 13 doctors to the graphic based prototype, but two of these tests partly failed in their quantitative measurements due to technical problems with Camtasia.

Table 1 – Quantitative measurements of performance

Group	N	Mean Time	Range Time	Mean Points	Range Points
Text	10	17:08	14:05-25:00	26.2	23 – 28
Graphic	11	15:38	13:00-17:55	28.1	26 - 30

Time in min:sec

Table 1 shows the quantitative measurements of the two groups solving the clinical questionnaire about the patient. In accordance with the evidence obtained in the project, this table should be displayed graphically to obtain the "non-reduced" picture of the distribution and pattern of performance for each group, but this would demand more space.

The previous IT experience of the doctors in the test was scored by the frequency of IT use, and the number of specific software packages used. The highest score was 22 and the lowest 2. No correlation between IT experience and time used or IT experience and point score in the two questionnaire experiments was found.

The user aspects of the two prototypes were analyzed with a semi-quantitative interview after each test. The text-based prototype was easy to understand, but difficult to navigate, in contrast to the graphics prototype ("you have to get used to this, but it is simple to work with").

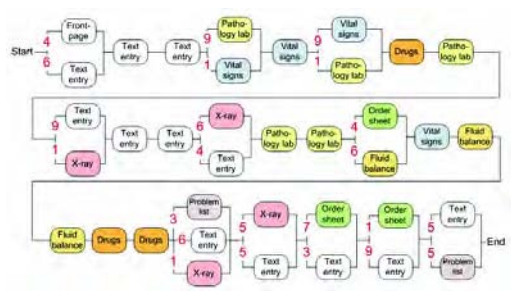


Figure 3 - Flow pattern of the text-based prototype, the number of participants branching is indicated by red numbers.

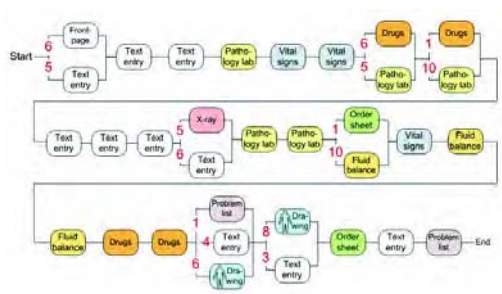


Figure 4 - Flow pattern of the text-based prototype, the number of participants branching is indicated by red numbers

Analysis of the flow and order of screens selected in the text based (10 doctors) and the graphics based (11 doctors) prototype respectively based on data obtained from Camtasia is seen in figures 3 and 4. The graphical prototype gave a simpler and more uniform flow and was generally evaluated higher. Several good ideas for features in future iterations were formulated by the doctors in the interview after the test.

Discussion

The aim of the present project was to develop a method for development of software for clinical use that was closer to the conventional way in which technology is matured, validated, and incorporated in everyday activity in the clinical domain.

Much of the development and subsequent adaptation of technology to clinical surroundings and clinical users in for example, clinical chemistry, molecular biology and diagnostic imaging has been done endogenously in the clinical organization, based on scientific (in contrast to commercial) sharing of knowledge, and furthermore driven by experimental methods. This implies that the clinicians are in control of the experiments and validation of results, with the scientific "responsibility" to colleagues and peer-review bodies. It will still be a cross disciplinary effort to develop clinical information systems and maturation of concepts to an operational level, as development of sufficiently advanced software platforms for hospital use will probably still be a commercial activity.

The general clinical problem used as a "case" in this paper was the overview of previous documentation in a patient's record in order to obtain a satisfactory understanding of the patients' history and present status, as the basis for new clinical (human-driven) reasoning. It can, of course, be questioned whether this can be mimicked by the ability to answer 30 specific questions about the patient using two different electronic interfaces. More experiments with more patients, cases and interfaces are needed to elucidate this aspect.

Using Microsoft Power Point as the tool for visualization of concepts is not advisable, since it becomes unpredictable when working with complex presentations containing many internal hyperlinks, but we used it because we could not find a better software package for the purpose at the time.

The validation experiments are crucial to the method, since it separates the developers from the feed back from end users in a sincere way, which, although it is cumbersome, will give more convincing results for use in the next iteration. It is of great importance to use structured and methodologically sound validation experiments to obtain evidence of suitable strength,

The graphics based prototype gave the highest score and the simplest flow of the two prototypes with the same information content. No statistical methods have been employed, since the differences are more qualitative than quantitative. The authors interpretation is, that the graphic prototype to a larger extend supported the users need to alternate between "the overview" and "the detail" without losing track when investigating a clinical problem. This is probably also the mechanism behind the more effective text reading of skimming and scanning and the *overview+detail view* of Hornbæk and Frøkjær (16).

On the basis of our literature studies and experiments, we have the following recommendations for interface design for clinical use. This advice will need further validation in future projects and experiments.

> Medicine is a visual, 3-dimensional discipline. This should be reflected in the interface.
>
> A rigorous, clinically intuitive structure in text and graphic displays, and the use of design elements with easily recognizable patterns is advised. The interface does not have to be simple, since the clinical user can quickly master a complicated design if it is "clinically logical", and there is no need to strive for a standard program layout – clinical functionality is more important.
>
> Browseability in text should be supported by different text presentations such as standard text, fisheye and overview + detail. Automatic highlighting of clinical terms can increase browseability. Filters for different types of entries could increase the acquisition of overview "building blocks".
>
> Automatic generation of problem lists, active medication and vital signs charts etc., and an anatomical view of the problem list (and other information) were rated as desirable features in our test group. Computer power should be used to ease the cognitive burden of the clinical user and will probably give a good "return on investment" (17). Opportunities for personal annotations and search facilities should be provided. A list of previous cases for junior doctors with notification facilities would aid personal reflection on performance and learning.
>
> Some indication of the complexity of the clinical problem and summaries of previous information should be easily displayed.

The browseability features of the paper record such as scattering documents over a desk top can only be partly recreated in the EHR by employing large or multiple displays. Gesture control of objects and object relations on a large display may be desirable in terms of "computer supported augmented clinical overview."

References

[1] Kanstrup A, Boye N, Nohr C. User-driven design considerations of m-learning services for residents - activation of a theoretical model of clinical knowledge. Medinfo2007, Brisbane 2007.

[2] Berg M. Working with protocols: a sociological view. Neth J Med 1996; 49(3):119-125.

[3] Berg M. Medical work and the computer-based patient record: a sociological perspective. Methods Inf Med 1998; 37(3):294-301.

[4] Berg M. Patient care information systems and health care work: a sociotechnical approach. Int J Med Inf 1999; 55(2):87-101.

[5] Evans D, Patel VL. Advanced Models of Cognition for Medical Training and Practice. Elsivier, 1991.
[6] Patel VL, Kaufman DR, Arocha JA, Kushniruk AW. Bridging theory and practice: cognitive science and medical informatics. Medinfo 1995; 8 Pt 2:1278-1282.
[7] Kushniruk AW, Patel VL. Cognitive computer-based video analysis: its application in assessing the usability of medical systems. Medinfo 1995; 8 Pt 2:1566-1569.
[8] Patel VL, Kushniruk AW. Interface design for health care environments: the role of cognitive science. Proc AMIA Symp 1998;29-37.
[9] Kushniruk AW, Kan MY, McKeown K, Klavans J, Jordan D, LaFlamme M et al. Usability evaluation of an experimental text summarization system and three search engines: implications for the reengineering of health care interfaces. Proc AMIA Symp 2002;420-424.
[10] Wulff H, Gøtzsche P. Rationel Klinik. 4 ed. Copenhagen: Munksgaard, 1997.
[11] Cleveland WS, McGill R. Graphical Perception and Graphical Methods for Analyzing Scientific Data. Science 1985; 229:828-833.
[12] Cleveland WS. Visualizing Data. Hobart Press, Summit, New Jersey, 1993.
[13] Cleveland WS. The Elements of Graphing Data. Revised edition ed. Hobart Press, Summi, New Jersey, 1994.
[14] Nygren E, Henriksson P. Reading the medical record. I. Analysis of physicians' ways of reading the medical record. Comput Methods Programs Biomed 1992; 39(1-2):1-12.
[15] Nygren E, Wyatt JC, Wright P. Helping clinicians to find data and avoid delays. Lancet 1998; 352(9138):1462-1466.
[16] Hornbæk K, Frøkjær E. Reading patterns and usability in visualizations of electronic documents. ACM Transactions on Computer-Human Interaction (TOCHI) 10 (2)[June 2003], 119-149. 2006. ACM press New York, USA. Ref Type: Electronic Citation
[17] Kaushal R, Jha AK, Franz C, Glaser J, Shetty KD, Jaggi T et al. Return on investment for a computerized physician order entry system. J Am Med Inform Assoc 2006; 13(3): 261-266.

Address for correspondence

Niels Boye, Regional Hospital of Randers,
Skovlyvej 2,
DK-8900 Randers
Denmark
E-mail: Niels@Boye.dk

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

User Interface Optimization for an Electronic Medical Record System

Kai Zheng[a], Rema Padman[b], Michael P. Johnson[b]

[a] *School of Public Health and School of Information, University of Michigan, U.S.A.*
[b] *H. John Heinz III School of Public Policy and Management, Carnegie Mellon University, U.S.A.*

Abstract

Many information technology-enabled healthcare applications have failed because their interfaces are difficult to use. Unfortunately, little attention has been paid in the health informatics community to designing effective user interfaces that are acceptable to healthcare professionals. This paper illustrates a method for improving application interface usability by applying sequential pattern analysis to analyze temporal event sequences recorded in an electronic medical record system. Such event sequences, or clickstreams, reflect clinicians' navigation patterns in their everyday interactions with the computer system. The identified patterns have been used by software developers to calibrate the user interface of the system, so that the within-application workflow is better aligned with clinicians' mental model of medical problem-solving. Such inferred patterns may also help to modify clinicians' suboptimal practice behavior components, as manifested through their actual usage of this point-of-care electronic system.

Keywords:

user-centered design; user interface design; sequential pattern analysis; human-computer interaction; usability assessment; data display

Introduction

Medical practice is a complex process. Large amount of data must be accessed, assembled, and analyzed at the point of care to inform proper medical decision-making. In the era of paper-based patient records, clinicians flip through stacks of paper charts to look for desired information. The use of electronic systems has greatly facilitated health data retrieval. However, it has also introduced new dimensions of problems. Two paper documents, for instance, can be laid out side by side for cross reference, while on a computer screen it is usually impractical to have two windows visible at the same time. How to preserve the easy "look-and-feel" of paper charts is a real challenge for software developers. In addition, poorly designed application navigation flow may also escalate learning effort, decrease productivity, and increase user errors [1, 2].

Lack of good user interfaces has been long recognized as a major impediment to the acceptance and routine use of clinical informatics applications [3]. Unfortunately, very few research studies have looked at design principles for building intuitive and effective healthcare user interfaces (UI); even fewer have validated the usability of existing UI design in realistic clinical settings. Consequently, "systems are created *ad hoc*, users are dissatisfied, and often systems are abandoned" [2].

The present study was motivated by these facts. The computer system in question, the *Clinical Reminder System* (CRS), is a "*lite*" electronic medical record system (EMR) that collects, stores, and manages a wide range of patient and clinical data [4, 5]. In addition to its regular EMR functionalities, CRS is also intended to improve quality of care by providing clinicians "just-in-time" alerts and advisories using evidence-based guidelines.

Since 2002, CRS has been deployed in an outpatient clinic at an urban hospital, and used by clinicians to treat patients in real time. While user, task, and representational analysis were performed during the software design phase with constant feedback by participating clinicians, its UI design was still critiqued after being routinely used in clinicians' everyday practice. In a user satisfaction survey following a 10-month field trial, users complained that the application's early user interface, shown in Figure 1, provided little guidance as to a desired workflow [4, 5]. As a result, user acceptance was not satisfactory, and the utilization rate of the system remained low [4, 5].

Although this UI reflected the best knowledge of developers and preferences of the client organization, the standard Windows-based layout was reported as "not aligned with our common practice styles". The horizontally arranged tabs, for example, did not reflect the preferred order of clinical information access. As a result, users expended substantial energy unnecessarily to adapt their practice to a UI design that they considered "uncomfortable".

Selected for best paper award.

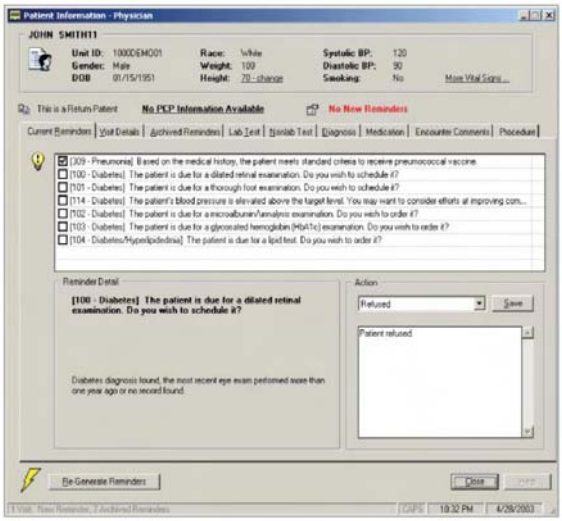


Figure 1 - An early user interface

To solve the identified UI flaws, the system was reengineered into a full web-based application. A screenshot of the new web interface is shown in Figure 2. Unique features of the web-enabled application provide tremendous promise for maximally preserving the "look-and-feel" of traditional paper charts. In the new design, for example, different features conveying different clinical information elements are no longer arranged in a tabular form, instead, they are displayed in the same workspace that can be easily navigated by mouse scroll wheels, simulating paper-flipping behavior. A navigation menu is also provided on an adjacent frame to enable fast switches across different features. However, it is not known whether this new design is consistent with, or represents an improvement upon, clinicians' typical workflow.

The study reported in this paper was therefore conducted to identify the preferred sequential order in which different features of the system are accessed. To learn clinicians' navigation behavior, this study uses a sequential pattern analysis method to analyze actual usage recorded in the computer logs that contain time stamped events. Actual usage data, unlike many software usability experiments, represent users' interaction with a system under real working conditions, rather than on contrived laboratory exercises.

Figure 2 - New user interface to be evaluated

Methods

Sequential pattern analysis

Sequential pattern analysis discovers hidden and recurring patterns within large sequences of events. It has been applied in a wide variety of domains such as web person specialization and page recommendation [6], HCI usability testing [7], and genetic sequence analysis [8]. In this study, a consecutive sequential pattern algorithm is employed to analyze the event sequences recorded in CRS. This algorithm detects consecutively occurring events that appear across different sessions. Such patterns, that represent adjacent feature accesses frequently occurring next to each other and in a given sequential order, are of particular interest to inform UI redesign.

Let s denote an even sequence by $<e_1, e_2, \ldots, e_n>$, where $e_j, j \in 1 \ldots n$, is an event that occurs at the j^{th} position in s. The consecutive sequential pattern algorithm finds a sequence $p <p_t, p_{t+1}, \ldots, p_{t+l}>$ that is a subset of s, which is also part of, or supported by, other sequences. The support for p is defined as the fraction of total sequences that supports p. When a sequence satisfies a certain minimum

support threshold, it is named a *Sequential Pattern*. The largest length sequential pattern that is not part of any other patterns is called a *Maximal Sequential Pattern*. The objective of the sequential pattern analysis is to find all such maximal sequential patterns.

When the minimum support is a constant for any given length, the most efficient algorithm starts with calculating support for all possible sequences composed of two consecutive events. When a sequence does not satisfy the minimum support, it is removed from further computation; otherwise, it is treated as a candidate sequence to compute support for larger length sequences. The algorithm stops when no larger length sequences based on a current candidate would satisfy the minimum support. The current candidate sequence is then chosen as a maximal sequential pattern.

Study site and data collection

In this study, 10 months of usage data were electronically collected from October 1, 2005 to August 1, 2006 and analyzed. These usage data were generated from the most recent web-enabled version of CRS. The system implementation was accomplished in the summer of 2005 and substantial training was provided afterwards.

The main CRS user population during the study period was composed of 40 first-, second-, and third year internal medicine residents. Residents who used the system for fewer than 5 patient encounters are excluded from the analysis. It is likely that such users' interactions with CRS do not reflect mature application usage. 30 active resident users were thus identified, whose system usage was recorded in 973 unique patient encounters.

Data analysis and results

Data preparation

Data preparation procedures were performed prior to the analysis. All events and their affiliated attributes, such as session ID and time stamp, were first collected from scattered data tables. The event type was then mapped based on a labeling schema, which is composed of distinct letter symbols. Table 1 lists all 17 main features[1] that the CRS application provides, ordered alphabetically by their labeling symbols[2]. The screenshot shown in Figure 2 illustrates the on-screen positions of each of the 17 major features.

Event sequences were then constructed. `HMMMYAD`, for instance, is a 7-length sequence composed of 7 events that occurred within a patient encounter, ordered chronologically by their time stamps. The resulting event sequences are further consolidated by collapsing repeating access to a same feature. For example the segment `MMM`, "prescribing multiple medications consecutively", is collapsed into one single event `M`. In this study only across-feature navigation is of interest, that is, "jumps" across different features.

Figure 3 shows the distribution of event sequence length after the collapsing operation. The sequences composed of 4 or less events are excluded from further data analysis because they provide little information in regard to sequential navigational patterns. This operation results in the loss of 6 additional users whose recorded sequence lengths are all below 5. After these data preparation procedures 473 event sequences are retained, generated by 24 distinct resident users. Distribution of number of sequences owned by each user is depicted in Figure 4. Several sample event sequences are shown below:

```
HMOMYXAM
GHXVHADADHA
HGYXADAOMYSX
OMRHFYXYXADADA
HXOPMOMOMOMODADAM
HSXDADADADADAMOMOMOMOMO
```

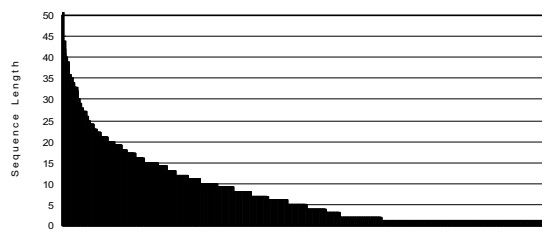


Figure 3 - Distribution of event sequence length

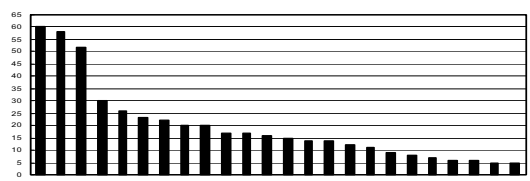


Figure 4 - Usage distribution among users

Frequency of feature access

Table 1 shows the aggregated proportion of feature accesses[3]. These proportions roughly represent how frequently each application feature was used. As shown in Table 1, among the 17 major features "Assessment and Plan", "Diagnosis", and "Medication" were most heavily used. Note that while "Encounter Memo" appears on top of screen, it was seldom accessed.

1 Feature that must be displayed in a certain position for legal reasons, such as patient's demographics always appearing on top of an encounter page, is excluded from the consideration of this study. Also excluded are non-actable or not yet activated features, for example "Reason for the Visit" that is entered by nurses when a patient calls to make an appointment.

2 A symbol letter is usually the first letter of a feature unless there is a conflict.

3 Repeating access to a same feature is counted only once.

Table 1 - Main features and overall frequency of access

Label	Feature	Proportion (%)
A	Assessment and Plan	21.18
B	Retaking BP	.34
D	Diagnosis	16.36
E	Medication Side Effects	.22
F	Family History	1.24
G	Allergies	1.88
H	History of Present Illness	7.26
L	Laboratory Test	3.58
M	Medication	14.53
O	Order	17.17
P	Procedure	.38
R	Encounter Memo	.44
S	Social History	2.85
T	Office Test	.62
V	Vaccine	.83
X	Physical Examination	6.69
Y	Review of Systems	4.43

Table 2 shows the results of the sequential pattern analysis. All maximal sequential patterns included in the table satisfy a minimum support threshold of 15%. These patterns are sorted by the level of support they received.

Table 2 - Maximal sequential pattern discovered

Maximal Sequential Pattern	Level of Support (%)
ADAD	51.16
DADA	43.97
XADA	40.17
OMOM	32.77
MOMO	29.39
YXAD	21.78
HS	19.03
OL	18.6
OMY	16.7
LO	15.64
HO	15.01

Some interesting sequential patterns emerge from Table 2. ADAD, appearing in 51.16% of all encounters, is the most salient pattern discovered, followed by a similar and partially overlapped pattern DADA, with 43.97% support. It indicates that the users of CRS frequently switched between the features "Assessment and Plan" and "Diagnosis". Similarly, users frequently switched between "Order" and "Medication", with 35.1% support for OMOM and 27.06% support for MOMO; and "Order" and "Laboratory Test", with 18.6% support for OMOM and 15.64% support for MOMO. A further examination found that A precedes D more often (89.18%) when a user entered the AD...AD or DA...DA segment. Similarly, O was usually accessed before M (72.57%), and before L (71.58%).

Supported by 40.17% of all encounters, ADA is preceded by X - "Physical Examination", and YXAD appears in 21.78% of time. This indicates that "Physical Examination" "Assessment and Plan" "Diagnosis" is a frequently traversed path, which is often preceded by accessing "Review of Systems". Further, OMY occurs in 16.7% of all sequences, indicating that OM - "Order" and "Medication" were often used before Y - "Review of Systems". HS - "History of Present Illness" then "Social History" and HS - "History of Present Illness" then "Order", are two other consecutive patterns with slightly smaller support, 19.03% and 15.01%, respectively.

An *ad hoc* within-sequence analysis was further conducted to detect sequence segment recurring within an encounter session. Results are shown in Table 3. The "Probability of Repeat" in Table 3 exhibits the probability of a two-length event segment recurring within a sequence. DA or AD - "Diagnosis" and "Assessment and Plan", OM or MO - "Order" and "Medication", and OL or LO - "Order" and "Laboratory Test", are three frequently repeating segments thus identified, which also confirm the cross sequence patterns of DADA, ADAD, OMOM, MOMO, OLOL, and LOLO. Because items in these reappearing sequence segments were usually accessed next to each other, they are hereby referred to as *Bundled Action.*

Table 3 - Recurring patterns within encounters

Sequential Pattern	Probability of Repeat (%)
AD	70.22
MO	64.98
OL	64.77
DA	64.35
OM	63.67
LO	51.35

The repeating access to bundled actions, however, blurs the boundary of "jumps" from a series of bundled action accesses to other features. For example the reappearing AD with varying length in the sequence HADAD...ADADXY impairs the analytical power for discovering whether there exists a pattern H-AD-Y that may help reveal interesting patterns at an overall level. Similar to collapsing repeating access to the same feature, repeating access to the same bundled action is further collapsed to count as one single occurrence. For example the HDAD...ADADAXY sequence is converted into HDAY to form a new, higher level sequence.

A second pass sequential pattern analysis was then conducted to analyze the event sequences obtained after this collapsing operation. ADO - "Assessment and Plan" to "Diagnosis" to "Order" is the only additional sequential pattern thus identified, supported by 15.64% of all encounters. This pattern indicates that after a user finished working on "Assessment and Plan" and "Diagnosis", he or she would switch to the "Order" section immediately to prescribe orders of new medications or laboratory tests.

Discussion

Based on the findings from analyzing actual usage data with sequential pattern analysis, several UI design principles can be arrived at:

- "Encounter Memo" should be properly relocated. This feature is less frequently used while occupying the most salient position in the current design;

- "Assessment and Plan", "Diagnosis", and "Medication" are the most frequently accessed features. They should be placed in the most salient positions on a computer screen;
- "Assessment and Plan" and "Diagnosis", "Order" and "Medication", and "Order" and "Laboratory Test" are *bundled actions*. They are usually accessed next to each other and often used multiple times within an encounter session. Navigation aids such as hyperlink shortcuts should be provided to facilitate these frequent feature switches;
- "Review of Systems", "Physician Examination", "Assessment and Plan", and "Diagnosis" should be presented adjacent to each other in this sequential order. Accesses to these four features often appear as a series of events occurring sequentially.

These design principles have been used in redesigning the existing user interface of CRS. Since the basic EMR functionalities that CRS provides are universal, these design principles may also be applicable to other electronic medical record systems.

Conclusions

Improving the UI design of an electronic medical record system can be successfully attained by analyzing the actual usage data recorded during its everyday use. The sequential patterns identified in this paper led to a set of design principles used in redesigning the application's user interface. These design principles mainly propose that different clinical information elements should be presented in the sequential order in which they are usually accessed, which reflects clinicians' mental model of medical problem-solving during patient encounters.

This study has a few limitations. First, actual usage data must be collected from a working system. Its current design, inevitably, may exert an influence on users' own working style. Second, the findings are derived from testing a single system with certain unique features. While the method and the results provide general insights into designing user interfaces for other types of health applications, they may not be used without careful customization. Finally, the user population of this study was mainly composed of internal medicine residents. The derived design pattern reflecting their practice style may not be generalizable to other clinical specialties.

References

[1] Bates DW, Kuperman GJ, Rittenberg E, Teich JM, Fiskio J, Ma'luf N, Onderdonk A, Wybenga D, Winkelman J, Brennan TA, Komaroff AL, and Tanasijevic M. A randomized trial of a computer-based intervention to reduce utilization of redundant laboratory tests. Am J Med 1999; 106(2):144-50

[2] Tang PC and Patel VL. Major issues in user interface design for health professional workstations: summary and recommendations. Int J Biomed Comput 1994; 34(1-4):139-48.

[3] Johnson CM, Johnson TR, and Zhang J. A user-centered framework for redesigning health care interfaces. J Biomed Inform 2005; 38(1): 75-87.

[4] Zheng K, Padman R, Johnson MP, Engberg J, and Diamond HS. An adoption study of a clinical reminder system in ambulatory care using a developmental trajectory approach. Medinfo. 2004; 11(Pt 2):1115-9.

[5] Zheng K, Padman R, Johnson MP, and Diamond HS. Understanding technology adoption in clinical care: clinician adoption behavior of a point-of-care reminder system. Int J Med Inform 2005; 74(7-8):535-43.

[6] Montgomery AL, Li SB, Srinivasan K, and Liechty JC. Modeling online browsing and path analysis using clickstream data. Marketing Science 2004; 23(4):579-95.

[7] Siochi AC and Ehrich RW. Computer analysis of user interfaces based on repetition in transcripts of user sessions, ACM Transactions on Information Systems 1991; 9(4):309-35.

[8] Eres R, Landau GM, and Parida L. Permutation pattern discovery in biosequences. J Comput Biol 2004; 11(6):1050-60

Address for correspondence

Kai Zheng, kzheng@umich.edu, M3531 SPH II, 109 S. Observatory, Ann Arbor, 48109-2029, U.S.A.

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

AdaRTE: Adaptable Dialogue Architecture and Runtime Engine. A New Architecture for Health-Care Dialogue Systems

L. M. Rojas-Barahona[a], T. Giorgino[a]

[a] *Laboratory for Biomedical Informatics, Dipartimento di Informatica e Sistemistica, Università di Pavia, Italy*

Abstract

Spoken dialogue systems have been increasingly employed to provide ubiquitous automated access via telephone to information and services for the non-Internet-connected public. In the health care context, dialogue systems have been successfully applied. Nevertheless, speech-based technology is not easy to implement because it requires a considerable development investment. The advent of VoiceXML for voice applications contributed to reduce the proliferation of incompatible dialogue interpreters, but introduced new complexity. As a response to these issues, we designed an architecture for dialogue representation and interpretation, AdaRTE, which allows developers to layout dialogue interactions through a high level formalism that offers both declarative and procedural features. AdaRTE aim is to provide a ground for deploying complex and adaptable dialogues whilst allows the experimentation and incremental adoption of innovative speech technologies. It provides the dynamic behavior of Augmented Transition Networks and enables the generation of different backends formats such as VoiceXML. It is especially targeted to the health care context, where a framework for easy dialogue deployment could reduce the barrier for a more widespread adoption of dialogue systems.

Keywords:

telemedicine, speech recognition software, ambulatory care information systems, telephone, computerized, chronic obstructive pulmonary disease, hypertension

Introduction

Dialogue technologies have been proven useful to provide the general public with access to telemedicine services. Several studies have discussed their advantages for chronic symptoms monitoring, interviews, counseling, education, etc. [1, 2]. Dialogue systems in health care context are deployed to complement traditional contact channels and have been used for several home-care interventions successfully [3-7]. Thus, they may be able to improve quality of service and communication in a cost-effective way.

In previous projects, customized technology was the response to vocal applications deployment. Available technology allowed implementation either via custom code, or proprietary dialogue-manager based solutions. Several dialogue manager based architectures have been devised in order to simplify complex programming present in custom-coded applications. In addition, the multitude of dialogue technology vendors naturally resulted in a proliferation of incompatible languages across vendors and platforms.

Recently, the concept of Voice Browser (VB) was introduced by W3C [8]. VBs foresee a dialogue-manager, which understand VoiceXML documents instead of a proprietary language. Despite providing a great deal of independence from speech recognition engine vendors, VoiceXML has serious shortcomings in allowing the reuse of components, database access and support of natural language processing (NLP) and multimodality [9]. Furthermore, the sequence of dialogue steps in VoiceXML is defined with a sort of form-filling mechanism and, as web based technologies, has to be generated dynamically by other code. In general, visual inspection and maintenance of scripts is less than straightforward.

The need for a leaner development methodology is especially evident when considering domains in which the manpower and the time-cost available for development are limited, such as health domain. In this paper, we present a novel architecture, AdaRTE devised in order to overcome the issues of the existing dialogue management methods. AdaRTE features were thought to reduce dialogue system development effort through re-use, support of augmented transition networks, adaptable decision takers and best practices adoption. We built AdaRTE, which implements these features for dialogue deployment, and we present the results obtained through the partial prototyping of two telephony-linked systems: the first inspired by the Chronic Obstructive Pulmonary Disease (COPD) care [5], and the second by the Homey dialogue system for hypertensive patient home management [6].

Our effort was mainly focused on health care dialogues systems, since our solution is especially targeted at offering low cost, standards-compliant deployment and experimentation through the incremental integration of other voice formalisms i.e. NLP based on lexicalized grammars.

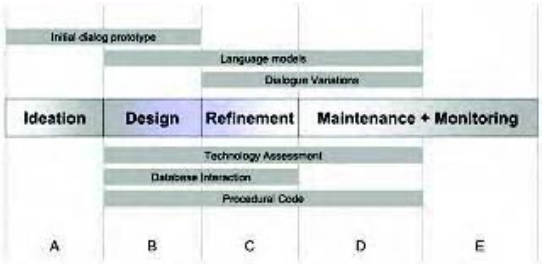


Figure 1 - Phases in the Homey development cycle

Background

A range of technologies are available for building dialogue systems. The simplest of these technologies is a linear script; others are state transition networks and plan-based dialogue systems [7]. Generally, the deployment of any of these techniques requires heavily scripted solutions. Additionally, there is not much information available about time and costs implied in dialogue systems development process. As a matter of fact, deployment of dialogue systems was considered art rather than engineering or science, because of the scarcity of standards. Commonly, the process of deploying dialogue systems was complicated, costly, time demanding and required speech technology experts. To give a specific example, in one previous projects, the EU-sponsored project Homey [6], the time spent in developing the technical aspects of the system has been rather long and the result was reusable only to a limited extent (Figure 1). The voice part of the Homey system required approximately one man-year worth for design and implementation. The Maintenance and Monitoring milestone targeted issues roughly grouped in the following areas: Incorporation of procedural code, database access, language models, increased variability of speech, prompts, modularization and issues related to ASR engine dependency.

More recently, on the other hand, the maturation of VoiceXML allows to deploy dialogue systems in a Web-based environment [8]. Its delivery contributed to reduce the proliferation of incompatible dialogue formalisms by offering one standard for voice applications, so that platform independence and simplicity of development are allow at the same time. Several dialogue systems deployed by using VoiceXML have been published lately. For instance, a dialogue for diabetes home monitoring was implemented by integrating VoiceXML with a Voice Service Provider (VSP), developing a visual user interface and the database backend in [10].

In spite of the advantages explained above, VoiceXML has inherent limitations which are well analyzed in [9, 11], such as its declarative and static structure, difficulty accessing remote resources (databases and ontologies) and lack of means for efficient and heavy computation. Furthermore, the strongest limit pointed out by the research community is that neither dynamic natural language understanding and generation nor multimodality is directly supported. As a consequence, a variety of extensions to VoiceXML has been proposed: for instance, DialogXML was applied to car telematics, in this approach the VB was extended to support NLP KANTOO generated grammars [12]. A prototype of an editor for creating VoiceXML documents is exposed in [13]. Despite the emerging VoiceXML-generative frameworks, we believe that a big effort should still be done in adapting dialogue systems best practices such as confirmation strategy, adaptability, mixed initiative, usable speech interfaces for users and graphical interfaces for developers, together with innovative speech solutions including NLP in VoiceXML based frameworks. [14]

Methods

Our proposed architecture, shown in figure 2, is primarily composed of a *dialogue interpreter, a runtime engine* and *an interface media realizer for backends generation.* A running system interacts with users which can be grouped in three main role categories: Application developers, patients and case managers, i.e. case manager nurses.

In order to enable rapid prototyping, every given dialogue should be developed in a graphical environment which will allow the layout of prompts, speech items, and updating their properties. The editor will represent and store the dialogue structure in a well-defined formalism. To simplify notation, this representation will also be called XML dialogue description from now on.

To cooperate with standards-based speech recognition software and respond to telephone-originated events, AdaRTE acts as a web server, dynamically generating VoiceXML code. The code is sent to the VB over HTTP and a local network connection. The VB, in turn connected to telephony hardware, will be in charge of interpreting documents generated according to user's interaction over the phone. The browser captures and recognizes the answers, and streams them back to AdaRTE through an HTTP *post* request.

Prompts, questions and other elements are the nodes (here named *blocks*) of an Augmented Transition Network (ATN) that specifies the flow of the conversation. Blocks, shown graphically in Figure 3, are represented in the description by XML tags. When the system is started, the XML dialogue description is read by AdaRTE which maintains an internal representation of the dialogue, and executes it when a call comes in. Consequently, it activates the dialogue blocks in sequence or according to a specific

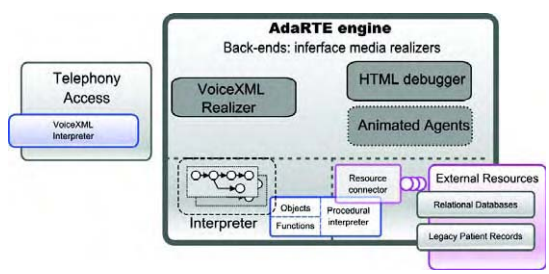


Figure 2 - AdaRTE architecture block diagram

criterion, constructs prompts, interprets the answers returned by the caller through the voice platform, and interacts with external resources as appropriate.

Usually, an ATN is associated with a context or a topic and here we call this structure *subdialogue*. The purpose of subdialogues is to help partitioning a complex application into modules. This contributes to structure the conversation layout for ease of maintenance by allowing reusing subdialogues within the same application, and makes it easy to reuse dialogue component blocks between applications (figure 4).

When a call is setup, the main subdialogue is retrieved and started; it can in its turn invoke other subdialogues, and so forth. If the execution flow reaches the end of the main subdialogue, the call is terminated. Subdialogues can also terminate unexpectedly if an exception occurs, and an exception handler is executed.

Blocks available for nesting within subdialogues are: prompt, question, script, decision, exception handler, promptset, placeholders, containers and subdialogue calling blocks.

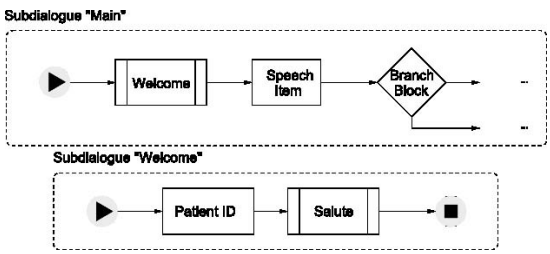


Figure 3 - Block-based dialogue description. Subdialogues are defined by the application developer (shown here as rounded dotted boxes), and can be invoked with a subdialogue calling block (shown in double-border)

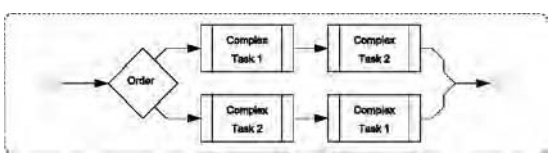


Figure 4 - Within-application reuse of blocks. Dialogue sequence can be rearranged without block duplication.

In addition, we grant the configuration of skip lists for the confirmation strategy related to questions. Containers permit designers to implement adaptability in the dialogue i.e. flexibility according to users experience with the system. Containers are used for common tasks in which one of several subdialogues is selected according to a specified policy (figure 5). Policies for activations of blocks inside containers could be: randomly, in sequence, ordered by call number and according to an externally defined schedule. Another policy could be generated by performing statistical tests to classify the level of experience of a user or even the likelihood of his/her encountering problems on specific parts of the dialogue.

Inclusion of procedural code is essential for flexibility, interoperability, and ease of programming. AdaRTE allows embedding snippets of code written in the ECMAScript standard language, into *script blocks*. The user-written code is run in a separate execution environment with extensive facilities and standard libraries. This also enables external resources access such as databases, ontologies, or any other commodity library.

Currently, the semantic recognition is implemented inside each question, and support the context-free grammar (CFG) formats offered by the VB [15]. However, we are working in the integration of a more elaborated semantic recognition solution by supporting NLP and lexicalized grammars which are more expressive than CFG formats.

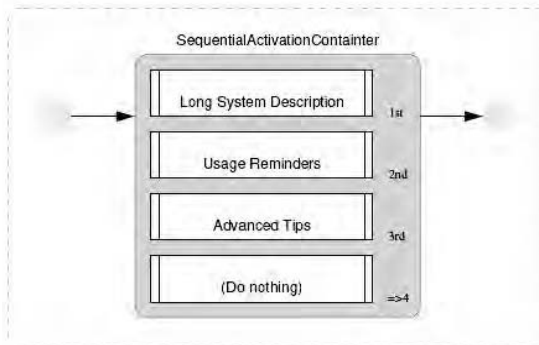


Figure 5 - Containers automate switching between homologous blocks. Switching happens according to a container-specific policy — in this case, only one contained block is activated per invocation, according to call number for that patient: a long system description is played on the first call, a brief reminder given on the second time he calls, and so on. Containers simplify the addition of variability to the dialogue

Results and discussion

The AdaRTE framework is currently in operation. It has been beta-tested with two realistic health care dialogue systems, derived by actual systems deployed and validated in the previous years. The first one is based on a prototype based on the TLC-COPD dialogue deployed in the past by the Boston MISU group and others [5]. For this specific example, we used Tellme Studio[1] as VSP. This pilot's deployment demanded less than two weeks of man effort. The fulfilled activities such as database schema definition and data preparation together with dialogue deployment are shown in figure 6a. This dialogue is executed in English language and uses keypad touch-tone (DTMF) interaction.

The second test case is the partial re-implementation of the Homey dialogue system. Homey had been deployed and evaluated in two Italian hospitals for the management of

1 Tellme Studio. https://studio.tellme.com/

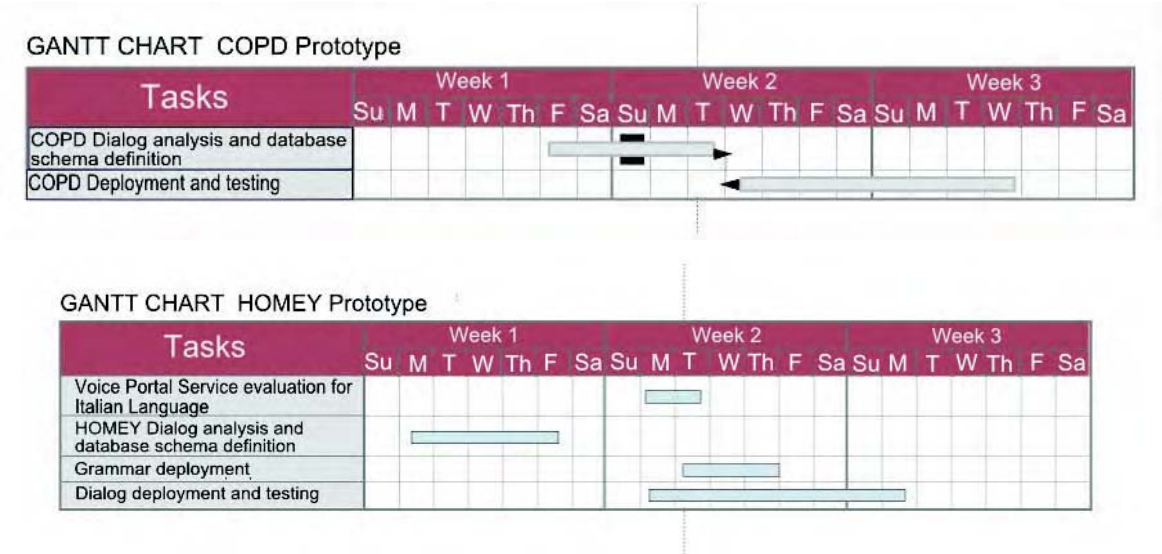


Figure 6 (a) Gantt diagram of the COPD dialogue pilot prototype. (b) Gantt diagram of the "Hypertension" pilot deployment.

hypertensive patients [6]. The system included an extensive Electronic Health Records system with storage of personal data and profiles, in order to support dialogue adaptivity. Re-engineering the system from the original proprietary dialogue manager to the AdaRTE architecture took approximately three weeks (eleven days of man effort). The development of this prototype involved the following activities: VSP evaluation, database definition and grammars and dialogue deployment (figure 6b). Unlike the TLC-COPD pilot, this system uses *speech* rather than DTMF input. We built grammars using the Nuance GSL language (Nuance 7) and SRGS grammar formats [15]. The language of the dialogue is Italian and the dialogue was deployed by using Voxpilot as VSP[2].

The expressiveness of the dialogue formalism yielded an important reduction of the time invested in developing these two prototypes. Examples of actual blocks' implementations are detailed in figures 7 and 8. Figure 7 shows the main subdialogue of the COPD partial implementation; an example of a script block which embeds a function that retrieve some patient's store data is shown in figure 8.

AdaRTE differs from other similar frameworks in that it is targeted towards the medical domain, which requires adaptable dialogs with complex structures and enquiry data collection tasks. Also, it offers a new level of flexibility to developers by allowing external resources access through the procedural features implemented inside script blocks, at the same time simple dialogs could be implemented by not expert authors. Finally, AdaRTE was thought to be a standard-compliant extensive architecture for the incremental adoption and experimentation of innovative speech technologies formalisms.

```
<script id ="2" next="3">
 <![CDATA[
        initalize_vars();
        function setPatientCode(pin){
              var conn = DBConnection.getConnection("lper");
              var statement = conn.createStatement();
              var resultSet = statement.executeQuery("SELECT nopaziente,
                    CONCAT(CONCAT(nome,' '),cognome) AS nomePaziente
                    FROM anagrafica WHERE cod_telefono like '" + pin + "'" );

              while (resultSet.next()) {
                    nopaziente = resultSet.getInt("nopaziente");
                    nomePaziente = resultSet.getString("nomePaziente");
              }
              ....
 ]]>
</script>
```

Figure 7 - Top-level dialog sequence (COPD example)

```
<subdef name="main">
  <start id="1" next="2"/>
      <subdialog id="2" next="3" name="identification"/>
      <if id="3" next="4" cond="followUpCALL == '1'">
            <else next="5"/>
      </if>
      <subdialog id="4" next="6" name="FollowUpCall"/>
      <subdialog id="5" next="6" name="dyspnea"/>
      <subdialog id="6" next="7" name="closingStatement"/>
      <catch event="dialog.finishCall" next="7">
          <start id ="1" next="2"/>
            <script id="2" next="3" >
              <![CDATA[
                finishCall = true;
              ]]>
            </script>
            <subdialog id="3" next="4" name="goodbyStatement"/>
          <end id="4"/>
        </catch>
    <end id="7"/>
</subdef>
```

Figure 8 - Procedural code in a script block

Future enhancements

Inclusion of spoken interfaces optimization techniques or *best practices* into custom-developed systems is not straightforward. A big advantage in using an interpretable and high-level dialogue representation language like the one proposed in this work is that such "dialogue practices" can be incorporated seamlessly into the underlying dia-

2 VoxBuilder. http://www.voxbuilder.com/

logue interpretation logic, removing the burden from the dialogue developer.

Furthermore, a complete project management support is foreseen, where a project involves a dialogue and its composing subdialogues, together with definition of templates. High level templates serve as guidelines in the development of abstract tasks, such as assessing the patient's psychological stage.

Currently, we have a strong commitment on the integration of a more elaborated semantic interpretation by integrating AdaRTE with a NLP application that supports lexicalized grammars to increase expressivity. In this way, not only recognition does not depend on the grammars supported by VBs, but also more natural dialogues will be supported improving the patient's perception of the dialogues.

Other components such as integration with Workflow Management Systems and facial expressions and gestures realizers must also be considered in future research.

Conclusion

We have presented an architecture for next-generation dialogue interpretations and successfully built an engine for easily dialogue deployment. AdaRTE supports the generation of the standard VoiceXML to communicate to VBs. As real-world test cases, we have reengineered two healthcare dialogue prototypes by using the novel architecture and showed that dialogue development and deployment times are remarkably optimized with respect to customized coding.

The AdaRTE system is foreseen not only as a reliable platform for dialogue deployment, but also as a framework for incorporating advanced features of speech recognizers, including increased support to adaptability, and natural language understanding and generation.

Acknowledgments

The authors thank Dr Robert Friedman (Medical Information Systems Unit, Boston, MA) for information on the TLC-COPD dialogues, and Timothy Bickmore (College of Computer and Information Science, Northeastern University, Boston, MA) for the discussions that led to the AdaRTE project.

References

[1] Migneault JP, Farzanfar R, Wright JA, Friedman RH. How to write health dialog for a talking computer. Journal of biomedical informatics. 2006 Oct;39(5):468-81.

[2] Piette JD. Interactive voice response systems in the diagnosis and management of chronic disease. Am J Manag Care. 2000;6:817-27.

[3] Corkrey R, Parkinson L. Interactive voice response: review of studies 1989-2000. *Behav Res Methods Instrum Comput*. 2002;34:342-53.

[4] Krishna S, Balas EA, Boren S, Maglaveras N. Patient acceptance of educational voice messages: a review of controlled clinical studies. *Methods Inf Med*. 2002;41:360-9.

[5] Young M, Sparrow D, Gottlieb D, Selim A, Friedman RH. A telephone-linked computer system for COPD care. Chest. 2001;119:1565-75.

[6] Giorgino T, Azzini I, Rognoni C, Quaglini S, Stefanelli M, Gretter R, et al. Automated spoken dialogue system for hypertensive patient home management. International Journal of Medical Informatics Elsevier. 2004;74:159-67.

[7] Bickmore T, Giorgino T, Picard R. Guest Editorial: Special Issue on Dialog Systems for Health Communication. Journal of Biomedical Informatics Elsevier. 2006;39:465-7.

[8] McGlashan S, Burnett DC, Carter J, Danielsen P, Ferrans J, Hunt A LB, Porter B, Rehor K, Tryphonas S. Voice Extensible Markup Language (VoiceXML) Version 2.0. 2004 [cited; Available from: http://www.w3.org/TR/voicexml20/

[9] Mittendorfer M, Niklfeld G, Winiwarter W. Evaluation of Intelligent Component Technologies for VoiceXML Applications; 2001.

[10] Black LA, McTear M, Black N, Harper R, Lemon M. Di@log by design: An Integrated Tele-Care Communication Infrastructure. In: Proceedings of the 26th Annual International Conference of the IEEE EMBS San Francisco 2004;2:3290-3.

[11] Mittendorfer M, Niklfeld G, Winiwarter W. Making the voice web smarter-integrating intelligent component technologies and VoiceXML. In: Proceedings of the Second International Conference on Web Information Systems Engineering 2002;2:126-31.

[12] Hataoka N, Obuchi Y, Mitamura T, Nyberg E. Robust speech dialog interface for car telematics service. In: First IEEE Consumer Communications and Networking Conference 2004:331 - 5.

[13] Hartman J, Vila JA. VoiceXML Builder: A workbench for investigation voice-based applications. In: Proceedings of the 31st Annual Frontiers in Education Conference, 2001 2001;3:S2C 6-9.

[14] Balentine B, Morgan DP. How to Build a Speech Recognition Application. A Style Guide for Telephony Dialogues. Second ed: EIG Press 2001.

[15] McGlashan S, Hunt A. Speech Recognition Grammar Specification Version 1.0. 2004 [cited; Available from: http://www.w3.org/TR/speech-grammar/

Address for correspondence

E-mail address: linamaria.rojas@unipv.it
(L. M. Rojas-Barahona).

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Multi-channel Physiological Sensing of Human Emotion: Insights into Emotion-Aware Computing using Affective Protocols, Avatars and Emotion Specifications

Panagiotis D. Bamidis, Andrej Luneski, Ana Vivas, Christos Papadelis, Nicos Maglaveras, Costas Pappas

Lab of Medical Informatics, The Medical School, Aristotle University of Thessaloniki, Greece

Abstract

This paper introduces a methodology for combining multi-channel psycho-physiological recordings of affective paradigms into a framework where the scientific results of such experiments are utilized in the human computer interaction context to model the computer's response based on the emotional context of the user and the situation. An affective protocol is described the results of which are expected to be combined with anthropomorphic avatars that enhance the man-machine interaction. The technological infrastructure of the later component is provided by means of XML specifications of signal descriptions and emotion recognition, as well as avatar behavior generator descriptions.

Keywords:

affective computing, emotion identification, specification, avatar technology.

Introduction

The interaction between humans and computers (HCI) has been a subject of research and discussion for a long time. The goal in the enhancement of HCI is getting it closer to the interaction between humans [1]. The essential human ability, and the one that nearly all the research is focused on, is human intelligence and its incorporation into computers. The HCI has experienced the introduction of facial and speech recognition, natural language recognition, as well as, software intelligent agents than can learn and reason on their actions [2]. However, until recently, emotions were rarely the topic mentioned in the discussions of human intelligence and their application in computers was not even considered. The reason was that emotions were considered as extraordinary human ability. Nowadays, however, there are many proofs of the importance of the emotions in the expression of intelligence [3]. The arguments put alongside the significance of emotions gave birth to a new area of "emotional intelligence", defined as "the capacity to understand emotional information and to reason with emotions" [3]. All the work that is done with computing and is related to, arises from or deliberately influences emotions on computers is called "affective computing" [4]. Affective computing has been a hot topic in recent years mostly because it introduced a new area in computing and therefore an increasing number of applications in that area. Until recently research has been mostly focused on monitoring user emotional reactions and trying to distinguish between different emotion categories such as fear, anger, sadness, happiness etc. The ability of recognizing human emotions requires the computer to monitor the user and based on certain parameters or conditions classify his/her emotional state. Since we are intending to copy the essence of human-human interaction, it is mandatory to identify the natural way of accomplishing this task by humans. Namely, human beings use several ways of communication and emotion recognition using their biological sensors of sight, touch and sound. Facial communication has proven to be of unique importance since facial expressions help in a great deal in identifying people's emotions [5]. Moreover, vocal communication as well as gesture based can be successful, to certain extend, in sensing and distinguishing between several classes of emotions.

Apart from using human natural senses, emotions can be recognized by monitoring psycho-physiological changes in the user [6, 7]. Naturally, for this to take place certain sensors have to be used recording heart pulse/activity (e.g ECG), galvanic skin response (GSR or SC), respiration rate, electroencephalogram (EEG) etc. This type of emotional reaction recording has been most attractive recently and it is the main focus in this project as well. While in the field of medicine the influence of emotions on the human health is of major concern, in human-computer interaction the focus is on understanding the user's emotions and improving the quality of software programs accordingly.

Understanding emotions is far from simple due to the complexity of human physiology. It would be so if each emotion category was characterized by a specific physiological pattern. However, certain emotions, even as different as love and fear, can cause similar effects on human physiology and, therefore, may be misrecognized. The emotion recognition process follows a sequence of steps even for the preparation of data before the recognition process. The initial step is to acquire the data from the user physiological signals. Following is the extraction of the key features of the signals [8]. The features form the basis of the comparison method. The feature extraction procedure can vary and depends on the goals of each specific project and therefore cannot be discussed in general. The final step in emotion recognition is performing

the classification of the emotional data into emotion categories based on specific classification techniques [9]. Mostly used pattern classification methods are the Hidden Markov Model [10], Fisher linear Projection [11], Support Vector Machines [9] etc. However, these techniques are appropriate for classification of large amounts of data into categories. Moreover, there is a clear need for collections of representative emotional signals in short samples of 20-50 seconds containing digital signal data, in a flat file format such as that used in the MIT-BIH file library [12]. There exists also a need for a more representative way of specifying an emotion into one data record. An XML file containing the required signal data was successfully used in [9], thereby providing evidence that an XML based representation of the emotion elements is a suitable format of data specification. Furthermore, the contents of the XML data record are of great importance for interconnection among different research results, platform independency and reusability purposes, which can be used in telemedicine, decision-making etc. [13]. Thus, it is essential to follow certain standards or widely accepted guidelines for the structure and names of the elements in the record. One related standard was introduced in [14]. Basically, it is an introduction of a markup language – ecgML for modeling and storing ECG data of patients based on XML representations.

Significant progress has been made in all the above fields and therefore new ideas arose which basically provide answers on the question "What can the computer do after recognising human emotions?". Lisetti and colleagues [15] have recently investigated the use of agent-centered modalities or modes (avatars), and multimodal feedback given to a system user. For example, an interface agent for an e-health system session can display empathy via an anthropomorphic avatar who adjusts its facial expressions and vocal intonation according to the user's emotional state, as the latter is depicted by the set of measurements.

In the light of the previous developments, the scope of this paper is twofold. First, to present a step-wise approach to the design of an experimental protocol that aims to enable multi-channel physiological sensing of a subject's emotion. The second goal is to prepare the theoretical and technological grounds for later direct adaptation of computer user interfaces guided by the elicitation and identification of emotions. The former goal is attained through the use of physiological sensors like EEG, ECG, and skin conductance, while the latter one is mainly driven by the notion of avatar technology and the preparation of the relevant technological platform that will exploit the results of the experimental analysis.

Material and methods

The AFFECTION project context

This piece of work is part of a collaborative project, called AFFECTION, between the Lab of Medical Informatics at the Medical School of the Aristotle University of Thessaloniki, Greece, and the Brain Science Institute of RIKEN in Japan. The project aims at creating a scientific foundation for the robust identification of human emotional states through fusion and correlation of data from a multi-sensor research framework. The emotion-related research findings will be subsequently incorporated within usability evaluation methodological frameworks to enable new objective/direct evaluation methodologies, as well as, new interface adaptation strategies. The project is envisaged to contribute to dynamic characterization and recognition of the subjects' emotional state upon interaction with computer systems. To achieve this, the project will employ multimodal recordings such as vocal expressions and physiological signals of the autonomic (heart rate, blood pressure, skin conductance etc) and central nervous system (EEG, MEG). Research will be carried along the following main stages: in the first, procedures, examples and baselines of multiple-channel emotional responding approaches are utilised to obtain fundamental inter-correlations of the various physiological measures in the light of behavioral data. The last stage involves the use of specific paradigms in order to incorporate the findings into a unified framework that ascertains the changing emotional state of the user and allows for adaptation of computer user interfaces based on the user and application context. A block diagram of the project is given in Figure 1.

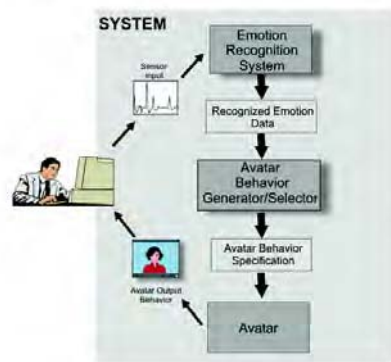


Figure 1 - A block diagram of the AFFECTION system

Affective protocol design

As already mentioned, emotions are complex phenomena, which include a wide range of observable behaviours, expressed feelings that are private and subjective, and changes in body states (distinctive somatic and autonomic responses). To achieve a first insight and start building on the scientific foundation of the AFFECTION project, the initial study aims at investigating the relationships between the different patterns of brain activation, autonomic responses, subjective experience and cognitive behavior related to emotionally-evocative photographs from the International Affective Picture System (IAPS) [16]. 50 healthy adults ranging in age from 18 to 40 years old form the participants' sample. Before the experiment, participants are updated and asked to sign a consent form. A full self evaluation questionnaire is given to them right afterwards, in order to have a (subjective) definition of the emotional state in general.

In Phase 1, we study the psychological and neurophysiological correlates of passive exposure to emotionally-

evocative stimuli. Four blocks of emotionally-evocative stimuli are selected from the IAPS, taking into account mainly two major dimensions, pleasure and arousal. Pictures are selected from the "Affective Space" as defined by the mean rating on the Valence (pleasure) and Arousal dimension. Each condition of "Affective Space" is manipulated between blocks, and the order of the Affective Space-block is counterbalanced across participants to avoid the order effect. In each trial, each stimulus is presented during 1 second. No inter-stimulus interval is scheduled for these blocks. There are 40 trials for each block, randomly selected from a larger set of pictures from each affective space condition. After each block, subjects fill in a questionnaire indicating how they feel right after this sequence block. In the second phase, the same pictures are used in a visuospatial attention paradigm, in order to investigate the effect of emotional processing on cognitive behavior. The independent variables are Affective Space block condition, and the visual field (left or right). On each trial, a central fixation cross appears for 500 ms followed by two pictures (one with high or low mean rating on valence and arousal and the other one with middle mean rating on both dimensions) for 500 ms. Then a target (a small asterisk) appears on the location of the emotional picture or in the location of the neutral picture (see Figure 2). The target remains until a response is made or until 2000 ms elapse. Participants are required to detect the appearance of the target; this can be done by pressing a key. During the experiment, recordings of the 10-20 EEG are conducted. In addition, simultaneous recordings of GSR, EOG, ECG are also taken.

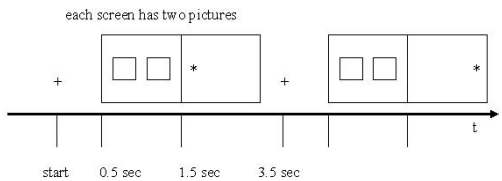


Figure 2 - Phase 2; block sequence of the affective protocol

The emotion recognition subsystem

Figure 3 shows a Logical View of the system Architecture.

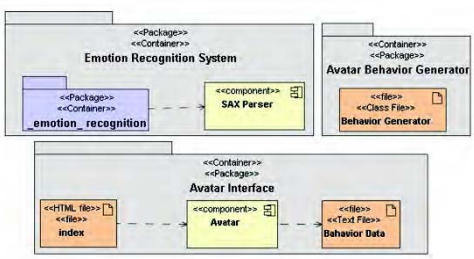


Figure 3 - Design model – logical view of the architecture

The emotion recognition system component is the most important element of the system since effective task execution of all other system components depends on it. It is the first component of the execution sequence and the component that interacts with the environment outside the system i.e. the component that accepts the input from the user physiological signals. As mentioned before, that input comes in the form of XML data specifications. There are two input requirements that this system needs. The first is the filename of the XML representation for the input emotion to be recognized and the other is actually the invocation of the recognition procedure. The operation of this subsystem can be divided into three main tasks, namely, reading user data; recognizing the emotion by comparison to its knowledge base; outputting the recognizing emotion into a proper format.

The emotion recognition system currently requires different types of physiological data in order to perform a successful recognition on the user emotion: Electromyogram (EMG), Blood volume pressure (BVP), Skin conductivity (SC), Respiration rate (RR), Electrocardiogram (ECG). It is envisaged to include EEG in the near future to fully accommodate the needs of the affective protocol analysed above.

Apart from the signal data on which the actual recognition method is applied, additional information is required by the system. This is: (i) User identification (e.g. personal information, possible medical information etc), (ii) Signal sample data (e.g. time length, measuring unit etc.), and (iii) Measurement information (recording time, date, place, temperature etc).

Since the input to the emotion recognition system is an XML file having all the above required data, an internal part of this system is a parsing functionality that extracts only the data needed for the recognition process.

Emotion recognition is the key feature of this system and the "reasoning" part of it. The reasoning is actually a comparison between the emotion data obtained from the user and the emotions that the system can recognize. The number of these emotions can be variable and depends merely on how many XML data files are stored, each representing one "classical" emotion such as fear, happiness, sadness etc (Figure 4). The comparison is done between each of those emotions and the input emotion data by comparing the four signals independently. A simple method of comparison is chosen at the time being, as the scope is to merely test the feasibility of the overall methodological approach. To be more precise, following an extensive literature review on emotion recognition using pattern recognition techniques, sample signals were extracted for several emotions (anger, fear etc.). These samples can be considered as characteristic signals of the appropriate emotions and further matching can be done by simply comparing any new signal to each of them. The comparison is done using a weighted version of the dot product (cosine matching) technique on data samples of the signal. For each signal type we chose a number of characteristic features like the mean value, the number of peaks, Average amplitude of the peaks etc.

Avatar behavior generator

This subsystem is quite simple. Its one and only goal is to generate a file with specific parameters that will instruct the avatar behavior. The reason for considering it as a separate component is its functionality. The reaction of the entire system to the input emotion depends on this component, and for experimentation purposes in this project it is created with a simple classification function. The basic idea of the classification function is to discover the level of emotional reaction and to connect that to the output behavior. For example, if the emotion recognition system gave as output that it recognized anger at user side and with 80% correctness or the user was feeling very angry, this component needs to find out which parameters are the appropriate for such a strong emotional reaction on the user side. This component takes two inputs from the emotion recognition system: the recognized emotion and the comparison value. There is much freedom in the decision for the output behavior. In different studies or experiments, the implementation of this component varies. In this first project demonstration, it was decided only to mimic the emotional reactions of the user, in the facial area, since the avatar represents only facial human-like characters. The output is a file that has a predefined structure, since the avatar is deployed into this system and has specific requirements that cannot be changed. The Haptek platform was used for the avatar creation. The avatar has to be embedded into a web page and thus there are two kinds of inputs. First, is the input that the user gives to the web page for running the avatar; the second input is the file containing the behavior parameters. The web page contains the avatar and one button for starting the avatar behavior.

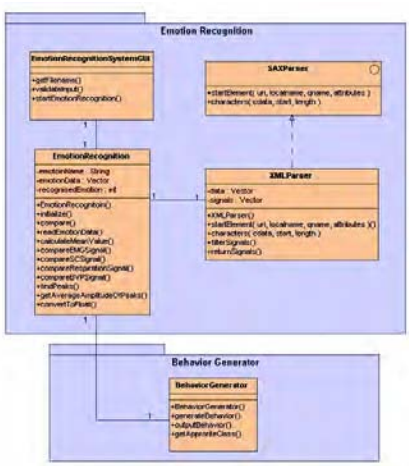


Figure 4 - Design model (class diagram) – emotion recognition container and behavior generator

Results

The emotions stored in the emotion recognition system, as well as, the input emotions are represented as XML files. The file is a modified representation of the standard for ECG data representation described in the literature review. The main element in the file is the EmotionRecord. It contains four records, each corresponding to the physiological signals, taken from the user emotion for example, EMG, BVP, SC, and RR in the current version. Furthermore, the record element consists of the record data (the main element) and some additional information such as the recording device, the recording date etc. Figure 5 shows a portion of the XML specification. An example of the comparison based on the features of the signals is given in Figure 6 only for the skin conductance case, since the implementation of the rest of the features is quite similar. The avatar implementation is a combination between the HTML page representing the interface with the user, the scripting files with the avatar code and the external data file for the avatar behavior. The appearance of the avatar is possible by installing the Haptek Player plug-in, which can run the Haptek HyperText commands as the interface between the HTML and the JavaScript functions. The behavior of the avatar and its output to the user is triggered by pressing a kind of "Run button" at the moment, but it is envisaged to be automatically triggered in future versions.

```
…
<EmotionRecord studyID="ECG00001">

      <Record type="EMG">
      <!--Recording device can be here-->
      <RecordData>
            <Channel>MLII</Channel>
            <Waveforms>
            <XValues>
            <XOffset dataType="time">00:00:00.000</XOffset>
            <Duration dataType="time">00:05:06.000</Duration>
            <SampleRate unit="Hz">20</SampleRate>
            </XValues>
            <YValues>
            <RealValue>
            <From dataType="time">00:00:00.000</From>
            <To dataType="time">00:00:10.000</To>
            <Data>1.15,2.50,2.10,2.15,2.10,2.15 … </Data>
            </RealValue>
                  </YValues>
            </Waveforms>
      </RecordData>
      </Record>
```

Figure 5 - XML specification of one record

```
public void compareSCSignal()
{
      Vector storedSCSignal = (Vector)emotion.elementAt(2);
      float x=0;
      int numberOfSCR = 0;
      int numberOfPeaks = 0;

      for(int i=1; i<(storedSCSignal.size()-1); i++)
      {
            if( (Float)storedSCSignal.elementAt(i) >
        (Float)storedSCSignal.elementAt(i-1) &&
                  (Float)storedSCSignal.elementAt(i) >
        (Float)storedSCSignal.elementAt(i+1))
            {
                  numberOfPeaks+=1;
                  if(checkForSCR(storedSCSignal,i) == true)
                  {
                        numberOfSCR+=1;
                  }
            }
      }
}
```

Figure 6 - The compareSCSignal method for SC features

Discussion

The intention of this paper was neither to introduce a new way of identifying an emotional signal, nor to put across a new affective protocol, but rather to carefully explain the individual methodological steps for the scientific exploitation of affective computing solutions. To this end, the

affective protocol design, currently under data collection, will provide the means to build up a knowledge base of emotional signals. One of the main strengths of this project was the idea for XML specifications of emotional data for the user. This can introduce new ways of emotion recognition by qualitative classification of emotions and not only by statistical quantitative experimentation. Furthermore, the contents of the XML data representation are represented by a modified standard for ECG patient data representation used elsewhere.

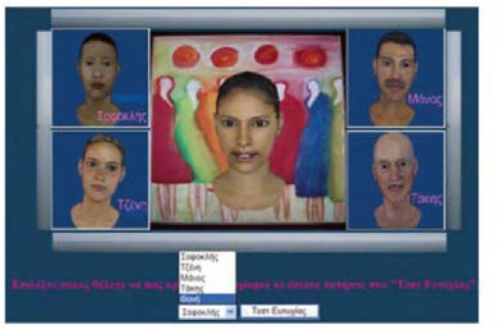


Figure 7 - Avatar appearance after emotion identification

The standard representation was modified to satisfy the needs of the emotion XML specification for this project. Obviously, the current XML specification of emotion data can be significantly improved by adding elements that will more accurately describe physiological signal data rather than only having numerical signal representations (e.g. EEG feature descriptions). Furthermore, the emotion recognition method based on the dot product comparison of the features can become more consistent so that it represents a proven classification of the emotions based on the results of the psychophysiological (affective) experiments. The generator of the avatar behavior is the component that can be also easily extended. Improvements can be envisaged in the "reasoning" method for generation of behavioral parameters. In specific, if the component is used in relation to larger amounts of emotion categories, it will be obviously more efficient and it can produce more classes of behavior parameters appropriate for the specific emotions. Finally, the avatar itself might go through several enhancements in its appearance to the user. The ultimate goal of embedding emotional awareness into the computer is to produce a system that can recognize emotions, and respond intelligently and appropriately in real-time, just like humans do. This paper provides a methodology for enabling the construction of more intelligent systems based on scientific reasoning and experimental results and not mere technological artifacts. The envisaged incorporation of the findings into a unified framework that ascertains the changing emotional state of the user and allows for adaptation of computer user interfaces based on the user and application context seems an exciting future prospect of this project. The wide variety of application areas that can be associated with such a system e.g. interactive games, learning systems, e-health/home care systems, etc call for a careful continuation of the project development.

Acknowledgments

This work has been benefited from a grant from the Greek GENERAL SECRETARIAT FOR RESEARCH & TECHNOLOGY.

References

[1] Reeves B, Nass C. The Media Equation: How People Treat Computers, Cambridge Univ. Press, 1996.
[2] Ghauoi C. Human computer interaction encyclopedia, Idea Group Inc., London, UK, 2006.
[3] Mayer JD, DiPaolo M, Salovey P. Perceiving affective content in ambiguous visual stimuli: a component of emotional intelligence. J Pers Assess 1990 ;54(3-4):772-81.
[4] Picard RW. Affective Computing. MIT Press, Cambridge, USA, 1997.
[5] Ekman P. Emotion in the Human Face, Cambridge University Press, New York, 2nd edition, 1982.
[6] Hone K. Empathic agents to reduce user frustration: the effects of varying agent characteristics. Interacting with Computers 2006; 18(2):227-45.
[7] Bamidis PD, Papadelis C, Kourtidou-Papadeli C, Pappas C, Vivas AB. Affective computing in the era of contemporary neurophysiology and health informatics. Interacting with Computers 2004; 16(4):715-21.
[8] Kim K, Bang S, Kim S. Emotion Recognition System Using Short-term Monitoring of Physiological Signals. Med. Biol. Eng. Comp. 2004; 42:419-27.
[9] Katsis CD, Ganiatsas G, Katertsidis N, Fotiadis DI. Emotional Understanding Using A Wearable Multisensorial Platform. In Bamidis PD et al (Eds), Improving the Quality of Health Information: an international perspective, Proc. iSHIMR 2005:188-97.
[10] Scheirer J, Fernandez R, Klein J, Picard RW. Frustrating the user on purpose: a step toward building an affective computer. Interacting with Computers 2002; 14(2):93-118.
[11] Healey J and Picard RW. Digital Processing of Affective Signals; In Proceedings of ICASSP, Seattle, 1998.
[12] MIT PhysioNet. PhysioNet: the research resource for complex physiologic signals. http://www.physionet.org/
[13] Varni A, Kemp B, Penzel T and Schlogl A. Standards for biomedical signal databases. IEEE Eng Med Biol Mag. 2001;20(3):33-7.
[14] Wang H, Azuaje F, Jung B, Black N. A markup language for electrocardiogram data acquisition and analysis (ecgML). BMC Medical Informatics and Decision Support 2003; 3(4).
[15] Lisetti C, Nasoz F, LeRouge C, Ozyer O, Alvarez K. Developing multimodal intelligent affective interfaces for tele-home health care. Int J Hum-Comput Stud 2003; 59:245-55.
[16] Bradley MM, Cuthbert BN, Lang PJ. Picture media and emotion: effects of a sustained affective context. Psychophysiology 1996; 33(6):662-70.

Address for correspondence

Panagiotis D. Bamidis,
Lab of Medical Informatics,
Medical School,
Aristotle University of Thessaloniki,
P.O. Box 323, 54124,
Thessaloniki, Greece.

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

A Framework for Cognitive Monitoring Using Computer Game Interactions

Holly B. Jimison, Ph.D.[a,b], Misha Pavel, Ph.D. [a,b], Payton Bissell, B.S. [b], James McKanna, B.S. [b]

[a] *Oregon Health & Science University, Portland, Oregon, USA*
[b] *Spry Learning Company, Portland, Oregon, USA*

Abstract

Many countries are faced with a rapidly increasing economic and social challenge of caring for their elderly population. Cognitive issues are at the forefront of the list of concerns. People over the age of 75 are at risk for medically related cognitive decline and confusion, and the early detection of cognitive problems would allow for more effective clinical intervention. However, standard cognitive assessments are not diagnostically sensitive and are performed infrequently. To address these issues, we have developed a set of adaptive computer games to monitor cognitive performance in a home environment. Assessment algorithms for various aspects of cognition are embedded in the games. The monitoring of these metrics allows us to detect within subject trends over time, providing a method for the early detection of cognitive decline. In addition, the real-time information on cognitive state is used to adapt the user interface to the needs of the individual user. In this paper we describe the software architecture and methodology for monitoring cognitive performance using data from natural computer interactions in a home setting.

Keywords:

Computer monitoring, cognitive assessment

Introduction

Cognitive performance is a key health concern of elders in the United States. In fact, maintaining cognitive health is often the most important factor in being able to age in place. Nearly 50% of all people over the age of 85 are found to have a measurable decline in cognitive function [1]. However, common clinical practice does not offer methods for detecting cognitive decline at an early stage, when therapies may be more effective. Recent research has demonstrated the importance of detecting cognitive decline in an early stage [2-4]. Some cognitive issues have immediately treatable causes, such cognitive disturbances due to medication interactions or short-term medical conditions. However, even with long-term conditions, such as dementia, there are many new therapies that researchers presume would have improved efficacy with earlier detection. In this paper we describe a framework for using unobtrusive computer interaction data to infer cognitive changes on the part of computer users. Frequent assessments allow us to detect relevant changes in various aspects of performance that can be used to adapt the user interface in real time and also provide a mechanism of early detection of cognitive problems.

Growing use of computers by elders

Elders are the fastest growing demographic of new computer users in the United States. In a recent survey conducted by the Pew Internet and American Life Project [5], they found that 22% of American adults over the age of 65 use the Internet. Interestingly, elders in this group are even more likely than other Internet users to go online and check email each day [5]. In addition, nearly 35% of elders who use a computer have played a game online, comparable to 39%, the average rate of computer game play for other age groups. Given this rapid growth of computer use by users at risk for cognitive problems, as well as the current large use of computers by the advancing wave of baby boomers, we have an important opportunity to collect and interpret naturalistic computer interaction data for diagnostic purposes. In this project on cognitive monitoring using computer interaction data, we have focused on the interpretation of interactions in computer games that we have specifically designed to probe cognitive performance.

Current methods of cognitive assessment

In standard clinical practice, cognitive screenings are usually performed only at advanced age or if there are already patient or family concerns about cognitive dysfunction. These screening tests, such as the Mini-Mental State Exam, the Kokmen Short Test of Mental Status, and the Memory Impairment Screen, can be performed in a physician's office, but are fairly coarse and not particularly useful for the early detection of problems [6]. More complete neuropsychological batteries can be performed to obtain more sensitive diagnostic information. These normally include measures of short-term and working memory, divided attention, motor speed, planning, and general executive function. In our project on designing computer games for cognitive monitoring, we attempted to incorporate proxies for the following standard tests of cognitive performance:

1. **Verbal Fluency -**The test is focused on semantic processing and recall from long term memory (LTM). The test procedure requires the participants to recall as many words as possible given a specific semantic category or one or more phonemic constraints.

Selected for best paper award.

2. **Word-List Acquisition and Delayed Recall -** The test is focused on learning and recall from short term memory (STM) as well as LTM. The test procedure requires the participants to learn and recall a list of words with three trials and then after an intervening task.
3. **Word list Recognition -** This is a test of the ability to recognize words previously presented during the Word-List Acquisition test. The participant is asked to discriminate between the words that were presented in the list from distractor words. Together with the Word-List Acquisition test, the recognition test can distinguish whether the "forgotten" items were truly lost or the memory trace was just too weak to support reliable recall.
4. **Trail-Making Test** - This test is focused on complex visual scanning, mental tracking and mental flexibility. The participants are asked to trace a sequence of digits and then a sequence of interposed digits and letters.
5. **Symbol Digit Modalities Test -** This test is used to assess the ability to sustain attention and to perform coding task. The participant is given a table associating a simple symbol with each digit and then is asked to assign a number to each of a long list of these symbols.
6. **Digit Span -** The focus of this test is working memory and sustained attention. The participant is asked to repeat a sequence of digits, starting with short sequences and then of increasing length. The following task is to do the same thing in reverse order.
7. **Finger tap test -** Although this test is focused on the speed of motor control, there is increasing evidence in the literature that this type of test is useful in predicting future decline in cognitive abilities. The participant in this test is asked to push a switch as many times as he can within a ten second interval. One feature of this test is that the results of the performance are insensitive to educational level and other demographic variables.

These neuropsychological tests are usually performed by trained psychologists and usually done no more frequently than once per year. One of the hallmarks of cognitive impairment is the increasing variability in performance. Infrequent assessments do not offer a mechanism to pick this up. In fact, the results of standard cognitive measures are clouded by a need to reference the performance metrics directly to population norms. Many cognitive tests are highly affected by differences in educational level, language abilities, etc.

In our work with monitoring computer game interactions to infer cognitive performance, we make use of these metrics of verbal fluency, short-term and working memory, planning abilities, and divided attention. However, we are able to make assessments every time an elder uses a computer. Although our computer assessments are less direct and less controlled than the standard tests, we have the benefit of multiple nearly continuous measures and can analyze within-subject trends. This substantially reduces unwanted confounding effects due to education, language abilities, and culture. In addition, we are able to characterize variability in performance over time, which in itself is a powerful indicator of cognitive function.

Materials and methods

Unobtrusive monitoring of computer game interactions

In our project on monitoring elders' computer interactions, we first performed a needs assessment to define elders' preferences for computer applications, games, and potential barriers to computer use. We used focus groups and surveys to help us define a set of features for an elder Web portal that we could use as a research environment to collect real-time interaction data. We also defined a set of enjoyable computer games that could be adapted for cognitive monitoring. To select the games for further development, we observed which features were most enjoyable and easily understood by elders and then also did a cognitive task analysis on each of the games to characterize its appropriateness for providing information on one of the cognitive dimensions described in the previous section on standard cognitive tests.

We currently monitor all keyboard and mouse interactions, both within game play, and in conventional computer applications. Each of the adaptive cognitive computer games for elders are designed to measure various aspects of the standard cognitive tests described in the previous section. For example, two of our computer games are designed to measure verbal fluency (e.g., ability to name as many animals as possible within 60 seconds).

Figure 1: A word jumble game where we measure the user's relative ability to find longer and more complex words from a set of 7 letters.

Figure 1 shows an example of a word jumble game, where the users are give a set of 7 letters and asked to generate as many words from that set as quickly as they can. They are given cues on the right of the screen to show how many words are possible. Entering a word using all 7 letters allows them to go onto the next round. Two basic metrics relating to verbal fluency can be generated from monitoring the users' interactions in this game: 1) speed of word generation and 2) the complexity of the words generated (defined by word length and frequency of use in the English language).

Figure 2: A word game where we measure the user's relative ability to find longer and more complex words in a difficult search environment.

Figure 2 shows another word game designed to measure verbal fluency, but with the additional task of search and planning. In this game, the user must connect adjacent letters to form words as quickly as possible. A higher score is given for longer words and for using highlighted letters. With the reference abilities of verbal fluency measured in the game shown in Figure 1, we are now able to quantify the increased performance requirements due to search and planning. In each of these games, the difficulty of the board layout is adapted to the skill of the user. Our adaptation algorithms keep the success rate at approximately 60%-80%, so that users are challenged, but not frustrated. This also gives us the best opportunity to measure performance, i.e., users' scores do not "top out" when the task is too easy or "bottom out" when the task is too difficult.

In these games, the user's creation of longer and more sophisticated words (against time and difficulty of available letters) we rate as having higher verbal fluency. We concentrate on monitoring relative performance (with respect to the user's baseline) to look for differences. This is likely to be a more sensitive measure that is less influenced by education and language abilities, and more influenced by cognitive changes.

For a direct measure of short-term and working memory, we adapted the Concentration card game to the computer, as shown in Figure 3. Users must remember the location of various cards they select and match pairs. Game difficulty is adapted based on number of cards and the cognitive difficulty of the matches. These range from simple shape and color matches to cognitively more difficult matches, such as matching a digital clock time with the analogue picture equivalent. For this game, we estimate memory ability using an adaptive memory buffer metric. This metric is defined by filtered estimate (similar to a moving average) of the maximum number of card flips back that a user can successfully remember seeing a target card.

We have designed other computer games to specifically test the remaining dimensions of cognition. Figure 4 shows a shape and color matching game that provides us with measures of planning (inferring the number of steps ahead a user would have to be able to plan in order to be successful). In this game we can also manipulate difficulty and provide added features to test memory and divided attention.

Figure 3: Example of a memory computer game.

Figure 4: Color and shape matching game that tests planning ability, memory and attention.

In addition to the games where we adapted activities that elders already found to be enjoyable, we also took standard tests like the Trail Making Test and adapted it to be fun. In this game, users use the mouse input device to select circles in numerical sequence, letter sequence, and mixed targets. This activity is similar to the Trail Making Test in requiring several dimensions of cognitive executive function, including visual search, attention and set switching. Each of these additional components can be extracted and assessed with repeated game use. These more frequent measures allow us to monitor within-subject trends and also monitor performance variability. Both of these features offer promise in being able to detect cognitive problems earlier, and potentially, more reliably.

Results

Most of our experience and testing of computer games for cognitive monitoring has come from our work with an implementation of the popular Solitaire game of FreeCell, as shown in Figure 5. We found that this game was by far the favorite with the elders that we interviewed and it was the first computer game we adapted for use in cognitive monitoring. In our research version, we compare user performance to our computer solver. The lower graph of Figure 5 shows the game difficulty starting at 82 moves to

optimal solution, with the lower line showing the computer solver's direct path to solution. The upper line shows the subject's moves going toward and away from best solution. We use the slope of the subjects performance as a measure of efficiency of play. In our early pilot work comparing FreeCell performance of cognitively healthy elders to those with diagnosed mild cognitive impairment, we were able to use the efficiency metric to distinguish the two groups.

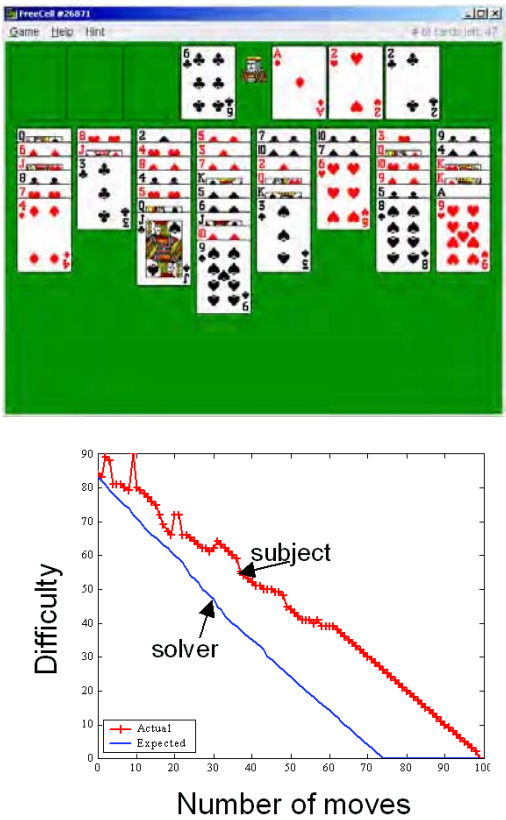


Figure 5: Sample game of FreeCell (Solitaire game requiring significant planning) and a diagram showing subject performance vs computer solver.

Table 1 shows the results of our early pilot tests to show the feasibility of monitoring computer interactions in the home. We monitored 12 elders in a local senior residential facility for a period of 3 weeks. Using conventional neuropsychometric tests described earlier, we found that 3 of the elderly subjects (mean age 80.2 +/- 8.0) had mild cognitive impairment. Using only data from their FreeCell performance we were able to distinguish cognitively healthy subjects from those with mild cognitive impairment. Interestingly, the variability of the measures over time was in itself a useful feature in classifying cognitive impairment.

Table 1: FreeCell efficiency cognitive metric scores for 9 cognitively health elders and 3 elders with mild cognitive impairment.

	Ave of Subjects' Ave Efficiency	**SD of Subjects' Ave Efficiency**	**Average of Subjects' SD Efficiency**
Normals	0.58	0.12	0.38
MCI	0.27	0.72	0.55

Subsequent to the pilot test of the FreeCell game, we deployed the full set of 9 adaptive cognitive computer games into the homes of 30 elders. The average age of our participants is 80.4 ± 6.0 years. Most are female (83%) and have an average level of 15.2 ± 2.7 years of education. In this evaluation, we have demonstrated that we are able to extract cognitive measures from routine game play on the part of elders. The play the games successfully and on debriefing, report that they enjoy the experience. The plot in Figure 6 shows their game usage patterns over a period of 3 months.

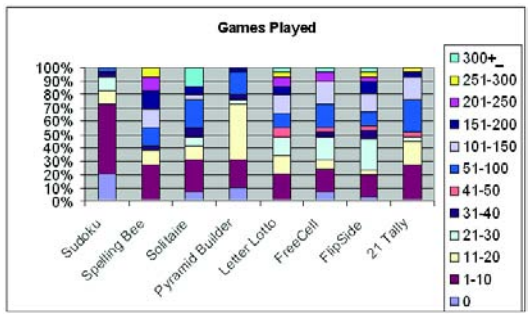


Figure 6: Computer game usage by 30 elders over a period of 3 months.

Software Architecture for Cognitive Monitoring

We have developed a rich set of tools for assessing cognitive performance based on the unobtrusive collection of computer interaction data. Our measures are based on keyboard and mouse interactions for both cognitive computer games and conventional applications. The measures include metrics of verbal fluency (word processing and word games), motor speed (login typing, game speed), memory, attention, planning and general executive function. Figure 7 shows our general software architecture for collecting and analyzing the monitoring data.

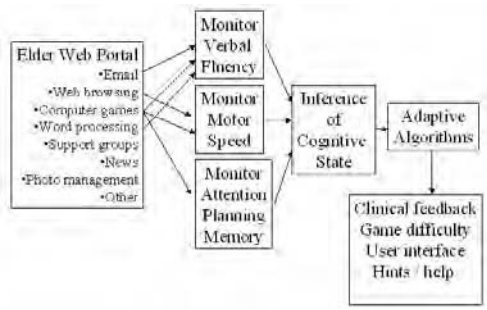


Figure 7: Overview of software architecture for cognitive monitoring.

Real-time analysis of game data takes place on the elder's local machine. This data is used to adapt the difficult of the ongoing computer game (in some cases where appropriate) and also used to adapt the level of difficulty for a user's upcoming games. We also use real-time analysis and feedback to tailor hints and help messages as part of the user interface. If we realize that a user is having memory problems or divided attention problems, we are then able to immediately adapt our user interface. Most importantly though, our work on cognitive monitoring is designed to provide clinical feedback to the elder. Based on the elder's preferences, he or she may choose to share this information with caregivers and clinicians.

Conclusion

We have demonstrated a software architecture for real-time unobtrusive monitoring of computer interactions for the purpose of inferring cognitive performance. This approach offers substantial benefits in being able to measure within subject changes over time in a natural setting. Our ability to detect trends in cognitive performance offers the possibility of early detection, both of future cognitive decline that could be treated early, and of near-term effects of medication interactions or more acute illnesses. Our early results demonstrate that elderly computer users enjoy the games and play them frequently. We also have early evidence of our metrics being able to distinguish between cognitively healthy elders and those with mild cognitive impairment. Our hope is that this monitoring information may be an inexpensive way of facilitating cognitive health management for elders, helping them maintain their quality of life and independence.

Acknowledgments

This work was supported by NIST ATP, the Intel Corporation, and by ORCATECH, Oregon's Roybal Center for Aging & Technology with a grant from NIA (P30AG024978).

References

[1] Garber AM. Long-term care, wealth, and health of the disabled elderly living in the community. In: Wise DA, ed. The Economics of Aging. Chicago: University of Chicago Press, 1989.

[2] Social Security Online, http://www.ssa.gov/OACT /TR/ TR04/ index.html, last update March 23, 2004.

[3] US Census Bureau, www.census.gov/population/socdemo / age.html, last revised: November 04, 2003.

[4] Kasper JD. Using the long-term care surveys: longitudinal and cross-sectional analyses of disabled older people. Proceedings of the 1987 Public Health Conference on Records and Statistics. DHHS Pub.No.88-1214. Hyattsville, MD: National Center for Health Statistics, 1988, pp. 353-358.

[5] Kiernan R, Meuller, J, Langston J, and Van Dyke C. The neurobehavioral cognitive status examination: a brief but differentiated approach to cognitive assessment. Annals of Internal Medicine 1987:107: 734-739.

[6] Petersen RC, Stevens JC, Ganguli M, Tangalos EG, Cummings JL, DeKosky ST. Practice parameter: early detection of dementia: Mild cognitive impairment (an evidence-based review): Report of the Quality Standards Subcommittee of the American Academy of Neurology. Neurology 2001; 56(9):1133-42.

[7] Chen P, Ratcliff G, Phil D, Belle S, Cauley J, De Kosky S, and Ganguli M. Cognitive tests that best discriminate between presymptomatic AD and those who remain nondemented. Neurology 2000: 55: 1847-1853.

[8] Callahan CM, Hendrie HC, and Tierney WM. Documentation and evaluation of cognitive impairment in elderly primary care patients. American College of Physicians 1995; 122: 422-429.

[9] Mitrushina MN, Boone KB, and D'Elia DF. Handbook of Normative Data for Neuroopsychological Assessment. Oxford University Press, 1999.

[10] Howieson d, Bowman S, Coehlo D, Lim S, and Kaye J. Natural history of cognitive decline in the old old. Neurology 2003; 60: 1489-1494.

[11] Geldmacher DS and Hills EC. Effect of stimulus number, target-to-distractor ratio and motor speed on visuospatial search quality following traumatic brain injury. Brain Injury 1997; 11(1):59-66.

Address for correspondence

Holly B Jimison, PhD
Department of Medical Informatics and Clinical Epidemiology
BICC 507, Oregon Health & Science University
3181 SW Sam Jackson Park Rd
Portland, Oregon 97239 USA
jimisonh@ohsu.edu

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Mobile Phone Computing for In-situ Cognitive Behavioral Therapy

Magnus Bang[a,b], Toomas Timpka[c], Henrik Eriksson[a], Einar Holm[d], Conny Nordin[e]

[a] *Department of Computer and Information Science, Linköping University, Sweden*
[b] *Interactive Institute, Stockholm, Sweden*
[c] *Department of Social Medicine and Public Health, Linköping University, Sweden*
[d] *Department Social and Economic Geography, Umeå University, Sweden*
[e] *Department of Neuroscience and Locomotion, Division of Psychiatry, Linköping University, Sweden*

Abstract

Cognitive behavioral therapy (CBT) for psychological disorders is becoming increasingly popular on the Internet. However, when using this workstation approach, components such as training and learning relaxation skills, problem solving, exposure exercises, and sleep management guidance must be done in the domestic environment. This paper describes design concepts for providing spatially explicit CBT with mobile phones. We reviewed and analyzed a set of treatment manuals to distinguish elements of CBT that can be improved and supported using mobile phone applications. The key advantage of mobile computing support in CBT is that multimedia can be applied to record, scale, and label anxiety-provoking situations where the need arises, which helps the CBT clients formulate and convey their thoughts and feelings to relatives and friends, as well as to therapists at subsequent treatment sessions.

Keywords:

mental health informatics, ubiquitous computing, cognitive behavioral therapy, social geography

Introduction

Psychiatric disorders are common in the industrialized world, which is illustrated by the observation that the cost of treating anxiety alone exceeds an estimated $40 billion annually in the United States [1]. Rehabilitation of individuals suffering from long-standing psychiatric and psychological disorders is often obstructed by insufficient coordination of the services provided by health care workers and public welfare agencies [2]. Thus there are strong incentives for developing self-support programs for this category of patients.

Cognitive behavioral therapy (CBT) is a widely acknowledged approach for treatment of numerous psychological disorders [3,4,5], and its efficacy has been demonstrated in several empirical studies and clinical trials [cf., 6,7].

CBT usually involves components such as motivation and education, relaxations skills, problem solving training, exposure exercises, and sleep management training. However, barriers related to accessibility and affordability have prevented many people from receiving this form of treatment [8]. In response to these problems, computer-mediated therapies have been introduced and used for more than a decade [9]. More recently, this strategy has also reached the Internet ("interapy") and definitely shows potential, particularly in terms of cost-effective distribution [10]. The online therapies typically contain self-help instructions as well as exercises that comprise either unstructured patient-therapist interactions or structured manual-based training with little, if any, therapist involvement [11,12].

A clear drawback of most of the current Internet-based therapies is that interaction is restricted to stationary computers. This may decrease the efficacy of training, cognitive restructuring, and learning, because the exercises become too abstract when performed in a decontextualized environment. Interestingly, a central tenet of CBT is self-managed recording of anxiety-provoking situations and exposure training conducted in real contexts. The basic hypothesis of our work is that, under appropriate circumstances, portable computer tools such as mobile phones have the potential to significantly improve both traditional CBT (*in situ*) and internet-based therapies.

As a step towards *using mobile phones as a component of in situ treatment,* we present a set of design requirements and concepts for 3G/4G mobile phones that emanate from an analysis of CBT manuals for generalized anxiety disorders (GADs) and information regarding the theory of social geography [13]. First, we outline the fundamental concepts of CBT, and thereafter we discuss our two approaches to design: ubiquitous computing and context-aware computing. We conclude our paper by discussing how modern mobile phone technologies like cameras and satellite positioning can be integrated into CBT self-help programs for clients with various psychiatric disorders such as agoraphobia and GAD.

Background

Cognitive behavioral therapy

CBT is used to treat mental disorders by attempting to modify negative belief systems and associated behaviors

[3,4,14]. The underlying concept is that distorted thoughts have a deleterious impact on emotions, and hence the objective of CBT sessions is to help people revise negative thinking and adopt new and healthier thoughts and actions. The adequacy of CBT for treating depression and generalized anxiety disorders is particularly well documented and acknowledged in the research community.

The components of CBT vary depending on the diagnosis and needs of the individual, but in all cases the general idea is to help the patient *identify and modify negative thinking patterns related to situations* that affect emotional responses. Counselors employ essentially three learning approaches in CBT: the process of *guided discovery using the Socratic method, thought diaries,* and *behavioral experiments*. Socratic questioning is a method to promote new alternative interpretations of situations. Once this process has started, the novel interpretations that emerge can be documented in a special diary [15] called a *dysfunctional thought record* (DTR). In the DTR, clients describe and keep track of common negative situations and the related emotions, automatic thoughts, and responses. Moreover, these qualities are scaled and outcomes of the situation are evaluated in light of the alternative thought patterns. The core idea is that the cognitive schema embedded in the diary will be internalized and automatically applied to situations in real life. The behavioral experiments in CBT aim to aid the clients in testing and validating held beliefs. For example, to treat people with a fear of wasps, counselors set up special experiments in which the phobia is put to the test by letting clients encounter the insects to ascertain whether the hypotheses are true (exposure therapy). Yet another important and integrated part of the treatment includes *reflective exercises and homework*. These assignments can even be done in the field, for instance to treat agoraphobia and social disorders.

Media in traditional CBT

Thus CBT is characterized by the use of a range of abstract methods and instruments such as the DTR, which are often accommodated in different types of technical artifacts. Tape recorders, pen and paper, thought diaries, activity schedules, and questionnaires are commonly used in CBT sessions and exercises [14]. Considering the example of management of phobias related to spiders and wasps, both models and living specimens of these arthropods are used in exposure treatment sessions. Counselors often give the client a ring binder at the onset of therapy to facilitate organization of materials as they are introduced sequentially during the treatment period. This binder contains the session outlines, the DTR, summaries, handouts, and homework instructions and exercises. Therefore, an important research question is how to replace the mentioned traditional tools with improved digital alternatives that preserve the cognitive schemata and their incorporated restructuring techniques.

Ubiquitous computing and mobile phones

Ubiquitous computing represents a shift from conventional desktop-based computing and solutions towards mobile computing and attempts to blend digitization into ordinary physical tools and environments [16]. The goal is to create a new class of user interfaces that more appropriately match our everyday activities and information needs and that do not require stationary work at a computer. A special area within this field is called context-aware computing [17]. According to Guanling and Kotz, mobile context-aware applications can identify and take advantage of contextual information in the physical environment, such as user location, time of day, nearby people and devices, and the user's activity [18]. Some examples of such applications are GPS-directed maps and special mobile games that utilize wireless Bluetooth as a proximity sensor to detect other gamers and active artifacts in the vicinity (social games).

Mobile phone applications are central components of ubiquitous and context-aware computing. Today, mobile phones are becoming increasingly sophisticated to include Internet access, cameras, web browsers, advanced media capabilities such as MP3, and features for GPS positioning (cf., Nokia N95). Some phones also have calendars and special fitness applications with motion sensors for use as a pedometer (cf., SonyEricsson W710i). Interestingly, it seems that mobile phones can provide suitable support for CBT treatment because they emphasize situatedness, mobility, thinking, and reflection.

Methods

We analyzed a set of manuals that are used by therapists to treat various psychiatric illnesses, particularly GADs [15,19, 20]. These manuals were chosen because our aim was to find general features of CBT to which mobile technology could be applied as a means of improving the traditional treatment components. We also used Hägerstrand's framework outlining 'social pockets of local order' as a foundation for provision of CBT *in situ* [13]. This framework departs from each person's residence and displays the different meeting places as 'social pockets'. With this framework, anxiety patients can be represented and followed in terms of, for example, movement patterns and changes in those patterns over time, as well as comparisons with normal populations. Below we describe how mobile phones can be used to specifically deliver CBT in this context.

Results

The digital dysfunctional thought record

The digital dysfunctional thought record (DDTR) is an electronic version of Beck's empirically validated DTR for recording negative situations and the related emotions and automatic thoughts on paper.

Figure 1 - Clients use the DDTR to record situations in which anxiety arises

Figure 1 shows a prototype of a DDTR for mobile phones that we implemented in Macromedia Flash Lite. Our version employs the camera of the mobile phone to document situations and places (tagged by geocode) that are experienced as unpleasant. It is also possible to add both text and audio input to comment on the pictures and short films. Scaling the anxiety level of the specific situation and media is done using the wheel on the phone (Beck scale 1–100).

A well-known problem with the CBT exercises and anxiety recording is that many patients do not have labels for their emotions, and they find it difficult to express their thoughts [14]. Researchers have suggested that clients should use *one word* to denote each type of feeling [21]. We support this approach in our application providing basic lists of definitions and synonyms for feelings that can be used during recording. From the perspective of CBT, there are some intrinsic problems with this method, which are discussed in the following section. In addition, it is imperative that the content of the DDTR can be shared with the therapist before each CBT session. Storing the media documentation as a single encrypted file makes transfer over the Internet convenient and secure. Naturally, an alternative is to transfer the files wirelessly via Bluetooth at the start of the face-to-face session. Nevertheless, we believe the main benefit of recording real situations and places for CBT is that it facilitates recall of negative thoughts, which makes it easier for the client to convey problems and situations to the therapist during the actual session.

Relaxation and sleep training

Learning relaxation techniques can benefit many patients, particularly those suffering from anxiety disorders such as a GAD. Relaxation training can reduce physiological arousal, and cognitive procedures can modify excessive worry. Basically, there are two relaxation techniques that are used to deal with anxiety in CBT: breathing training and progressive muscle relaxation [20, 22]. These components are often introduced before the actual cognitive restructuring sessions. Programs for training relaxation skills via the mobile phone are promising. For example, a therapist can give verbal instructions on how to breathe and record them on audiotape or a CD for later use by the client. It would be a relatively straightforward task to transfer this information to a mobile phone equipped with an MP3 player. There are several benefits associated with having the relaxation media pervasively available to allow meditation wherever the client prefers.

Supporting daily routines

Many psychological disorders are associated with reduced organizational skills and problems with memory recall, as a combined result of the illness and the medication. This problem leads to difficulties in managing normal events and even in remembering the visits to the therapist. Mobile phone technologies include several aids (e.g., calendars that have alarms and provide reminders) that can be available everywhere and can help CBT patients structure their days. These applications are not currently applied in CBT, but they can easily be introduced for that purpose.

Social and spatial isolation are also connected with psychological disorders. According to the theory of social pockets, activity patterns and spatial movement from home are indicators of the severity of illness. Context-aware applications that track movements in space and also the number of calls and messages that are made could constitute a useful scale for implicitly tracking progress. Special phones with step-counter and GPS motion-tracking software are already on the market. These phones measure walking distance and speed and calories burned during the day, and thus they can be used to develop an application that combines motion sensing and a calendar with alarms to aid patients with dysfunctional activity patterns. For example, a context-aware application could be designed to monitor sleeping and activity patterns and possibly also to enforce daily routines and—in cooperation with the counselor—promote a more active lifestyle.

Distraction support

Distraction techniques are often used in CBT to take a patient's mind off his/her problems and thereby lower anxiety levels. Physical activity, mental imaging, and doing arithmetic are ways of blocking the focus on feelings and directing the patient's thoughts to real-world activities and tasks. Special media such as images and certain interactive games can obviously be incorporated into the phone to offer a distracting effect. However, many therapists feel that such techniques should only be applied solely as a first-aid approach, because they can be used to avoid emotions and negative thoughts, which can have an impact on the therapeutic outcome.

Therefore, before integrating such techniques into mobile applications, studies should be conducted to evaluate short and long-term effects.

Session recording

According to Willis and Sanders [14], audio recording of counseling sessions is an invaluable strategy. In short,

giving clients the means to listen to previous sessions is helpful, because it reinforces what has been discussed and gives client useful feedback (memory). We feel that future CBT applications for mobile phones should include such basic recording software to allow patients to listen to their therapy sessions at any time and anywhere.

Discussion

Mobile digital technologies can be applied to supply spatially explicit CBT *in situ*. We believe that an appropriate starting point is to support key practices in CBT. For instance, mobile phones can help patients by making it possible to record and spatially and temporally organize anxiety-provoking situations, and also to facilitate the labeling of thoughts and feelings. As mentioned above, cognitive restructuring is a basic principle of CBT and the traditional tools such as the DTR has been developed specifically to support this. The idea is that the CBT clients will internalize these schemas and automatically apply them even after conclusion of the CBT period and throughout life. Thus, the question arises as to whether new digital versions of the CBT tools, such as our DDTR, encompass the necessary cognitive properties. The traditional diaries and scales used to record situations, as well as the schemata they embody for cognitive restructuring, have been empirically tested for decades. Therefore, efforts must be made to validate the mobile digital CBT tools by applying the same procedures used to test the traditional CBT tools.

Our DDTR has a dictionary of terms for labeling feelings, and this is included to assist users in finding appropriate words to describe and audio record their experiences. However, it is possible that this dictionary can inappropriately guide the user's choice of words. If that is the case, it could lead to false labeling, which would naturally impair the therapy. Clearly, more research is needed in this area.

We are currently developing a set of phone-based applications to support the practice of CBT, and we are also performing a small study to address the above-mentioned question regarding directed labeling. We are conducting that work from the perspective of participatory design, so that the CBT clients themselves can influence the construction of the applications.

Conclusion

In this paper, we outline a set of tools that can be part of a client mobile multimedia application to support CBT. Specifically, we describe a new way to document and label the anxiety-provoking situations that are part of regular CBT treatment. The diaries that are traditionally used to record situations, as well as the cognitive restructuring schemata they embody, represent evidence-based clinical technologies. Moving to modern media in CBT, as we suggest here, will require comprehensive evaluation. Nonetheless, we believe that new media can have a profound impact on CBT, in both a positive and a negative sense. Thus we strongly recommend that implementation of novel applications and designs for mobile CBT should be accompanied by controlled trials that assess the efficacy of these innovative strategies.

References

[1] Greenberg PE, Sisitsky T, Kessler RC, Finkelstein SN, Berndt ER, Davidson JR, Ballenger JC, Fyer AJ. The economic burden of anxiety disorders in the 1990s. J Clin Psychiatry 1999 60(7):427-35.

[2] Timpka T, Bang M, Delblanco T, Walker J. Information infrastructure for inter-organizational mental health services: An actor network theory analysis of psychiatric rehabilitation. Journal of Biomedical Informatics. In Press.

[3] Beck AT (ed). Cognitive therapy of depression. New York: Guilford Press 1979.

[4] Beck J. The past and future of cognitive therapy. Psychother Pract Res.1997; 6: 276-284.

[5] Myhr G, Payne K. Cost-effectiveness of cognitive-behavioural therapy for mental disorders: implications for public health care funding policy in Canada. Can J Psychiatry 2006;10:662-70.

[6] Mohman J, Gorenstein EE, Kleber M, de Jesus M, Gorman JM, Papp JA. Standard and enhanced cognitive behavior therapy for late-life generalized anxiety disorder: Two pilot investigations. American Journal of Geriatric Psychiatry 2004;11:24-32.

[7] Brown TA, O'Leary TA, Barlow DH. Generalized anxiety disorder. In: Barlow DH, Ed. Clinical handbook of psychological disorders. NewYork: Guilford, 2001.

[8] National Institute of Health. *Treatment of panic disorder.* Paper presented at the NIH Consensus development conference consensus statement (Volume 9), Bethesda, MD.

[9] Marks I, Shaw S, Parkin R. Computer-aided treatments of mental health problems. Clinical Psychology: Science and Practice 1998;5:151-170.

[10] Lange A, Schrieken B, van de Ven JP, Bredeweg B, Emmelkamp PM, van der Kolk, Lydsdottir L, Massaro, M, Reuvers A. "Interapy": The effects of a short protocolled treatment of posttraumatic stress and pathological grief through the internet. Behavioural and Cognitive Psychotherapy 2000;28:175-192.

[11] Carlbring P, Westling BE, Ljungstrand P, Ekselius L, Andersson G. Treatment of Panic Disorder Via the Internet: A Randomized Trial of a Self-Help Program. Behaviour Therapy 2001; 32(4):751-764.

[12] Andersson G, Carlbring P. Editorial – special issue: Internet and cognitive behaviour therapy: new opportunities for treatment and assessment. Cognitive Behaviour Therapy 2003;32:97–99.

[13] Hagerstrand T. Time-geography: focus on the corporeality of men, society and environment. In: The Science and Praxis of Complexity. New York: United Nations University, 1985.

[14] Willis F, Sanders D. Cognitive Therapy: Transforming the Image. London: Sage Publications, 1997.

[15] Beck JS. Cognitive Therapy: Basics and Beyond. New York: Guilford Press 1995.

[16] Weiser M. Hot Topics: Ubiquitous Computing. IEEE Computer, October 1993.

[17] Shilit BN., Adams, NI, Want R. Context-Aware Computing Applications. In: Proceedings of the Workshop on Mobile Computing Systems and Applications. Santa Cruz: IEEE Computer Society 1994: 85-90.

[18] Guanling C, Kotz D. A Survey of Context-Aware Mobile Computing Research. Technical Report TR2000-381, Department of Computer Science, Dartmouth College, 2000.

[19] Dick LP, Gallagher-Thompson D, Coon DW. Cognitive-Behavioral Therapy for Late-Life Depression: A therapist manual. Palo Alto, CA: Older Adult and Family Center, Veterans Affairs Palo Alto Health Care Center, 1995.
[20] Stanley MA, Diefenbach GJ, Hopko DR. Cognitive behavioral treatment for older adults with generalized anxiety disorder. Behavior Modification 2004, 28 (1), 73-117.
[21] Greenberger D, Padesky C. Mind Over Mood. New York Guilford Press 1995.
[22] Bernstein, DA, Borkovec TD. Progressive relaxation training. Champaign, IL: Research Press 1973.

Acknowledgments

This research was supported by a grant from the Health Informatics program at VINNOVA.

Address for correspondence

Dr. Magnus Bang, E-mail: magba@ida.liu.se
Department of Computer and Information Science
Linkoping University, SE-581 83 Linkoping, Sweden

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Methods for Measuring the Impact of Health Information Technologies on Clinicians' Patterns of Work and Communication

Johanna I Westbrook[a], Amanda Ampt[a], Margaret Williamson[b], Ken Nguyen[a] Leanne Kearney[a]

[a]*Health Informatics Research & Evaluation Unit, Faculty of Health Sciences, University of Sydney, Australia ,*
[b] *National Prescribing Service, Australia*

Abstract

Evidence regarding how health information technologies influence clinical work patterns and support efficient practices is limited. Traditional paper-based data collection methods are unable to capture clinical work complexity and communication patterns. Our objective was to design and test an electronic data collection tool for work measurement studies which would allow efficient, accurate and reliable data collection, and capture work complexity.

We developed software on a personal digital assistant (PDA) which captures details of nurses' work; what task, with whom, and with what; multi-tasking; interruptions and event duration. During field-testing over seven months across four hospital wards, fifty-two nurses were observed for 250 hours. Inter-rater reliability scores were maintained at over 85%. Only 1% of tasks did not match the classification developed. Over 40% of nurses' time was spent in direct care or professional communication, with 11.8% in multi-tasking. Nurses were interrupted approximately every 49 minutes. One quarter of interruptions occurred while nurses were preparing or administering medications. This approach produces data which provides greater insights into patterns of clinician's work than has previously been possible.

Keywords:

computerised order entry systems; nursing staff; medication systems; time and motion studies; computers, handheld; observation; interdisciplinary communication

Introduction

Evidence regarding how health information technologies influence patterns of clinical work and support efficient work practices is limited. A recent systematic review [1]uncovered 23 studies since 1984 which examined the impact of system use on clinicians' (doctor and nurses') time. These studies in general adopted either work sampling or time and motion methods. Only six (26%) examined work on general wards, all in US hospitals, while the remainder focused on specialized settings (e.g., ICU and general practice). Overall, studies which compared electronic with paper systems and calculated task time per patient or consultation, reported that computer use increased time required to complete tasks. Studies which examined task time across multiple patients or working shifts found computer use was more time efficient than paper-based systems [1].

Studies of changes in work distribution and communication patterns following system use are less prevalent, but do include evidence of changes. For example in the US, where nurses usually transcribe handwritten medication orders, CPOE eliminates this task [2, 3]. Following the introduction of a CPOE system clinicians may sequence work differently. As Callen et al [4] found, clinicians reported " ..if you are waiting for something on the computer you go and do something else." (p648) This may result in, for example, batching the ordering of patients' tests to one time of the day. Shu et. al. [5] found interns in a US hospital spent more time alone and less time with other doctors after system introduction. A French hospital study [6] of doctor-nurse communication around medications showed that a CPOE system, in comparison with paper-based medication records, resulted in a move from synchronous communication to asynchronous. This introduced opportunities for misunderstandings and increased the extent to which nurses had to make assumptions about orders. Carpenter et. al. [7] also reported a tendency for doctors in a US hospital to talk with nurses less about medication orders following system implementation.

Understanding these shifts in patterns of communication between clinicians are important as poor communication wastes time, threatens patient care and may be one of the chief culprits behind preventable adverse events in clinical practice [8]. Any potential negative consequences of changes in communication patterns may be more than offset by the improvements in information exchange provided by having legible, easily accessible information which computerized systems afford clinicians. However until we have better quality data about how systems disrupt existing patterns of clinical work and communication we cannot move to re-design work practices or systems in ways which avoid any possible negative outcomes.

This research agenda needs to continue to progress beyond answering the question on whether the use of a computer save a clinician time to questions about how patterns of work are re-arranged in response to the introduction of health technologies. Where time is released, or additional time consumed, how do clinicians re-distribute their time amongst work tasks? What amount of variation exists among different clinical sub-groups and do work tasks get

Selected for best paper award.

re-distributed across groups? For example, if senior clinicians are found to spend less time in patient ordering following computerisation, is this because the system is efficient or because they have re-allocated this task, either explicitly or implicitly, to their junior colleagues? We need to examine how system use interferes with communication processes and as Gorman et. al. [9] suggest, ensure that such systems *".. facilitate care without interfering with or eliminating aspects of the process that are essential to high reliability performance in the face of urgency, uncertainty and interruptions"* (p383)

We require studies which investigate whether changes in patterns of clinical work result in improved care delivery, patient outcomes and the work experiences of health professionals. These questions require a multi-method approach, and work measurement studies form an important component of such investigation. Researchers should be able to build upon previous work undertaken in different settings and countries. This requires standardization of measurement approaches and the adoption of valid and reliable measurement tools. Major factors identified for the paucity of evidence in this area are the limitations and varieties of methods used [1, 10, 11], the difficulties of capturing the non-linear and interruptive nature of clinical work, and the lack of consistency in the application of rigorous research methods.

Our objective was to design, build and test an electronic data collection tool for use in work measurement studies which would allow efficient, accurate and reliable data collection while also capturing a greater level of work complexity than previous paper-based methods have allowed. Results from a small number of previous studies [12, 13] suggest that the use of handheld computers (personal digital assistants – PDAs) may be useful for this task but researchers have presented minimal information about application and reliability issues relating to these tools. Building upon our previous work designing a paper-based, multi-dimensional work classification tool for nurses [11], we sought to investigate how much additional detail and task complexity data we could add using a PDA without reducing data accuracy or reliability. In this paper we detail the development of these methods and tool, outline the data collection process using the PDA in an observational study of fifty-two nurses undertaking everyday work activities over 250 hours in wards of an academic hospital, and provide a summary of the results. This comprised the first stage in a study to measure the impact of a commercial electronic medication management system (e-MMS) on doctors' and nurses' work and communication patterns.

Using these baseline data we aimed to examine how nurses distributed their time across major work tasks and communication events, and how patterns vary, for example by nurse classification. We also sought to measure the extent of multi-tasking and interruptions to nurses' work.

Materials and methods

Design of a multi-dimensional work task classification

Our first objective was to develop a work task classification system which would be incorporated in the PDA data collection tool. As a basis for this we used a multi-dimensional work measurement classification which we had applied in a paper-based work-sampling study of nurses and showed high levels of inter-rater reliability and face validity [11]. We extended the classification to contain greater detail about work tasks which previous literature indicated may be most susceptible to change following system introduction and thus would address our long-term aim to use the PDA in a study to examine the impact of an e-MMS on clinicians' patterns of work and communication.

The classification captures the complexity of nursing tasks in terms of: i) the activity being undertaken, ii) other people involved in the task, and iii) the method of task performance. The ten broad categories of work tasks and definitions are listed in Table 1.

Table 1 - Broad work task categories

Work Task	Definition
Direct Care	Tasks directly involved with patient care, eg direct communication with pt. &/or family, bathing, applying dressings etc.
Indirect Care	All tasks indirectly related to patient care (eg reviewing results, planning care)
Medication Tasks	All tasks associated with medication, includes preparation, administration, documentation, discussion & clarification
Documentation	Documentation (paper and electronic), excludes medication documentation
Professional Communication	All non-medication related communication with another health professional, includes ward & patient hand overs
Ward Related Activities	Ward activities, includes coordinating beds & staffing
In Transit	Time between tasks and between patients
Supervision	Supervising others, including students
Social	All non work communication, meal breaks
Other	Any other task not included above

Using a PDA allowed us to add further detail to some work tasks by using drop down submenus. Figure 1 shows the 10 broad work task categories and Figure 2 one of the submenus for medication-related tasks. Further we wished to capture instances of multi-tasking (undertaking two tasks at the same time) and interruptions. This had not been feasible with a paper version. Interruptions were defined as the ceasing of one task in order to attend to another task. For example, a nurse who stops administering a drug in order to answer a question from a colleague. If the nurse

continues to administer the medication while answering her/his colleague the activity would be recorded as multi-tasking.

Data collection tool

We designed a work measurement data collection tool using an HP iPAQ rx3000 Pocket PC running Windows Mobile 2003 with a SD card expansion slot. The PDA program was developed in Visual C# using the .NET Compact Framework v1.0. Using ActiveSync enabled the use of SQL and the synchronization of multiple PDA databases with one central database on a PC. Figure 1 shows the interface design with the 10 broad mutually exclusive work task categories listed on the left. Each task is automatically time-stamped. On the top right hand side the data collector records who else is involved in the task, and in the bottom right of the screen any tools/equipment that are used. Selection from any one of the three sections will immediately start a new time stamp – thus if the data collector is aware a new activity has started as the nurse is now with another professional, but is unsure what the activity is on initial observation, she/he can select another entry point, such as who the nurse is with which might be easier to distinguish.

If the observed subject is interrupted or is multi-tasking, the observer hits the "Add" button. If the subject is multi-tasking then the observer will continue to record all work tasks until the multi-tasking ends (End Multi). If the subject is interrupted then the interruption button is selected and the interrupted task appears as a tag on the bottom of the screen. When the subject returns to the task the observer is able to pick up this pending task. This allows, for example, calculation of the time from interruption until final completion of the task. The 'Ignore' button is used as a quick way for the observer to delete data recorded in error.

All recorded data are placed in a database which is stored on the SD card. At the end of an observation session the

Figure 1 - PDA work measurement data collection tool

PDA is synchronised with a computer running Microsoft SQL Server using merge replication. SQL Server replication is commonly described by using the publisher/subscriber metaphor. A database server that makes data available for replication (source server) is referred to as the publisher; a collection of one or more database objects that are enabled for replication is called a publication.

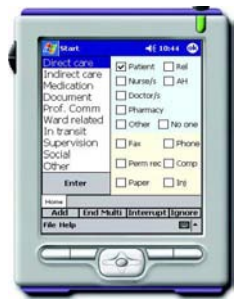


Figure 2 - Drop down submenu displayed

One or more servers that receive data and/or transactions from the publisher are called subscribers (the PDAs, in this case). Replication is managed by the system database, which by default is called distribution. A distribution database—which can reside on the publisher, subscriber, or on a separate server or computer—is created when replication is configured.

The server that hosts the distribution database is referred to as the distribution server or distributor. Merge replication combines data from multiple sources into a single central database. This allows multiple observers to use different PDAs for data collection. On synchronization, data from each PDA is then transferred to the central database.

An extraction program was created to extract the data from the distributor. This was written in Visual C# using the .NET Framework v1.1. This allowed data to be imported into a statistical package such as SPSS for analysis. The data collection tool, synchronization and extraction programs underwent considerable field trials. Once these were completed we undertook a full-scale study to examine nurses' patterns of work and communication tasks. This constituted the first stage in a study to investigate changes in nurses' work patterns following the implementation of a commercial electronic medication management system.

The work measurement study was conducted over seven months (July 2005-February 2006) in four wards (respiratory, renal/vascular and two geriatric), at a major academic teaching hospital in Sydney, Australia. Data were collected between the hours of 7am – 7pm, Monday to Fridays.

Procedures and participants

Nurses on the four wards were invited to participate at ward information sessions. Following signed consent nurses were assigned a study identification number, and demographic information regarding their age, nurse classification (e.g., enrolled nurse, registered nurse-new graduate, registered nurse, clinical nurses specialist etc.), and length of experience was collected. The observer shadowed nurse participants for an average of one hour blocks, noting all work tasks performed using the PDA to record data. When the participant nurse engaged with patients, visitors, or other health professionals, the nurse was asked to introduce the observer and seek permission to continue. Alternatively the observer would identify themselves. The

study was approved by the ethics committees of the hospital and the University of NSW.

Sample size calculations, based upon our previous work-sampling studies [11],were made which determined the number and times of hours of observation per nurse category required. Our intention was to be able to detect differences in the broad work categories pre and post system implementation. As each category had its own mean and standard deviation, it was decided to use medication-related tasks to determine the sample size – this category had the largest standard deviation and is the most directly related to the proposed system change. We determined that at alpha=0.05 and with power of 80%, 226 hours of observation were required to detect a three minute difference in the proportion of time spent in medication-related tasks pre and post system introduction using a two-tailed t-test. We achieved 250 hours of data collection during the pre-implementation phase.

Data collector training

All observers were clinically experienced registered nurses. The initial training was delivered by one of the authors (AA), who had experience with previous work measurement studies [9]. Areas covered included outline of study purpose and methodology, instruction in PDA use with explanation of definitions and practice scenarios, introduction to study wards, practice sessions on the wards followed by inter-rater reliability testing. Once the second data collector was trained (LK), she then delivered all subsequent training to three more data collectors. Approximately fifteen hours was required for each data collector to be trained.

Inter-rater reliability and data analysis

Following training of each new data collector, inter-rater reliability tests were conducted, as well as at other random times. Two data collectors simultaneously but independently observed a nurse and recorded work tasks on their own PDA for approximately 45 minutes. Data were then analysed and compared. New collectors did not start collecting data until overall percentage time in tasks was in agreement with a more experienced collector over 85% of the time. All data collectors maintained this level of agreement, (range 85%-98%).

Descriptive statistics with 95% confidence intervals were calculated. For this paper, differences in task time distribution by wards and nurse classification were analysed using SPSS.

Results

Work task distribution

Fifty-two nurses over four wards were observed for a total of 250 hours and 20 minutes. During this time 15,533 tasks were recorded. Figure 3 shows the distribution of observed time spent in different work tasks. Direct care and professional communication accounted for 41% of total task time.

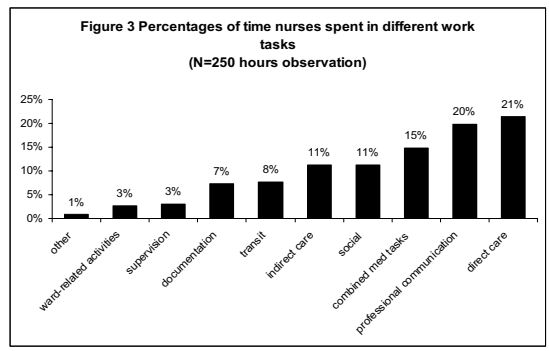


Figure 3 - Percentages of time nurses spent in different work tasks

The pattern of task time distribution differed between enrolled nurses and registered nurses. As Figure 4 shows enrolled nurses spent a greater proportion of their time in direct care activities while registered nurses spent comparatively more time in medication tasks and professional communication.

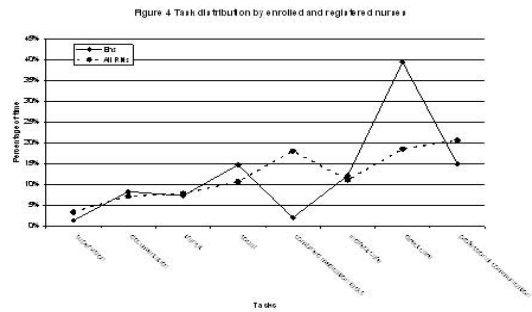


Figure 4 - Task distribution by enrolled and registered nurse

We examined task time distribution by wards and found considerable similarity in the pattern of time distribution despite differences in the ward specialties. As Figure 5 shows, for nearly all task categories the 95% confidence intervals overlapped.

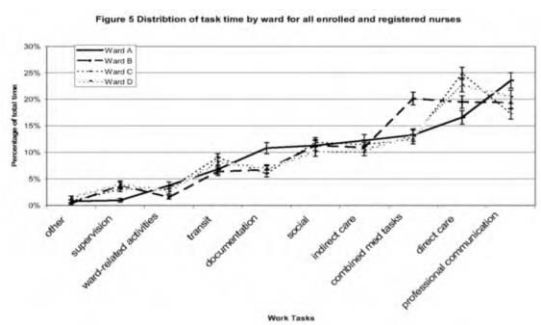


Figure 5 - Distribution of task time by ward for all enrolled and registered nurses

Multi-tasking and interruptions

The total task time recorded was 279 hours 51 minutes, compared to 250 hours 21 minutes of observed time, which revealed that nurses spent 11.8% (29hrs 30 minutes) of observed time multi-tasking in which two or more tasks were performed in parallel. On average nurses were found to multi-task for approximately one hour and two minutes every shift. In addition, on average, nurses were interrupted every 49 minutes. One quarter of all interruptions occurred while nurses were undertaking medication-related tasks and a further 23% of interruptions occurred while nurses were documenting.

Professional communication

Overall nurses spent 20% of observed time in professional communication which equates to approx. 12 minutes/ hour. The majority (77%) of this time was spent communicating with other nurses. Only 8%, a mean of 8 minutes per shift (of which 45mins is meal break) was spent talking with doctors.

Discussion

Our aim was to design research methods and a tool to allow efficient, reliable and accurate data collection about clinicians' patterns of work and communication, which could potentially be used across specialties, professional groups, settings and countries. The observational methods and the PDA data collection tool proved to be a reliable means for collecting complex, multi-dimensional data about nurses' work and communication patterns. High levels of inter-rater reliability were achieved with an average of 15 hours of training. Training, recruitment and 250 hours of data collection were undertaken by the 1.5 full-time equivalent staff over approximately 10 months. Thus, we found it to be a relatively time-efficient method, particularly as data are entered electronically in real-time reducing transcription errors. The work task classification and accompanying definitions dealt well with the array of nurses' observed work tasks with less than 1% of all tasks being categorized in the other category.

Our initial analysis comparing ENs and RNs provides some validation of the data collected. For example, ENs are legally restricted in their performance of many medication-related tasks and this was demonstrated in the data. Further examination of work tasks by nurse classification will provide a useful means in the second stage of the research by which to examine how tasks may be redistributed across nursing classifications following the introduction of the electronic medication management system.

The results demonstrated the greater detail of data about work and communication patterns which can be obtained using the PDA tool compared to traditional paper-based methods. A good example was new data about multi-tasking and interruptions. Most previous studies of interruptions to clinical work have not distinguished between interruptions and multi-tasking. This tool provides a useful way of doing this.

Further, more detailed analyses, such as which tasks are interrupted and which tasks are more likely to be interruptors are possible. Interestingly we found that the greatest proportion of interruptions occurred while nurses were undertaking medication tasks. Such a finding raises the question of the role of interruptions in contributing to medication preparation and administration errors. We are currently investigating this issue further. The data cannot reveal whether interruptions had negative or positive effects. For example, the outcome of an interruption to stop administering a drug in error versus an interruption which occurs in the middle of a drug dose calculation will be quite different.

The PDA allows for detailed capture of patterns of work tasks and communication. This information, along with identification of tasks and situations that may lead to cognitive overload through multi-tasking and interruption, is essential in understanding how clinical information systems may impact upon the quality and safety of care.

Acknowledgments

J Forbes, A McLaren, K Gatwood for assistance with data collection. This research was funded by an ARC grant in partnership with NSW Health. JW is supported by an NHMRC Fellowship. We thank the hospital & nurses who participated in the study.

References

[1] Poissant L, Pereira J, Tamblyn R, Kawasumi Y. The impact of electronic health records on time efficiency of physicians and nurses: A systematic review. Journal of American Medical Informatics Association 2005;12(5):505-516.

[2] Ball MJ, Weaver C, Abbott PA. Enabling technologies promise to revitalize the role of nursing in an era of patient safety. International Journal of Medical Informatics 2003;69(1):29-38.

[3] Kaushal R, Jha A, Franz C, Glaser J, Shetty K, Jaggi T, et al. Return on investmant for a computerized physician order entry system. Journal of American Medical Informatics Association 2006;13(3):261-266.

[4] Callen J. The people and organisational issues which impact on the diffusion of IT systems in healthcare facilities [PhD]. Sydney: University of New South Wales; 2006.

[5] Shu K, Boyle D, Spurr C, Horsky J, Heiman H, O'Connor P, et al. Comparison of time spent writing orders on paper with computerized physician order entry. In: Patel Vea, editor. Medinfo; 2001; London: IOS Press; 2001. p. 1207-1211.

[6] Beuscart-Zephir M, Pelayo S, Anceaux F, Meaux J, Degroisse M, Degoulet P. Impact of CPOE on doctor-nurse cooperation for the medication ordering and administration process. International Journal of Medical Informatics 2005;74:629-641.

[7] Carpenter J, Gorman P. What's so special about medications: A pharamcist's observations from the POE study. In: Proceedings of American Medical Informatics Association Symposium 2001; 2001; 2001.

[8] Gosbee J. Communication among health professionals. British Medical Journal 1998;316:642.

[9] Gorman P, Lavelle M, Ash J. Order creation and communication in healthcare. Methods of Information in Medicine 2003;42:376-384.

[10] Burke T, McKee J, et al. A comparison of time-and-motion and self-reporting methods of work measurement. Journal of Nursing Administration 2000;30(3):118-125.

[11] Ampt A, Westbrook J, Creswick N, Mallock N. Comparison of self-reported and observational work sampling techniques to measure nurses' work. Journal of Health Services Research and Policy 2007;In press.

[12] Asaro P. Sychronized time-motion study in the emergency department using a handheld computer application. In: Proceedings of the World Congress on Medical Informatics (MedInfo); 2004; San Francisco: ISO Press; 2004. p. 701-705.

[13] Starren J, Chan S, Tahil F, White T. When seconds are counted: Tools formobile, high resolution time-motion studies. In: Proceedings of American Medical Informatics Symposium; 2000; 2000. p. 833-837.

Address for correspondence

Professor Johanna I Westbrook
email - J.Westbrook@usyd.edu.au

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Enhancing User Acceptance of Mandated Mobile Health Information Systems: The *e*POC (*electronic* Point-Of-Care Project) Experience

Lois Burgess[a], Jason Sargent[a]

[a] *Centre for Health Service Development, University of Wollongong, Australia*

Abstract

From a clinical perspective, the use of mobile technologies, such as Personal Digital Assistants (PDAs) within hospital environments is not new. A paradigm shift however is underway towards the acceptance and utility of these systems within mobile-based healthcare environments. Introducing new technologies and associated work practices has intrinsic risks which must be addressed. This paper contends that intervening to address user concerns as they arise throughout the system development lifecycle will lead to greater levels of user acceptance, while ultimately enhancing the deliverability of a system that provides a best fit with end user needs. It is envisaged this research will lead to the development of a formalised user acceptance framework based on an agile approach to user acceptance measurement. The results of an ongoing study of user perceptions towards a mandated electronic point-of-care information system in the Northern Illawarra Ambulatory Care Team (TACT) are presented.

Keywords:

ambulatory care, handheld computers, point-of-care, information systems, user acceptance

Introduction

A considered, measured transition from Health to eHealth is currently underway. While the move towards an electronic health record (EHR) is often the focus of discussion in this regard, it is the myriad in-situ electronic clinical information systems under development in which [The] effective use of Information Management and Technology and the implementation of core standards, applications and support networking will promote an environment where necessary information about a patient will be available to clinicians when and where it is needed [1]. For small, site-specific IS development projects to leverage on the advantages that this presents, the authors of this paper contend that agile systems development methodologies should be engaged from project initiation to enhance end-user acceptance of the delivered system. The decision to do so enables a best fit system for the end-user (clinician), the provider, and most importantly the patient. This is particularly important for mandated technology implementations. In scenarios of mandated technology implementation, impressions may exist whereby any form of consultative input from end-users is inherently removed; leaving end-users disillusioned with the technology. Moreover, where the intended end-users of a system are traditionally averse to technology through entrenched paper-based work practices (such as community health workers) the process of managing system change bears a considerable determination in system implementation success. This paper outlines an agile approach to end-user acceptance, taken to proactively address issues surrounding transition from traditional paper-based systems to mandated electronic systems in a community health setting with historically low levels of technology acceptance and use. The ePOC (electronic point-of-care) project is used to demonstrate how higher levels of user acceptance of mobile health information systems are attainable, even in mandated implementation environments.

The *electronic* Point of Care Project

The *electronic* Point-of-Care Personal Digital Assistant project (ePOC) is a multi-phase, collaborative research and development, Australian Research Council 3yr funded project, comprising a research team from the University of Wollongong, Flinders University, the University of South Australia and Pen Computer Systems Pty Ltd. The project client is The Ambulatory Care Team (TACT) Northern Illawarra which is supported by South Eastern Sydney and Illawarra Area Health Service (SESIAHS).

TACT employ 21 staff members consisting of 3 Doctors (including a Medical Director), 1 Nurse Unit Manager, 13 nurses (4 full-time and 9 Part-time), 2 Pharmacists (part time), 1 Physiotherapist and 1 COPD coordinator[1]). TACT Northern Illawarra[2] provide outpatient healthcare delivery; where patients are given a choice of having treatment in their usual place of residence (including aged care facilities) or other locations as an alternative to hospital. Community-based health services within New South Wales, Australia deliver over eight million occasions of service per annum provided by more than 7,000 clinicians from more than eight hundred and fifty health service loca-

1 Chronic Obstructive Pulmonary Disease Specialist.
2 Northern Illawarra comprises Wollongong, Shellharbour and Kiama Local Government Areas (LGAs)

tions. The cost of this service is estimated to be almost $450 Million dollars per annum [2].

In essence, the ePOC project is developing an integrated Ambulatory Care information system, deployed on a Personal Digital Assistant (PDA) platform. A PDA based point-of-care system for TACT is significant in that it will provide for *collection*, *delivery and exchange* of timely information (both text and images) at the point-of-care leading to a more efficient health care system. Current systems at TACT are paper-based and are limited to what the healthcare worker can effectively carry. The key advantages of a PDA system are its high mobility and flexibility in matching complex healthcare workflow requirements as well as immediate updating of healthcare records at the point-of-care. PDAs as wireless deployment platforms for mobile-based *hospital* clinical information have proven to be among the most cost effective ways to improve patient care quality and reduce medical data collection errors. Medical professionals empowered by information make better decisions while at the patient's bedside [3]. The challenge now is to extend information empowerment of clinicians from a patients (hospital) bedside to (home) point-of-care. It is within this project context that a new approach to technology acceptance, aimed at enhancing user acceptance of mobile health systems, utilising concepts drawn from agile systems development methodologies has been developed and is currently being tested and validated.

Theoretical basis of the study

The theoretical basis of this study centres on Dynamic Systems Development Methodology (DSDM) [4] and the Unified Theory of Acceptance and Use of Technology (UTAUT) [5]. From the outset, the research approach adopted by the ePOC project has centered on a consultative, open approach with all project team members; in particular, project team health area managers (Clinical, Research and IT) and the intended end users (TACT Doctors, Nurses and Para-Health Professionals) of the PDA application. To facilitate this approach, the ePOC project incorporates a number of research and development phases. According to Davis and Venkatesh, taking an iterative, consultative approach helps focus upon the identification, correction and prevention of requirements errors that have been introduced in the original specification of requirements [ibid].

The Dynamic Systems Development Methodology (DSDM)

Dynamic Systems Development Methodology is a framework supported by its continuous user involvement in an iterative development and incremental approach which is responsive to changing requirements, in order to develop a system that meets the business needs on time and on budget. It is one of a number of agile methods for developing software and forms part of the Agile Alliance. DSDM was developed in the United Kingdom in the 1990s by a consortium of vendors and experts in the field of Information System (IS) development, the DSDM Consortium. [6]

There are nine principles of DSDM [ibid]. While all nine are noteworthy, this paper focuses on the following six:

- Active user involvement.
- A focus on frequent delivery of products.
- Iterative and incremental development to ensure convergence on an accurate business solution.
- Reversible changes during development.
- Integrated testing throughout the life cycle.
- Collaboration and cooperation between all stakeholders.

The Unified Theory of Acceptance and Use of Technology (UTAT)

The Unified Theory of Acceptance and Use of Technology (UTAUT) was developed through a review and consolidation of the constructs of eight models that earlier research had employed to explain IS usage behaviour (theory of reasoned action, technology acceptance model, motivational model, theory of planned behavior, a combined theory of planned behavior/technology acceptance model, model of PC utilization, innovation diffusion theory, and social cognitive theory) [7].

A fundamental differentiation between variants of technology acceptance models and the approach proposed in this current research is the factor of time in which review of the information system being studied occurs. TAM [8], [9] and the extended Technology Acceptance Model (eTAM) [10] are tools which evaluate the perceived ease of use of technology application adoption (such as PDAs) however, the time when the process of managing an information system adoption occurs is made by reviewing prior actual adoptions, investigating variance of perceptions and applying the results to subsequent implementations. Liang et als study of usage of eTAM to predict actual PDA usage among healthcare professionals is such an example of this traditional review past implementations, apply findings to new/next implementation technology acceptance approach [11]. The approach taken in the ePOC project addresses the issues surrounding traditional approaches to user acceptance measurement by mapping concepts from agile systems development (DSDM) to user acceptance (UTAUT). This means that user acceptance is measured at each stage of development of the proposed system, enabling the project team to intervene at regular stages throughout the system development lifecycle to ensure that user requirements are being met and therefore increase the likelihood of user acceptance of the delivered system.

This approach is consistent with Toleman et als assertion that perception takes place through a sequence of stages through which a potential adopter of an innovation passes before accepting the innovation [12]. The ePOC project sought to gain end-user perceptions of a proposed new electronic point-of-care system, highlighting points where intervention by systems designers and project managers may be required to ensure the highest possible levels of user acceptance and support for the proposed system. In addition, more informal interaction was encouraged through a discussion forum on the project website and

notice board at the research site. Clinicians are actively encouraged to raise questions and/or concerns regarding the proposed system. Depending on the nature and imperative of the concern, the project team will either provide an immediate response or schedule it for discussion at a proceeding workshop. This enables intervention at critical points in the system development lifecycle.

The overall aim of this process for the ePOC project is to enhance user acceptance, the goal being to achieve an acceptance level (UA) as close as possible to 100%.

Methodology

User Surveys and Focus Groups are the predominant methods of data collection utilised in the ePOC project. This paper presents the results of five (5) surveys and the focus groups.

Development of the primary survey instrument

Klines Groupware Adoption Scale [13], [14] was adapted to meet the needs of the project. Klines original five subscales (EOU, Technical Support, Training, Work Needs Met, System Capabilities and Consultation), with a seventh subscale Commitment were included in the questionnaire. An eighth subscale, Persuasion was included in the questionnaire after the first survey.

The initial pre-implementation questionnaire was distributed to TACT five months (into the 3yr project) after project initiation, but well before the introduction of a prototype PDA device into the unit. The aim of the questionnaire was twofold; firstly to gauge the levels of perception by intended end-users towards to ePOC project and PDA device as the platform of choice and secondly, to validate the survey instrument. Results of the survey indicated significant numbers of TACT staff took a neutral stance on many questions (that is, responded Neither Agree nor Disagree). A focus group was conducted shortly after to attempt to understand the reasons behind this response. The outcome of the focus group revealed the assumption by the project team that TACT staff fully understood the aim of the ePOC project, their role in the project and that staff were receiving comprehensive updates on the project was not the case. Information passed on to the TACT Medical Director and Nurse Unit Manager at project committee meetings had not been conveyed to staff at TACT.

It was deemed imperative to the success of the ePOC project that staff at the unit be involved throughout the project and that management at the unit demonstrate a positive attitude and commitment to the proposed system. This outcome triggered a response by the project team to adopt a more concerted, consultative approach to system design, development and project management. An agile model of project management was adopted with the aim of increasing user acceptance throughout system development. This was facilitated through a series of surveys, focus groups, discussion sessions, newsletters, seminars and hands-on exercises with the PDA device scheduled to coincide with the release of each module (sub component) of the ePOC system. The approach taken is depicted in Figure 1.

An example of a module for ePOC is Clinical Observations. This module replicates current paper-based data entry capture of a patients clinical observations (such as Pulse, Respiration Rate, Blood Pressure and Blood Sugar Level, plus others) electronically on a PDA screen. The Clinical Observation module was developed by taking into account existing paper-based information requirements then user interface design considerations and workflow improvements. The prototype ePOC clinical Observation module was then tested in a field trial to determine if the module met expectations at point-of-care. TACT staff were surveyed and the results analysed to determine if, as a result of the project teams intervention and the new approach taken had enhanced end-user acceptance. Consider the Clinical Observation as Module 1 in Figure 1.

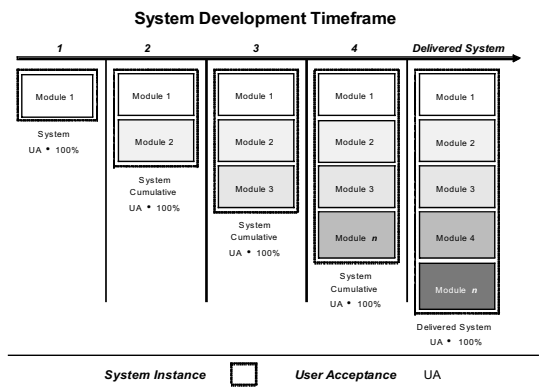


Figure 1 - ePOC Approach to System Acceptance

This process was repeated for subsequent modules; Peripheral Line Checklist, Electronic Protocols and MIMS on PDA. MIMS - an acronym for Monthly Index of Medical Specialties is an online prescribing guide for doctors, nurse prescribers and pharmacists. Significant findings with regards to each of the sub-scales of the survey instrument are discussed in the results section. Additionally, aggregated results of each of the five surveys and the sequencing of each in respect to points of intervention by the project team to clarify identified problems with each of the modules, impact on work practices and transfer of concepts onto the PDA device by TACT staff are shown in Table 1 on the following page.

Results

Results are categorised according to Klines subscales and the items included in the instrument. Aggregated results from the five surveys (questionnaires) are detailed in Table 1 and discussed below. Due to page limits imposed, discussion of results is limited to aggregated results across the survey subscales. Results of individual scale items are available on request. The first two surveys were conducted to assess user perceptions of the proposed Ambulatory Care Information System. Due to the fact that a large num-

Table 1- Aggregated Results (Positive Perception Levels)

	Ease of Use	Training	Technical Support	Consultation	Work Needs Met	System Capabilities	Commitment	Persuasion
Questionnaire 1								
Project & PDA	71%	84%	86%	76%	72%	62%	87%	N/A
Intervention 1								
Questionnaire 2								
Project & PDA	89%	98%	100%	93%	86%	86%	100%	87%
Field Trial 1								
Questionnaire 3								
Module 1	83%	54%	83%	50%	61%	71%	84%	67%
Intervention 2								
Questionnaire 4								
Module 2	94%	88%	100%	100%	89%	91%	100%	87%
Questionnaire 5								
Module 3	90%	91%	96%	90%	84%	88%	93%	90%

ber of TACT staff responded with a neutral response (neither agree/disagree) to all of the subscales in the first survey, a project champion (Medical Director, TACT) was appointed to drive the project at the client site. The purpose of this intervention was to determine if the appointment of a project champion would have a positive or negative effect on user perceptions of the proposed system. A seventh subscale, Persuasion was included in the questionnaire for the second survey to measure the impact of the intervention. The results of the second survey indicate that user perceptions across all subscales increased significantly, therefore demonstrating a positive outcome.

Following the second survey, TACT staff were provided with PDAs and given time to familiarise themselves with the devices prior to implementation of the first system modules, Clinical Observations and Peripheral Line Checklist. A field trial followed, after which a third survey and focus group were conducted. The results of the field trial indicate that staff acceptance of the system *decreased* after they had used it in the field. The focus group revealed a number of design issues related to the system and the user interface. The issues of concern raised by staff during the focus group were considered by the development team and the modules modified. Feedback from intervention 2 indicated that users were much more satisfied with the Clinical Observations module after their concerns over the design of the module and the user interface were addressed. Lessons learned form the previous survey and intervention were incorporated into the redesign of Module 2, Electronic Protocols (standardized work practices/ checklists based upon prescribed treatment of common diseases, for example steps to be undertaken in the treatment of Pneumonia or Cellulitis). Following implementation of the module onto the PDA device, staff again surveyed. The results of this survey record the highest levels of acceptance of all previous surveys providing further support for the approach to user acceptance taken by the project team.

The results of the fifth survey (Module 3, MIMS on PDA) also exhibit high levels of user acceptance across all 8 subscales, with increases in acceptance in the areas of training and persuasion. The other 6 subscales were down slightly on the previous survey. Notwithstanding, the results are promising. A focus group is planned for the near future to get feedback from users on the issues that emerged from the survey. This will enable further intervention and refinement of Module 3 before its final implementation on the PDA device. As a result of the feedback from staff at the unit, the project development team is currently working on graphical representation of the Clinical Observations and Peripheral Line Checklist modules. The surveys and focus groups will be repeated once the modules have been modified.

Conclusion

The results of the surveys and focus groups clearly demonstrate the viability of the approach to user acceptance taken in the ePOC project. Intervention by the project team at identified points in the development of the system has clearly resulted in increasing levels of user acceptance. Involving users in systems design has demonstrated benefits. Taking this concept a step further and applying it to user acceptance measurement throughout the system development lifecycle ensures that users will not only be more satisfied with the functionality of the system, but also its utility. Measuring user acceptance throughout the development of a system rather than post-implementation provides developers with an opportunity to address users concerns at the point where they arise, making it easier and more cost effective to modify system components. While research into user acceptance of the Ambulatory Care system is continuing, the results so far are more than pleasing. Future research will further develop the user acceptance framework and test and validate it in an enterprise-wide system development and implementation. It is conceivable that this approach can be applied to any system implementation that utilises a component-based system design and development methodology.

Acknowledgments

This research is derived from A PDA (Personal Digital Assistant) based Point of Care (POC) e-Health Solution for Ambulatory Care Services, an Australian Research Council Linkage Grant based at the Centre for Health Service Development, Wollongong University. The support of Pen Computer Systems, The Ambulatory Care Team (TACT) Northern Illawarra, South Eastern Sydney Illawarra Area Health Service (SESIAHS), Flinders University and the University of South Australia is acknowledged. Further information regarding the ePOC PDA project can be found on the ePOC projects website at the following URL:

http://www.itacs.uow.edu.au/cear/ehealth/ePOC/

References

[1] NSW Health Strategy for the Electronic Health Record. Report of the Health Information Management Implementation Coordination Group. February 2004. Available as PDF Online: <URL: http:// www.health.nsw.gov.au/iasd/chime/newspdf/ehr_doc.pdf > Last Accessed 11 October, 2006.

[2] NSW Health Community Health Information Management Enterprise: Overview: What is CHIME? Available Online: <URL http://www.chime.gov.au/iasd/chime/chime/ overview/index.html >Last accessed October 7, 2006.

[3] MedPDA.net, Medical PDA Reviews: Doctors using PDAs make less mistakes? October 6, 2004. Online <URL: http:// www.medpda.net/archives/001050.html > Last accessed September 19, 2006.

[4] DSDM Consortium. History. Available Online: <URL: http://www.dsdm.org/about/history.asp > Last Accessed 01 December, 2006.

[5] Venkatesh V, Morris, MG, Gordon B, and Davis F. User acceptance of information technology: toward a unified view. MIS Quarterly 2003: 27(3): 425-54.

[6] Dynamic Systems Development Methodology, Wikipedia. Available Online <URL: http://en.wikipedia.org/wiki/ Dynamic_Systems_Development_Method > Last Accessed November 2, 2006.

[7] Wade M and Schneberger S. Theories used in IS research: Unified Theory of Acceptance and Use of Technology. York University, Canada. Available Online: <URL: http://www. istheory.yorku.ca/UTAUT.htm > Last Accessed 03 November, 2006.

[8] Bagozzi R P, Davis F D and Warshaw P R. Development and test of a theory of technological learning and usage. Human Relations, 1992: 45(7): 660-686.

[9] Davis F D, Bagozzi R P and Warshaw P R. User acceptance of computer technology: A comparison of two theoretical models. Management Science, 1989: 35, 982-1003.

[10] Venkatesh V and Davis FD. A theoretical extension of the technology acceptance model: four longitudinal field studies. Management Science. 2000: 46(2): 186-204.

[11] Liang H, Xue Y and Byrd T. PDA usage in healthcare professionals: testing an extended technology acceptance model. International Journal of Mobile Communications 2003: 1(4): 372389.

[12] Toleman M, Ally M and Darroch, F. Aligning Adoption Theory with Agile System Development Methodologies. PACIS2004, July 8-11, Shanghai, China. 2004. Available Online: <URL: http://www.sci.usq.edu.au/staff/markt/ papers/PACIS2004.pdf#search='Aligning%20 Adoption%20Theory%20with%20Agile%20System%20De velopment%20Methodologies' > Last accessed September 29, 2006.

[13] Kline TJB. The Groupware Adoption Scale: a measure of employee acceptance. Human Systems Management, 2001: 20: 59-62.

[14] Kline TJB and Gardiner H. The successful adoption of groupware: Perceptions of the users. Human Systems Management, 1997: 16: 301-306.

Address for correspondence

Dr Lois Burgess
Centre for Health Service Development
University of Wollongong
Northfields Ave,
2522, Wollongong, NSW
Australia

Email: lburgess@uow.edu.au

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

When Usage and User Satisfaction Differ: The Case of an Electronic Discharge Summary

Thomas Bürkle[a], Philip A. Engel[b]

[a] *Department of Medical Informatics, University Erlangen-Nürnberg, Germany*
[b] *ADMED GmbH, Cologne, Germany*

Abstract

We describe the results of a longitudinal study regarding system use and user satisfaction before and after introduction of an electronic discharge letter application in a pediatric intensive care unit (PICU) of a German university hospital. The new discharge letter application is part of the hospital information system (HIS). The study covered an eleven month time period and used system logs as well as questionnaires including a modified questionnaire of user interaction satisfaction QUIS. We used methods which are comparable to a previous study examining a HIS based discharge letter in three departments of an Austrian hospital. In comparison we found out that user satisfaction was lower in our case. Interestingly, we noticed that in our case this was mirrored by an increasing use of the new discharge letter although there was no pressure to switch to the new HIS based discharge letter application.

Keywords:

evaluation, clinical information system, discharge letter

Introduction

Evaluation of information and communication technology (ICT) which is used in a healthcare environment, is required to avoid potential side effects, because ICT can be inappropriately specified, have functional errors, be unreliable, user-unfriendly, ill-functioning or the environment may not be properly prepared to accommodate the ICT in the working processes [1]. The choice of evaluation methods and technologies is wide [2] and depends on available resources, goals of the evaluation and type of the technology to be examined [3]. ICT changes the ways how persons perform activities and influences the working environment over a long period of time. Within this study we were interested in usability and benefit of a ward specific discharge letter application within a hospital information system. Expectation of the users, extent of system use and user satisfaction were the parameters we measured before, during and after the introduction of the discharge letter application. Therefore we selected an interventional study design with three survey periods in order to display the development of the mentioned parameters over time. We used methods derived from other studies [4,5,6 and mainly 7,8] in order to compare our results with the results of other sites and researchers. Our study design tends towards a subjectivist approach [9], but we used classical observation methods only for a small part and relied mostly on questionnaires.

Materials and methods

The Münster University Hospital pediatric ICU

The pediatric intensive care unit PICU of Münster University Hospital has 16 beds and cares for all pediatric departments like pediatric oncology, pediatric cardiology and general pediatrics. A majority of the patients are premature babies and cardiosurgical cases. The PICU has 2.25 senior physician positions, 9.5 physician positions and 37 nursing positions. Nurses and physicians work in three shifts. In 2003 450 patients have been treated with a mean stay of 10.2 days per patient.

The hospital information system

Münster University Hospital introduced the commercial hospital information system Orbis® in 2000 [10], which is today sold by Agfa® and used in more than 400 German hospitals. Orbis® supports the development of user specific applications with a generator tool [11]. Technically Orbis® uses a document based approach. Electronic forms such as a lab result report, a radiology request or a discharge summary can be designed with the generator tool and may look similar to a paper based equivalent. At Münster University Hospital a variety of applications have been developed using the generator tool [12-14]. From our experience, the standard discharge letter which comes with Orbis® is not used in any department because it supplies a form which is almost empty and does not support the user with precompleted data from the patient record. Therefore we developed specific discharge letters for a variety of departments using the generator tool [11]. A ward specific discharge letter is based upon the typical discharge letter setup of the respective ward, it mirrors its structure and subdivision and supplies much prefilled patient information such as recent lab values, discharge diagnoses, last medication etc.

The study

A longitudinal single group non randomized intervention study using questionnaires and system logs was established with 3 survey periods. The intervention consisted of introducing a ward specific discharge letter application within the Orbis® HIS for the pediatric ICU in September 2003. The Orbis® system logged timestamps of first opening and last editing of each discharge letter. This data was analyzed in the study. The use of the new application was not mandatory, physicians could choose to use the old fashioned method of writing discharge letters in MS Word® on the same PC and to store them in a common directory. Discharge letters traditionally written in MS Word® couldn't be monitored reliably, as they were written based upon a previous letter and often subsequently stored under the old filename. Thus they were not considered for the study.

Three different types of questionnaires were distributed at three different points in time, namely before intervention in July 2003 (t1), five months after intervention in January 2004 (t2) and ten months after intervention in June 2004 (t3). Observation objects were all persons writing discharge letters, namely all physicians working regularly at the pediatric ICU in the respective questionnaire period as well as the ward secretary. While formal staffing counts 2.25 senior physician and 9.5 physician positions plus 1 secretary, there is considerable fluctuation due to physicians rotating among the various pediatric departments and some physicians hold part time positions only. We had a return of 14 questionnaires in the first round, 17 in January 2004 and 18 in June 2004. Questionnaires were marked with a numeric code to assure that answers of the same person could be traced through all 3 sampling periods. The follow up group which completed all 3 questionnaires had n=10 participants (9 physicians and 1 secretary)

The questionnaires comprised between 5 and 6 pages and were split into 4 chapters. Parts of the questionnaire were derived from previous studies on discharge letter applications [4,5,6,7,8]. In the first chapter we grouped questions regarding demographic data such as age, profession and time working at the PICU. The second chapter concerned expectations from (at t1) respectively experiences (at t2 and t3) with the electronic discharge letter. Statements were to be answered with yes or no. Besides it contained, similar to Ammenwerth and Kaiser [7,8], a part where the user could assess different activities like writing, modifying and searching discharge letters on a five point Likert scale. The third chapter, available only at t2 and t3, contained 17 questions of the modified and abbreviated QUIS questionnaire [5,6,7,8], which was also measured with a five point Likert scale. For the original QUIS refer to [17]. The fourth chapter contained questions regarding usability of other HIS functions. In the results' section of this paper we will only refer to the first three chapters. The questionnaire was each time complemented by open questions for free commenting. Any question could be skipped by marking "no answer" to avoid bias. Answers on the five point Likert scale have been converted to numerical values between 1 (do not agree at all) and 5 (do fully agree). All statistic results are descriptive only. We will also report some observational results which were made when the discharge letter workflow on the PICU was examined.

Results

Demographics

Our respondents in the follow up group (n=10) had a median age of 37 years and had been working 1-3 years (median) at the PICU at t2. At t3 the median stay time at the PICU had shifted to 4-10 years. One person had been working for over 10 years at the PICU. All ten had previous experience with computers. Of a total of 14 respondents at t1 twelve owned a PC themselves.

System use

Figure 1 shows the system use in terms of cumulated PICU discharge letters written in Orbis® during the study. A total of 568 discharge letters were recorded in the study period between August 13th, 2003 and June 29th, 2004 (48 weeks or 336 days). The corresponding graph shows a steady rate of approx. 7.04 discharge letters per week until January 04, afterwards we noted an increase to 15.3 letters per week.

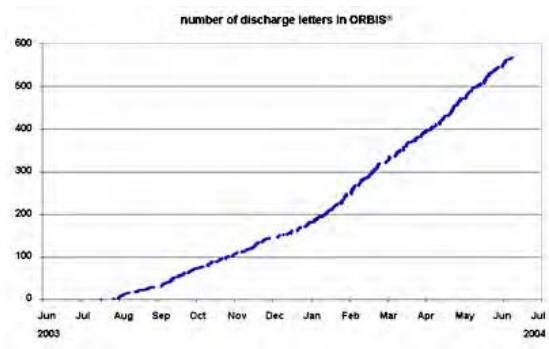


Figure 1 - system use: letters written with Orbis®

Expectations and experience

The expectations and experiences chapter was part of all three questionnaires. It comprised four questions to compare the expected benefits with those realized when the application was active (see also [4]). At t1 the question would e.g. read "I can imagine that the Orbis® discharge letter will be time saving for me" while at t2 and t3 we asked "Did you save time using the Orbis® discharge letter?". In a similar fashion the other three questions examined how easy it was to learn the use of the new application, if the application would make work easier and if it would improve the quality of patient care. All four questions had to be answered either with "yes", "no" or "Don't know". Results are given in table 1 for the follow up group with n=10 participants. The table uses an index of approval which is calculated as the relation between the number of positive answers and all answers. If all 10 participants would have answered positively, this index would

have been 1.0, if all users would have expressed a negative opinion it would have turned out 0.0.

Table 1 - Expectations and Experiences, index of approval

Index of approval	expected	experienced	
	t1	t2	t3
Easy learning	0.5	0.8	0.7
Easier work	0.9	0.6	0.7
Time saving	0.9	0.4	0.5
Better quality	0.7	0.0	0.1
Total	0.8	0.5	0.5

For time saving and easier working we measured high expectation values of 0.9 at t1, a slightly lower value for quality improvement and an indecisive (0.5) opinion regarding ease of learning. After system introduction at t2 and t3 the experienced values were found to be much lower (near 0.5) both for time saving and workload improvement while learning the use of the discharge letter application turned out to be easier than expected. At t3 experienced workload improvement scored a little higher at 0.7 but didn't reach the 0.9 at t1 which reflects the high expectation in the new application. Hardly any participant noticed an improvement in quality of patient care at t2 and t3.

Results of QUIS

Table 2 demonstrates the results gained with the modified QUIS [5,6,7,8]. QUIS was only included at t2 and t3, n is the total number of valid responses received for the respective question and varies for t2 (17 returned questionnaires) and t3 (18 returned questionnaires).

Table 2 - QUIS results at t2 and t3. Median, 25% Quantil and 75% Quantil and (for comparison) weighted mean values

No	Question	t2					t3				
		n	Q0.25	**Q0.5**	Q0.75	X	n	Q0.25	**Q0.5**	Q0.75	X
c11	Extra work with no visible benefit	16	4.0	**4.0**	4.3	3.71	18	2.3	**4.0**	4.0	3.33
c12	Easy to learn	17	3.0	**4.0**	4.0	3.47	18	3.3	**4.0**	4.0	3.67
c13	Others have benefit but not me	16	3.8	**4.0**	4.0	3.41	17	3.0	**3.0**	4.0	2.94
c14	Essential functionality is missing	16	2.0	**2.0**	3.0	2.18	18	3.0	**4.0**	4.8	3.72
c15	Well adaptable to needs	15	2.5	**3.0**	3.5	2.65	17	2.0	**3.0**	3.0	2.67
c16	Use is simple and self explaining	17	2.0	**3.0**	4.0	3.06	18	2.0	**3.0**	4.0	3.00
c17	Effort of use is appropriate	15	2.0	**2.0**	4.0	2.47	16	2.0	**2.5**	4.0	2.50
c18	Letters and results faster available	17	3.0	**4.0**	4.0	3.47	18	2.0	**3.5**	4.0	3.17
c19	Helpful persons to turn to available	16	2.0	**3.0**	4.0	3.00	17	2.0	**3.0**	3.0	2.56
c110	User training is unsatisfactorily	17	2.0	**3.0**	4.0	2.94	14	2.0	**3.0**	4.0	2.22
c111	My needs were respected	16	3.0	**3.0**	4.0	3.00	16	1.8	**3.0**	3.3	2.33
c112	User surface is nonuniform	15	2.0	**3.0**	4.0	2.76	14	2.3	**3.0**	4.0	2.44
c113	Makes information transfer easier	16	2.0	**3.5**	4.0	2.88	15	3.0	**4.0**	4.0	2.89
c114	Unreliable system	16	3.0	**4.0**	4.0	3.29	15	3.0	**4.0**	4.0	3.00
c115	Makes letter writing easier	12	2.0	**3.0**	4.0	2.00	13	2.0	**3.0**	3.0	2.00
c116	Is helpful for clinical research	11	1.0	**2.0**	4.0	1.53	14	2.0	**2.0**	2.0	1.56
c117	Few chances to do things wrongly	16	2.0	**3.0**	4.0	2.88	16	2.0	**2.0**	3.0	2.22
	Summary user interaction satisfaction					3.04					2.89

The table contains median, 0.25 Quantil and 0.75 Quantil as well as mean values (for later comparison with [8]). Values of six negatively phrased questions have been reversed to achieve a constant scale between 1 (do not agree at all) to 5 (do fully agree) for all QUIS questions. In order to compensate for questions which were not answered by all participants, mean values have been weighted according to [8] for the calculation of the summary user interaction satisfaction.

The results show a remarkable decrease in user satisfaction for several statements. E.g. statement c14 reads "There are essential functionalities missing in the discharge letter application". We note an agreement of 2="do not agree" for this statement at t2, five months after system usage. However at t3 after ten months system use, most users agree with this statement at 4="do agree". Question c117 reads "There are few things which can be made wrong when using the system". At t2 this statement was seen neutral at 3. Five months later users disagreed at 2="do not agree".

Accordingly the summary user interaction satisfaction score, calculated according to [8] on the base of weighted mean values, shows a decline from 3.04 at t2 to 2.89 at t3.

In table 3 we demonstrate the results concerning the assessment of different activities connected with the discharge letter (part of chapter 2 of the questionnaire). Results in a converted five point Likert scale are given for the follow up group with n=10.

Table 3 - median user satisfaction discharge letter activities

Activity	User satisfaction		
	t1	t2	t3
Write new letter	2.0	3.0	3.0
Modify letter	3.0	3.0	3.5
Sign letter	4.0	3.0	3.0
Search letter	2.5	4.0	4.0

Apart from the sign procedure for letters, results at t2 and t3 after introduction of the Orbis® discharge letter are equal or better in comparison to results that were received when writing discharge letters in MS Word®. Most values are slightly positive at or above the median value 3="don't know". Searching a discharge letter seems to have improved with values of 2.5 before and 4.0 afterwards.

Discussion

The results of this study indicate an increased usage of a HIS integrated discharge letter function within a pediatric intensive care unit although there was no pressure to switch from the previous method of writing discharge letters in MS Word®. Even at the end of the study, some very complicated discharge letters were still written in MS Word® but the increasing amount of Orbis® discharge letters clearly indicates that physicians decided to preferably use the HIS function. The total number of 568 discharge letters in 336 days written in Orbis® correlates fairly well with the average number of patients treated per year in this unit considering the fact that for most patients only one discharge letter is written. Furthermore, on request of the pediatricians the discharge letter function has in the meanwhile been spread to other pediatric wards successfully.

This increased usage is in contrast to a decrease in user satisfaction with the new electronic system measured with the QUIS part of the questionnaire. Expectations which had been set into the new application have not been met (see table 1). On the other hand user responses with regard to performing different activities such as writing, modifying and searching discharge letters seem to be more positive at t2 and t3 (values between 3 and 4 on a Likert scale).

The study is investigative rather than confirming, no hypotheses have been set or confirmed. The study design tends to be subjectivistic. From an objectivistic viewpoint there are some weak points in study design. It was non randomized and went over a long time period of 11 months. An influence of increasing workload and other factors cannot be excluded. We have a low number of study participants although we did include all staff concerned with the discharge letter application in the PICU. We enforced a high questionnaire return rate with the help of a senior PICU physician who assisted in distributing and collecting the questionnaires. This could have lead to bias in non responders or persons who are indecisive. We tried to minimize this effect by offering a no response opportunity for each question. Due to the small number of participants we did not do a formal evaluation with test-retest or split half methods for reliability and we do not have a formal gold standard to assess validity for the questionnaires. We may however state that the modified QUIS part (the original can be found in [17]) was evaluated and used in an identical fashion for the assessment of discharge letter functions by Ohmann, Boy and colleagues [5,6] and then again in a study by Ammenwerth and Kaiser [7,8]. The latter determined Cronbachs Alpha for reliability of this part with an excellent value of 0.9257 (max would be 1.0) and found good correlation of 0.78 between general satisfaction and QUIS determined satisfaction as an indicator for reasonable validity. Furthermore Ammenwerth and Kaiser [8] used similar questions for the assessment of activities performed with the new application (table 3). Other techniques (table 1) were used with good success in a previous study by ourselves [4]. While QUIS values were calculated on top of all respondents at t2 and t3 respectively, results given in table 1 and table 3 were calculated for the follow up group with n=10 only. However we did compare some of those questions for the follow up group and all residents at times t1, t2 and t3 exemplarily and found only small differences in the respective satisfaction values.

When looking at results of our study we may state in comparison to Ammenwerth and Kaiser [8] that our respondents were less happy with the discharge letter function. In our case summary user interaction satisfaction (weighted mean value) was between 3.04 at t2 and 2.89 at t3 whereas Ammenwerth and Kaiser measured higher values between 2.80 (neurology department) and 3.73 (transplantation surgery) among physicians in different units and 3.23 (transplantation surgery) as well as an extraordinary high value of 4.18 (neurology) among secretaries. In their publication [8] activities such as writing, modifying and searching discharge letters range among 1.67 (modifying letters in neurology) to 5.0 (secretaries searching letters in transplantation surgery) with an average of 2.83 for neurology and 4.17 in internal medicine. Our values from table 3 compare best to the neurology department which is least happy in the Ammenwerth and Kaiser study.

Ammenwerth and Kaiser explain differences between departments with organisational aspects in writing discharge letters. They assume that the department which previously had the best discharge letter workflow (neurology) reported the highest additional effort to switch to the new discharge letter and scored the lowest user interaction satisfaction. When examining the answers to open questions, we found comments that the layout in the Orbis® discharge letters was less favorable than in Word, that users complained about too much data (e.g. lab values) which was included automatically in the Orbis® discharge letter and that those letters were not specific enough. During observation we noticed that writing a PICU discharge letter at Münster university hospital is a fairly complicated

process where several physicians work on one letter during several days. In open questions some physicians complained that the long lasting procedure of assembling the discharge letter leads to overtime work which demonstrates a certain degree of general dissatisfaction with the process [19].

These observations could be compatible with a situation where a complicated workflow (namely writing a PICU discharge letter) was well accomplished with an established method (copying an old discharge letter of another patient in MS Word® and modifying it) and the introduction of a new method (writing discharge letters in Orbis®) was not perceived as an organizational improvement. But then, why did the physicians, without being forced to do so, switch to the new method for most cases and did even ask to spread the new method to more pediatric wards?

Conclusion

A subjectivistic study design like the one presented here will not confirm or deny any hypothesis but rather intends to focus on interesting aspects which may then be examined with objectivistic or further subjectivistic methods. In our case it revealed an interesting discrepancy between system use and user satisfaction. More studies will be required to examine differences in workflow before and after system introduction as well as other external influencing factors which can explain for the unexpected result.

Acknowledgements

We would like to thank all colleagues of the Münster University Hospital PICU, our special thanks go to Dr. I. Hörnig-Franz und Mrs. A. Kulms.

References

[1] Ammenwerth E, Brender J, Nykänen P, Prokosch HU, Rigby M, Talmon J. Visions and strategies to improve evaluation of health information systems - Reflections and lessons based on the HIS-EVAL workshop in Innsbruck. Int J Med Inf. 2004;73:479-491.

[2] Friedmann CP, Wyatt JC. Evaluation methods in Biomedical Informatics. Springer Science and Business Media New York 2006.

[3] Bürkle T, Ammenwerth E, Prokosch HU, Dudeck J. Evaluation of clinical Information Systems - What can be evaluated and what cannot? J Eval Clin Pract. 2001 Nov;7(4):373-385.

[4] Bürkle, T, Kuch R, Prokosch HU, Dudeck J. Stepwise Evaluation of Information Systems in an University Hospital. Methods Inf Med 1999;38:9-15.

[5] Ohmann, C, Boy O, Yang Q. A systematic approach to the assessment of user satisfaction with healthcare systems: Constructs, models and instruments. Stud Health Technol Inform. 1997;43 Pt B:781-785.

[6] Boy O, Ohmann C, Aust B, Eich HP, Koller M, Knode O, Nolte U. Systematic evaluation of user satisfaction of physicians with a hospital information system - initial results. Stud Health Technol Inform. 2000;77:518-22.

[7] Ammenwerth E, Kaiser F, Wilhelmy I, Hofer S. Evaluation of user acceptance of information systems in health care - the value of questionnaires. Stud Health Technol Inform. 2003;95:643-648.

[8] Ammenwerth E, Kaiser F. Evaluation der Benutzerzufriedenheit mit der rechnergestützten Arztbrief- und Befundschreibung. UMIT Schriftenreihe 1/2002, Innsbruck 2002.

[9] Moehr JR. Evaluation: salvation or nemesis of medical informatics? Comp Biol Med 2002;32:113-125.

[10] Prokosch HU, Bürkle T, Frankewitsch T, Ganslandt T, Gieshoidt K, Janßen U, Müller M, Lange M, Lüders D, Saile F, Vaclavik M, Veltmann U. Vom Krankenhauskommunikationssystem zur Elektronischen Krankenakte: Eine Statusbeschreibung zum KIS-Ausbau am Universitätsklinikum der Westfälischen-Wilhelms-Universität Münster. Informatik, Biometrie und Epidemiologie in Medizin und Biologie 2001;32:252.

[11] Kuhn KA, Lenz R, Elstner T, Siegele H, Moll R. Experiences with a generator tool for building clinical application modules. Methods Inf Med 2003;42(1):37-44.

[12] Bürkle T, Ganslandt T, Tübergen D, Menzel J, Kucharzik T, Neumann K, Schlüter S, Müller M, Veltmann U, Prokosch HU. Guideline Based Structured Documentation: The Final Goal? Stud Health Technol Inform 2002;90: 256-261.

[13] Müller ML, Bürkle T, Irps S, Roeder N, Prokosch HU. The diagnosis related groups enhanced electronic medical record. Int J Med Inf. 2003;70(2-3):221-228.

[14] Kunze U, Ganslandt T, Preusser P, Bürkle T, Prokosch HU, Senninger N, Haier J. Ein interdisziplinäres Modell zur onkologischen Dokumentation in einem Klinikinformationssystem. Zentralbl Chir. 2004;129(1):10-13.

[15] Chin J. Development of a tool measuring user satisfaction of the human-computer interface. In Chi'88 Conf. Proceedings: Human factors in computing. Association for Computing Machinery New York 1988:213-218.

[16] Ammenwerth, E, Gräber S, Herrmann G, Bürkle T, König J. Evaluation of health ISs-Problems and challenges. Int J Med Inf. 2003;71(2/3):125-135.

[17] Engel PA. EDV im Krankenhaus: Evaluation der Einführung der elektronischen Arztbriefschreibung am UKM. Inaugural Dissertation Universität Münster 2004. Available under http://miami.uni-muenster.de/servlets/DerivateServlet/Derivate-2177/01_diss_engel.pdf

Address for correspondence

Thomas Bürkle, Lehrstuhl für Medizinische Informatik, Krankenhausstrasse 12, 91054 Erlangen, Germany.
thomas.buerkle@informatik.med.uni-erlangen.de

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Mapping Clinicians' Perceptions about Computerized Protocol Use to an IT Implementation Framework

Shobha Phansalkar[a], Katherine A. Sward[b], Charlene R. Weir[a,d], Alan H. Morris[c]

[a] *Department of Biomedical Informatics, University of Utah, Salt Lake City, Utah, U.S.A*
[b] *College of Nursing, University of Utah, Salt Lake City, Utah, U.S.A*
[c] *Director of Research, LDS Hospital, Pulmonary Division, Salt Lake City, Utah, U.S.A*
[d] *Associate Director, Education and Evaluation, VA Salt Lake City GRECC*

Abstract

Previous studies have described the determinants of successful information technology (IT) implementation. In 2003, Kukafka et al. integrated several theoretical perspectives and proposed a framework for IT implementation. This framework is applicable to IT implementation in general but lacks the identification of factors affecting adoption, which are specific to the technology under consideration. We developed and validated a model that specifically identifies factors associated with clinicians' adoption of computerized protocols. The purpose of this paper is to identify the relations between the specific factors associated with intention to use computerized protocols and the high level variables that constitute the framework proposed by Kukafka et al. Incorporation of a specific model into a general schema for IT implementation allows implementers to assess the specific individual, organizational and environmental changes required to bring about successful implementation of computerized protocols. An understanding of clinicians' perceptions specific to the technology in use will allow its seamless integration into an organization's healthcare IT plan. Strategic planning requires enhancing the framework with additional detail related to the specific technology under consideration.

Keywords:

computerized protocols, clinician perceptions, implementation, clinical protocols [MeSH], decision making [MeSH], factor analysis [MeSH]

Introduction

Computerized protocols are knowledge-based systems used to provide decision-support to clinicians, for management of patients. Computerized protocol assisted decision support can range from receiving guidelines for disease management[1] to receiving individualized, detailed and explicit instructions, delivered at the patient's bedside, that are driven by dynamic patient-specific data[2]. Our use of the term 'computerized protocols', in this study, refers to explicit protocols that are capable of providing the latter, more specific support.

Explicit computerized protocols are capable of providing decision-support in a variety of complex clinical situations [3-6] at the bedside of the patient [2, 3]. Despite their advantages, resistance to adoption of evidence-based medicine, also labeled as 'clinical inertia', has impeded the implementation of protocol based decision support systems[4, 5]. Focusing on the technological aspects of information systems rather than their behavioral, organizational or attitudinal impacts has been identified as a cause for information system failure[6, 7]. To counter this resistance several behavioral models have been used for guiding implementation strategies[8]. Grimshaw and colleagues suggested the exploration of theory-based evaluations alongside randomized trials of dissemination and implementation strategies[9]. Such theoretical approaches not only provide a framework for understanding specific determinants of provider behavior but also allow exploration of causal mechanisms of adoption.

In 2003, Kukafka et al. proposed an IT implementation framework that integrated variables from prominent behavioral theories and models[10]. This multi-factor model can help in understanding the determinants involved in implementation and successful adoption of information systems in a health-care organization[10]. This framework is, however, applicable to IT implementation in general and previous studies have expressed the need for research on factors affecting adoption that are specific to the technology under consideration, so as to improve their predictive ability [11-13]. There is a paucity of literature that specifically identifies the cognitive and attitudinal factors associated with clinicians' adoption of computerized protocols [5, 14]

Previously, we developed a multivariate model of clinicians' behavioral intention to use computerized protocols; and we developed and validated a survey instrument, based on the model, that measures clinicians' behavioral intention to use computerized protocols [15]. This was a task-specific model focused on computerized protocol use, rather than information systems in general. The purpose of this paper is to identify the relationships between the specific factors associated with intention to use computerized protocols and the more general, high level framework proposed by Kukafka et al. Incorporation of the specific

model into a general schema for IT implementation places the specific model within a larger context and allows implementers to assess the specific individual, organizational and environmental changes required to bring about successful implementation of computerized protocols.

Methods

Development and validation of a model predicting clinicians' intention to use computerized protocols

A 'Grounded Theory' approach [16] was adopted for the development of a model predicting clinicians' intention to use computerized protocols [15]. This approach is used extensively in the social sciences and is based on an inductive rather than deductive approach to theory development. We conducted semi-structured interviews of clinicians with extensive experience in the use and development of explicit computerized protocols at LDS Hospital, in Salt Lake City, Utah. Five physicians, 3 nurses and 6 respiratory therapists were interviewed. The results of this work are published elsewhere[17] but we will summarize them briefly here. Three reviewers examined the transcripts looking for themes. After substantial group discussion, 39 themes were identified. The themes were then reduced and categorized into constructs derived from Expectancy-Value Theories[18], especially the Theory of Planned Behavior[19] and Intrinsic Motivation theories[20]. Inter-rater reliability for matching themes to constructs, was measured using Cohen's kappa ($k = 0.48$) and consensus was attained following discussion.

The constructs of the model for predicting computerized protocol use were subsequently used to develop items for a survey instrument. For each of the 8 constructs, we constructed up to 5 items. We administered the instrument to 240 physicians, nurses and respiratory therapists from University of Utah Hospital, Veterans Affairs Medical Center, Salt Lake City, and Intermountain Healthcare (LDS Hospital and Cottonwood Hospital). Factor analysis identified nine factors that accounted for 66% of the total variance cumulatively. Factors identified were: Beliefs regarding Self-Efficacy, Environmental Support, Role Relevance, Work Importance, Beliefs regarding Control, Attitude towards Information Quality, Social Pressure, Culture, and Behavioral Intention. The strongest predictor was Beliefs regarding Self-Efficacy, which accounted for 26% of the total variance of intention to use explicit computerized protocols. Results supported the reliability and construct validity of the instrument[15].

Mapping constructs predicting computerized protocol use to the IT implementation framework

The validated factors identified above were mapped to the IT implementation framework proposed by Kukafka et al. The IT implementation framework is adapted from Green and Kreuter's PRECEDE and PROCEED model [21]. Kukafka et al. proposed the adoption and application of the original model to IT implementation. This model identifies educational and organizational strategies for IT implementation. Underlying these strategies are three categories of factors, namely predisposing, reinforcing and enabling factors, which if modified, will be most likely to result in behavior change, i.e. adoption of the technology.

Results

The constructs from the model for predicting computerized protocol use [15] were mapped to the Phase 4 (Educational and Organizational) assessment in the IT Implementation Framework. Outlined below is the mapping of the specific constructs related to computerized protocol use to the Phase 4 assessment factors in the IT implementation framework. We have also included the definitions for the specific constructs.

- **Predisposing Factors**: These are factors related to the characteristics of an individual that motivates behavior change.
  - ***Work Importance***: An individual's tendency to orientate and value work in general.
  - ***Role Relevance***: The degree to which the computerized guidelines are relevant to one's perceived role at work.
  - ***Beliefs regarding Control***: The belief that that the behavior in question is in the control of the person.
  - ***Beliefs regarding Self-Efficacy***: The degree to which one believes that one has the skill to effectively engage in the behavior.
  - ***Attitudes towards Information Quality***: The attitude of the individual towards the quality of information that drives the logic of the computerized protocol.
- **Reinforcing Factors**: These are extrinsic factors in the form of rewards or punishments that are anticipated as a consequence of behavior change.
  - ***Social Pressure***: The perceived pressure that an individual might face from peers or supervisors that might facilitate or impede the behavior change.
- **Enabling Factors**: These are characteristics of the environment that facilitate the change by providing the skills or resources needed to bring about the behavior.
  - ***Culture***: This refers to the specific norms of the organization and whether the behavior change is in keeping with the goals of the organization itself.
  - ***Environmental Support***: The degree to which the environment is perceived as supportive to bring about the behavior change.

The specific relationships of the constructs to the overarching factors from the IT implementation framework are detailed in Figure 1. Figure 1 also illustrates how the specific model predicting behavior change fits into the realm of the general schema for IT implementation.

Discussion

Understanding user adoption of information systems is crucial for successful implementation. To facilitate this understanding several behavioral models have been employed for the adoption of information systems. A rich body of literature, incorporating models from various

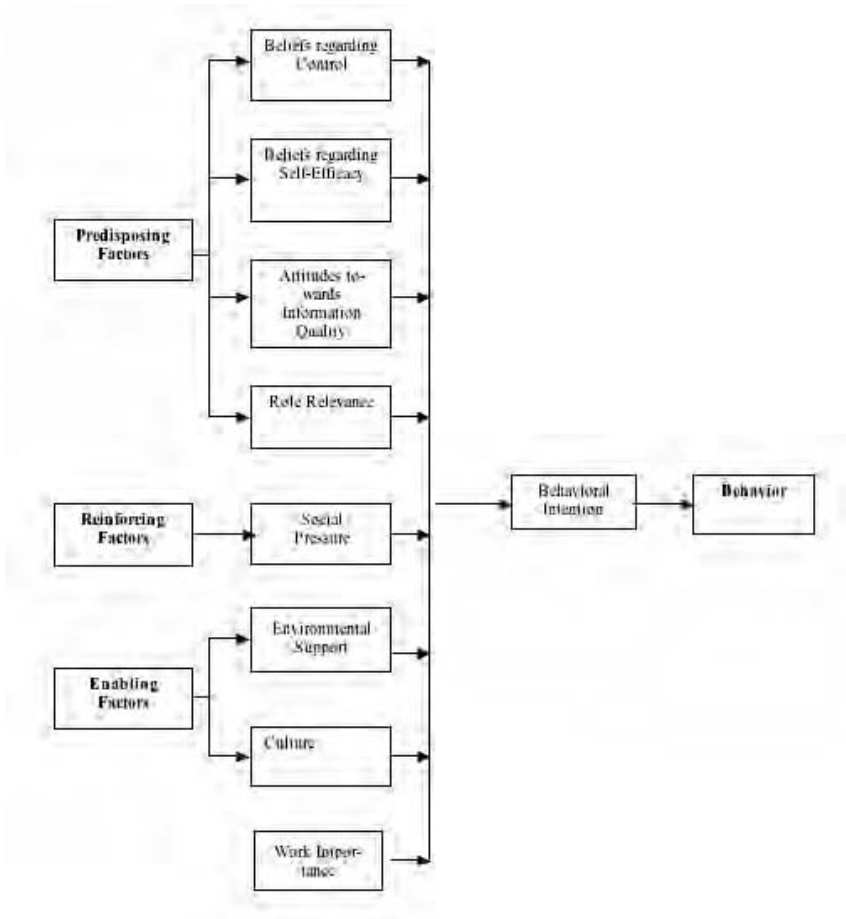


Figure 1 - Mapping of constructs related to computerized protocol use to the IT implementation framework

domains such as social and cognitive sciences, psychology, and the business sciences, has emerged. The constructs proposed in each of these models allow system implementers to understand the barriers to end-user adoption and design strategies for successful implementation of these systems.

Kukafka et al.[10] synthesized this literature and proposed an integrated multi-level framework for IT implementation. This framework is the first complex, multi-dimensional model that takes into account a variety of individual as well as organizational factors. Additionally this model identifies the factors that influence behaviors linked to IT use.

We developed and validated the model and the corresponding survey instrument specific to computerized protocol use. We then assessed how our model fits into the general schema for IT implementation, as proposed by Kukafka et al. Two constructs of Education and Policy Organization from the Kakafka model did not map onto any constructs in our model. For each of the predisposing, reinforcing and enabling factors, we identified specific constructs related to computerized protocol use.

The framework proposed by Kukafka et al. is designed as a foundation for driving IT implementation efforts in general. This IT implementation framework is designed to help IT implementers assess the individual and organizational changes needed for installment of a new innovation. However, identification of specific constructs related to the use of the technology under consideration need to be developed. These specific constructs allow IT implementers to make judgments about how the specific technology would fit into their healthcare IT plan. Also, identification of changes related to a specific technology would result in being able to make informed judgments about the timeline for implementation for a new technology. The IT framework serves as a guideline but needs explicit underlying constructs in order to make it readily usable. Our model for predicting computerized protocol use adds rich detail to the IT implementation framework, and can help explain, how specific predisposing, reinforcing, and enabling factors determine behavior related to use of computerized protocols.

This paper identifies how the IT implementation framework can be adapted for use with computerized protocol implementation. Identification of constructs related to

other technologies and how these can fit into the high-level factors proposed in the framework need to be assessed.

Conclusion

Incorporation of a specific model into a general schema for IT implementation allows implementers to assess the specific individual, organizational and environmental changes required to bring about successful implementation of computerized protocols. An understanding of clinicians' perceptions specific to the technology in use will allow its seamless integration into an organization's healthcare IT plan. Strategic planning requires enhancing the framework with additional detail related to the specific technology under consideration.

Acknowledgments

I acknowledge the contributions of Dr. Homer Warner for valuable help and discussions that led to the development of the model and this paper. Supported in part: NIH/ NHLBI ARDS Network: No1-HR-46062.

References

[1] A. Woodward, M. Wallymahmed, J. Wilding, G. Gill. Successful cardiovascular risk reduction in Type 2 diabetes by nurse-led care using an open clinical algorithm. Diabet Med 2006;23(7):780-787.

[2] A. H. Morris. Rational use of computerized protocols in the intensive care unit. Crit Care 2001;5(5):249-254.

[3] M. Vogelzang, F. Zijlstra, M. W. Nijsten. Design and implementation of GRIP: a computerized glucose control system at a surgical intensive care unit. BMC Med Inform Decis Mak 2005;5:38.

[4] R. Grol, J. Grimshaw. From best evidence to best practice: effective implementation of change in patients' care. Lancet 2003;362(9391):1225-1230.

[5] L. S. Phillips, W. T. Branch, C. B. Cook, J. P. Doyle, I. M. El-Kebbi, D. L. Gallina, C. D. Miller, D. C. Ziemer, C. S. Barnes. Clinical inertia. Ann Intern Med 2001;135(9):825-834.

[6] S. A. Davis, R. P. Bostrom. Training end users: an experimental investigation of the roles of the computer interface and training methods. . MIS Quarterly 1993;17:61-85.

[7] G. Gery. Granting three wishes through Performance-Centered Design. Communication of the ACM 1997(July).

[8] S. E. McLean, L. A. Jensen, D. G. Schroeder, N. R. Gibney, N. M. Skjodt. Improving adherence to a mechanical ventilation weaning protocol for critically ill adults: outcomes after an implementation program. Am J Crit Care 2006;15(3):299-309.

[9] J. M. Grimshaw, M. P. Eccles, A. E. Walker, R. E. Thomas. Changing physicians' behavior: what works and thoughts on getting more things to work. J Contin Educ Health Prof 2002;22(4):237-243.

[10] R. Kukafka, S. B. Johnson, A. Linfante, J. P. Allegrante. Grounding a new information technology implementation framework in behavioral science: a systematic analysis of the literature on IT use. J Biomed Inform 2003;36(3):218-227.

[11] T. Williams, C. May, F. Mair, M. Mort, L. Gask. Normative models of health technology assessment and the social production of evidence about telehealth care. Health Policy 2003;64(1):39-54.

[12] L. M. Hilz. The informatics nurse specialist as change agent. Application of innovation-diffusion theory. Comput Nurs 2000;18(6):272-278; quiz 279-281.

[13] V. Venkatesh, F. D. Davis. A theoretical extension of the technology acceptance model: Four longitudinal field studies. . Management Science 2000;46(2):186-204.

[14] F. D. Davis. Perceived usefulness, perceived ease of use, and user acceptance of information technology. . MIS Quarterly 1989;13(3):319-340.

[15] S. Phansalkar, C. R. Weir, A. H. Morris, H. R. Warner. Clinicians' perceptions about use of computerized protocols: A multicenter study. Int J Med Inform. 2007;Accepted, In Press.

[16] A. Strauss, J. Corbin. Basics of qualitative research: Grounded theory procedures and techniques. Newbury Park: Sage; 1990.

[17] S. Satsangi, C. R. Weir, A. H. Morris, W. H.R. Cognitive Evaluation of the Predictors of Use of Computerized Protocols by Clinicians. In: Proceedings of the American Medical Informatics Association; 2003; Washington D.C.: Hanley & Belfus; 2003.

[18] M. Fishbein. Attitude and the prediction of behaviour. New York: Wiley. ; 1967.

[19] M. Fishbein, I. Ajzen. Belief, attitude, intention, and behavior: An introduction to theory and research. Reading, MA: Addison-Wesley; 1975.

[20] E. L. Deci. Intrinsic Motivation. New York: Plenum; 1975.

[21] L. Green, M. Kreuter. Health Promotion Planning. 2nd edition ed. Mountain View: Mayfield Publishing Co.; 1991.

Address for correspondence

Shobha Phansalkar, RPh, MS
Department of Biomedical Informatics,
University of Utah School of Medicine
26 South 2000 East, Suite 5700 HSEB
Salt Lake City, UT 84112-5750, USA
E-mail: shobha.phansalkar@hsc.utah.edu
Tel: +1-801-582-1565 Extn: 2221, Fax: +1-801-584-5640

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

e-Health in Scotland: Setting a Baseline for Stakeholder Alignment

Sharon Levy, Anne Casey, Alison Wallis

Royal College of Nursing Dundee /London/ Lothian, UK

Abstract

Gaining knowledge of nurses' attitudes towards and understanding of various aspects of the Scottish e-Health programme is vital for needed 'stakeholder alignment'. This paper is focused on the Scottish results from a large on-line survey carried out in 2006, across the UK. Key findings, identified through analysis of both qualitative and quantitative data, are discussed. Results suggest that overall there is willingness and enthusiasm to engage and to see the modernisation of the NHS in Scotland underpinned by advanced and effective IT systems. It also shows that nurses have clear ideas about how information technology could help them and their patients in delivering safe care that enhances the visibility of the nursing contribution to care outcomes. It is argued that results help in setting a base-line from which to judge the success or otherwise of the needed e-Health clinical change management programme within clinical settings.

Keywords:

e-Health, nursing, Scotland, stakeholders alignment

Introduction

Moving away from a reactive, crisis-management, acute-orientated care towards anticipatory, preventative and continuous care which is integrated and supports faster, safer, more efficient, patient-centred services is the required direction of travel set for the National Health Service (NHS) in Scotland [1]. e-Health, the National Programme for the management of information and technology in the Scottish NHS was noted as a key driver to achieving such vision, in line with recommendation made in the Kerr review [2]. That report suggested that the Scottish Executive refocus its e-Health strategy and appoint a visible and high profile leadership so as to ensure clinical buy-in. The Scottish Executive appointed, in late 2006, a full time e-Health Nursing, Midwifery and Allied Health Professions (NMAHP) lead to promote and support the needs of the largest group of professional care providers in the health service.

One of the key priorities for the NMAHP lead is to achieve what the Auditor General for Scotland has called 'stakeholders alignment'. This phrase is used in the published national audit report on management of IT in the Scottish health service [3], and is amongst the 10 'leading practices' set to support effective implementation of clinical IT systems. The report also notes that the Scottish Executive is taking steps to promote its e-Health strategy, "in line with good practice", and yet there is no signposting to available data that set a baseline from which to judge the success or otherwise of the needed stakeholders alignment. This paper aimed to fill this apparent gap, through reporting on the analysis of the Scottish data extracted from the largest UK nursing on-line IT survey undertaken to date [4]. It is argued that the presented results should be used to identify a development plan as well as to ascertain current levels of knowledge and engagement of the nursing professions with the e-Health national programme in Scotland. Adopting the findings as an agreed base line will enable the programme to chart future progress of their alignment with the nurses, health visitors and midwives who provide care for the population of Scotland.

Method

The method chosen to conduct our survey was an on-line format and we used the Nursix.com portal, which provided immediate quantitative data analysis and feedback. The questionnaire which was piloted and used in early 2004 [5] and developed further in 2005 [6] included 7 demographic questions, 23 multiple choice and 1 open ended question for free text input. The 2006 study commenced on 31st of May and lasted for 34 days. Participants were asked to take part in the study through direct e-mail approach (n = 50,000), promotion on the public part of the Royal College of Nursing (RCN) web site, the RCN on-line discussion zone and RCN Bulletin (circulated to all 392,000 members). The results from analysis of completed questionnaires (n = 4453) makes this study the largest nursing e-Health survey in the UK to date. Amongst the total UK sample were 415 Scottish nurses (9%) and this paper is focused on analysis of their response.

Despite the fact that the proportion of Scottish nurses within the sample population is similar to their proportion within the RCN membership, we are not claiming the sample is representative of the nursing, midwifery and health-visiting workforce in Scotland. This was a convenience sample of self-selected RCN members who are, nevertheless, more likely to be interested in and aware of NHS e-Health developments. After all, subjects willingly completed the survey which relied on a degree of competence in using a computer and the internet. This makes the reported findings about these nurses' level of e-Health awareness even more significant.

Results

In response to the question '*How much information have you had about NHS IT developments?*' 35% of respondents from Scotland felt they had fully or reasonably adequate information. This is a slight increase of awareness amongst subjects compared with 20% in 2004 and 29% in 2005. Nevertheless, 38% had inadequate information and a further 28% stated they had no information about these developments at all. A similar picture is noted with awareness of integrated electronic health record developments. Most (72%) nurses had inadequate or no information about such development at all. Interestingly, 42% of respondents said that e-Health developments were a 'Very Important' or 'Important' priority at their place of work with a further 31% who were 'unsure' about the corporate priority that e-Health gets within NHS Scotland. Nevertheless, one in three (32%) of the Scottish nurses who responded to the survey felt that spending large amount of money in the pursuit of e-Health was a poor (24%) or very poor use of NHS resources.

When asked how important is consultation with individual practicing clinicians about new NHS IT developments in Scotland, the majority of nurses believe that consultation was very important (55%) or fairly important (30%). However, when asked what consultation they have had about integrated electronic health care records, 69% told us that their views were not sought. It is worth noting that only 1% of respondents explicitly indicated that they do not wish to be consulted.

Despite many having inadequate information about proposed e-Health developments, many respondents (56%) believe that integrated electronic health care records will improve clinical care with 49% agreeing that using electronic patient health records will lead to significant (31%) or slight improvements to their nursing practice. Thirty percent of the nurses were 'unsure' about the effect and impact of such developments on their practice.

The final questions regarding integrated electronic health care records were set to identify possible effects on the confidentiality of care records and perceived impact on relationship with patients. Thirty nine per cent of respondents believe that there would be little or no effect on confidentiality of care records. However, a significant minority (31%) believe that the use of electronic records may pose a threat to confidentiality. Although 46% of these Scottish nurses felt that electronic records will have no or little effect on the therapeutic relationship with patients, 38% suggested that such development will have a beneficial impact.

Training and support needs of nursing staff who are expected to use new Scottish e-Health systems and products were also sought. As was noted in results from the 2004 and 2005 surveys, the vast majority (95%) of respondents indicated that training is 'Very Important' (83%) or 'Fairly Important' (7%) to the success of the Scottish e-Health programme. When asked how much training they had received in working time in the last six months, 71% had no training at all in working time (compared with 68% in 2005 and 69% in 2004). The majority (68%) of Scottish subjects also felt that around the clock technical support is essential to the success of the integrated electronic health care records with a further 20% indicating it was important or fairly important (6%).

Access to hardware was also included in the survey: respondents were asked how many people share a computer in their immediate clinical area. Interestingly, Scotland had the worst ratios for hardware with 15% needing to share access to a computer with more that 30 other colleagues with a further 9% sharing it with more than 20 people and 11% with more than 10 others.

Apart from gathering quantitative data, the online survey also gave participants an opportunity to add free text regarding 'anything else they wish to note'. A quarter of those who completed the questionnaire opted to add qualitative information and thus provided a personal reflection on different aspects of e-Health as it relates to their practice area and experience. The main themes to emerge from this qualitative data were **communications, access, training and support**.

Many nurses felt that they have too little information on current developments, with those who care for clients in the community developing a sense of abandonment: "Working as Community Palliative Care nurses, employed by a voluntary organisation but working as sole practitioners in a geographical area, we are not routinely copied into NHS information and miss out on new developments…NHS Trusts do not remember we are there!"

Community staff also seem to be unable to access training: "Staff within the community setting are all basically self learning…. We require more training and more computers and those we have do not allow us to contribute to patient records which would be very beneficial to the patient, nurse and doctor. I feel there is not enough emphasis placed on community involvement."

However, even those who work in acute settings report difficulties in accessing training: "Training for some nurses with no IT skills is essential if this is to be implemented efficiently and effectively. Many managers view IT skills for other departments rather than nursing..." "Not enough protected time to be trained, not enough access to computers, our one computer [situated in the ward office] is a work tool and as such is always busy, there is always somebody needing access to it, to check emails, send emails, make documents, compile records/files.." "In my clinical area, there is no computer access. Access is only available for more senior managers."

There are those who are fearful that nurses are driving down the 'high tech low touch' route: "I feel the Health service is becoming so impersonal, we are heading for supermarket health care and that is so sad, because we as nurses have so much more to give." Yet there are others who argue that nurses must embrace the 21st century: "Many people believe that nursing will suffer, but realistically a lot of the repetitive work carried out by nursing staff should be minimised, for example, how many times does a patient get asked their name and date of birth ???"

The great majority of nurses believe that e-Health will provide better and safer care: "The introduction of electronic patient records will ensure a more seamless and holistic approach to patient care particularly in the out of hours periods when often problem arise. The more information that is available the better and quicker the outcome of care can be made."

Being a very practical profession, nurses are aware of problems as well as possible ways to address the challenges ahead: "Systems need to be shared between Health and Social work and other partners who have access to providing any input to the patient, or they are no use. There will still be duplication of information and extra work if they are not". The striking finding is that many want to be involved, informed and engaged in e-Health developments so that both patients and clinicians are able to benefit from safe and efficient systems that support optimal care.

Discussion

Delivering 21st century care that is safe, effective and evidence based, relies in part on efficient use of clinical tools that deliver reliable, accurate, and timely information. Information and communications technologies (ICT) have the potential, in combination with organisational modernisation, to revolutionise the way care is provided. The Scottish e-Health strategy [7] is focusing on delivery of a common ICT system that will one day replace paper records. The strategy is also promising to connect every clinician to a secure health information network. There is a commitment to enabling healthcare professionals to access best practice guidance and knowledge at any point of care. However, the provision of modern healthcare in Scotland will require not only new infrastructure but also new thinking about practice and new skills to maximise the use of the clinical information superhighway.

Results from our on-line survey suggest that nurses in Scotland are not yet adequately prepared for the e-Health vision and thus unable to utilise or maximise the opportunities that ICT offers the profession. Despite the fact that the majority of Scottish nurses who responded to the survey see the potential of e-Health for both service users and care providers, they are as yet unable to integrate the use of ICT into their routine practice. Nurses do not get the right amount of training, nor are their views sought when systems are being designed, built, tested and implemented. The risks to the e-Health programme are grave. Poorly designed systems that do not fit with the requirements of the clinicians and with clinical workflow lead to inadequate implementations that can contribute to unsafe practice. Moreover, excluding the rich and holistic elements of nursing gives rise to task orientated products that restrict core element of professional care and affect the care outcome.

Closer collaboration between the professional bodies and the Scottish executive is urgently needed to provide a range of solutions to the needs identified by nurses in this survey. These include:

- Provide e-Health resources including adequate Information, advice and guidance
- Provide e-Health educational resources including a range of learning 'products', events and Continuous Professional Development (CPD) opportunities for e-Health
- Devise an educational policy that embeds e-Health competencies in both undergraduate and post graduate courses
- Resolve e-Health workforce issues so that the contribution nurses have on care outcome becomes visible
- Facilitate e-Health evaluation, research and development from the perspective of the nursing profession
- Grow and maintain e-Health capacity and leadership in nursing.
- Moreover, Scottish nursing must be an integral part of a UK initiative to develop an infrastructure and process for getting consensus on, maintaining and getting conformance on record content standards: the tools, charts, record structures etc that reflect best, evidence based nursing practice. The achievement of appropriate and agreed nursing content in Scottish electronic health record will ensure that the people of Scotland get the care from nurses that they expect and deserve.

Conclusion

As attitudes to computers were found to be a cardinal factor for effective use of IT in the work place [7], understanding nurses' attitudes towards many aspects of the Scottish e-Health programme is vital for planning effective change management within the clinical environment. This survey captured the views of a large number of Scottish nurses (n=415), albeit self selected. More research could be undertaken to see whether these findings are supported using a wider, more representative sample of nursing professionals. However, this paper clearly demonstrates that many nurses are keen to be involved in developments and implementation of all new information and record systems that have nursing components to demonstrate the unique value nursing brings to the clinical encounter.

Acknowledgments

The authors wish to thank those RCN members who were willing to share their views and aspiration with them.

References

[1] Scottish Executive Health Department, Delivering for Health, Edinburgh, Scottish Executive 2005.
[2] Scottish Executive Health Department, Building a Health Service Fit for the Future Edinburgh, Scottish Executive 2005
[3] Audit Scotland Informed to care: managing IT to deliver information in the NHS in Scotland, Edinburgh Audit Scotland 2006
[4] Royal College of Nursing, IT Survey 2006: Nurses and NHS IT developments, London, Royal College of Nursing June 2006 available at http://www.rcn.org.uk/publications/pdf/nurses_and_NHS_IT_developments_survey_2006.pdf

[5] Royal College of Nursing, Nurses and NHS IT developments: results of an online survey by Nursix.com, London, Royal College of Nursing; February 2004 http://www.rcn.org.uk/downloads/research/nurses-it-devs-survey.doc
[6] Royal College of Nursing, A year on: nurses and NHS IT developments London, July 2005; http://www.rcn.org.uk/downloads/press/nurses-it-devs-survey2005.doc
[7] National eHealth/IMT strategy 2004-8
[8] Jayasuriya, R. and Caputi, P. Computer attitude and computer anxiety in nursing: Validation of an instrument using an Australian sample. *Computers in Nursing*. 1996; 14(6): pp.340-345.

Address for correspondence

Sharon Levy, EPICentre, 9 Bell Street, Dundee, DD1 1HG Scotland.

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Usability of Institutional Cancer Web Sites: an Italian Case Study

M. Cristina Mazzoleni[a], Raffaella Butera[a,b], Franco Corbella[c], Vittoria Balcet[c], Enrico Masenga[b]

[a] *IRCCS Fondazione Maugeri, Pavia, Italy*
[b] *ProgettareWeb.it, Italy*
[c] *IRCCS Policlinico San Matteo, Pavia, Italy*

Abstract

In order to evaluate if and to what extent Italian speaking cancer patients can benefit from information available on cancer web sites, an "in vitro" usability (ISO definition) study has been carried out. It investigated the usability of the web sites of the most representative Italian Institutions in the oncological field for the adult patients needing to find information about head and neck cancer. Specific evaluation criteria from the literature were used. The results point out some problems about accessibility, in line with other studies, and about the usefulness of the contents, in particular in the web sites of care delivery institutions: a grey present situation, but there are already grounds for significant improvement. Institutions and organizations must not waste the opportunity of being valuable sources in order to build the so called "informed patient," and the usability of their web sites could make the difference.

Keywords:

internet, medical informatics, usability, patient education.

Introduction

A more informed patient is unanimously recognized as a determinant for the improvement of both the population health status and the effectiveness of the healthcare systems. Patient information/education still represents a challenge, and the web is one of the emerging tools to achieve this goal. Initially, the major concerns dealt with the *reliability* of health information available on the web for the general public, but the discussion is still open [1]. In the last fifteen years many empirical studies, the great majority dealing with Anglophone web sites, have been carried out in order to evaluate the *quality* of information for consumers. In a patient/consumer-centered perspective *usability* is the next step and could/should be the keyword and the focus, provided that the content is "good."

The document *ISO 9241-11 (1998) Guidance on Usability* by the International Organization for Standardization defines usability as "the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use". This definition implies the *usefulness* concept (*achieve specified goals with effectiveness*), while another authoritative figure - Jacob Nielsen - considers usability as a part of usefulness. The two constructs are anyway evidently strictly bound. Both of them refer to the potential benefit that the users can get from the evaluated object, in terms of result - to meet a need - and necessary effort to achieve it.

The present paper deals with an "in vitro" usability (ISO definition) study having for *object* the web sites of the most representative Italian Institutions in the oncological field; for *users* the patients, adults; for *need* to find information about head and neck cancer; for *context* a standard PC with the most popular operating system (Windows XP), browser (Internet Explorer 6.0), screen resolution (1240x768 and 800x600), with no optional plug-in. The study was aimed at evaluating if and to what extent Italian speaking cancer patients can really benefit from cancer web sites for a more active participation to the care process.

Materials and methods

Sample construction

To identify the most representative healthcare institutions in the oncological field, the list of the members of a national network of excellence named "Alleanza contro il Cancro" was used. To this first group some National Patient Associations and the Italian Association of Oncology Physicians were added because of their popularity and relevance. The choice of restricting the sample of web sites only to the ones published by authoritative well known institutions had the implicit assumption that the provided information was reliable.

Evaluation criteria

Evaluation criteria were based on the "Research-based Web Design & Usability Guidelines" [2] developed by the U.S. Department of Health and Human Services. These guidelines are primarily addressed to web site managers and are aimed at "creating better and more usable health and human services web sites". The approach taken to produce these guidelines is compliant with the traditional model used to build guidelines in the health care domain, supporting each statement with scores (range 1-5) related to both relative importance (relative importance score, RIS) and strength of evidence (strength of evidence score, SES). All of the items with RIS equal to 5 were taken into consideration, independently from the SES value (Table

1). The items were numbered consecutively (ID); the reference to the original guidelines items are listed in the second column of the table.

Table 1 - Items with Relative Importance Score equal to 5

ID	GL #	Guideline item heading		SES
1	1:1	Provide useful content		5
-	1:2	Establish user requirements		4
-	1:3	Understand and meet user's expectations		3
-	1:4	Involve users in establishing user requirements		3
2	2:1	Do not display unsolicited windows or graphics		3
3	3:1	Comply with section 508	Comply with W3C WCAG 1.0	2
	3:2	Design forms for users using assistive technology		2
4	3:3	Do not use colour alone to convey information		4
5	5:1	Enable access to the homepage		3
6	5:2	Show all major options on the homepage		2
7	5:3	Create a positive first impression of your site		4
8	6:1	Avoid cluttered displays		3
9	6:2	Place important items consistently		4
10	6:3	Place important items at top center		4
11	8:1	Eliminate horizontal scrolling		4
12	9:1	Use clear category labels		4
13	10:1	Use meaningful link labels		4
14	13:1	Distinguish required and optional data entry fields		3
15	13:2	Label pushbuttons clearly		2
16	15:1	Make action sequences clear		4
17	16:1	Organize information clearly		4
18	16:2	Facilitate scanning		4
19	16:3	Ensure that necessary information is displayed		2
20	17:1	Ensure usable search results		3
21	17:2	Design search engines to search the entire site		3

The items 1:2, 1:3, 1:4, addressing specifically the development process and were excluded. The items 3:1 and 3:2 were transformed into a single item "Comply with the World Wide Web (W3C) Consortium Accessibility Initiative Guideline (WCAG) 1.0" keeping the same meaning. The compliance was tested using an automatic evaluator [3]: web sites were considered compliant if W3C WCAG 1.0 Priority 1 level checkpoints were satisfied. For all of the mentioned items, two observers marked each web site as compliant or not compliant. In case of discordant judgments, a common result was obtained through a further joint inspection.

In order to evaluate web sites compliance with the item 1:1 ("Provide useful content", the only one with both RIS and SES equal to 5) usefulness was defined as the capacity of meeting the patient's information need. The results of a review [4] about information needs of cancer patients were used. In the quoted review the needs are divided into 10 categories and in 64 subcategories, after having analyzed the results of 91 articles published between 1980 and 2003. From the original list proposed by the authors of the review, the 7 categories quoted in at least 25% of the papers were extracted and considered in our analysis as primary patient's needs: disease-specific, treatment, prognosis, rehabilitation, coping, impact on interpersonal/social relations, and consequences on body image and sexuality. The 41 subcategories with a frequency of occurrence within the category grater than 5% (see the Results section for the complete list) were analyzed in detail. The presence of information about the selected subcategories and specifically related to head and neck cancers was checked by the two observers. For each web site, a score was attributed to each subcategory (subcategory score, SS) in case of presence of specific information (SS=1), presence in general but not related to head and neck cancers (SS=0.5), or absence of information (SS=0), respectively. These results were summarized at category level: each category was given a category score (CS) equal to 1 in case of presence of information (SS greater than 0) in more than 50% of the subcategories, 0 otherwise.

Results

The sample

Fourteen web sites were identified. One of them required optional software for displaying Java applets in order to start navigation: being out of the test constraints, this site

was excluded, due to its total inaccessibility in a standard situation. The final sample was composed of eight web sites managed by research and care delivery institutions (RCI) and five by patients/physicians/research associations (PPRA). The list of evaluated web sites is available form the authors on request.

Compliance with items 2 to 21

Table 2 shows the percentages of web sites compliant with the usability requisites listed in Table 1, with item 1 ("Provide useful content") excluded. As to item 14, no data entry was required in any of the tested pages: no evaluation, consequently, was possible. Search functionality (items 20 and 21) was present in 11 out of 13 web sites: the percentages shown are referred only to the reduced sample. Search results were very poorly displayed (item 20), with only one exception in which the retrieved information was provided in a very effective way. In general evaluated web sites provided only basic functionality. As to the accessibility requirements, 38% (5 out of 13) of the home pages were sealed as not compliant with the W3C WCAG 1.0 Priority 1 level. Within this group of non accessible sites, three belong to PPRA.

Table 2 - Web sites compliance with usability requisites 2-21

Requisite	% (*)
2. Do not display unsolicited windows or graphics	93
3. Comply with the W3C WCAG 1.0	61
4. Do not use color alone to convey information	100
5. Enable access to the homepage	93
6. Show all major options on the homepage	100
7. Create a positive first impression of your site	100
8. Avoid cluttered displays	100
9. Place important items consistently	100
10. Place important items at top center	93
11. Eliminate horizontal scrolling	100
12. Use clear category labels	100
13. Use meaningful link labels	100
14. Distinguish required and optional data entry fields	-
15. Label pushbuttons clearly	100
16. Make action sequences clear	100
17. Organize information clearly	100
18. Facilitate scanning	93
19. Ensure that necessary information is displayed	86
20. Ensure usable search results	16
21. Design search engines to search the entire site	100

* percentage of compliant web sites

Compliance with item 1 "Provide useful content"

Table 3 shows the distribution of subcategory scores (SS) for RCI and PPRA web sites. Limiting the analysis to information specifically related to head and neck cancer (SS=1), only 10 subcategories out of 41 were present in more than 6 web sites.

Figure 1 shows the comparison between the content available on RCI and PPRA web sites. The bars represent, for each category, the percentage of web sites in which at least 50% of the subcategories were present.

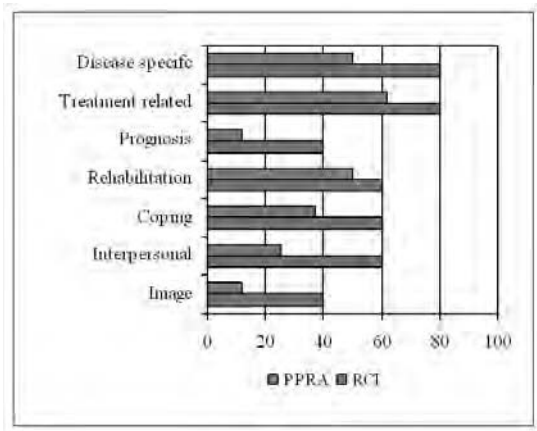


Figure 1 - Percentages of PPRA and RCI web sites with CS=1

Discussion

There are several methods to analyze the usability of a software, and hence of web sites. Some tools have been developed and validated to assess the subjective perception of the users [5,6], but not many experiments have been carried out to evaluate usability according to well defined criteria and guidelines. The "Research-based Web Design & Usability Guidelines" proved to be a good tool to inspect the usability of health related web sites, joining technical and content-related aspects. The choice of considering only the requisites with RIS equal to 5 has led to the exclusion of other items such as those related to readability and comprehensibility of the texts (RIS equal to 4). These requisites, when dealing with health related web sites, are in our opinion even more critical than others quoted in Table 1. A readability test is ongoing for a sample of retrieved texts. Fairly good results are expected, in accordance with a previous study performed in 2003 [7], in

which 75% of the analyzed texts were easily readable for Italian people who have attended compulsory education.

Overall, all of the web sites inspected met the great majority of the usability basic requisites: this could be due to the great simplicity of most of them in terms of structure and provided functionality. However, it should not be forgotten that one web site was excluded from the sample due to the impossibility to display it in the standard PC configuration defined for our study. As regards accessibility, although automatic evaluation is quite a rough method, it is suitable for identifying at least inaccessible pages. The results indicate a lack of attention paid to web accessibility, despite a recent law for the public administration web sites and many recommendations published in the grey literature. Although referring only to a small sample, the situation depicted in this study is comparable with that reported in other studies [8,9,10] where the percentages of non accessible sites were respectively 60%, 66% and 65%.

As to requisite 1 "Provide useful content", the assessment took into account only the breadth of topics' coverage because of the a-priori assumption that the information came from authoritative institutions, and hence was reliable and accurate, if present, by definition. The fact that an institutional web site has necessarily to deal with all of the topics can be questionable. A patient develops, during the care process, trust in and familiarity with a particular institution: in our opinion he/she could expect the web site of this institution to be comprehensive and to be an effective tool to easily find the needed information. This doesn't mean that a single web site has to provide, by itself, all of the needed information, but it should provide a way to reach it. For this reason in the present study a topic was considered as being present both in the case of direct publication and in the case of delivery via a link to a specialized external web site.

The results point out that, globally, web sites provide more frequently information related to the medical culture (category DS, Disease specific; category TR, Treatment; category R, Rehabilitation). At the same time, the least present subcategories were the ones related to the prognosis (P), a topic that is ranked quite high in the patient's needs list.

Table 3 - Cancer patient's information needs and number of sites, out of thirteen, with subcategory score (SS) 0, 0.5 and 1

Category	Subcategories	SS=0			SS=0.5			SS=1		
		rci	**ppra**	**tot**	**rci**	**ppra**	**tot**	**rci**	**ppra**	**tot**
Disease-specific (DS)	Type of cancer/nature of disease	4	1	**5**	0	0	**0**	4	4	**8**
	Aetiology and course of disease	4	1	**5**	0	0	**0**	4	4	**8**
	Physical effects of disease	4	1	**5**	0	0	**0**	4	4	**8**
	Specific diagnosis information	4	1	**5**	0	0	**0**	4	4	**8**
	Stage of disease	4	1	**5**	0	0	**0**	4	4	**8**
	Symptoms of cancer/management of symptoms	4	1	**5**	0	0	**0**	4	4	**8**
	Where to get information about specific diagnosis	4	2	**6**	0	0	**0**	4	3	**7**
Treatment-related (TR)	Side effects, risks and benefits of treatment	3	2	**5**	2	1	**3**	3	2	**5**
	Available treatments/treatment options	4	1	**5**	1	1	**2**	3	3	**6**
	Treatment plan/description/logistical info	4	1	**5**	0	0	**2**	3	3	**6**
	Tests and procedures involved in treatment	4	1	**5**	1	1	**2**	3	3	**6**
	Reducing side effects of treatment	5	4	**9**	1	1	**2**	2	0	**2**
	Alternative or complementary treatments	5	2	**7**	1	1	**2**	2	2	**4**
Prognosis (P)	Chance of cure	6	2	**8**	0	0	**0**	2	3	**5**
	Life span or survival rate	7	2	**9**	0	0	**0**	1	3	**4**
	Recurrence of cancer	7	2	**9**	0	0	**0**	1	3	**4**
	Spread of disease or metastasis	7	3	**10**	0	0	**0**	1	2	**3**
	Expectations for future health condition	8	4	**12**	0	0	**0**	0	1	**1**
	Effect on life plan or long term goals	8	4	**12**	0	0	**0**	0	1	**1**
	Outcome of no treatment or delayed treatment	8	5	**13**	0	0	**0**	0	0	**0**

Category	Subcategories	SS=0			SS=0.5			SS=1		
		rci	ppra	tot	rci	ppra	tot	rci	ppra	tot
Rehabilitation (R)	Self care issues or home care during recovery	5	3	**8**	0	0	**0**	3	2	**5**
	Nutrition during recovery	5	3	**8**	0	0	**0**	3	2	**5**
	Immediate post-treatment follow-up care	5	4	**9**	0	0	**0**	3	1	**4**
	Long-term side effects of cancer or treatment	6	4	**10**	0	0	**0**	2	1	**3**
	Recognizing/preventing treatment complications	5	4	**9**	0	0	**0**	3	1	**4**
	Recovery time	5	3	**8**	0	0	**0**	3	2	**5**
	Where to get medical supplies/equipment	4	3	**7**	0	0	**0**	4	2	**6**
	Maintaining physical health or physical activity	5	2	**7**	0	0	**0**	3	3	**6**
	Prevention and early detection	2	0	**2**	0	0	**0**	6	5	**11**
	Maintaining psychological health	2	0	**2**	0	0	**0**	6	5	**11**
	Health behavior and promotion	2	0	**2**	0	0	**0**	6	5	**11**
Coping (C)	Emotional reactions/support, coping	5	2	**7**	1	0	**1**	2	3	**5**
	Community counseling or support	5	2	**7**	1	0	**1**	2	3	**5**
	Support groups	6	2	**8**	1	0	**1**	1	3	**4**
	Support from other patients	6	2	**8**	1	0	**1**	1	3	**4**
Interpersonal / social (IS)	Effect on family, friends, or caregivers	5	2	**7**	1	0	**1**	2	3	**5**
	Effect on social life or leisure	5	2	**7**	0	0	**0**	3	3	**6**
	Risk of disease for family members	8	4	**12**	0	0	**0**	0	1	**1**
	Effect on employment or work life	6	2	**8**	0	0	**0**	2	3	**5**
Image / sexuality (I)	Sexuality	7	3	**10**	0	0	**0**	1	2	**3**
	Physical appearance/physical attractiveness	7	3	**10**	0	0	**0**	1	2	**3**

The limited presence of this kind of information could reflect the caution usually used in Italy to deal with such problems outside the face to face patient-physician relationship. Cultural differences may play an important role [11] in determining the content of health related web sites. Moreover, within category P, one subcategory is completely absent: no information at all is provided as to the "outcome of no treatment or delayed treatment". The difference between RCI and PPRA web sites as regards content coverage is quite evident from the graph of Figure 1. Among the RCI web sites, three out of eight gave no information or minimal information about a very limited number of topics, neither directly or indirectly through links to other web sites. This is the reason why in Figure 1 the percentages of RCI web sites are never higher than 62%. The highest percentages were reached in those cases (categories Disease specific, Treatment, Rehabilitation) in which at least one RCI web site provides information through a link to the web site of a PPRA. Among the PPRA web sites, only one out of five did not provide health related information. The comparison with the results of other studies aimed at inspecting the completeness of health related information on web sites is not easy, due to differences in the used methods, in the sample building procedure, and in specific disease. In many papers the *quality* of the content has been evaluated comparing the information provided with the content of "golden standards" like published guidelines and recommendations. In the present study we made an attempt to evaluate the *usefulness* of the content, starting from the patient's information needs. In a study [11] some percentages are reported as to the presence of treatment related topics in samples (any kind of web sites, not institutional only, gathered through search engines) of English and German web sites: information about the most frequent therapy options (category TR in the needs schema) are present in 90% and 69% of the English and German samples respectively, compared to a figure of 69% (9/13) of the PPRA plus RCI total sample in our study. In [12] a description of US Children's Hospitals web sites is provided: the conclusion referred to the contents "users *[....]* would be disappointed by most of the sites" indicates a failure of institutional web sites in meeting the users information needs. In another paper [13] about Norwegian hospital web sites, health related contents are not even quoted, fact from which the absence of useful content, as defined in our study, could be derived.

Independently from the numbers, a certain disappointment as to the usefulness of the contents in the Italian Cancer Institutions web sites was unavoidable. A gleam of hope appeared when, surfing MEDLINE, a paper titled "More information more choice: an Italian database for oncology

patients" [14] was retrieved. It describes a "library for the patients" that was developed and put on line within a nation wide project named Azalea. Partners of this project were most of the institutions whose web sites have been inspected in the present study. The service, since August 2006, is no longer available due to reasons independent from its quality. It has been available for more than 2 years, with an increasing number of accesses (up to 20.000 unique visitors per month in the last trimester of presence on line). More than 2000 documents were available, dealing with most of the topics listed in Table 3, and integrated into the database after expert validation. Last but not least, a reassuring seal of WACG compliance was present in the homepage. The hope is that this valuable effort will not have been wasted.

Conclusions

When dealing with patients looking for information on the web, the usability of Institutional web site should be taken into great account, since, as stated by the Internet Healthcare Coalition, people are expected to use web sites managed by institutions or organizations in which they have confidence. The results reported in this paper point out some important usability problems in particular about accessibility as to patient associations web sites, and about the provision of useful content, as to the web sites of care delivery institutions. The present situation of the Italian Cancer Institutions web sites is grey, but luckily there are already the grounds, and maybe even more, for a significant improvement. Institutions and organizations must not waste the opportunity to be valuable sources in order to build the so called "informed patient," and the usability of their web sites could make the difference.

Acknowledgments

The authors thank Dr. Ivana Truccolo and Dr. Katia Bianchet of the Biblioteca Scientifica - Centro di Riferimento Oncologico Aviano IRCCS, and Dr. Ing Marco Pagani of CBIM, Pavia for having provided the data about the Azalea project.

References

[1] Adams SA, DeBont AA. Notions of reliability: considering the importance of difference in guiding patients to health care Web sites. Methods Inf Med. 2003;42(4):307-12.
[2] www.usability.gov/pdfs/guidelines_book.pdf
[3] www.webxact.watchfire.com
[4] Rutten LJ, et al. Information needs and sources of information among cancer patients: a systematic review of research (1980-2003). Patient Educ Couns. 2005;57(3): 250-61.
[5] Nahm ES, Resnick B, Mills ME. Development and pilot-testing of the perceived health Web site usability questionnature (PHWSUQ) for older adults. Stud Health TechnolInform. 2006;122:39-43.
[6] Kirakowski J, Corbett M: SUMI: the software measurement inventory. Brit J Educ Technol. 1993;24:210-212.
[7] Mazzoleni MC, et al. Italian speaking health consumers and the offer of information on the internet. Proceedings of "Cancer on the net", New York, 2003
[8] O'Grady L. Accessibility compliance rates of consumer-oriented Canadian health care Web sites. Med Inform Internet Med. 2005;30(4):287-95.
[9] Chiang MF, Starren J. Evaluation of consumer health website accessibility by users with sensory and physical disabilities. Medinfo. 2004;11(Pt 2):1128-32.
[10] Mancini C, Zedda M, Barbaro A. Health information in Italian public health websites: moving from inaccessibility to accessibility. Health Info Libr J. 2005;22(4):276-85.
[11] Weissenberger C, et al. Breast cancer: patient information needs reflected in English and German web sites. Br J Cancer. 2004;91(8):1482-7.
[12] Kind T, et al. Do the leading children's hospitals have quality web sites? A description of children's hospital web sites. J Med Internet Res. 2004;6(2):e20.
[13] Norem J, Moen A. The websites of Norwegian hospitals: do they meet national guidelines and patient's expectations? J Telemed Telecare. 2004; 10(5):272-6.
[14] Bogliolo A, et al. More information, more choice: an Italian database for oncology patients. Ann Oncol. 2005;16(12):1962-7.

Address for correspondence

M. Cristina Mazzoleni - Fondazione Salvatore Maugeri, IRCCS
Via Ferrata 8,
27100 Pavia,
Italy
email: cmazzoleni@fsm.it

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Investigating Internet Use by Mental Health Service Users: Interview Study

John Powell[a], Aileen Clarke[a]

[a] *Health Sciences Research Institute, Warwick Medical School, University of Warwick, United Kingdom*

Abstract

The internet is an increasingly important source of mental health-related information, and has the potential to be harnessed as a tool to support self-care and informed decision-making. Yet little is known about the motivations and attitudes of users. We therefore undertook a qualitative interview study with a purposive sample of mental health service users with internet experience, to explore issues with respect to mental health-related internet use. One of the prime motivations for online mental health seekers was to find experiential information from other people with similar problems. This information allowed users to know they were not alone, and to instill hope that others in the same situation had recovered. Benefits of the internet as an information source included convenience, privacy and anonymity. Problems related more to misuse of the internet rather than concerns over inaccuracy. Such qualitative work is important in an emerging research area to understand internet use better.

Keywords:

internet, mental health, qualitative research

Introduction

The internet is playing an increasingly significant role in health information-seeking [1]. At the same time health systems in developed countries are witnessing increasing consumer involvement in healthcare, with the recognition of the importance of health literacy, the developing role of the informed expert patient, and the increasing importance of self-management in chronic disease [2]. These developments can be facilitated and enhanced using e-health tools. An important first step to achieve this is to understand the motivations and attitudes of e-health users.

A recent UK study has shown that over 10% of the general population has used the internet for mental health information [3]. Online mental health interventions such as internet-based cognitive behavioural therapy are increasingly being developed [4]. In a qualitative study to investigate mental health information needs (reported elsewhere [5]) we asked additional questions related to mental health-related internet use to explore the issues in using this increasingly important medium. In particular we investigated the advantages and disadvantages of the internet as a source of mental health information, and explored user motivations and concerns.

Materials and methods

Recruitment

In this qualitative study, we undertook a series of in-depth interviews with adult mental health service users. As investigating internet use was one of our key aims we deliberately over-sampled individuals who had direct experience of using the internet for mental health information. Participants were identified and recruited purposively through various means. In primary care and secondary care settings in Oxfordshire, UK, recruitment used information sheets displayed in healthcare settings and given to potential participants by general practitioners and psychiatric staff. In addition, study advertisements were sent to mental health user organizations, and posted on a consumer health information website. We included participants aged 18 or over, with personal experience of mental health problems, and who had recent (or current) experience of mental health services. Participants were excluded if they were unable to take part in a one hour interview due to ill health.

Procedure

One interviewer (JP) conducted all interviews in person or via the telephone. A topic guide was used based on a review of the literature. Written consent was given and all interviews were audio-recorded and transcribed. Open-ended questions and follow-up prompts relating to internet use were used to explore issues related to online mental health information seeking. Questions inquired about benefits and problems of internet use for mental health information, and motivations for online information seeking. National Health Service ethics committee approval was given for the study.

Analysis

A grounded approach was used to identify themes and subthemes through a process of familiarization with the data, open coding, axial coding, and selective coding [6]. Two investigators (JP, AC) undertook the analysis. Themes related to general mental health information seeking have been described elsewhere [5]; this current paper presents the findings specifically with respect to mental health-related internet use.

Results

Participants

Thirty-six participants were purposively recruited and all consented to be interviewed. There were 25 females and 11 males from different points in the mental health system representing a variety of occupational backgrounds and with an age range from 25 to 64 years. We intentionally over-sampled individuals with some experience of using the internet (32 of 36 interviewees). This was not a representative sample, but a purposive one designed to explore specific issues.

Themes

Benefits of the internet

The first theme concerned the advantages of the internet as a source of mental health information. These can be summarised as anonymity, privacy, convenience, accessibility and empowerment.

Interviewees reported valuing the "unobtrusive" nature of the internet - both in terms of the anonymity it affords users and also the way that you can participate without interacting.

As interviewee 22 (a 38 year old social scientist) put it

> "It's such an unobtrusive, discrete way of doing it [finding information]. I mean you can do it very privately without even, having to talk to another person."

Interviewee 30 (a 52 year old building contractor) also described the advantages of the internet:

> "I think another advantage is that you can do it in the privacy of your own home because if you are conscious of the stigma or have difficulties with speaking about what you have with other people then, you know you can do it completely in your own privacy."

Interviewee 36 (a 34 year old shop manager) described the convenience (not having to go out), the possibility of interacting with other people in a similar situation (who may not be available in "real life"), and the fact that one can "eavesdrop" and not conform to the normal social rules of "real life" - for example by just leaving:

> "[Its] easy to find information, so much easier than say… I don't use a library, or books. You don't have to go out. And other people, I think, the sort of support from other people, there is always someone who has had your problem, whatever it is you know. Which is amazing. Its nice to know that you are not alone, the silly little things you see coming on line and saying, thank God I found this great pal, I'm not alone anymore, you know, what they thought was some obscure condition that no one else had because maybe they are the only person in their city that has got it or something. There is always loads of other people on the internet. And also you can just leave incidentally if you want to which you can't do in a real situation. … You can just sit there and eavesdrop (laugh) … but you couldn't really do that in a real situation, you can't just stand at a corner of the room and listen, people won't let you."

Interviewee 15 (a 30 year old freelance journalist) contrasts her experience of seeking help from her general practitioner and with that of internet help-seeking.

> "I tend not to be very assertive in things to do with myself and not, you know, not want to take up GP's time so I find it very difficult to ask for information that focuses on myself, whereas if I sit at the internet, you know, I am anonymous, nobody would, uhh, I am not taking up anyone's time and its just a lot easier, a lot easier and I can do it when I am ready to do it and I don't have to wait and get stressed about it. Also I don't have to worry about, if I ask a question someone is going to say 'well, why are you asking' you know, whatever and I don't get the third degree. So that's why I prefer to do it [on the] internet rather than in person."

There also seem to be particular benefits of the anonymous nature of the internet for people with mental health problems, as interviewee 32 (a 48 year old retail manager) explains.

> "From my point of view its [using the internet] because I cannot interact in a social group. That's my worst problem, one of my worst problems is being able to interact with a group of people. I find it absolutely impossible anymore and I hide myself away, I am becoming a hermit I would say. But I do go, I do go out, I have to go out to the doctors, I have to take my wife to the shop, I don't go in the shop, but she goes, but, you know, so its anonymity, I haven't said that right but you know what I mean, because they can't see me, if they could see me, I would stay away."

The interviews also demonstrated the benefit of the internet in improving access to information. Two elements were seen as beneficial: access to other people with the same condition was greatly facilitated by the global nature of the internet, and secondly, access to expert knowledge was valued and seen as empowering. For example interviewee 1 (a 58 year old civil servant) explained:

> "If patients can actually get information off the internet, they've got some, umm, something to argue with the doctor about, I think in terms of empowering people, being without information is disempowering and this evens up the power between the doctor and the patient and the doctors may think you are a nuisance simply because you arguing with me and there is an element of the doctor knowing best, but umm, I mean, that it, that is my view and certainly gathering information in the last few days felt quite armed to go in and sort out the psychiatrist simply because we got this information"

Problems of the internet

Regarding any particular problems with using the internet as a source of healthcare information, it was interesting that the accuracy of online information was less of a concern for interviewees than misuse of the internet. People recognised that there are websites with poor or inaccurate

information but this was not seen as a major issue. Interviewees acknowledged the unregulated nature of the internet and that some of the information was “silicon snake oil” (interviewee 11, a 43 year old mature student), but they seemed to have confidence in their ability to discern accurate sources and not be taken in by the “quacks” or “cranks”. Individuals had developed their own strategies for dealing with inaccurate information, and had learned to trust certain websites - usually those with identities that they trusted in the "real" world - such as sites run by the NHS, the BBC or major mental health charities. Misuse with disruptive or malicious intent was seen as more problematic. Interviewee 14 (a 43 year old care worker) commented on her fears of who might be using chatrooms for victims of abuse:

> "I have come across [websites] where they have got like chat rooms for people who have been abused and things like that and mental health stuff and I wouldn’t go in there because I am thinking 'ooh', you know, I don’t know whether it would be full of genuine people or whether it would be full of, you know, if there is going to be people in there that want to hear about people who have been abused and get a kick out of it."

Interviewee 16 (a 33 year old care worker) and interviewee 32 (a 48 year old retail manager) describe the disruption of websites they were using:

> "Then, over a period of time, the site seemed to be taken over by some very young teenagers particularly, some at boarding schools and everything, and they were using the chat rooms and the message boards making threats saying, 'I am going to kill myself now' and then they would sign off and things like that, and to me that wasn’t helpful … I know the internet is all about free speech and free opinion and everything but for somebody, yeah, I mean, I won’t deny, I, I have made three serious attempts on my life and a number of others and, I, I spent nearly three months in a coma and I have got liver damage and I live with that permanently and to then find sites on telling you how to commit suicide, I do find it somewhat distasteful (pause) and you, you had that on NetDoctor [a health information website] people were asking the best way to kill themselves and I can't deal with that, and a lot of people couldn’t deal with that."

> "Unfortunately you do get people who come on there that are not ill. You can tell they are not 'cause they start arguing, not arguing but leaving nasty comments and like I have just said there was one person not long [ago] who come on and say, 'you are not depressed, you are just this, that and the other, get your self together'. …When people come on and are being nasty like that, I don’t like that."

The major concern about using the internet for health information for our interviewees therefore concerned disruptive online behaviour, rather than poor quality information. A further minor complaint from our UK respondents concerned the predominance of US websites found when searching for mental health information, and the fact that the material on these sites was often not suitable for a UK context.

Hearing about other people’s experience online

The importance of hearing about other people’s experience of mental health problems, and using the internet to find these was the prominent motivation emerging from the interviews. In particular the internet met the needs of users to know that they were not alone with their problems (characterised as “universality”) and to know that others in a similar situation have been able to get better (characterised as “hope”). Interviewees also valued how the interactivity of the internet allowed them to obtain understanding and empathy from others in their situation.

Knowing one is not alone was seen as reassuring, and also it helped the individual reject their notion of 'madness'. As interviewee 2 (a 30-year old biochemist) put it:

> "I think it has helped in that, I understand that I am not the only person to act and behave in the way I do, Certainly since joining [name of website] its just so nice to read what people write, thank goodness I am not as mad as I thought I was."

In our interviews madness was seen as something individual, idiosyncratic, an extreme deviance from normality. "Some kind of alien" as interviewee 15 put it. Whereas if you have something that other people also have, then you no longer see yourself as 'mad'. As illustrated by interviewee 31 (a 55 year old occupational therapist):

> "I think the most helpful thing was to know that some other people had similar feelings and I wasn’t going mad."

The benefit of not only knowing that one is not alone, but also that other people have had the same problem and got better is illustrated by interviewee 13 (a 27 year old teacher):

> "As I got a bit braver and kind of went back to work and so on, I did the odd search on the internet and I would read other people's experiences of schizophrenia and it was often by their families, or sometimes it was by the person themselves, and just hearing that somebody had recovered would mean so much to me because when I was still kind of in the recovering process, I was heavily depressed thinking that I was never going to be the same person that I had been before, I became ill, and reading experiences where people had recovered, it was such a boost, because you thought well, if they can do it, you know, I will be damned if I can’t."

Not only did individuals want to know that they were not alone and that others had got better, but they also wanted to interact with others, because only other people who have been through the same experiences as them know "what it's like". Interviewee 34 (a 25 year old unemployed man) who had depression talked about his experience of using an internet bulletin board, and the value of interacting with others who also had this diagnosis:

> "They can understand and know what you are going through as well. … If you have not been through the experience you don't understand what it is."

Interviewee 5 (a 37 year old teacher) described how finding people who knew what she was going through on the internet had been a "lifeline":

> "It has been a lifeline literally… It makes a difference between being lonely and afraid and unable to do anything and being able to turn on a metal box and at the other end of it, there's people that know exactly what you are going through and they can support you through it because they have been through it and they have come through the other side or they are going through it still."

Finally, interviewee 16 (a 33 year old care worker) described how contact with peers in the same situation on the internet allowed her to fill an information gap that her doctor was unable to meet:

> "You can reach the end of your doctor's knowledge and then you can go online, and you can talk to other people who have also been treated for years and years and years and they can help you to come up with new ideas … For me, personal information is the most useful, you know, what it was like to take this particular drug, or what its like to have a particular condition because the person who has got it or who is taking that drug they can describe what it is really like and sometimes it then kind of makes sense. You may have seen that in the official description but you didn't really understand exactly what it meant."

Other motivations

Aside from finding experiential information, the other main motivations concerned undertaking personal research into the condition. Prime areas for research were the causes of illness, alternative diagnoses, and treatment options. Internet research (often in association with searching other sources) was frequently undertaken in response to a lack of information available from the health service. Interviewee 11 (a 43 year old mature student) explained:

> "Nobody told me anything, I have researched it myself. I knew a fair amount and I got that information from work, but, nobody offered me any information, nobody pointed me in the right direction I should say … I came across something in a book ... That's Borna disease virus. I didn't know whether this was anything relevant … so I then did some research through the internet. I ran a search on the name of the disease."

The following quote from interviewee 14 (a 43 year old care worker illustrates how this personal research links to the reported benefit of empowerment described earlier:

> "I diagnosed myself with fibromyalgia, my doctors hadn't done it and I did take stuff from the internet and leaflets and said, I think that's what I have got, you have not listened to me all these years, you have just said 'oh yes, we think you are depressed you know' … and this doctor actually read it and my symptoms and said 'oh you are very clever … you have diagnosed yourself' … and this was just information … We went to see the specialist and he said to me, the doctor has diagnosed.. and I said 'no I did it' and he said 'how' and I said 'I got information off the internet, looked up symptoms that I was suffering and then went to a library and got some information and then a girl by pure chance had a leaflet' … and apparently, I have just recently found out that depression is part of fibromyalgia."

Discussion

Studies in areas other than mental health have identified the benefits of the internet for health consumers of anonymity and convenience of access [7;8]. Our interview analysis supports these previous findings particularly around the advantages of "privacy". Privacy in this case encompasses both anonymity and the private access that people have to the internet in their own homes and it is important for the avoidance of stigma. The internet offers advantages in acting as a medium of mass communication whilst allowing for individual interaction. It provides both a public and a private space for information-seeking from a variety of sources while allowing the concealment of individual identity.

Our finding that users trust certain websites and that these tend to be related to organisations they would trust in the real world is also supported by other work [9;10]. A focus group study in eight European countries found that participants often reported a feeling of being overwhelmed by the volume of information on the internet, and had concerns about information quality [11]. While we found that individuals did express reservations about the reliability of online information, they were actually more concerned with internet misuse than with untrustworthy information. This is supported by a US population survey which showed that 81% of internet users expect to find reliable information about health or medical conditions online [12]. Indeed 46% of internet users in this survey said they would use the internet as the first source next time they needed reliable medical information, compared with 47% who would contact a medical professional. This is interesting in the context of the overwhelming volume of articles published in the biomedical literature expressing concern at the quality of online information.

Previous work has identified that patients exchange information with each other online [13;14], and there are emerging findings of the value of hearing other people's experience on the internet. In two qualitative studies re-analysing illness narratives collected for the DIPEx project, Ziebland and colleagues have shown that cancer patients use the internet (amongst other reasons) to find experiential information from other patients [7;9]. Hardey has discussed how his studies of interviews with internet users and analysis of internet-based illness narratives show that individuals use the internet both for finding out and displaying personal experiences as well as for professional information and advice, and how the sharing of experiences is part of a wider shift in the relationship between lay and medical expertise [15].

Conclusion

The internet is a valued source of information on mental health issues and users describe its benefits, particularly concerning privacy. It is meeting the need for a private space to discuss mental health issues. There are concerns about the internet; related more to misuse than to inaccuracy. In particular the internet is a source of information about other people's experience of illness, providing universality and hope. Healthcare providers wishing to harness the internet as a mental health resource need to take account of the motivations of users and their perceptions of the risks and benefits of mental health-related internet use.

Acknowledgments

Dr Powell was funded by a National Health Service Research and Development Programme Training Fellowship in Health Services Research.

References

[1] Baker L, Wagner TH, Singer S, Bundorf MK. Use of the internet and e-mail for health care information: results from a national survey. *JAMA* 2003;289:2400-6.
[2] Gray JAM. The Resourceful Patient. Oxford: eRosetta Press, 2002.
[3] Powell J, Clarke A. Internet information-seeking in mental health: population survey. *Br J Psychiatry* 2006;189:273-7.
[4] Christensen H, Griffiths KM, Jorm AF. Delivering interventions for depression by using the internet: randomised controlled trial. *BMJ* 2004;328:265.
[5] Powell J, Clarke A. Information in mental health: qualitative study of mental health service users. *Health Expect* 2006;9:359-65.
[6] Strauss A, Corbin J. Basics of Qualitative Research. Techniques and procedures for developing grounded theory. 2nd ed. London: Sage, 1998.
[7] Ziebland S, Chapple A, Dumelow C, Evans J, Prinjha S, Rozmovits L. How the internet affects patients' experience of cancer: a qualitative study. *BMJ* 2004;328:564.
[8] Umefjord G, Petersson G, Hamberg K. Reasons for consulting a doctor on the Internet: Web survey of users of an Ask the Doctor service. *J Med Internet Res* 2003;5:e26.
[9] Rozmovits L , Ziebland S. What do patients with prostate or breast cancer want from an Internet site? A qualitative study of information needs. *Patient Educ Couns* 2004;53:57-64.
[10] Dutta-Bergman M. Trusted online sources of health information: differences in demographics, health beliefs, and health-information orientation. *J Med Internet Res* 2003;5:e21.
[11] Coulter A, Magee H. The European Patient of the Future. Maidenhead: Open University Press, 2003.
[12] Horrigan JB, Rainie L. The Pew Internet & American Life Project. Counting on the Internet. Washington, DC: The Pew Internet & American Life Project, 2002.
[13] Feenberg AL, Licht JM, Kane KP, Moran K, Smith RA. The Online Patient Meeting. *J Neurol Sci* 1996;139:129-31.
[14] Hoch DB, Norris D, Lester JE, Marcus AD. Information Exchange in an Epilepsy Forum on the World Wide Web. *Seizure* 1999;8:30-4.
[15] Hardey M. E-Health: The internet and transformation of patients into consumers and producers of health knowledge. *Information, Communication and Society* 2002;4:388-405.

Address for correspondence

Dr John Powell,
Associate Clinical Professor, Health Sciences Research Institute, Warwick Medical School, University of Warwick,
Gibbet Hill Road, Coventry CV4 7AL, United Kingdom.
Telephone: +44(0)2476 574883.
Email: john.powell@warwick.ac.uk.

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Text Characteristics of Clinical Reports and Their Implications for the Readability of Personal Health Records

Qing Zeng-Treitler[a], Hyeoneui Kim[a], Sergey Goryachev[a], Alla Keselman[b] Laura Slaughter[c] Catherine Arnott Smith[d]

[a]*Decision Systems Group ,Brigham and Women's Hospital, Harvard Medical School, U.S.A.*
[b]*Lister Hill National Center for Biomedical Communications, National Library of Medicine, U.S.A.*
[c]*Center for Shared Decision Making and Nursing Research, Rikshospitalet-Radiumhospitalet Medical Center, Norway.*
[d]*School of Library and Information Studies, University of Wisconsin-Madison, U.S.A.*

Abstract

Through personal health record applications (PHR), consumers are gaining access to their electronic health records (EHR). A new challenge is to make the content of these records comprehensible to consumers. To address this challenge, we analyzed the text unit length, syntactic and semantic characteristics of three sets of health texts: clinical reports from EHR, known difficult materials and easy-to-read materials. Our findings suggest that EHR texts are more different from easy texts and more similar to difficult texts in terms of syntactic and semantic characteristics, and EHR texts are more similar to easy texts and different from difficult texts in regard to text unit length features. Since commonly used readability formulas focus more on text unit length characteristics, this study points to the need to tackle syntactic and semantic issues in the effort to measure and improve PHR readability.

Keywords:

consumer health, readability, personal health record, consumer health vocabulary, natural language processing

Introduction

Increasingly, consumers are taking an active role in their own health care by accessing and contributing to their personal health records (PHR). This role has become widely recognized by health care organizations and policy makers. One major source of PHR content is the institutional electronic health records (EHR), which are complex documents created by health care professionals for medical, legal, financial and administrative purposes. The organization, syntax, vocabulary and underlying conceptual knowledge employed by medical records may not be easily comprehended by lay people. For consumers with an average level of health literacy, understanding EHR content is challenging: Is "negative x-ray finding" good or bad? What does "FH of MI" mean? Which section(s) of the discharge summary describe my treatment plan?

For the PHR to fully realize its potential in helping consumers to manage complex health data and to facilitate informed decision making and self-care, its content should be easily understandable to consumers. A prominent panel of fellows of the American College of Medical Informatics recently published a white paper [1] stating "In order to be useful to the patient, the PHR must present data and accompanying tools in ways that enable the individual to understand and to act on the information contained in the record…..Both terminology and data presentation must be adapted to the individual using the PHR, so that they realize optimal benefits."

Considerable readability issues exist for today's PHRs which typically contain selected portions of an EHR and are aimed at a rather educated user group [2, 3]. The need to make the EHR information comprehensible will be more critical as an increasingly diverse patient population gains access to increasingly comprehensive records.

We have embarked on a project to translate EHR information into intelligible structure and plain language for PHR users. The goal of translation requires us to first understand and measure the readability of EHR information. This paper presents an analysis of EHR text characteristics as one of the initial steps toward text translation.

Background

Although more than a few health-specific literacy tests such as the Test of Functional Literacy in Adults (TOHFLA) have been developed [4], practically no health-specific readability measure is available. Recognizing the potential limitations of existing general-purpose readability measurements, we and other researchers began to examine various characteristics of health texts.

Our previous studies focused on the vocabulary aspect and resulted in the development of term and concept familiarity estimation methods [5]. In evaluation studies, our predicted term and concept familiarity was shown to be well correlated with actual consumer vocabulary knowledge and comprehension [6, 7] and outperformed the word length and word list techniques employed by the general-purpose readability formulas [5].

In a 2006 report, Rosemblat and colleagues examined what text features health communication experts use to determine the readability of consumer-oriented health texts [8]. The two significant factors they identified were "vocabulary" (i.e. number of words that are likely to be familiar to readers) and "main point" (i.e. ability of readers to identify

and understand the "take home" message). In this study, the presence and absence of these factors in the texts were established subjectively by the experts.

Also in 2006, Leroy and colleagues published a study that analyzed and compared the text characteristics of four types of documents: easy and difficult WebMD documents, patient blogs, and patient educational material, for surface and content-based metrics [9]. The easy and difficult WebMD documents were determined using the Flesch-Kincaid formula. They found a number of syntactic and semantic similarities and differences: for example, the easy WebMD pages are the most similar to patient blogs in terms of vocabulary difficulty.

No previous study has examined the readability-related characteristics of clinical reports in EHR systems, though a study by Chapman et al. did apply the Flesch-Kincaid formula to a set of dictated and transcribed x-ray reports [10]. While it was not the authors' finding, we observed from the results of Chapman's study that the readability measure (Flesch-Kincaid) was greatly underestimating the difficulty of these reports: the average grade level of these reports was reported to be 7.6. Based on our experience of natural language processing (NLP) of radiology reports, they are often difficult for non-clinician researchers with graduate school education (equivalent to grade level 18 and above) to comprehend.

Materials and methods

Materials

We collected three sets of health documents: EHR reports, difficult texts and easy texts.

The first set contains 40 EHR reports randomly selected from the clinical data repository of the Brigham and Women's Hospital and Massachusetts General Hospital (Boston, MA, U.S.A.). We retrieved 10 outpatient clinic notes and 10 discharge summaries from each institution. The reports cover topics such as chief complaint, history of illness, laboratory finding, treatment, and discharge plan. The medical diagnoses which appeared in the reports included common disease such as asthma, diabetes mellitus, pneumonia, and osteoarthritis. The average length of the documents is 3374 characters.

The second set is 40 abstracts of scientific journal papers randomly retrieved from MEDLINE (www.pubmed.org). The majority of journals indexed for MEDLINE are intended for a readership of researchers and clinicians, and typically require substantial background knowledge in specialty areas (e.g. molecular biology or nephrology) to understand. The abstracts, thus, are good examples of materials that are difficult for lay health consumers. This set of documents included various topics such as abdominal pain, asthma, hypertension, and paranoid schizophrenia. The average length of the documents is 1801 characters – abstracts are short by nature.

The third set is a convenience sample of 40 easy-to-read documents. We collected 27 (self-labeled) easy-to-read documents from multiple high-qualify consumer health Web sites: 21 from the Food and Drug Administration (www.fda.gov), 4 from the National Institute of Mental Health (www.nimh.nih.gov), and 2 from the National Institute on Alcohol Abuse and Alcoholism (www.niaaa.nih.gov). We also selected 13 records from the Reuters Health (http://www.reutershealth.com). The topics covered by these easy-to-read materials varied as well, including allergy, heart attack, breast feeding, alcoholism, and depression. The average length of the documents is 4101 characters.

Methods

Each document was processed by HITEx – a suite of open-source NLP tools that we have developed [11]. Each document was tokenized, split into sentences, and had part-of-speech (POS) tags assigned. Noun phrases were subsequently extracted and mapped to the Open-Access Collaborative (OAC) consumer health vocabulary[1].

For each parsed document, we first calculated the total number of characters, words, sentences and paragraphs. We considered a word to be any token that does not contain punctuation symbols. Paragraphs were defined depending on the document style: In the EHR reports we used, paragraphs are separated by a blank line; in the easy text sample, they are marked by line breaks. We then calculated the average word length (i.e., number of characters per word), average sentence length (i.e., number of words per sentence), and average paragraph length (i.e., number of sentences per paragraph).

Next, we calculated the frequency distribution of POS categories in each document. For the purpose of statistical analysis, some less frequent POS categories were merged (e.g., all punctuation categories were merged into one), reducing the total number of categories from 30 to 13.

Thirdly, we calculated the average term and concept familiarity scores for each document. These scores were obtained from the OAC consumer health vocabulary[1]. The OAC vocabulary provides three scores: a frequency-based term score (derived from term occurrence data), a context-based term score (derived from term co-occurrence data) and a context-based concept score (derived from concept co-occurrence data). The term scores reflect the string (surface)-level difficulty for consumers and the concept scores reflect the concept-level difficulty for consumers [6]. The scores have the range between 0 and 1, with 1 indicting perfect consumer familiarity (i.e., the easiest) and 0 indicting complete consumer unfamiliarity (i.e., the most difficult). We used the scores to gauge the semantic complexity of the contents.

Some terms did not map to OAC and not all OAC terms had the three scores assigned yet. When we calculated the weighted averages of the scores, these out-of-dictionary terms and terms with missing scores were excluded.

Finally, we calculated the Flesch-Kincaid grade level [12] for every document using a Microsoft Word™ built-in function.

In statistical analysis, mean and 25% and 75% quintiles of each text characteristic were first calculated. We then tested whether the EHR set shares the same text character-

1 More detailed explanations and related publications of this vocabulary and the term/concept familiarity scores it provides can be found on the Consumer Health Vocabulary Initiative's Web site (www.consumerhealthvocab.org).

istics with either the difficult or the easy document set. The distributions of the characteristics were examined using the Shapiro-Wilkins W test ($p < .001$) to assess normality. When distributions are not normal, differences in distributions were tested using the Wilcox rank sum test. The text characteristics with normal distributions were tested for differences in means using the t-test.

Results

The text characteristics of the three sample sets (EHR, difficult, and easy text) were different in many aspects. The mean and 25%/75% quantiles of the text characteristics are reported in Table 1-3.

Table 1 - Means and differences in text unit length characteristics

	EHR	Easy	Difficult
Average # of Characters per Word^	4.97 (4.78, 5.23)	4.71* (4.40, 5.02)	5.53* (5.28, 5.83)
Average Num of Words per Sentence	13.46 (10.32, 16.35)	16.93 (12.38, 21.89)	17.60** (14.53, 20.41)
Average Num of Sentences per Para–graph	3.57 (1.96, 3.41)	1.84 (1.45, 2.0)	15.50** (13.0, 18.0)

Lower and upper quantiles are in parentheses.
* compared to EMR, $p < .05$ ** compared to EHR, $p < .0001$
^ tested for means using t-test.

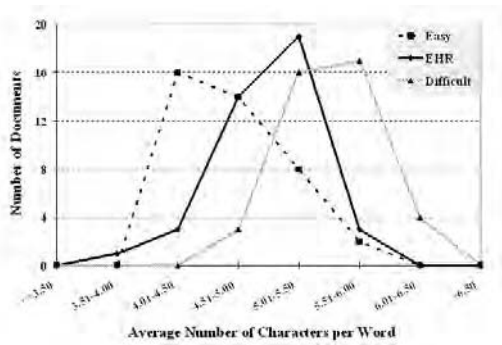


Figure 1 - Distributions of the num of characters per word

On the text unit length level, the EHR sample falls between the easy and difficult texts in terms of word length (Figure 1); it has the shortest sentence length (Figure 2), which is not statistically different from that of the easy texts; it also has very few sentences per paragraph (Figure 3), which is not statistically different from that of the easy texts while being very different from that of the difficult texts. The short average word and sentence lengths of the EHR sample are largely due to the use of abbreviations and incomplete sentences.

On the syntactic level, the EHR sample differs from the easy texts statistically in most of the POS categories (Table 2). It does have some similarity with the difficult texts, for example, both have high proper noun usage and less verb and adverb usage.

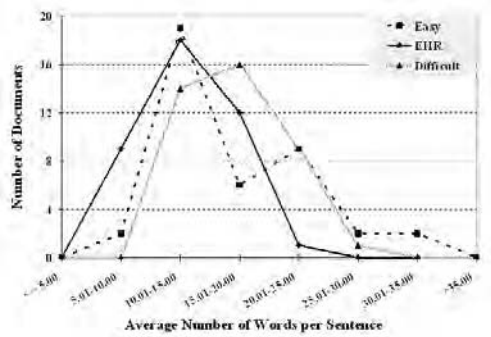


Figure 2 - Distributions of the number of words per sentence

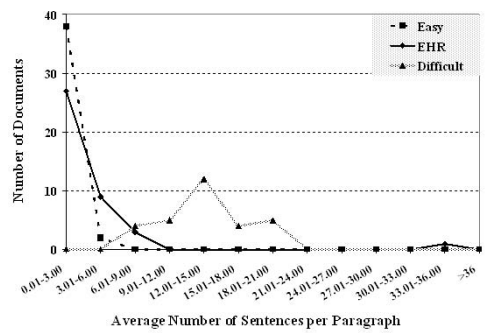


Figure 3 - Distributions of the number of sentences per paragraph.

Table 2 - Means and differences in syntactic characteristics (parts of speech categories per sentence)

	EHR	Easy	Difficult
Verb	1.62 (1.31, 2.01)	2.70** (2.09, 3.21)	1.93* (1.41, 2.29)
Noun^	3.30 (2.27, 4.08)	4.75** (3.29, 6.08)	5.48** (4.47, 6.19)
Proper Noun	3.32 (1.62, 4.50)	1.27** (0.51, 1.71)	3.01 (2.0, 4.0)
Punctuation	2.44 (1.77, 2.70)	2.37 (1.69, 3.17)	3.39** (2.63, 4.0)
Pronoun	0.36 (0.15, 1.56)	0.65** (0.41, 0.87)	0.09** (0, 0.15)

	EHR	Easy	Difficult
Adverb	0.33 (0.16, 0.51)	0.58** (0.44, 0.67)	0.40 (0.24, 0.47)
Adjective^	0.87 (0.54, 1.16)	1.41** (0.95, 1.67)	1.76** (1.43, 2.02)
Particle	2.02 (1.33, 2.53)	3.12 (2.16, 3.94)	3.09** (2.48, 4.70)
Determiner	0.03 (0, 0.06)	0.10** (0.04, 0.12)	0.05 (0, 0.07)
Proposition	0.02 (0, 0.04)	0.03 (0, 0.05)	0.02 (0, 0.04)
Symbol	0.09 (0, 0.08)	0.00** (0, 0)	0.06 (0, 0.06)
Modal	0.09 (0.02, 0.10)	0.32** (0.25, 0.39)	0.09 (0, 0.13)
Possessive	0.35 (0.06, 0.53)	0.47 (0.27, 0.62)	0.05** (0, 0.10)

Lower and upper quantiles are in parentheses.
* compared to EMR, p <.05 ** compared to EHR, p <.0001
^ tested for means using t-test.

Table 3 - Means and differences in semantic characteristics

	EHR	Easy	Difficult
Average Context-based Term Scores	0.60 (0.58, 0.66)	0.72** (0.68, 0.80)	0.62 (0.53, 0.70)
Average Frequency-based Term Scores^	0.63 (0.59, 0.67)	0.77** (0.74, 0.80)	0.67* (0.63, 0.71)
Average Context-based Concept Scores^	0.66 (0.65, 0.68)	0.73** (0.71, 0.75)	0.68 (0.65, 0.71)

Lower and upper quantiles are in parentheses.
* compared to EMR, p <.05 ** compared to EHR, p <.0001
^ tested for means using t-test.

On the semantic level, EHR's mean familiarity scores are the lowest in all three metrics (Table 3). The scores range from 0 to 1, terms/concepts with lower scores are considered to be less familiar to consumers and more difficult. Statistically significant differences were found between EHR and easy texts in every metric, while EHR and difficult texts only differ on the frequency-based term scores. The score distribution curves of the EHR and difficult texts overlap to a large extent (Figures 4-6).

It is a common belief that the EHR is not easy for consumers to understand and there is ample empirical evidence supporting the view [1]. The comparison of EHR text characteristics with the characteristics of the difficult and easy health text samples suggest that the syntactic and semantic characteristics are the key to explain EHR's low readability. On the other hand, the text unit length features (word, sentence and paragraph lengths) failed to account for the difficult of EHR texts: EHR texts are similar to easy materials regarding these features.

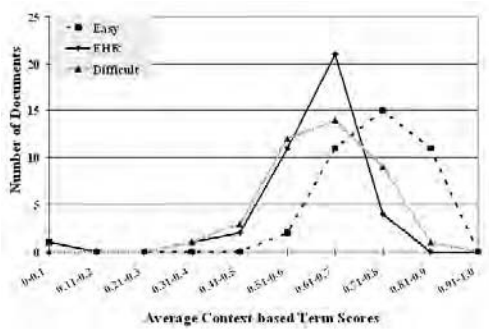


Figure 4 - Distributions of the average context-based term scores

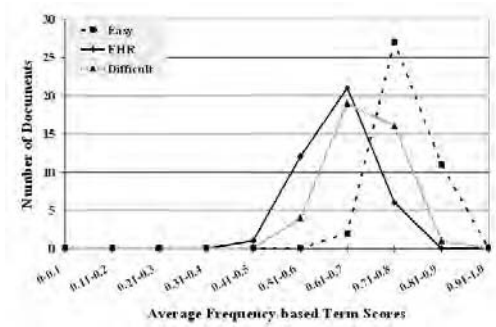


Figure 5 - Distributions of the average frequency-based term scores

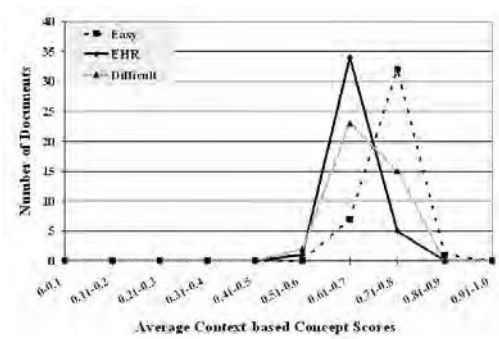


Figure 6 - Distributions of the average context-based concept scores

Table 4 - Means and differences in readability scores

	EHR	Easy	Difficult
Flesch-Kincaid Grade Levels	9.68 (8.55, 11.50)	8.23* (5.65, 12)	11.98** (12.0, 12.0)

The Flesch-Kincaid levels of the three samples also differ statistically (Table 4). The average grade of 8.23 is probably an accurate assessment of the easy texts. The 11.98 grade level assigned to the difficult texts is an underestimation, however, can be partially blamed on the abstract style. The 9.68 grade level of EHR, though, is clearly inaccurate.

Discussion

Although the clinical reports in EHRs are primarily written and read by health professionals, consumers are gaining access to them through the proliferation of PHRs and the readability of EHR reports for consumers has been recognized as a problem. This paper presents an analysis of the text unit length, syntactic, and semantic characteristic of EHR texts, and their implications for PHR readability.

Our statistical analysis indicates that EHR texts are more different from known easy texts and more similar to known difficult texts on the syntactic and semantic levels, while EHR texts are more similar to easy texts and different from difficult texts on the text unit length level. On the other hand, the commonly used readability formulas focus on text unit length rather than syntactic and semantic features, which we believe is the main cause of Flesch-Kincaid formula's inaccurate assessment of the difficulty of the EHR reports. To measure and improve EHR readability for the PHR audience, syntactic and semantic characteristics must be taken into consideration.

One may argue that it is a bit farfetched to compare the text unit length characteristics of EHR and easy texts, since they are obviously different types of documents. Please note that most of the readability formulas which are commonly used by biomedical researchers for a wide range of text materials are based on text unit length. The similarity between EHR and easy texts in terms of unit length points to the limitations of the unit length-based measurements.

Some limitations of this study are: Although the text samples we used are comparable in size to a couple of the related studies described in the Background section, they are small. MEDLINE abstracts are good examples of difficult health texts; however, they do not provide a diverse representation of difficult health texts. In follow-up studies, full length articles will be used as well.

Our results suggest that there is a need for an EHR specific readability measure and we plan to develop such a metric. We are also interested in exploring the role syntactic and semantic characteristics play in other health texts that consumers are exposed to and validate our findings through user studies. The ultimate goal of our research is to translate EHR texts into a lay-friendly language for PHRs.

Acknowledgments

This work is supported by the NIH grant R01 LM07222 and by the Intramural Research Program of the NIH, NLM.

References

[1] Tang PC, Ash JS, Bates DW, Overhage JM, Sands DZ. Personal health records: definitions, benefits, and strategies for overcoming barriers to adoption. J Am Med Inform Assoc 2006;13(2):121-6.

[2] Hassol A, Walker JM, Kidder D, Rokita K, Young D, Pierdon S, Deitz D, Kuck S, Ortiz E. Patient experiences and attitudes about access to a patient electronic health care record and linked web messaging. J Am Med Inform Assoc 2004;11(6):505-13.

[3] Winkelman WJ, Leonard KJ, Rossos PG. Patient-perceived usefulness of online electronic medical records: employing grounded theory in the development of information and communication technologies for use by patients living with chronic illness. J Am Med Inform Assoc 2005;12(3): 306-14.

[4] Baker D, Williams M, Parker R, Gazmarian J, Nurss J. Development of a Brief Test to Measure Functional Health Literacy. Patient Education and Counseling 1999;38:33-42.

[5] Zeng Q, Kim E, Crowell J, Tse T. A Text Corpora-Based Estimation of the Familiarity of Health Terminology. Lecture Notes in Computer Science 2005;3745 / 2005:184-192.

[6] Keselman A, Tse T, Crowell J, Browne A, Ngo L, Zeng Q. Assessing Consumer Health Vocabulary Familiarity: An Exploratory Study. In: MEDNET; 2006; Toronto, CA; 2006.

[7] Keselman A, Massengalea L, Ngo L, Browne A, Zeng Q. The Effect of User Factors on Consumer Familiarity with Health Terms: Using Gender as a Proxy for Background Knowledge about Gender-Specific Illnesses. In: International Symposium on Biological and Medical Data Analysis; 2006; Thessaloniki, Greece; 2006. In press.

[8] Rosemblat G, Logan R, Tse T, Graham L. How Do Text Features Affect Readability? Expert Evaluations on Consumer Health Web Site Text. In: MEDNET; 2006; Toronto, CA; 2006. In press.

[9] Leroy G, Eryilmaz E, Laroya BT. Health Information Text Characteristics. In: American Medical Informatics Association (AMIA) Annual Symposium; 2006; Washington DC; 2006. In press.

[10] Chapman WW, Aronsky D, Fiszman M, Haug PJ. Contribution of a speech recognition system to a computerized pneumonia guideline in the emergency department. Proc AMIA Symp 2000:131-5.

[11] Zeng QT, Goryachev S, Weiss S, Sordo M, Murphy SN, Lazarus R. Extracting principal diagnosis, co-morbidity and smoking status for asthma research: evaluation of a natural language processing system. BMC Med Inform Decis Mak 2006;6:30.

[12] Zakaluk BL, Samuels SJ. Readability: Its Past, Present, and Future: Intl Reading Assn; 1988.

Address for correspondence

Qing Zeng-Treitler, Ph.D.
Decision Systems Group
Brigham and Women's Hospital
Harvard Medical School
75 Francis Street Boston, MA 02115
Email: qzeng@dsg.harvard.edu

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Generic Screen Representations for Future Proof Systems – Is It Possible? Two-model approach to a generic GUI

Helma van der Linden[a], Thilo Schuler[b], Rong Chen[c], Jan Talmon[a]

[a] *Medical Informatics, University Maastricht, The Netherlands*
[b] *Department of Medical Informatics, University of Freiburg, Germany*
[c] *Department of Biomedical Engineering, Linköping University, Sweden*

Abstract

Semantic interoperability should not only cover system interpretation of incoming information, but should be extended to include screen representation. This article describes a two-model approach to generate a screen representation for archetype-based information, which is inspired by the two-model approach used by openEHR for their archetypes. It provides a separation between software-related display knowledge and domain-related display knowledge and is designed with reuse of components in mind. This approach leads to a flexible GUI that can adapt not only to information structures that are not predefined within the receiving system and display them in a meaningful way, but also to novel ways of displaying the information.

We are working on a proof of concept implementation to validate the approach.

Keywords:

computerized medical record, electronic health record, user-computer interface, openEHR, archetypes, HL7

Introduction

Patient mobility is increasing over the last decade varying from "shopping around for care" in different health care organizations to prolonged stays abroad with the increasing need of care. This results in fragmented patient-related health information distributed across different systems.

A current approach for future health information systems is to create a virtual health record that integrates the fragmented information by connecting distributed systems and presenting them as one. The underlying principle is information exchange based on standardized messages. The two major approaches in this respect are HL7 v3[1] [1] and CEN/TC251 13606 [2, 3] combined with openEHR archetypes[2] [4].

With ongoing efforts towards harmonization of the best of both frameworks it will be possible to see virtual health records come into reality within some decades.

In this article we assume the existence of such an environment where virtual health records exist and information exchange is not limited to the systems of a single organization, but expanded to incorporate other organizations on a regional or maybe even global scale. In this environment it is possible to retrieve or receive patient information of which the structure has not been known before. Indeed, as advocated by openEHR, true future-proof electronic health record systems will be able to accommodate new medical concepts without the need for redevelopment. The key idea behind archetypes is to express new information structures as a combination of predefined classes.

The openEHR Foundation has currently the only architecture that allows handling of unknown information structures. We therefore focus on openEHR archetypes in this article. This does not imply that archetypes are the only means of exchanging information.

Scenario

A GP suspects that the patient suffers from a hereditary disease and refers the patient to the genetics clinic where tests will be done to confirm or reject his suspicion. Unfortunately, his suspicion is confirmed and the GP receives a discharge letter that contains a summary, the lab results and a family tree.

When we assume that the discharge letter is a structured message containing the data structures with the relevant data rather than a formatted display document, the question arises how the information of the discharge letter should be displayed on the GP's screen and in particular the information that is normally not part of the GP's system (i.e. the family tree).

The article discusses an approach to display new information structures on a user's screen using as much display knowledge as available.

Background

ISO 18308 defines semantic interoperability as the ability for information shared by systems to be understood at the

1 In this article HL7 will refer to the new v3 standard in its latest form.
2 In this article we will refer to the CEN 13606/openEHR archetypes as openEHR or archetypes for brevity.

level of formally defined domain concepts so that the information is computer processable by the receiving system [5].

Both HL7 and openEHR support semantic interoperability at two levels: at the data structure level and at the domain level.

At the data structure level, medical concepts are described using predefined data structures. This ensures that the information exchanged is complete (i.e. it contains all relevant data and metadata) and can be parsed, stored and subsequently retrieved. At the domain level metadata such as the code and coding scheme are used to avoid ambiguity in understanding.

The PropeR project has revealed that semantic interoperability is not simply a matter of interoperability between systems, but also between user and system. To refer to the scenario, even if the GP's system is capable of storing and subsequently retrieving the fully structured family tree, if there is no suitable screen representation, it is very difficult for the GP to correctly interpret the information.

Van der Meijden [6] and van Ginneken [7] have already discussed the difference between data entry and data retrieval with respect to the screen representation. Since it is logical to assume that data entry is only done in the local system, the issue of undefined data structures does not occur during data entry as we may assume that the local system is designed to support the specific user tasks in the application domain. The approach we present here will primarily be focused on data consultation and not on data entry.

Methods

In the context of the PropeR project [8, 9] we built a web based EHR system based on a simplified version of archetypes. We focused on the implementation of a domain-agnostic system and the strict separation between archetypes and screen representations.

We followed a similar approach by researching the feasibility of generating a GUI based on openEHR archetypes [10].

The lessons learned in both projects were combined to develop a more generic approach that can handle the situation we discussed before.

Results

Presentation level interoperability

In our view displaying information in an unambiguous way that supports the user's work processes, requires three types of knowledge:

- Knowledge of the information to display;
- Knowledge of the way a user is accustomed to view information;
- Knowledge of the device that is used to display the information.

Information-related presentation knowledge

At the lowest level this refers to the display of the data types that are used to construct the information structure: numbers are displayed differently than text. This is however, not sufficient. Even the example of a simple blood pressure shows that a higher level of knowledge is necessary to correctly display a blood pressure; that is in the common form of two numbers separated by a slash. A graphic tree form would best represent a family tree.

Localized presentation knowledge

Displaying information can be subject to local customs, varying from the local language and the local date format to preferred units (e.g. mg/dl vs. µmol/l) and coding schemes. There are also personal differences in what the best way of information presentation is with different reasons such as learned behavior or different cognition strengths (e.g. visual, textual).

Device-related presentation knowledge

The current trend towards ubiquitous computing has produced a large range of devices capable of sending and retrieving information ranging from desktop computer and laptops to tablet pcs, pda's and smartphones. While each modern model contains a webbrowser, and thus an abstraction from the underlying device, the supported functionality and the screen size places extra constrains on the presentation.

These different types of knowledge are often hard coded into the GUI of the client application. This makes it very hard to display incoming information from a different domain.

A two-model approach to generic GUI generation

Given the premise that future-proof systems are also capable of displaying information from other domains, it is necessary that these systems contain domain-agnostic screen representational functionality.

From the PropeRWeb application we learned that screen representation knowledge, however low-level, should not be incorporated in the archetype definition [11]. Not only does it introduce two different kinds of knowledge (medical domain knowledge and presentation knowledge) in a single model, but it is also common knowledge that a single data type, especially numerical, can be displayed in different ways, for example as a single number or as a table or graph.

In our approach we distinguish two models: a display oriented model (the GUI model) that defines widgets as screen presentation units and a domain oriented model (the content model) that defines content units which create meaningful presentations using widgets. The first model is the realm of the GUI designer, while domain experts use the second model.

Localized presentation knowledge is defined in profiles and views. This results in four sets of presentation units, which are shown in figure 1: widgets, content units, views and profiles.

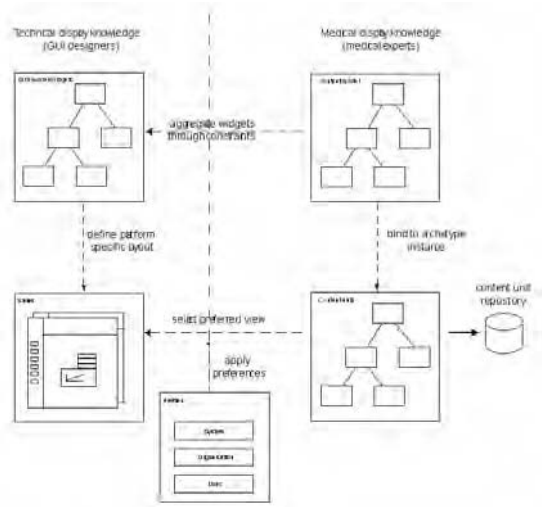


Figure 1 - Two-model approach to a generic GUI

All presentation units follow object-oriented design in which a specific unit inherits characteristics from a more generic unit. This improves consistency and flexibility. It is also valid to define multiple units for the corresponding information unit.

GUI model

The building blocks of a GUI are widgets. A widget is a display unit that contains presentation knowledge for a single data type. These widgets can be mapped to classes in the Reference Model of openEHR. Two types of widgets exist: data-oriented widgets such as "text", "image" and "number" and layout-oriented widgets such as "list" and "table".

The GUI model defines generic widgets that are converted to specific versions in the underlying system by using views.

Content model

Content units are defined using a content unit definition language. They are a semantic aggregation of widgets. They can be regarded as the display counterpart of archetypes. Content units, like archetypes can include other content units. Content units specify the binding to the information in the archetype instance as well as an established layout, such as the X/Y format of a blood pressure. Note that this layout only specifies relative positions of the included widgets and/or content units.

At the top level a content unit matches a COMPOSITION. These high-level content units are called *documents*. Different documents can be designed to reflect the differences in users' roles.

Content units can also include calculations, e.g. the total score of a test. They can also include normal ranges for semantic interpretation of the value. For example: the value of a body mass index can be color-coded based on the semantic interpretation (e.g. "normal" is green, "obese" is red).

Like archetypes content units are stored in a content unit repository. Since they are a semantic, platform independent representation of an archetype, they can be shared in the same way archetypes are sharable.

Views

Views can be regarded as implementations of content units customized for the device or application that will be displaying the information. Views can also include other GUI artifacts such as navigation bars.

A view is focused on presentation of the content and therefore part of the GUI designers' realm.

Profiles

The focus of profiles is the conversion of the information to match the user's expectations and thus avoid interpretation errors.

A profile contains preferences at various levels that modify the presentation of the information. There are three levels:

- *System level*. This level contains generic preferences that should always be applied, e.g. language, date format, metric vs. imperial system etc.
- *Local level*. This level contains generic preferences that are organization or location specific and are more domain-related. These preferences include preferred units and preferred terminologies.
- This level can also include role-based preferences that refer to role-based documents.
- *User level*. This level contains specific user related preferences that can modify the preferred view for a certain type of COMPOSITION e.g. if the user prefers graphs to tables.

Presentation generation

The process of generating the presentation is based on the pipeline concept. A pipeline can be compared to an assembly line where material arrives in a certain form, which is then processed by various stages along the line and finally delivered as a complete product. Adding or removing stages delivers a different product without affecting the other stages. A successful implementation of the pipeline concept can be found in Apache Cocoon [12].

A pipeline offers a component-based approach to the transformation of information. By adding or removing transformation components, the end result can change without affecting the other components. Different pipelines can implement different functionalities while sharing components.

In our approach an openEHR composition enters the pipeline. This composition can be the result of a query for information or the result of a notification of new information. In both cases the composition can contain instances of unknown archetypes.

Transformations handle device selection, presentation units selection, profile application and the final rendering of the selected view.

Figure 2 shows the generation process.

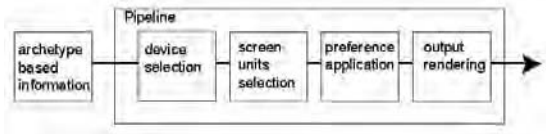


Figure 2 - Generation of view

There are several advantages to this solution:

- First and foremost this approach offers the flexibility of defining specific, optimized screen representations for known information structures, while providing the means to generate useable screen representations of unknown information structures.
- By separating the various types of presentation knowledge into distinct models, it is possible to separate pure GUI knowledge from medical display knowledge, thus honoring the two-model approach and promoting reuse.
- This approach is flexible enough to build a role-based GUI. Nurses and physicians can see the same information but optimally presented for their specific needs, while the only difference in development might be a document definition.
- The evolution of medical knowledge will always create new data types and new archetypes, which would lead to new display representations. Our approach ensures that as much of the available display knowledge can be reused.
- New and novel ways to view EHR information are a topic of ongoing research. By adhering to the proposed approach these views can benefit from the available display knowledge that is already expressed in content units. [13]
- Both HL7 and openEHR archetypes are using a limited set of predefined data types with an ongoing effort to harmonize the sets between the two parties. By describing one or more widgets for each data type it should be possible to provide a meaningful display of the information, without incorporating presentation knowledge in the information structure. This means the current archetype or message specifications need not be extended and the number of widgets is not very large.
- The information can be converted to match local and user preferences such as preferred coding scheme, language, units and more.
- A fallback mechanism is used to select a more generic representation in the absence of a specific one.
- Standardized content units could be shared between systems; the same way archetypes can be shared. This increments the intelligence and usability of the system.
- The pipeline approach not only allows reuse of components, but also offers flexibility in adding functionality by a simple addition of pipelines.

This approach complements the openEHR architecture where templates are used to create a higher-level composition by constraining and ordering archetypes. In contrast, the openEHR templates are used to *create* an information structure, while the approach proposed in this paper is used to *display* the information.

HL7 focuses on message exchange only and therefore considers this problem to be part of the receiving system's domain. However, given the similarities in structure between archetypes and messages we believe this approach is equally useful in that realm.

There are also disadvantages:

- Higher-level, specific views can only exist for predefined information. New information or information from different domains will fall back to a more basic representation.
- A repository, equal to that for archetypes, is necessary for the various presentation units.
- A mechanism for retrieving an appropriate screen representation for the current archetype is necessary, since the most appropriate selection is based on multiple parameters, described earlier, that cannot be stored in the archetype instance.

The advantages of having flexible GUI interfaces outweigh the disadvantages. By incrementally defining screen representations that can be built on top of each other, there is less duplication of work in building a GUI. A higher-level screen representation allows the user to better interpret the presented information thus leading to more efficient and more reliable information exchange. Screen representations for new information structures can be added to the system without major redevelopment of the application.

Currently we are working on a proof of concept using the Apache Cocoon web application framework [12] to build a web application that can display instances of various openEHR archetypes based on the approach described here.

The Apache Cocoon web application framework is a generic open source framework that is heavily based on the concept of separation of concerns to define strict distinctions between model and view. It implements the pipeline concept and also provides a set of generic widgets and an XML-based language to define what we call views. Since Cocoon excels in processing XML it is a good candidate to build a generic generated web based GUI using the approach that we have presented before. A first version will be presented in the openEHR workshop of Medinfo 2007.

Related work

A similar approach is developed by Ocean Informatics and implemented in their EhrView [14]. The EhrView application modifies the information through a series of XSLT stylesheets. These stylesheets are selected by matching the archetypes names in the composition. The matching process selects the most specific stylesheet available in a repository.

The EhrView application does not separate content related modeling from software related modeling and it is only defined for one type of device: a regular screen of a desk-

top or laptop pc. It also offers limited options to adjust to local or user preferences.

Fiala et al. [15] have described a component-based approach for adaptive web documents that influenced our approach. They too make a distinction between content-related and display related presentation knowledge and they also used the pipeline concept to define web document generation. However, their focus is on adapting information presentation to user preferences and devices. The information is known and defined in advance and there is no method to handle unknown information structures or describe conversions to preferred units.

Conclusion

Semantic interoperability does not stop when information from one system can be successfully understood and/or incorporated in another system. It is also necessary to provide a screen representation that gives the user of the receiving system a clear understanding of the new information.

In this article we described an approach that extends the two-model approach that is currently used by openEHR by a similar approach for the GUI.

We argued that this approach leads to a flexible GUI that can adapt to information structures that are not predefined and still display them in a meaningful, higher-level way. We are currently working on a proof of concept implementation.

Acknowledgments

We would like to thank the team from Ocean Informatics who kindly provided the information we needed. We would also like to thank Patrick Ahles for his valuable insights.

References

[1] Health Level 7 Inc. [last visited 21 Mar 2007]; Available from: http://www.hl7.org

[2] Health Informatics – Electronic health record communication - Part 1: Reference Model: CEN/TC251; 2005 March 2005. Report No.: prEN13606-1:2005:E.

[3] Health Informatics – Electronic health record communication - Part 2: Archetypes: CEN/TC251; 2005 December 2005. Report No.: prEN13606-2:2005:E.

[4] openEHR. [last visited 21 Mar 2007]; Available from: http://www.openehr.org

[5] ANSI. ISO/TS 18308 Health Informatics - Requirements for an Electronic Health Record Architecture: ISO; 2003 2003-04-23.

[6] van der Meijden MJ, Tange HJ, Boiten J, Troost J, Hasman A. An experimental electronic patient record for stroke patients. Part 2: system description. Int J Med Inform. 2000 Sep;58-59:127-40.

[7] van Ginneken AM. Considerations for the representation of meta-data for the support of structured data entry. Methods Inf Med. 2003;42(3):226-35.

[8] van der Linden H, Boers G, Tange H, Talmon J, Hasman A. PropeR: a multi disciplinary EPR system. Int J Med Inform. 2003 Jul;70(2-3):149-60.

[9] van der Linden H, Talmon J, Tange H, Grimson J, Hasman A. PropeR revisited. Int J Med Inform. 2005 Mar;74(2-4):235-44.

[10] Schuler T, Garde S, Heard S, Beale T. Towards Automatically Generating Graphical User Interfaces from openEHR Archetypes. Stud Health Technol Inform. 2006;124:221-6.

[11] van der Linden H, Grimson J, Tange H, Talmon J, Hasman A. Archetypes: the PropeR way. Medinfo. 2004;11(Pt 2):1110-4.

[12] Apache Cocoon Web Application Framework. [last visited 21 Mar 2007]; Available from: http://cocoon.apache.org

[13] Sundvall E, Nyström M, Forss M, Chen R, Petersson H, Åhlfeldt H. Graphical Overview and Navigation of Electronic Health Records in a prototyping environment using Google Earth and openEHR Archetypes. MedInfo 2007; 2007; Brisbane: accepted for publication; 2007.

[14] Ocean Informatics Pty L. EhrView. [last visited 21 Mar 2007]; Available from: http://www.oceaninformatics.biz/products/EhrView.html

[15] Fiala Z, Hinz M, Meißner K, Wehner F. A Component-based Approach for Adaptive, Dynamic Web Documents. Journal of Web Engineering. 2003 September 2003;2 (1&2):058-73.

Address for correspondence

Helma van der Linden
Medical Informatics
University Maastricht
POBOX 616
6200 MD Maastricht
The Netherlands
hvanderlinden@mi.unimaas.nl

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Speech Recognition in Dental Software Systems: Features and Functionality

Jeannie Yuhaniak Irwin, Shawn Fernando, Titus Schleyer, Heiko Spallek

Center for Dental Informatics, Department of Biomedical Informatics, University of Pittsburgh, Pittsburgh, PA United States

Abstract

Speech recognition allows clinicians a hands-free option for interacting with computers, which is important for dentists who have difficulty using a keyboard and a mouse when working with patients. While roughly 13% of all general dentists with computers at chairside use speech recognition for data entry, 16% have tried and discontinued using this technology. In this study, researches explored the speech recognition features and functionality of four dental software applications. For each system, the documentation as well as the working program was evaluated to determine speech recognition capabilities. A comparison checklist was created to highlight each program's speech functionality. Next, after the development of charting scripts, feasibility user tests were conducted to determine if performance comparisons could be made across systems. While four systems were evaluated in the feature comparison, only two of the systems were reviewed during the feasibility user tests. Results show that current speech functionality, instead of being intuitive, is directly comparable to using a mouse. Further, systems require memorizing an enormous amount of specific terminology opposed to using natural language. User testing is a feasible way to measure the performance of speech recognition across systems and will be conducted in the near future. Overall, limited speech functionality reduces the ability of clinicians to interact directly with the computer during clinical care. This can hinder the benefits of electronic patient records and clinical decision support systems.

Keywords:

Dental Informatics, Speech Recognition Software, User-Computer Interface, Medical Informatics Applications, Practice Management, Dental

Introduction

During care, dental clinicians have difficulty using a keyboard and a mouse, primarily because of infection control concerns but also because they are constantly using their hands for procedures and their office space and setup make it difficult to have the keyboard and/or mouse in close proximity. A solution to this problem, which is being employed in medicine [1-3], is the use of speech recognition applications to interact with the clinical computer. A recent study published on computing in clinical dental care offers insight into the adoption and use of speech applications in general dentistry [4]. Thirteen percent of all offices surveyed used speech input; however, 16% tried and discontinued using the technology [4]. Those who discontinued using speech did so because of technical problems with speech recognition (57%), lower efficiency compared to other data entry methods (13%), usability problems (9%), and other issues (22%) [4]. It is clear that there may be significant barriers to using the speech modules of current dental systems. Currently, dental speech applications typically implement command-and-control functionality as well as the transcription of free text [5]. The command-and-control functionality supports two types of activities: (1) navigating within the application (for instance to select a specific patient) and (2) entering structured data in text fields, list boxes, radio buttons and checkboxes [5]. Transcription is used primarily for dictating progress notes, surgical reports and similar textual information [5].

To date, there is no comprehensive overview or evaluation of the currently available speech recognition products in general dentistry. This study is significant because it is not only a comparative analysis of the four software applications, but it will also be an informative starting point for researchers interested in the benefits of speech modules as well as the development and design of future dental speech interfaces.

Materials and methods

Speech recognition features comparison

To compare the speech functionality of the software applications, the research team first acquired full working versions of the four practice management systems (PMS). Based on findings from a recent study, these four systems make up approximately 80% of the current practice management market in the United States [4]. The researchers reviewed each program's user manual to determine the system's speech features and functions. Then each system was installed according to its default installation configuration to allow the researchers to explore the programs' speech functionality. Researchers manually tested and used all components of speech functionality within each system. Lastly, the software vendors were contacted to answer any specific questions regarding the system's speech functionality and features. For example, a call to one of the companies was made when it was not clear what

speech engine was used in the system. A comparison checklist was created to highlight each program's technical details pertaining to speech input, including which features were present/absent as well as the program's limitations.

Feasibility user tests

In the second portion of this study, feasibility user tests were conducted to determine if the performance of speech recognition could be evaluated across systems. User tests have the potential to evaluate the efficiency, effectiveness, and user satisfaction of tested systems [6-8]. While four systems were evaluated in the feature comparison, only two of the systems were reviewed during the feasibility user tests. In our study, each user completed a task (efficiency and effectiveness) using one of the programs and then filled out a questionnaire (satisfaction) about their experience with the system.

To develop the task, a simulated intraoral patient record was created which contained a wide range of findings specific to testing the clinical speech capabilities of each system. Explicitly, the task was to chart 18 different hard and soft tissue findings via speech. During the feature comparison phase of this study, it became clear that to use the speech features of these systems, the user must have not only an in-depth working knowledge of the program, but must learn many specific and sometimes complex commands to interact via voice. Because of the complexity and specific command knowledge needed to complete the charting task via speech, participants could not be expected to complete the tasks on their own. Therefore, a step-by-step script was created for each of the two programs, which the participant used to test the system. As classic user testing does not involve using verbatim scripts [6], this unseen obstacle was one of the reasons feasibility tests were conducted on only two systems before performing an entire set of user tests. After development of the scripts, each was sent to their corresponding manufacturer to be evaluated for correctness and efficiency. Each manufacturer reviewed and edited its individualized script and then returned it to us with any changes in how the task should be completed via voice in their system.

To conduct the feasibility user tests, a computer (Windows XP, 1.5GHz Intel Pentium 4 processor, and 256MB of RAM) was equipped with an extra 80GB hard drive. This was necessary to store images of the machine with each software package installed. Norton Ghost 2003 (Symantec, Cupertino, CA) was used to make an image of the machine at baseline (with a fresh XP install). The first program and corresponding speech module were installed in default mode and configured with a patient family and a provider. Each system was used in its default installation configuration, that is, the speech interface for entering intraoral findings were not customized in any way in order to avoid "tuning" the program in preparation for entering the simulated patient. An image of the machine was then made with Ghost and the "first Program" image was stored. This was repeated for both systems. The result was three images, one of the machine at baseline, one for program one and one for program two. Before each user test the fresh installation image for the program being tested was restored. Using fresh images each time eliminated factors such as other user's voice profile existing, or other user's task results interfering with the next test.

Participants included three undergraduate students and one faculty member who work in our Center for Dental Informatics, at the School of Dental Medicine, University of Pittsburgh. The only criterion to participate in the study was a lack of experience with any of the speech features of the clinical charting interface of the PMS.

Four feasibility user tests were conducted, each user tested one of the two systems. Two users tested System One and two users tested System Two.

All programs that were tested required users to learn the speech aspect of the system via a brief training session. Each user was to test one of the two systems; therefore they only had to complete the one training session for their assigned program. The training sessions for the systems had minor differences, but generally, they were each approximately 20 minutes in length and required the user to read pre-determined words and sentences that appeared on the screen. Each user was supervised during his/her training session to assist with problems and questions. Assisting the participant during training ensured that the head-set microphone was adjusted properly and that the user was speaking optimally for the task. Successful completion of the training session was required to take part in the task evaluation.

To start the session, a background questionnaire was administered to each participant. The questionnaire was a modified version of a validated tool that measures dental students' use of, knowledge about, and attitudes towards computers [9]. In future user testing studies, the questionnaire will be used to determine any variance of the results based on users' age, sex, native language (English or non-English), prior computer experience, and affinity towards computers.

Next, each participant was randomly assigned to test one of the programs and asked to complete the required training for that program as described above. The participant was then asked to read the script to chart the 18 different hard and soft tissue findings via speech. Each participant was given the individualized script for the software program they were testing and asked to read the script verbatim. During the task completion, if the system's response resulted in being off script (e.g. the chart exited), the observer interrupted the participant, corrected the problem, and had the participant begin again either where they left off (if possible with out redoing steps) or start on the next finding. If the system did not respond at all, the participant was asked to repeat the command two more times (a total of three) and then asked to either move on to the next command, or if that was not possible, move on to the next finding (e.g. if the system does not select a tooth, the user will move to the next finding). If the system charted an error that did not result in being significantly off script (e.g. selecting the wrong tooth), the participant was asked to ignore the error and continue. When the task was complete, the user was asked to turn off the microphone.

During each session, two observers took hand-written notes and, to supplement data collection, the entire session for each participant was video recorded to capture the screen, including mouse clicks and audio.

Following completion of the script, each participant was asked to complete a user satisfaction questionnaire. The questionnaire contained 27 items answered via a 7-point Likert scale, and was based on the Subjective Assessment of Speech System Interfaces (SASSI) project [10]. The validated questionnaire by Hone and Graham is broken down into six main factors which can help predict a user's satisfaction with speech-based systems: system response accuracy, likeability, cognitive demand, annoyance, habitability, and speed [10].

To determine the efficiency and effectiveness of the systems, the following items were calculated: Time to complete the training and the time to complete the script (adjusted for off script actions). To evaluate effectiveness, all errors were recorded, including if there were any findings that were not charted because of detrimental errors (e.g. the chart closes). Errors included a detrimental error, which was defined as a time when the system's response resulted in the participant being completely off script. A repeated command error was recorded if the system did not respond and the participant had to repeat the command. In that case, each repeat was documented. A record was kept of which word was repeated. A wrong response error was documented if the system's response differed from the participant's input. Also, in this case, a record was kept out the system's response (e.g., did it choose another tooth, did it select distal instead of mesial, etc.). Lastly, if the system responded but the participant did not say anything, this was documented as an insertion error. Again, in this case, a record was kept about the system's response. To determine overall user satisfaction, the mean score for the user satisfaction questionnaires was calculated.

Results

Features and functions comparison

Table 1 shows which features and functions in each of the four systems could be completed via voice. Systems One and Two used Microsoft Speech Recognition Engine (Microsoft, Redmond, Wash.), whereas Systems Three and Four used the default speech engine installed on the computer as long as it had SAPI 4.0 or 5.0 program files. SAPI stands for Speech Application Programming Interface (Microsoft, Redmond, Wash). System One is the only program that allowed free text dictation into a "clinical notes" area, and this was done via a Dragon NaturallySpeaking Engine (Nuance Communications, Burlington, MA). The training sessions for all of the systems were documented to take approximately 5-10 minutes, and to use the free text dictation of System One, an extra 30 minutes of training was necessary. All of the systems had training sessions similar to the Microsoft Speech Recognition training. All of the systems allowed a user to complete extra training if necessary, and Systems Two and Three allowed the user to train with specific dental terms. Next, none of the programs allowed for naturally spoken text, i.e. all required specific speech commands; and the number of possible speech commands for each program were approximately 573 for System One, 140 for System Two, 41 for System Three, and 53 for System Four. Only System One allowed the use of the international communications alphabet (alpha, bravo) to assist with speech related interactions. All systems had the ability to provide audio confirmation of a given command, but only Systems Two and Three gave complete visual confirmation of commands. Systems One and Four did not provide visual conformation for some actions.

Table 1 - Functions that can be completed via voice

	Systems			
	One	**Two**	**Three**	**Four**
hard tissue charting	Yes	Some	No	No
periodontal charting	Yes	Yes	Yes	Some
dictate raw clinical notes	Yes	No	No	No
chart existing and proposed findings	Yes	Yes	No	Some
select tooth surface	Yes	Some	Yes	Some
select patient	Yes	Some	No	No
open chart	Yes	Yes	Some	No
select items from list via name shown	Yes	Some	No	No
navigate through chart ("next", "move down two", etc.)	Yes	Yes	Some	Some
use all displayed options and buttons	Yes	Some	Some	Some
access menus, buttons, pop-ups, and checkboxes	Yes	Yes	Some	Some
undo last command	Yes	Yes	Yes	Some
clear/delete entries	Yes	Yes	Some	Some
start and stop listening	Yes	Yes	Yes	Yes

Feasibility user tests

Initializing the computer with the images of each system and configuring each system with default settings and a generic patient family and provider took multiple attempts and a lot of time to perfect. Also, after the first user-test, it was determined that the format of the script - how the information was presented to the participant, as well as the general instructions - affected how well the participant could successfully complete the task. It was decided that because these were feasibility user tests, small aspects of the test would be changed after each test to eventually discover the optimum way to conduct the user tests in the future.

While user tests were done on only two systems, the scripts were for all four. These scripts which documented exactly how many and what steps were necessary to complete the task in each program, were in themselves major

findings for our study; see Table 2 for a comparison of the scripts. The total number of commands in each script were as follows: System One, 114; System Two, 92; System Three, 42; and System Four, 47; Systems Three and Four were only able to chart the periodontal findings via voice. Therefore, Systems One and Two had more than twice as many commands because they charted the hard tissue findings as well as the periodontal findings. In three of the scripts (Two, Three and Four) it was necessary for the user to utilize the mouse or keyboard at some point during the task completion, i.e. the charting could not be done via voice alone. It is also important to note that even though Systems Three and Four had no voice functionality for hard tissue charting, some hard tissue charting (charting missing teeth) was necessary to complete the periodontal part of the task.

Table 2 - Comparison of commands necessary to complete the charting task as documented in the scripts

	Systems			
	One	**Two**	**Three**	**Four**
total number of commands in script	114	92	42	47
total number of voice commands in script	69 (H) 45 (P)	41 (H) 45 (P)	0 (H) 38 (P)	0 (H) 39 (P)
total number of mouse/keyboard commands in script	0 (H) 0 (P)	5 (H) 1 (P)	4 (H) 0 (P)	3 (H) 5 (P)
percent completed with voice alone	100	93	90	82

(H) – Hard Tissue Charting, (P) – Periodontal Charting

The scripts demonstrate that in all systems, the commands are very specific and are directly comparable to using a mouse. For example, as opposed to being able to say "existing ML composite on tooth nine", the following had to be said (the quotation marks indicate that the command is spoken to the system): "select 9", "restorative", "move down 1", "ok", "mesial", "Lingual", "ok", "existing"(excerpt from script for System One).

Four user tests were conducted, two each on Systems One and Two. The first user tested System One, and because it was the first test, there were problems with the script format and the instructions, and there were many errors during the test. Therefore, changes for the next test were made and the data from the first user were discarded. The final results are based on three user tests, one with System One and two with System Two.

For System One, it took the user 11 minutes and 8 seconds to complete the training and 5 minutes and 20 seconds (adjusted) to complete the charting task (script). There were eleven repeated commands, the most frequent being the word "ok" which was repeated seven times. There was one detrimental error in which the system exited when the command was "existing". There were three wrong response errors, an example of one being when the system selected tooth 30 when the command was "select 3". Lastly there were three insertion errors during the pocket depth charting; numbers were inserted that were not said in a command.

For System Two, which two users tested, the average time to complete training was 9 minutes and 1 second and the average time (adjusted) to complete the script was 9 minutes and 13 seconds. There were an average of sixteen repeated commands, the most frequent being the word "ok", which needed to be repeated on average five times. There was one detrimental error in each test. For example, the user said "quick pick menu 11", and the system opened the patient history form. There were two wrong responses in each test, for example when the user said "3" for a pocket depth, the system thought it was selecting tooth three. There were no insertion errors in either test. The video recordings were used to verify the observed results.

As the user satisfaction questionnaire was based on a 7 point Likert scale, with the most positive answers being scored as a seven and the most negative answers being scored as a one the best possible score a system could receive is 189. The satisfaction score for System One (one user) was 104 and the average score for System Two was 77 (individual scores of 27 and 127).

Discussion

The results from this study show that current clinical software systems for dentists are attempting to accommodate speech recognition as a means of interaction, but the current systems have many limitations which may hinder their use. As shown in the feature comparison, speech functionality varies across all systems, with two of the systems not having the ability to complete hard tissue charting via voice, which was more than half of the common charting task. The fact that scripts had to be developed to conduct the user tests shows that these systems are not designed to be used without prior understanding of the software and the memorizing of or easy access to an enormous amount of specific terminology. The requirement of prior knowledge of the system is understandable due to the complexity of these programs; however if the commands and interaction with the system could be done with a more natural vocabulary, ease of use would be significantly improved. The scripts also show that three of the systems required at least one keyboard or mouse command to complete the charting task, which defeats the purpose of using speech recognition as a hands-free way to interact with the computer. If a clinician were to use one of these systems, she would still have to de-glove to interact with the computer, and the keyboard and mouse would still need to be easily accessible. This limited speech functionality has the potential to reduce the ability of clinicians to interact directly with the computer during clinical care, and may be the main reason 16% of dentists that tried using the technology later discontinued [4].

The feasibility user tests were only conducted with three participants; therefore it is impossible to make any generalizations or comparisons with the results. However, it is

clear that data can be collected in this manner for performance comparisons across the software packages. The average number of detrimental errors per system and the most frequently repeated words per system may provide insight into issues that can be resolved in future designs of dental speech systems. More training in specific dental vocabulary or more intuitive speech engines may be design suggestions resulting from future user tests.

Full scale usability testing will be conducted in the near future; however based on the feasibility tests, some minor changes in methodology will be made. For example, the individual charting tasks on each script will be grouped into findings for easier reading and a more straightforward task flow. Better instructions will be given to the users. Originally, because the participant was to read a script verbatim, very little instruction was given. After just one test, it was apparent more instructions and explanations were necessary. Next, the user satisfaction questionnaire will be modified. After reviewing the range of results from the questionnaire, it was discovered that having a script read verbatim is not conducive to enabling the participants to answer the questions asked on the validated satisfaction questionnaire. For the future tests, a new open-ended satisfaction survey will be develop to better asses user satisfaction based on the given task. Lastly, conducting and optimizing the user tests exposed the researchers to more speech related features and functionality of the systems that may have been overlooked during the first part of this study. Hence, more detail will be added to the feature and functionality comparison list after the completion of multiple user tests.

There are certainly limitations to this study. By using a verbatim script task, the dynamics of classic user testing are changed. More attention must be given to what is actually being compared across systems during this altered user test. Even though user tests will give a good indication of efficiency across all systems (i.e., which system completes the task in the least amount of time) unless these times are compared to charting the same information via a keyboard and mouse and ultimately to hand-written charting, we can never make accurate conclusions about efficacy compared to other data entry/retrieval methods.

Conclusions

This study shows that clinical speech functionality in current dental systems is somewhat cumbersome and poorly designed. This limited speech functionality has the potential to reduce the ability of clinicians to interact directly with the computer during clinical care. In the future, dentistry will see the influx and be able to reap the benefits of decision support tools and shared electronic medical records [11]. However, unless better speech functionality is implemented, the benefits and effectiveness of any electronic patient records and clinical decision support systems may be greatly impeded.

Acknowledgments

We gratefully acknowledge Thankam Thyvalikakath and Miguel Humberto Torres-Urquidy for their input on this project. We would also like to thank the NLM and NIDCR for their funding support 5T15LM007059-179.

References

[1] White K. Speech Recognition Implementation in Radiology. Pediatric Radiology 2005: 35(3): 841-846.

[2] Zafar A, Overhage M, McDonald CJ. Continuous speech recognition for clinicians. *J Am Med Inform Assoc* 1999: 6(3):195 -204.

[3] Grasso M. The long-term adoption of speech recognition in medical applications. In: Proceedings of the 16th IEEE Symposium on Computer-Based Medical Systems; 2003: p. 257-62.

[4] Schleyer T, Thyvalikakath T, Spallek H, Torres-Urquidy M, Hernanadez P, Yuhaniak J. Clinical computing in general dentistry. *JAMIA* 2006: 13(3): 344-352.

[5] Drevenstedt G, McDonald J, Drevenstedt L. The role of speech-activated technology in today's dental practice. *JADA* 2005: 136(2): 157-161.

[6] Nielsen J. Usability Engineering. San Francisco, CA: Morgan Kaufmann; 1994.

[7] Tomko, S. Speech Graffiti: Assessing the User Experience. CMU LTI Tech Report CMU-LTI-04-185 (2004). www.cs.cmu.edu/~stef/papers/mthesis.ps

[8] Harris TK and Rosenfeld RA. A universal speech interface for appliances. In: Proceedings of the 8th International Conference on Spoken Language Processing (ICSLP'04). Jeju Island, South Korea; 2004.

[9] Schleyer TK, Torres-Urquidy H, Straja S. Validation of an instrument to measure dental students' use of, knowledge about, and attitudes towards computers. *J Dent Educ*. 2001: 65(9): 883-91

[10] Hone KS, and Graham R. Towards a tool for the subjective assessment of speech system interfaces (SASSI). Natural Language Engineering 2000: 6(3-4): 287-303.

[11] Schleyer TK. Why integration is key for dental office technology. *J Am Dent Assoc*. 2004: 135(Suppl): 4S–9S.

Address for correspondence

Jeannie Yuhaniak, Department of Biomedical Informatics, Center for Dental Informatics, University of Pittsburgh, 333 Salk Hall, 3501 Terrace Street, Pittsburgh, PA 15261, USA. jeannie@dental.pitt.edu

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Assessing the Impact of Recording Quality Target Data on the GP Consultation Using Multi-Channel Video

Maigaelle V. Moulene[a], Simon de Lusignan[a], George Freeman[a], Jeremy van Vlymen[a], Ian Sheeler[a] Andrew Singleton[a], Pushpa Kumarapeli[a]

[a] *Primary Care Informatics, Division of Community Health Sciences, St. George' - University of London, United Kingdom*

Abstract

Background: In the UK routinely collected computerized clinical data is used to assess progress towards financially incentivised quality targets for chronic disease management including hypertension. Objective: To develop a method for assessing the impact of recording quality target data in the clinical consultation. Methods: Raters were trained how to rate a multi-channel video of a simulated clinical consultation for interaction between actors, computer use, non-verbal communication. Results: 25% of consultation time is computer use and a median of 4 to 5 items were coded per consultation mainly items related to the hypertension quality target. Intraclass correlation coefficient showed good inter-rater reliability (>0.9; p<0.001). Conclusion: We have successfully piloted a novel technique for observing the influence of the computer on the consultation. Despite increasing computer use to record quality target data the over whelming proportion of the consultation remains doctor patient interaction.

Keywords:

video recording, consultations, computer, primary care, observation, medical records system, computerized.

Introduction

Chronic disease management is an increasing priority for general practice and in the UK financially incentivised quality targets have been introduced to raise the standard of chronic disease management [1]. As the population ages, people live longer but with an increasing burden of chronic disease such as hypertension. In the UK the Quality and Outcomes Framework (QOF) provides standards for the management of chronic disease. For example, in hypertension, blood pressure must be measured every 9 months and, ideally, kept below a target of 150/90. Performance of practices against these targets is measured using routinely collected computer data recorded as part of normal consultations. Reports are automatically generated from within the GP computer system based on the number of people with a diagnostic code for any of the chronic diseases and whether their disease management achieves the necessary target. Financially rewarded quality points are awarded based on achieving the target moderated by the prevalence of the condition in the practice compared with the national average. Management of hypertension attracts 105 points, each point worth £120 [177 Euros], to a practice of 10,000 patients, a total in excess of 18,000 Euros for each year the quality target is achieved. There is, therefore, considerable financial incentive for chronic disease data to be "coded," i.e. recorded as structured data, within the clinical consultation.

Video observation, using a single camera, is a well established method of assessing clinical competence in GP. Yet, there are still many limitations to single channel video, it is hard to interpret body language of the consulter or patient and it lacks information on how the computer is being used [2]. We have developed multi-channel methods to overcome the limitations of single channel video [3]. Three channel video uses several cameras capturing more detailed views of the doctor and the patient and includes a recording of how the computer is being used in the consultation, to allow computer analysis of the consultation and more in-depth analysis of the patient - doctor interaction. However, the shortcomings of this method were: the time taken and subjective nature of evaluation of the consultation; the difficulty in interpreting the patient's body language; and the high cost of making professional standard videos because analogue video at that time did not allow precise time sequence mapping. We have overcome the problems of cost using modern budget digital cameras; the addition of a fourth video channel looking at the patients body language; and, although we have made progress in using pattern recognition software to make automated records of the consultation this requires much more development if it is to record the subtleties of the clinical consultation [4]. We therefore used the four-channel refinement of our multi-channel video method [5] to analyse the time spent using the computer and how this might impact on the consultation; using simulated consultations for hypertension.

Methods

We carried out a literature review using standard bibliographic data bases restricting our search to articles published after 1990. We visited four GP practices to identify how clinical data are currently entered into the four main computer systems: EMIS PCS, EMIS LV, IPS Vision and iSoft Synergy. These four systems account for over

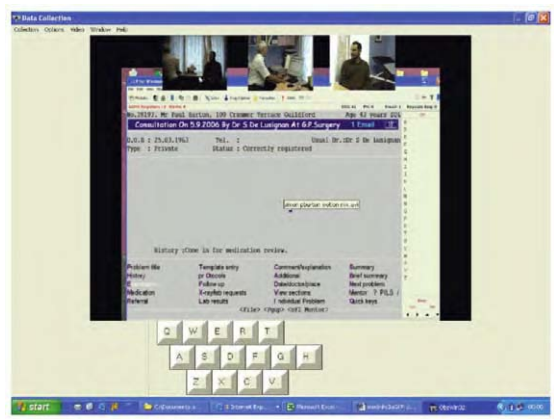
90% of GP clinical computer systems used in England. All these brands of GP computer system use the 5-byte version of the Read codes to record structured data. They also all have an integrated report tool which automatically reports anonymised data about their progress towards achieving quality targets. We spoke to experienced practitioners and discovered how chronic disease management data were entered into each system. We asked in detail about how the financially incentivised quality target (QOF) data are entered into each system and collected screen shots of each step in the process. We used the latter to train our raters how to recognize coding and recording of QOF target data.

We filmed simulated consultations with three of the GP computer systems: EMIS LV, EMIS PCS and IPS Vision in a real or simulated setting depending on the availability of the clinical computer system at our institution. We filmed three experienced GPs, all with over 15 years experience and one trainee GP. The GPs consulted the computer system they used in their clinical practice, limiting our choice to three out of the four systems initially observed; two of the GPs consulted using EMIS LV. The patients were simulated by members of staff experienced in role playing patients for clinical exams. We filmed a series of consultations to allow the GP to become familiar with the setup; however, we only used the consultation about hypertension for this analysis. Each clinician was presented with the same simulated patient for a review of their hypertension; their past history was identically loaded in each computer system.

We used three standard video cameras (Sony DCR HC45E DV) to record the consultation. One video camera recorded the doctor's head and upper body (to capture the clinicians direction of gaze and body language), the second the patient's upper body (we find that capturing the patients hands is important in interpreting their body language) and the third is an overview of the whole consultation (but excluding the examination couch). The fourth video channel was recorded using Camtasia screen capture software which records the computer screen and data entered into the clinical computer system. We mixed the separate video channels, so they could be displayed simultaneously, in-house. The separate video recordings were transferred from the mini DV cameras into Final Cut Studio editing software running on a G5 Mac to do this. The first step after transfer is to synchronise the images and audio feeds for each consultation. Next we created a layered video composition in Final Cut's compositing application (Motion). This composition displayed all the GP's computer usage and the camera angles in one screen. Finally this composition was rendered into AVI (Audio Video Interleave) format and recorded on to DVD.

We used ObsWin to analyse the consultations [6]. It integrates video files and uses keys on a keyboard where each key is a variable and acts as a 'stopwatch'. The raters had to watch the consultation video three times. We set up ObsWin so that one row of keys was used on each successive observation of the consultation (Fig 1). The first viewing used the lower row of keys to explore the interaction between the actors (including the computer): computer-doctor interaction (**Z,**), doctor-computer interaction (**X,**), doctor-patient interaction (**C**), patient-doctor interaction (**V**). The second run through used the second row of keys and measured data entry times: referrals (**A**), clinical coding excluding QOF target codes (**S**), quality target (QOF) coding (**D**), free text (**F**), prescribing (**G**), new prescribing needed to achieve a QOF target (**H**). The final screening used the top row of keys to measure body language occurring in the consultation: eye contact (**Q**), doctor looking at computer while speaking to patient (**W**), examination (**E**), patient speaking to doctor while doctor looking at computer screen (**R**) and Silent Time(**T**). Eye contact (**Q**) and silent time (**T**) are important non-verbal cues the former establishes rapport, shows the patient the doctor is engaged and trying to understand patient perspective, the latter also provides an important non-verbal cue [7]. We measured the time when the doctor was using the computer or looking at the computer while speaking (**W**) and when the patient speaks (**R**) to detect interference in the consultation. We could record examination (**E**) of blood pressure, but the examination couch was not under observation by our cameras thus other examinations (**E**) were not recorded and our simulated patients would have declined such examination. We derived information about the impact of coding from a combination of these variables. **C**+**V** represented the total verbal communication; **S**+**D** was the time spent on coding entries.

Figure 1 – ObsWin with integrated multi-channel videos



The raters were trained using an instruction manual [8] and special training videos. The raters were six volunteer biomedical informatics students from the authors' institution. They, MVM and an additional postgraduate student rated the consultations. Four or five raters rated every consultation. All the rating was done using seconds, results are presented as minutes and seconds; i.e. min:sec. We used intraclass correlation coefficient to test inter-rater reliability using SPSS version 14.0 Reliability Analysis intraclass correlation program.

Results

Clinical data entry

All four computer systems (EMIS LV, EMIS PCS, IPS Vision, iSoft) had similar methods for coding data even though they look quite different [9]. In all the systems problem titles are either coded from a picking list, or reselected if an existing problem. They also all had some type of standard form to speed up coding of chronic disease management; although they had different names and slight differences in functionality e.g. Templates and protocols in EMIS; SOPHIES in the iSoft system. These forms prompted the clinician to record all the clinically relevant data. All the systems also had prompting screens identifying the data items missing from patients' records and needed for the quality targets. The systems also allow free-text to be recorded. The principal differences between the computer systems were in appearance and the density of text. EMIS LV had a single screen through which data was entered and had the fewest lines of text while IPS Vision has the most with five windows providing access to the clinical record [10].

General observations about the consultations

The four hypertension monitoring consultations took between 7 and 11 minutes; blood pressure measurement was common to all the consultations and issuing of a repeat prescription for antihypertensive medication took place in all but one. In one consultation no items were coded the only use of the computer was to issue a repeat prescription. All the other clinicians coded between four and seven items in the consultation: blood pressure measurements, problem title or diagnosis and smoking status were the commonest and all form part of the quality target data requirements. The times taken on each aspect of the consultation are shown in Table 1. No referrals were made

Table 1 – Rating consultations: four simulated consultations using EMIS PCS, EMIS LV, iSoft and Vision

Consultation characteristic	Key **	EMIS PCS		EMIS LV (1)		EMIS LV (2)		IPS Vision		Summary	
		Time	%	Time	%	Time	%	Time	%	Median	IQR
General observations											
Duration		6:55		11:26		10:10		8:36		**9:13**	
Coded entries		0		5		7		4		**4.5**	
Time/ code		0		0:19		0:18		0:19		**0:19**	
First view: ***Interaction between actors (C+V = total verbal communication)***											
Computer-Dr	**Z**	0:03	0%	0:06	0%	0:11	0%	0:05	0%	**0:06**	0:03
Dr-computer	**X**	0:51	12%	2:42	24%	4:04	40%	2:24	28%	**2:30**	1:03
Dr-Patient	**C**	3:19	47%	3:36	32%	6:18	61%	3:36	41%	**3:36**	0:39
Patient-Dr	**V**	2:35	37%	4:30	40%	2:48	27%	2:24	29%	**3:42**	0:36
Second view: ***Data entry times (S +D = Total coding time)***											
Coding	**S**	0	0%	0	0%	0:49	10%	0:11	2%	**0:30**	0:19
QOF-code	**D**	0	0%	1:36	14%	1:18	13%	1:06	13%	**1:18**	0:18
Free-text	**F**	0	0%	1:06	9%	0:19	3%	0:41	8%	**0:41**	0:23
Prescribing	**G**	0:48	11%	0	0%	0:54	8%	0:36	7%	**0:48**	0:05
Third view: ***Body language in consultation (Q+T), Examination (E) and computer interference (W+R)***											
Eye contact	**Q**	4:18	61%	6:00	53%	4:30	43%	3:12	37%	**4:24**	0:54
Dr Comp&speak	**W**	0:29	7%	0:17	2%	2:18	22%	0:46	9%	**0:38**	0:40
Examination	**E**	1:12	17%	1:42	15%	0:58	9%	1:36	19%	**1:24**	0:30
Silent time	**T**	1:08	16%	2:00	17%	0:29	4%	0:58	11%	**1:00**	0:27
Pt-Dr&Comp	**R**	0:03	0%	0:18	2%	0:43	7%	0:38	7%	**0:28**	0:26
Reliability test: ***Intraclass correlation coefficient (ICC)***											
ICC Value		0.962		0.926		0.931		0.896			
95% CI		0.895 – 0.991		0.833 – 0.978		0.854 – 0.977		0.783 – 0.962			

Key: **Letters in column 2 are the ObsWin keys used to record elements of the consultation. Time = minutes: seconds
Z = Computer-doctor interaction; X = Doctor-computer; C = Doctor-patient; V= Patient-doctor interaction
S = Coding; D = Quality target coding; F = Free-text entry time; G = Prescribing time using the computer
Q= Eye contact; W = Doctor using computer and speaking; R = Patient speaking to doctor, while doctor uses computer;
95% CI = 95% Confidence Intervals.

(**A**) in any of the consultations and no new medications were started specifically to achieve the quality targets (**H**) so these lines are not shown in the table. The total times on each activity can add up to more than 100% as more than one activity maybe going on simultaneously.

ObsWin analysis

The majority of the consultation time spent on doctor patient interaction and communication. The patient speaking to the doctor varied between 2:42 minutes to 4:30 minutes (26-40%) of the consultation and the doctor speaking to the patient varied from 1:54 min to 6:18 min (32% to 61%). In three consultations the doctor spoke more than the patient and in one consultation the patient spoke more than the doctor. The doctor and patient spoke to each other for 70 to 88% of the consultation

Computer data entry was divided between entering coded data, free text and prescribing; referral was not needed in this consultation. The shortest and longest times using the computer were 51 seconds (12%) and 4:04 minutes (40%) using the computer. The other two consulters spent 2:24 minutes (27%) and 2:42 minutes (21%) on the computer. Coding time varied from 1:17 minutes to 2:06 minutes. It took between 18 to 19 seconds entering each coded item. Only three out of 16 items coded were for items not required for the hypertension quality target. The three of four doctors who made a free text record entered data for a mean of 42 seconds (median 41, range 19 to 66 seconds.) They recorded between one and four lines of free text, typing from four to 60 characters per line of text. Repeat prescription issue time was between 30secs and one minute.

All the GPs spent between six and eight and a half minutes on the consultation aside from computer use. There was no relationship between time spent on the computer and time taken to complete the consultation. The consulter who spent most time on the computer (40%) entered 7 coded entries into the computer and wrote one line of free text. This consulter still spent 5:54 minutes on the rest of the consultation. Another consulter spent 2:24 minutes on the computer (27%) coding four entries and typing three lines of free text. The rest of the consultation lasted 6:12 minutes. The last consulter who spent 2:42minutes (21%) of the consultation on the computer coded five entries and typed four lines of free text. This consulter spent 8:36minutes on the rest of the consultation. Finally the consulter who only used the computer 59 seconds for prescribing did not code any entries but spent 6:09 minutes on the rest of the consultation.

Not using the computer in the consultation appeared to allow more time for non-verbal communication and more computer use was associated with the need to communicate during computer use. Three to six minutes of the consultation time was spent on eye contact. Silent time was generally between 29 seconds to two minutes of the consultation. The consulter who spent the most time on the computer, also spent the most time speaking to the patient while using the computer, 2:18 minutes, whereas the other consulters only spoke to the patient 17 to 46 seconds of the time while using the computer. This consulter had proportionally the least amount of eye contact with the patient throughout the consultation (37%), whereas the consulter who spent the least time on the computer had the most eye contact during the consultation (61%). Patients speaking to the doctor while the doctor used the computer varied; but generally if the GP used the computer more the patient spoke to them more during computer use.

Examination time, in this case taking the patient's blood pressure, lasted 58 seconds to 1:42 minutes (9-17% of the consultation).

The type of computer system used did not seem to make a difference in the time spent entering coded data, the time spent entering coded data was very similar. Although their interface with the clinician varies greatly, this appeared to make little difference to the clinician using them who knew how to navigate to the appropriate parts of the program.

Reliability

We found good inter-rater reliability using intraclass correlation coefficient. The intraclass correlation coefficient of the consultations based on four or five observers ranged from 0.962 (95% CI 0.895-0.991) to 0.896 (95% CI 0.854 – 0.977). All these observations are significant at the ($p<0.001$). Most of the raters assessed two or three consultations. These estimates suggest that this method has a high inter-rater reliability.

Discussion

This technique to demonstrate how the computer now occupies around a quarter of consulting time, that most of our consulting GPs are coding 4 to 5 items and that these are predominantly data items required for the relevant quality target. We found that patient doctor verbal communication still remains the majority of time spent on the consultation, although the total time spent on the computer varied considerably between consulters. The type of computer system used did not have an impact on how much time was spent on clinical or quality target coding, clinician factors appeared to be much more important.

We have successfully piloted a novel but reliable tool for the analysis of the clinical consultation. We have demonstrated that it is feasible to combine: a carefully developed training package; multi-channel video and the use of ObsWin to achieve reliable results largely using student observers. Although it took at least two hours to analyze one consultation this is much less time consuming and more reliable than using a stop-watch and on screen timers to calculate the time taken as we have used in previous investigations [3,4]. Clinicians who participated in this experiment found the detailed observation of their behavior provided valuable feedback.

If this method shows similar reliability when used by other teams, it could provide a method for comparing the impact of quality targets on the clinical consultation as well as a mechanism for comparing different GP computer systems. It would allow the time taken to retrieve information, time

taken in data entry and the extent to which the computer appears to interfere with the consulting process to be analyzed. With larger samples, statistically significant differences might be apparent between consultations, which did and did not include quality target data collection and varying computer systems, both overall and in relation to other specific elements of the consultation (e.g. time taken to prescribe or refer.) It is possible that structured feedback might have a positive influence on consultation outcomes.

There are research methods and technical limitations to this study. This is a pilot study using simulated consultations with only a small sample size and a limited number of raters. Hypertension is a common disease that GPs frequently encounter and the 18-19 second consultation time reflects their ability to manage it. The consultations lasted from 7 to 11 minutes. UK GPs generally have 10 minute booked consultations and the consulting GPs said they felt that these consultations reflected their usual practice. Showing reliability with a small number of consultations and student raters is both a strength and weakness of this study. The method should be reliable in a larger sample, but has not as yet been tested in a wider range of consultations. If the tool is used in a wider range of consultations it may require clinicians to rate the consultations as it may not be possible to train non-experts in how to recognize coding in a wide range of clinical contexts. Some variables used in the consultation were easy to measure while others were more difficult. Use of keyboard and speaking time were the easiest except for very brief episodes. The raters found inferring whether there was eye contact and whether the GP was being provided information by the computer the most difficult elements of the consultation to assess. We did not test two of the items in the scale as the relevant activity did not take place in the consultation. The rating scale might be improved by adding a specific item for use of on-line information items or decision support.

The coding time in our study (18-19 seconds) is consistent with that reported for using Read Codes. This is consistent with the one previous UK study, though about half the time reported from a study from USA. The UK study measured the average time spent on coding as 14-27 seconds when using Clinical Terms and 18-49 seconds using Read Codes [11]. Cimino et al. studied the coding of data in an outpatient setting using videotaping as well and found that data entry time averaged 40.4 seconds per item [12]. Our results maybe shorter because hypertension is a very common chronic disease which GPs are faced with on a regular basis and the GP computer systems provide special data entry forms for this condition.

Further research is needed to see if this technique is reliable in other clinical contexts – both different clinical consultations and other specialties to compare whether consultations which involve quality target data recording are different from those without and to explore differences between clinical computer systems.

Conclusion

This novel multi-channel video technique is a reliable method for measuring patient – doctor - computer interaction; time spent on clinical and quality target coding; and other activities during the GP consultation. Computer use varied widely depending on the consulter's computing style, but the average time spent on the computer in these simulated hypertension consultations was 25%. Whilst entering quality target data and clinical coding took 18-19 seconds for each coded data item and 4 to 5 items were entered per consultation, patient - doctor communication, including verbal and non-verbal communication, still forms the overwhelming majority of the GP consultation time.

Acknowledgments

Informatics students for rating the consultations; Nigel Hague for clinical coding advice; Toby Fallon and Eben Wilson for technical help and Vickie Wood for reviewing the manuscript.

References

[1] Roland M. Linking physicians pay to quality of care a major experiment in the UK. N Engl J Med 2004; 351(14): 1448-54.

[2] Coleman T. Using video-recorded consultations for research in primary care: advantages and limitations. Fam Pract. 2000;17(5):422-7.

[3] Theadom A, de Lusignan S, Wilson E, Chan C. Using three-channel video to evaluate the impact of the use of the computer on the patient-centredness of the general practice consultation. Informatics in Primary Care 2003 11: 149-56

[4] Sheeler I, Koczan P, Wallage W, de Lusignan S. Low cost method for three-channel video assessment of the consultation. Informatics in Primary Care *(in press)* 2006

[5] Leong A, Koczan P, de Lusignan S, Sheeler I. A framework for comparing video methods used to assess the clinical consultation: a qualitative study. Med Inform Internet Med. 2006;31(4):255-65.

[6] Martin, N., Oliver, C., & Hall, S. ObsWin: Observational Data Collection & Analysis for Windows. London: Antam 2003.

[7] Byrne PS, Long BEL. Doctors Talking to Patients. Exeter: The Royal College of General Practitioners, 1984.

[8] Primary Care Informatics, St George's. Instruction Manual – How to measure clinical coding during the GP consultation. http://www.gpinformatics.org/reference/GPcoding.htm

[9] Primary Care Informatics, St George's. Recognising structured data and Screenshots of EMIS PCS, EMIS LV, iSoft, IPS Vision. http://www.gpinformatics.org/reference/data_recording.htm

[10] Jordan K, Porcheret M, Croft P. Quality coding in general practice computerized medical records: a systematic review. Family Practice 2004;4:396-412.

[11] Brown P, Warmington V, Laurence M, Prevost AT. Randomised crossover trial comparing the performance of Clinical Terms Version 3 and Read Codes 5 byte set coding schemes in general practice. BMJ 2003; 326:1127

[12] Cimino JJ, Patel VL, Kushniruk AW. Studying the human-computer-terminology interface. J Am Med Inform Assoc 2001;8:163-73.

Address for correspondence

Simon de Lusignan, Community Health Sciences, Hunter Wing, St. George's - University of London, London, SW17 0RE, UK. slusigna@sgul.ac.uk

Chapter 8.

Sustainability

This page intentionally left blank

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Sustainable Health Systems: Addressing Three Key Areas

Prajesh N. Chhanabhai, Alec Holt, George Benwell

Department of Information Science, University of Otago, New Zealand

Abstract

In the modern context sustainable health systems are being developed using the newest technological and communication technologies. This is proving to be a great success for the growth of Health Informatics and healthcare improvement. However this revolution is not being reached by a lot of the world population. This paper will address the importance of closing the Digital Divide, Empowerment of health consumers and the importance of converging communications. Key areas in the development of a truly sustainable health system

Keywords:

digital divide, empowerment, internet use, developing countries., communication technologies

Introduction

Health and Information Technology is an area that is recognised globally as being the next technological revolution. This is being helped by new technologies that have been devised to facilitate this transition e.g. Healthphone [1]. However this growth is only being appreciated by those that can access, afford and use it. There is a large population group that this Technological revolution is still leaving out. It is imperative to understand these before a country can build any truly sustainable health system.

This paper will highlight the issues that are faced by developing countries in Africa and may indicate that simply introducing an electronic medium will not solve problems related to the dissemination of health information. This paper advocates introducing technology but also paying attention to its effect and the importance in understanding cultural paradigms when it is introduced.

Background

The use of technology is growing in all areas of health communication, including consumer, and healthcare provider education, interaction as well as in the areas of decision and social support, health promotion, knowledge transfer and the delivery of services. This has lead to the development of a number of new fields such as captology the study of computers as persuasive tools and eHealth the integration of health care delivery and information delivery through computer based technologies[2]. The development of such areas is an indication of the immense potential for health communication efforts on a global scale [2].

Despite this development, the adaptation and integration of information technology in the health sector is unfolding at a slower rate than their counterparts in the finance and commerce sectors [2]. Currently most developments that include the health consumer are run by for-profit eHealth companies. These companies currently utilise the Internet and web related technologies to run their organisations. The most common focus of these organisations is to provide tools, solutions, products or services that aid some aspect of clinical care or eCommerce. However, as most of these organisations are for-profit, their impact on areas and communities that are economically challenged is minimal or non-existent. As it currently stands these facilities can only be accessed by those that have access to the Internet and the ability to pay for such services, hence segregating a large number of the population in many countries and in a majority of the cases this is the population that needs the greatest need for eHealth services [3].

According to Kreps many of the people who are at most risk from serious health conditions come from underserved populations, populations that are generally made up of individuals who are of low socioeconomic status, possess low level of health literacy and are members of marginalised ethnic and minority groups [4]. These underserved and vulnerable populations often have limited access to relevant health information especially information that is otherwise easily available over the Internet. This is one of the symptoms of the Digital Divide, however within the health sector the Digital Divide as a more specialised problem. Many of the characteristics that identify those on the have not side of the Digital Divide also apply to those who suffer from the negative effects of health disparities. While information and knowledge are not guarantors of good health care decisions and adherence to recommended health behaviour, their ease of availability has shown to contribute to them [5][6]. This has been recognised by the White House, who in their *Healthy People 2010* report indicated that health communication through the use of computer technologies is a means of bridging the digital health divide [7].

Table 1 - World Internet Usage and Population Statistics. Highlighting the impact of the Digital Divide in Africa. Adapted from Internet Usage Statistics The Big Picture [9]

World Regions	Population (2006 Est.)	Population % of World	Internet Usage, Latest Data	% Population (Penetration)	Usage% of World
Asia	3,667,774,066	56.4 %	394,872,213	10.8 %	36.4 %
Europe	807,289,020	12.4 %	308,712,903	38.2 %	28.4 %
North America	331,473,276	5.1 %	229,138,706	69.1 %	21.1 %
Latin America/Caribbean	553,908,632	8.5 %	83,368,209	15.1 %	7.7 %
Africa	915,210,928	14.1 %	32,765,700	3.6 %	3.0 %
Middle East	190,084,161	2.9 %	19,028,400	10.0 %	1.8 %
Oceania / Australia	33,956,977	0.5 %	18,364,772	54.1 %	1.7 %

Digital divide

Information is critical for the development of any system. This is of crucial importance in health systems, in which the transfer of correct information maybe the difference between curing a patient and killing a patient. Information flow can be regarded as one of the most important factors for improving health systems especially in under-developed resource poor countries. The importance is even greater than worrying about the availability of infrastructure, the increasing of health workers and the distribution of funds, because without the correct information flow none of these other factors can function effectively or efficiently, and in most cases even in fthe correct settings. Access to information is essential in developing countries as it allows health policy makers in these countries to understand their current deficiencies thus enabling them to develop practical ways in which to solve these deficiencies. Information can thus be seen as a form of empowerment however this empowerment can only be useful if it comes from correct information, this is where the dissemination of information is important [8].

Advances in information and communication technologies have inherently made the distribution of health information globally seem effortless. The technologies in particular the Internet, allows information to be made available the instant it is produced. The information that is transmitted by these technologies can allow users to choose the exact type of information they want and can view the information in a number of different mediums. This information can then be accessed by a multiple range of users, be it a farmer in rural Africa or a health policy government official. However, this is a utopian idea as it is rare for a farmer in rural Africa to have access to the Internet.

The information gap between the developed world and developing countries is currently widening, this is aided by the growth of the Digital Divide. The impact of the Digital Divide is so large that the [10][8][11] have all commented that it is more dramatic than any other inequity in health or income. This fact is frightening as the development of information and communication technologies was prophesized has being one of the solutions for these other inequities. The extent of the Digital divide can be gathered from the following statistics. Africa has a population of over 900 million which makes up 14.1% of the world population. Only 32 million of the entire population have access to the Internet, with a penetration percentage of 3.6%. Egypt, Morocco, Nigeria and South Africa make up 60% of the total Internet users in Africa. Table 1 shows the global Internet use by region. The table highlights that despite having the second largest percentage of the worlds population, the Internet use in Africa can be regarded as being the lowest per head. According to the United Nations Development Program There are more Internet hosts in New York than in continental Africa more hosts in Finland than in Latin America and the Caribbean [12]

These statistics show that the digital divide is growing larger, and in the health context this is contrary to the aim of medical research that is being produced. There is now numerous research into how the Internet and such pervasive communication technologies will play an important role in improving health outcomes. There is also the concern that if this technology is introduced or made available in developing countries that it will not be used [13]. Norman and Skinner have reported in their paper that in the United States of America and Canada alone over 40% of adults have low basic literacy levels, thus e-health resources are likely to be inaccessible to large segments of the population [14].

The question that then begs to be answered is if the technology is available in such countries and is not being used then how effective will it be in places like Africa? Using information technology does not mean just placing the infrastructure and expecting the health consumer to buy into it. Using Information technology in the ehealth context requires the user to have a certain level of e-health

literacy, the ability to read, use computers, search for information and to understand the information in context [14]. In addition to these fundamental factors there is also the need to address the cultural dimensions of change. In developing countries like Africa, more than any other place, there are very strong ties to the traditional beliefs and these need to be factored into the development of a good health information dissemination system.

Empowerment

Traditionally the healthcare consumer has been the least consumer like and the least informed [15]. Protection from social stigma as well as the feeling that patients would get more sick once they knew their medical condition were the reasons used by many traditional physicians to discourage empowerment With expanding populations and the increase in occurrences of epidemics, medicine has become more scientific and thus medical knowledge has started to become available to the lay public [15]. In the last twenty years the emphasis has changed from cure of health conditions to prevention, with an emphasis on health and wellness [16]. According to Amatayakul *Patients have become interested in making **choices** for themselves about their physicians, treatments and lifestyles* [15], this can be clearly observed in the change of terminology from medical care to healthcare. The term medical care focused primarily on processes administered by a physician, whereas healthcare encompasses a broader range of services and procedures [17].

The emergence of communication technologies and the incentives in the health sector to include consumers in their operations are some of the factors in increasing the importance of the consumer in the healthcare setting [17]. The biggest factor in cementing the role of the patient as a consumer is through the growth and the innovative capabilities of technology. The increasing availability of interactive information has enabled many services to be made available online.

boards and bulletin boards has allowed individuals to share experiences of specific diseases and treatments. This has introduced another dimension to the healthcare industry where consumers are more knowledgeable and understanding of the terminology and procedures that are used in the health sector.

According to Eysenbach, initially the technology had been looking at development and growth through the eyes of the medical professional, with the drive towards consumerism[18]. This has changed and has seen the birth of consumer health informatics. Consumer health informatics is defined by Eysenbach as *the branch of medical informatics that analyses consumers' needs for information; studies and implements methods of making information accessible to consumers; and models and integrates consumers' preferences into medical information systems*. [18]. This definition agrees with Amatayakuls statement that the *principle of consumer health informatics is that of empowering individuals to play a greater role in their own healthcare and to be active participants in the decisions that affect their healthcare* [15]. Figure 1 shows how consumer health informatics has shifted from individuals just having a choice to being empowered. Essentially, all health information can now be used to benefit the consumer of healthcare.

Empowerment will only occur if the consumers themselves are allowed to interact with the healthcare system. An interaction in which they do not receive just limited feedback, but one that promotes two way feedback, benefiting all parties. This idea has been adopted by the European Union, who realised that the greatest interaction would occur if patients had access to their records [18]. In October 1998 the European Union required that each of their member countries passed legislation that would ensure consumers in those countries to have access to all their health records [18].

In the developing world context, this empowerment will be a much harder process to implement due to the nature of lifestyles and beliefs in such countries. As was mentioned earlier it is the shift in cultural paradigms that needs to be addressed. Empowerment does not have to occur with only the use of technology. The important factor is getting the information out in a manner that is both timely and comprehensible to the lay public.

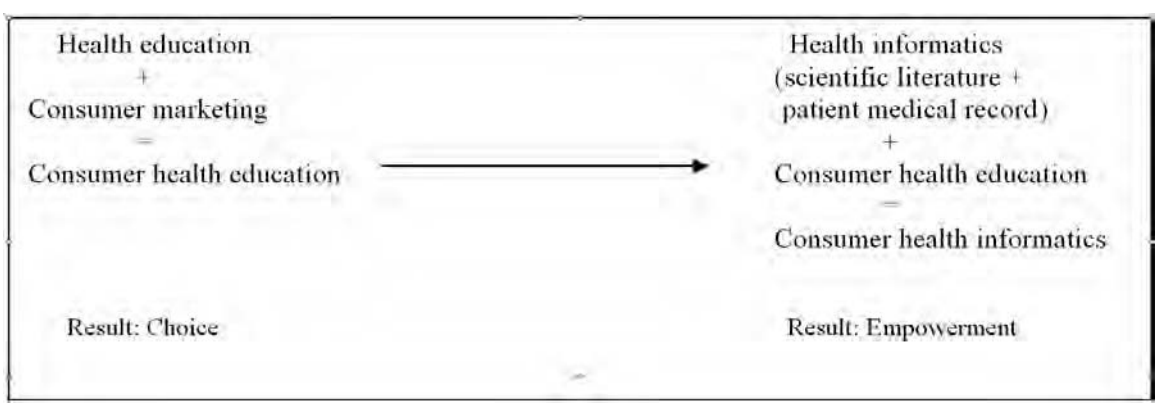


Figure 1 - The shift from an individual having a choice to being empowered. Adapted from Amatayakul [15]

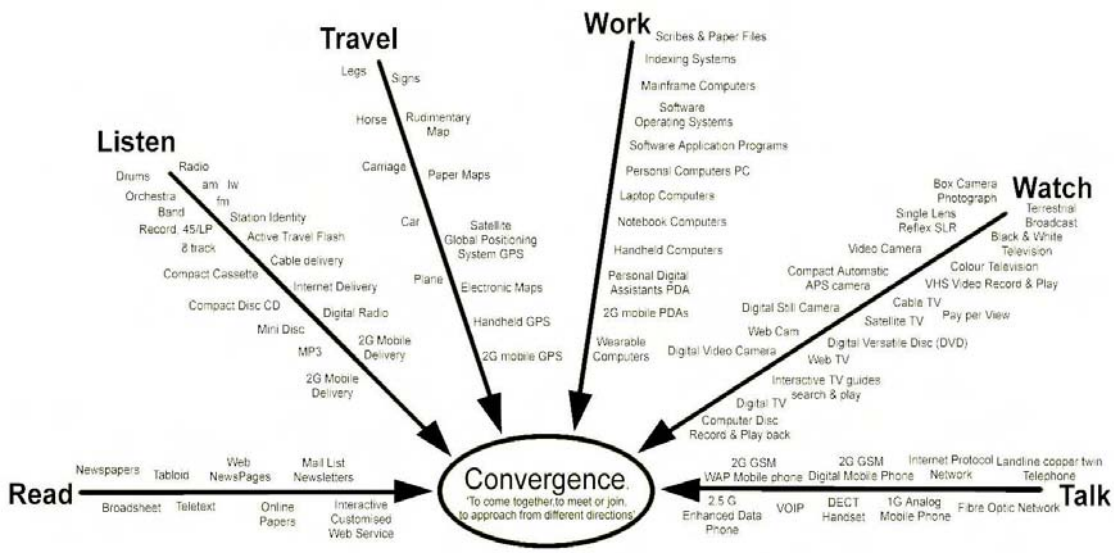


Figure 2 - Diagrammatic representation of the convergence of communications

A good example of this is the difference between the HIV/AIDS policy in Uganda and Botswana. Both countries are known to have some of the higher HIV/AIDS infection rates in the world. Both countries established different schemes to bring down the infection rates. The success rates in both countries have differed considerably.

The main documented reason was due to the condom promotion. According to Allen and Heald, in Botswana condom promotion provoked antipathy from church groups, local healers, parents and chiefs. These all being very focused on the cultural shift and the failure in understanding why this was being done, they had not been informed correctly as why the process was occurring. Conversely in Uganda condoms were not initially introduced, rather the presidents role in being vocal about the problem had an effect of inspiring local councillors, chiefs and church leaders to take this as a directive to educate their people about changing their sexual behaviour [19]. This program is now being attempted in Botswana. This example indicates the importance of disseminating health information via the right channels to obtain the best possible result that was aimed for.

Empowerment is feared by many in the medical field as care givers feel that this knowledge can harm rather than hurt the patient. However, in this case empowerment is through education. With a global focus on the Internet and its capabilities, researchers must not forget those areas that cannot harness these technologies. Rather, it is imperative to find a way that empower, through education via the correct channels

Converging communications

It is thus imperative to examining technology based health communication to understand who and what the communication is designed for. The goals of health promotion and disease prevention communication efforts are to help health consumers and information seekers gain knowledge about health issues and improve health. The goals of communication of health care delivery are to treat illness, maintain or improve health among patients and increase cost and delivery efficiencies[23]. Health communication efforts are designed to improve lifestyle behaviours, reduce risk factors for disease, increase compliance with a medication or treatment plan, better self manage a condition, provide social support or provide help with decision making procedures[24].

Through new technologies, health promotion and disease prevention interventions are being delivered successfully on-line, on CD-Rom, over the telephone, through handheld computers (PDAs), and via other technologies (Figure 2) for a variety of topics including weight control, injury prevention, smoking cessation, nutrition promotion, and medication compliance [20][21][22]. New technologies also allow health care delivery to transfer its model of care into a model of telemedicine, consisting of, but not limited to, telephone, video, and e-mail consultations, e-prescribing, claims processing, physician Web portals, and Electronic Health Records [23][2][24]

Figure 2 describes emerging technologies and trends that are singularly powerful. Their convergence could shift basic paradigms in health and health care. Potential examples of such converging applications include wireless, subcellular biosensors that monitor individual health parameters in real-time; techniques for meta-analyses of genetic, biophysical, and behavioural information to inform development of personalised health interventions including therapies; and tailored, broadband, interactive multimedia health communications that occur, irrespective of economic background[24].

The merging of consumer informatics and health communication have a combined effect of focusing on how communication methods will have an impact on consumer decisions. Consumer informatics aims to shift public knowledge, motivations, and attitudes towards clinical behaviours and with the adoption of health communication strategies this will yield more interactive, flexible and multidimensional healthcare tools.

Conclusion

This paper has highlighted the fact that Information technology is important in the health sector. Its role is growing in leaps and bounds. However, its growth will be hampered by its lack of accessibility in key areas. These areas lie in the developing world, primarily in Africa. It is also these areas that have the lowest Internet access rates and the highest rates of deaths by infectious diseases [25]. The Internet may not be the solution in dissemating health information in economic and poverty stricken areas, rather some of its methodologies can be implemented in other ways to provide optimal solutions.

In order to build sustainable health systems, it is important to understand why current systems will not work in specific areas, and to be able to adapt these systems. These adapted systems will then be more productive and effective in the areas that need it the most. One way of achieving this goal is to look at the converging of various technologies. This will include both electronic and non-electronic means to achieve the communication framework that will allow the right health information to be disseminated to the right people.

References

[1] Cuzens S. The power of cooperation. Bright. 2006:17:28- 31.

[2] Suggs SL. A 10-year retrospective of research in new technologies for health communication. Journal of Health Communication.2006:11: 61-77.

[3] Eng TR. The ehealth landscape: a terrain map of emerging information and communication technologies in health and health care. The Informatics Review. 2001 [Online] Available http://www.informatics-review.com/thoughts/rwjf.html

[4] Kreps GL. Disseminating relevant health information to underserved audiences: implications of the Digital Divide pilot projects. Journal of the Medical Library Association. 2001: 93: 68-73.

[5] Tang PC. and Lansky D. The missing link: bridging the patient-provider health information gap. Health Tracking.2005: 24: 290-1295.

[6] Harris L, Dresser C and Kreps GL. E-Health as dialogue: communication and quality of cancer care. Dialogue Systems for Health Communication. 2004 [Online] Available http://www.ccs.neu.edu/home/bickmore/dshc/kreps.pdf

[7] Department of Health and Human Services. Healthy People 2010. 2000 [Online] Available http://www.cdc.gov/nchs/about/otheract/hpdata2010/abouthp.htm

[8] Godlee F, Horton R and Smith R. Global Information Flow. Biomed Central News and Views. 2000:1:4-6.

[9] Internet Usage Statistics- The Big Picture. 2006 [Online] Available http://www.internetworldstats.com/stats.htm

[10] [10] Fors M. What the United Nations can do closing the Digital Divide. UN Chronicle Online Edition. 2005 [Online] Available http://www.un.org/Pubs/chronicle/2003/issue4/0403p31.asp

[11] Edejer TTT. Disseminating health information to developing countries: the role of the Internet. BMJ. 2000:321:797-800.

[12] United Nations Development Programme. Report of the Meeting of the high-level panel of experts on information and communication technology. 2000 [Online] Available www.un.org/documents/ecosoc/docs/2000/e2000-55.pdf

[13] Jensen M. African Internet connectivity. 2006 [Online] Available http://www3.sn.apc.org/africa/

[14] Norman CD and Skinner HA. eHEALS: the ehealth Literacy Scale. J Med Internet Res 2006:8:(4):e27. [Online] Available http://www. Jmir. Org/2006/4/e27/

[15] Amatayakul M. EHRs and the consumer: a new opportunity. In Murphy GF, Hanken MA and Waters KA eds. Electronic Health Records Changing the Vision. Philadelphia: WB Saunders Company: 1999: 26-68.

[16] Mongerson P. A patients perspective of medical informatics. Journal of American Medical Informatics Association. 1995:2 (2): 79-84.

[17] Friedman M. How to cure health care. Public Interest. 2001:142 (Winter).3-30.

[18] Eysenbach G. Recent advances: Consumer health informatics. BMJ.2000:320. 1713-16.

[19] Allen T and Heald S. HIV/AIDS policy in Africa: What has worked in Uganda and what has failed in Botswana? J. Int. Dev. 2004:16:1141-1154.

[20] Anderson DG. and Stenzel C. Internet patient care applications in ambulatory care. Journal of Ambulatory Care Management. 2001: 24:138.

[21] Dijkstra A, De Vries H, and Roijackers J. Computerized tailored feedback to change cognitive determinants of smoking: A Dutch field experiment. Health Education Research. 1998:13:197206.

[22] Friedman RH, Kazis, LE., Jette A, Smith MB., Stollerman J, Torgerson J, and Carey K. A telecommunications system for monitoring and counselling patients with hypertension. Impact on medication adherence and blood pressure control. American Journal of Hypertension. 1996: 9:285292.

[23] Chhanabhai P. EHRs: Fear of breach? The New Zealand publics opinion. Unpublished Masters thesis. University of Otago. 2006.

[24] HIFA2015. Communications with members of the HIFA2015 forum [Online] Available http://www.dgroups.org/groups/hifa2015/

[25] World Health Organisation. World Health Statistics 2006. 2006 [Online] Available http://www.who.int/whosis/en/index.html

Address for correspondence

Prajesh Chhanabhai
Department of Information Science
University of Otago
PO Box 56
Dunedin
New Zealand
Email: pchhanabhai@infoscience.otago.ac.nz

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Locating Nursing Classification Schemes within Health Information Strategies for New Zealand

Shona K. Wilson[a], Jan Duke[b],

[a] *PhD Student, Graduate School of Nursing, Midwifery and Health, Victoria University of Wellington, New Zealand*
[b] *Professor of Nursing and Head ,Graduate School of Nursing, Midwifery and Health, Victoria University of Wellington, New Zealand*

Abstract

The potential to use classification schemes to describe and measure nursing in a country that has previously not used them as a part of practice is fraught with issues. Such is the case for New Zealand. Without nursing specific classification scheme use in the information systems of day to day function, nursing cannot validate what it does and the difference it makes to health outcomes for New Zealanders.

The local use of valid and reliable classification schemes as tools to capture locally generated data that is able to be used as quality data needs to be considered alongside the national use of reliable clinical reference tools that are consistent with international standards. This may make the difference to the potential for significant contribution of nursing practice specific data to health information collections in preference to a 'one fits all' approach to user interface nursing classification scheme adoption at a local level. Tensions between a top-down approach and a locally based bottoms-up practice based approach and associated issues provide the core to this paper.

Keywords:

nursing classification schemes, national health information strategies, New Zealand

Introduction

There seems to be little doubt in the international world of informatics that the collection of nursing specific data can contribute to the measurement and validation of nursing practice [1]. However how to ensure that data can ultimately contribute to national and international health knowledge bases in a manner that will be sustainable over time remains emergent. The fabric of information collection, storage, retrieval, communication for optimising use is complex and involves many layers of key stakeholders ranging from users, developers, administrators and educators through to decision and policy makers. Unless the data collected is high quality data as defined by Kerr, (2006) and fits within provider based generic and national information systems frameworks, the information and knowledge it will ultimately provide will be compromised [2]. A tension arises between the nature and frequency of the data collected at the user interface and the multiple uses for each piece of datum within the health sector through to a policy level.

Contemporary information systems use by nurses in New Zealand appears to be largely focussed on information retrieval rather than data input. This means that although the retrieved information may contribute to nursing decision making it will not represent the actions and outcomes that are considered to be the basic elements of nursing practice. Therefore for nursing the discipline specific data-information-knowledge continuum remains unachievable. For nursing specific data to be useful it must represent nursing practice within clinical information systems in a way that is internationally, nationally and locally comparable [3] and yet satisfy the perceived need within the work force. Despite legislative and professional need for clear, concise, timely, accurate and current client record keeping by nurses processes remain problematical within the profession. Most documentation is as hardcopy files despite the presence of information systems in practice. Alternatively some free text descriptions exist within systems that do not contribute to wider knowledge repositories. The problem with free text is that it cannot be readily classified within the context of quality classification of health related phenomena.

At the same time there seems to be recognition of the need for measurement of health outcomes [4]. This tends to result in the application of a variety of tools being created for data collection at a local level. Many of the tools do not appear to have a theoretical or research base and most if not all do not appear to be integrated into clinically based information systems. The result is that existing clinical information systems do not and cannot meet the basic needs of nursing practice and consequently cannot contribute to the health information collections at a national level. To do this nursing must begin to collect the activities and outcomes of nursing practice using a reliable and validated discipline specific classification scheme that is integrated within the clinically based electronic client record [3]. In turn the electronic client record must contribute in some way to wider health information collections.

One issue for New Zealand is that clinical information systems are not yet in widespread use amongst major health providers although Patient Management Systems (PMS)

are. The next generation in systems development for the collection of clinically based data, the clinical information system (CIS), will provide the avenue for health professionals, including nurses to capture their daily activities and generate aggregate health outcome measurement [3]. Success of this depends upon a high level of participation at the user level to input data at the point where care is delivered.

Nurses appear to not know about or understand the differences between management/administrative systems and clinically based information systems. Nor do they appear to discern clearly between medical and nursing models of care delivery. Concurrently clinically based systems development seems to rely upon clinician input. When this happens nurses working at the clinical face tend to opt for transposition of what they already know into the technological world, often resulting in the mechanisation of the familiar paper based world. As classification schemes for nursing are not part of the paper nursing world in New Zealand subsequent implementation is unlikely to be driven by clinicians. Success is dependent upon nurse users in large numbers accurately recording the activities of practice to produce quality data, reliable information and knowledge. To do this they must understanding the meaning of each user interface category.

A solution may be to use a nursing classification scheme at the user interface that has high reliability and validity test scores. A problem is that such classification schemes have been developed and tested outside New Zealand and may not necessarily fit the nursing context of the country outside the country of origin [5]. Every change that occurs at the user interface has the potential to reduce the reliability of the tool and subsequently the quality of the data collected and the potential to be useful beyond the local context.

National trends in health information collection

Each country while acknowledging the place in and influences of the global information community has differences within the way health care is funded and provided that impact upon potential for quality data collection within information systems. The more competitive commercial approach of systems development may well carry with it a trade-off in quality of data collection than the more collaborative approach of a top down nationally supported model. Clearly for successful implementation of nursing classification schemes to collect nursing specific quality data within clinically based information systems an approach is needed that enables local, regional and national collections. The collections must be able to compare like data and make relationships between that data to validate the activities and outcomes for nursing in the context of health information collections. Then the difference that nursing makes to the health outcomes for New Zealanders will be able to be measured.

New Zealand as a country with a small population and a single government has some advantages over larger more complex societies for health information technology implementation. Health information strategy and implementation is a government direction resulting in the widespread public consultation and uptake of the Working to Add Value through E-information (WAVE) (2000) [6] document and more latterly the New Zealand Health Information Strategy (HIS-NZ), (2005) [7]. The government led approach to national health information collection has to date resulted in the implementation of a National Health Index (NHI), which categorises the demographic information about each member of the population, who accesses the health care system. Each person is subsequently registered with a alpha-numerical unique identifier that is used by all providers of health care and disability services. The NHI identifier and associated information is used to help with the planning, co-ordination and provision of related services across New Zealand. It is also intended to improve the flow and sharing of health information across providers and locations. The Medical Warnings System (MWS), associated with the NHI, warns healthcare providers of any known risk factors that may be significant when making decisions about care. The providers with electronic access can access the national NHI and MWS repository and automatically populate their PMS accordingly. The NHI and MWS do not contain any clinical information. A National Minimum Data Set (NMDS) for hospital events is available from a national level for providers of hospital services. No NMDS has yet been developed for use by primary and tertiary providers. Consumers of health care can request and receive any information about them that is stored within any of the databases that are held at a national level associated with the NHI [8].

A current development toward the national collection of health information include the establishment of a national Health Provider Index (HPI) as a register of health providers and organisations to enable role based secure electronic access to national repositories [8] Concurrently there is a public consultative process taking place to establish key direction strategy for information collection in the primary health care sector.

Together the NHI, MWS and HPI provide the basic elements of the international development of National Nursing Minimum Data Sets (NNMDS) [9] except for the inclusion of standardised nursing terminologies. The challenge will be to find a find a way to integrate these elements with nursing classification schemes implemented at a local level.

PMS are now established within the secondary care sector, which is predominantly provided by publicly funded District Health Boards (DHBs). Many of the PMS are supported by CIS, which may provide the avenue for clinician based field data input.

However the concept of an Electronic Health Record (EHR) that captures longitudinal lifespan health details of the population and that is accessible regardless of location or provider remains the ‘holy grail’ of health data collection and flow for New Zealand [10]. This may be concerned with two significant factors. The primary factor is that the supply of information systems for the health sector is driven by

commercial companies. In tandem, standards and recommendations from the government are intended as a guideline and are not mandatory. All the government can do to encourage vendors to adhere to any associated standards is to provide incentives for vendor buy-in.

The introduction of commerce in the health sector is a relatively new phenomenon to New Zealand. There was a highly regulated mostly publicly funded health sector until the mid-80's when a process of de-regulation and revision of the health care system occurred. This introduced a more business supply/demand model and choices to access private sector health care began. Since that time the presence of commerce in health care delivery has mushroomed and in particular to do with health information technology.

Health information systems vendors have valued clinical input to the planning and design phases of the Systems Development Life Cycle (SDLC) [11] for the systems developments which have resulted in some systems being developed that provide a framework for the system. This enables service providers to configure according to their own unique needs. In some instances this has resulted in a variety of developments within common systems with little or no interoperability between them. Local needs do not always concur with regional and national needs and local systems can adopt a life of their own.

Most significantly other than the NHI and MWS there is a lack of standardisation at the local level, so that when any EHR becomes a reality, the ability to compare like elements of data will be reduced.

The development and introduction of standards for the collection of meta-data at a national level [12] will hopefully be taken up and adopted by the process of government led public consultation. However that is likely to take some time, while aficionados of health care development continue to develop as a business means. This has an impact upon the potential to collect nursing specific data. Classification schemes for the collection of the basic elements of nursing practice exist internationally. Consistently they have been developed to be used at the clinical nurse user interface [13]. Terms and associated definitions are accordingly attuned to the user interface. If in this climate of health data collection movement any are integrated within existing systems at a local level without cognisance of the national movement the potential for descriptions and measurements for nursing beyond that level will be negated.

Nursing classification schemes

For any nursing classification scheme implemented in nursing to gather quality data in New Zealand the scheme must be simple to use, attract large numbers of nurses as users and suit the context of the service. Like most countries nursing practice in New Zealand is delivered throughout diverse settings alongside disparate health information systems. It is doubtful that a single reliable and valid scheme will suit all nursing settings [13]. This paper proposes that a variety of schemes can be implemented to suit each unique setting, whilst retaining quality of data. The solution is to implement known reliable and valid schemata that are an exact coding match to cross-map with multi-disciplinary clinical reference terminologies. This broadens the scope of choice of a classification scheme to implement and guarantees a return of quality data.

The issue with this is that to date New Zealand has not committed to a national clinical reference terminology (such as SNOMED CT). A solution would be to use the International Classification for Nursing Practice (ICNP) as a clinical reference terminology ensuring that the terminology used at a local level is an exact coding match. When a national multi-disciplinary clinical reference terminology is implemented it should then cross-map accordingly. Negotiations between SNOWMED CT and ICNP began in April [14] and are currently taking place to ensure that there will be a cross-map available in due course.

Currently some of the major nursing classification schemes that have been validated as reliable and useful by research in practice internationally are validated as cross-mapping to SNOMED CT directly, without the use of an intermediary nursing specific clinical reference terminology [15]. However without the use of a nursing specific clinical reference terminology in New Zealand the potential for nursing data to be compared across international nursing borders will be reduced. But can nursing in New Zealand wait until there is an exact cross-map between user interface languages, the discipline specific clinical reference terminology (ICNP) as outlined as International Standard (ISO) 18104 [16] and the multi-disciplinary clinical reference terminology? When that occurs, a small country like New Zealand must then be able to purchase the terminologies within the global market and implement in a resource scarce health care sector.

Compounding these issues is that nursing classification schemes as they exist elsewhere have not been adequately tested to see if they describe the basic elements of nursing practice in the context of New Zealand. Do the terms and definitions within the classification schemes accurately summarise the nuances and differences in practice between New Zealand and the country of the terminology's origin? Clearly the need to begin to collect the nursing specific data outweighs the temptation to develop a local classification scheme which would take time and risk compromising the potential to accord an exact cross-mapping with international reference terminologies. Analysis and evaluation as testing of any terminology in the local context would appear to be an imperative in the cycle of development and implementation as suggested within the SDLC [13]. Unfortunately historically it appears that this important evaluative step is often omitted in New Zealand.

Like other countries New Zealand lacks the nurses who are adequately prepared to begin this journey. Whilst there is a small core of nursing informatics enthusiasts the call for their expertise is shared between academic research and education, systems development, local clinical information needs and nationally led strategic approaches. Such personnel need in-depth understanding of the top down

approach, insight into the local nursing systems development approach, the ability to identify with international trends and location of the issues within the socio-cultural context for resource allocation in New Zealand.

Is there a solution?

The way forward for nursing in New Zealand to begin to collect discipline specific data that will ultimately describe and measure the basic elements of nursing practice needs to be carefully planned in consideration of these contextual issues. To date there is not much available theory driven nursing informatics research literature available to inform such an undertaking as the focus has been historically driven by problem solution and subsequent systems development [11]. The work of Effken (2003) provides a conceptual framework toward organising informatics research based upon the constructs that make up the meta-paradigm of nursing. The aim of her framework is to provide a means to develop nursing informatics theory development as a research endeavour. The work locates the SDLC within the major constructs of nursing practice and emphasises the need for ongoing analysis throughout the cycle. Effken (2003) invites nursing academics and researchers to apply the main principles of the framework to nursing informatics theory development internationally and so validate its utility. This may provide a theoretical grounding for the implementation of classification schemes set in the context of the body of nursing knowledge that is New Zealand nursing practice.

The time has come to move beyond the dualistic methods of having to choose between the top-down and bottom-up approaches. A more hybrid approach that considers and values the issues of each may provide the solution for nursing classification scheme implementation in New Zealand. Acknowledging that a universal nursing classification scheme is unlikely to fit the diverse settings in New Zealand, implementation can begin by selecting a nursing classification scheme that is:

- valid and reliable internationally for the collection of quality data
- an exact cross map with international clinical reference terminologies, and
- matches a particular nursing service practice setting in New Zealand

If all implementation at local level adheres to these three basic criteria for implementation then the data collected and stored will be able to be used in many ways. It will be able to be used at a local level for care planning, measurement of service delivery, and begin to describe the difference nursing makes to the local population targeted by the service. It will also ensure that ultimately the data can contribute to other repositories as they are developed, such as regional and national EHR repositories and any professional nursing database that may evolve in the future.

Different classification schemes can be used in different settings as long as they have the exact cross map at coding level of CIS with the clinical reference nursing specific and multi-disciplinary clinical reference terminologies that are in use internationally. Some classification schemes lend themselves more to a particular service than others. For instance the Omaha System appears to be a good match for any New Zealand Mental Health Nursing Service. Mental Health in New Zealand has Health of the Nation Outcome Scale (HoNOS) as mandatory for the collection of Mental Health data [17]. There appears to be a high level of synergy between the Omaha System and HoNOS in the identification of problems as well as the language and structure of outcome scales. While a match between the two has not yet been tested, this does mean that nurses would recognise the similarities and so may be better placed to adopt the Omaha System in practice. The Omaha System has also been evaluated and demonstrates a high level of utility for Mental Health [18]. Another classification scheme may have greater match in another area such as acute nursing service but as long both have an exact match with the overarching clinical reference terminology, the data collected will be able to be used across local, regional, national and international settings.

Conclusion

The central argument for this paper is that for nursing practice to be able to validate what difference it makes to the health of New Zealanders and contribute to repositories concerned with the body of nursing knowledge and development, New Zealand needs to consider the introduction of both the nursing specific clinical reference terminology (e.g. ICNP) and a multi-disciplinary clinical reference terminology (e.g. SNOMED CT) at a national level. These reference terminologies would need to integrate with the NHI, MWS and HPI, as part of national health information trends. There is also a need to ensure that any implementation within any system at a local level is an exact match to the clinical reference terminologies. In turn an exact match between coding of the system and terms and definition of the validated and reliable nursing classification scheme needs to be ensured. If this occurs, then in time all the data that is collected at the user interface will ultimately contribute to the national health information collections of the future as well as to the national and international body of nursing knowledge.

Acknowledgements

The Foundation of Science Research and Technology, New Zealand, for a scholarship to undertake this study

References

[1] Harris, M. R., Graves, J.R. Solbrig, H.R. Elkin, P.L. and Chute, G.C. Embedded Structures and Representations of Nursing Knowledge. J Amer Med Inform Assoc 2000

[2] Kerr, K. (2006) The institutionalisation of data quality in the New Zealand Health Sector. 2006 Unpublished thesis at University of Auckland, New Zealand.

[3] Androwich, I.M., Bickford, C.J., Button, P.S., Hunter, K.M., Murphy, J. & Sensmeier, J. (2003) Clinical Information Systems: A Framework for Reaching the Vision. Washington, D.C: American Nurses Publishing.

[4] New Zealand Health Information Service. (2006). About NZHIS. Online at: http://www.nzhis.govt.nz/aboutnzhis.html [Retrieved 02 October, 2006]
[5] Clark, J. (1998) The International Classification for Nursing Practice Project Online Journal of Issues in Nursing. Available at htttp://www.nursingworld.org/ojin/tpc7/tpc7_3.htm [Retrieved 27 July, 2006]
[6] Ministry of Health. From strategy to reality: the WAVE project. Report of the WAVE Advisory Board to the Director-General of Health. Wellington, New Zealand: Ministry of Health; 2001.
[7] Health Information Strategy Steering Committee. Health Information Strategy for New Zealand, 2005. Online at: http://www.moh.govt.nz/moh.nsf/0/1912064EEFEC8EBCCC2570430003DAD1/$File/health-information-strategy.pdf [Retrieved 06 June, 2006)
[8] New Zealand Health Information Service. National Health Index, nd. Online at: http://www.nzhis.govt.nz/nhi/brochure.html#08 [Retrieved 20 June, 2006]
[9] MIA-NI. Nursing Matters: International Nursing Minimum Data Set (i-NMDS), 2004. Online at http://www.icn.ch/matters_i-NMDS_print.htm Retrieved [06 June, 2005)
[10] Kerr K. The electronic health record in New Zealand. Health Care and Informatics Online. 2004 Online at: http://hcro.enigma.co.nz/website/index.cfm?fuseaction=articledisplay&FeatureID=040304. [Retrieved 25 July, 2005].
[11] Effken, J.A., (2003) An Organizing Framework for Nursing Informatics Research in Computers in Nursing Vol.21 No. 6 pp 316-325. Nov/Dec 2003
[12] [12] New Zealand Ministry of Health: Health Information Standards Organisation (HISO) Online at: http://www.moh.govt.nz/moh.nsf/f872666357c511eb4c25666d000c8888/9a18e48fca47e93dcc256e2200071d6d?OpenDocument [Retrieved 26 November, 2006]
[13] McCormick KA., Jones CB. Is one taxonomy needed for health care vocabularies and classifications? Online Journal of Issues in Nursing 30 September 1998. http://www.nursingworld.org/ojin/tpc7/tpc7_2.htm [Retrieved 06 April, 2005]
[14] SNOMED International: ICN and SNOMED Further Collaboration. 2006. Online at: http://www.snomed.org/news/documents/icn.pdf [Retrieved 20 November, 2006]
[15] SNOMED International Nursing Working Group n.d. Mapping. Online at: http://www.snomed.org/clinical/SNOMEDMapping.html [Retrieved 06 June, 2006]
[16] International Organisation for Standardisation. ISO 18104: 2003. Health Informatics – Integration of a Reference Terminology for Nursing. Online at: http://www.iso.org/iso/en/Catalogue DetailPage.Catalogue Detail?CSNUMBER=33309 [Retrieved 01 August, 2006]
[17] The National Centre of Mental Health Research and Workforce Development: Te Pou o Te Whakaaro Nui. HoNOS Family of Measures, n.d. Online at: http://www.tepou.co.nz/page/tepou_83.php Retrieved 06 October, 2006
[18] Barrera C, Machanga M, Connolly PM, & Yoder M. Nursing care makes a difference. Application of the Omaha System. Outcomes Management Journal, 2003 Oct-Dec;7(4):181-5.

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

SHARE, from Vision to Road Map: Technical Steps

Mark Olive [a], Hanene Rahmouni [a], Tony Solomonides [a], Vincent Breton [b], Yannick Legré [b], Ignacio Blanquer [c] and Vicente Hernandez [c]

[a] *CCCS / CEMS Faculty / UWE, Bristol / Coldharbour Lane / Bristol BS16 1QY / UK*
[b] *LPC / CNRS-IN2P3 / Campus de Cézeaux / 63177 Aubière Cedex / France*
[c] *Universidad Politécnica de Valencia / Camino de Vera s/n / E-46022 Valencia / Spain*

Abstract.

We present the 'HealthGrid' initiative and briefly review work carried out in various European healthgrid projects. We report on joint work with numerous European collaborators. Since the European Commission's Information Society Technologies programme funded the first grid-based health and medical projects, the HealthGrid movement has flourished in Europe. Many projects have now been completed and 'HealthGrid' consulted a number of experts to compile and publish a 'White Paper' which establishes the foundations, potential scope and prospects of an approach to health informatics based on a grid infrastructure. With a second generation of projects now funded, the EC has commissioned the SHARE Project, a study to define a research roadmap for a 'healthgrid for Europe' as the preferred infrastructure for medical and health care projects in the European Research Area. The project explores the ways in which the healthgrid approach supports modern trends both in research in biomedicine and in healthcare, such as evidence-based practice and information integration.

Keywords:

healthgrid, e-health, grid applications

The HealthGrid initiative

'Grid' has been identified as one of the key technologies to support the European Research Area. The impact of this concept is expected to reach far beyond eScience, to eBusiness, eGovernment, and eHealth, but a major challenge is to take the technology out of the laboratory to the citizen. A *healthgrid* is an environment in which medical data can be stored and made available to all actors in the healthcare system, doctors, allied professions, healthcare centres, administrators and, of course, patients and citizens in general. Such an environment has to offer all appropriate guarantees in terms of data protection, respect for ethics and observance of regulations; it has to support the notion of 'duty of care' and may have to deal with 'freedom of information' issues. Working across member states, it may have to support negotiation and policy bridging.

Pioneering projects in the application of grid technologies to the health area have been completed, and the technology to address high level requirements in a grid environment has been under development and making good progress. Because these projects had a finite lifetime and the vision required a sustained effort over an extended period, and besides because there was an obvious need for these projects to cross-fertilise, the 'HealthGrid initiative', represented by the HealthGrid association (http://www.healthgrid.org), was launched to bring the necessary long-term continuity. Its goal is to encourage and support collaboration between autonomous projects in such a way as to ensure that requirements really are met and that the wheel, so to speak, is not re-invented repeatedly at the expense of other necessary work.

Writing about the healthgrid initiative very soon after its inception, this community identified a number of objectives [1]: identification of potential business models for medical grid applications; feedback to the grid development community on the requirements of the pilot applications deployed by the European projects; development of a systematic picture of the broad and specific requirements of physicians and other health workers when interacting with grid applications; dialogue with clinicians and those involved in medical research and grid development to determine potential pilots; interaction with clinicians and researchers to gain feedback from the pilots; interaction with all relevant parties concerning legal and ethical issues identified by the pilots; dissemination to the wider biomedical community on the outcome of the pilots; interaction and exchange of results with similar groups worldwide; and the formulation and specification of potential new applications with the help of the end user communities.

The grid concept is rooted in the physical sciences and these considerations were not a central concern to general grid developers. Even today these requirements are not a priority for developers, even though they have been fed through to the middleware services community. Thus HealthGrid identified the need for a specialist middleware layer, between the generic grid infrastructure and the medical or health applications.

Selected for best paper award.

Among data related requirements, the need for suitable access to biological and medical image data arose in several early projects, but for the most part these are present in other fields of application also. Looking to security requirements, most of these are special to the medical field: anonymous or private login to public and private databases; guaranteed privacy, including anonymisation, pseudonymisation and encryption as necessary; legal requirements, especially in relation to data protection, and dynamic negotiation of security and trust policies while applications remain live. Medical applications also require access to small data subsets, like image slices and model geometry. At the (batch) job level, medical applications need an understanding of job failure and means to retrieve the situation.

The white paper: from grid to HealthGrid

The next step for the HealthGrid community was to try to systematise the concepts, requirements, scope and possibilities of grid technology in the life sciences. The White Paper [2] defines the concept of a healthgrid more precisely than before: ... grid infrastructures comprising applications, services or middleware components that deal with the specific problems arising in the processing of biomedical data.

The ultimate goal for eHealth in Europe may be the creation of a single healthgrid incorporating a 'principle of subsidiarity' for independent nodes of the healthgrid as a means of implementing all the legal, ethical, regulatory and negotiation requirements. We may anticipate, however, the development path to proceed through specific healthgrids with perhaps rudimentary inter-grid interaction/interoperational capabilities. We may therefore identify a need to map future research and advice on research policy, so as to bring diverse initiatives to the point of convergence.

Healthgrid applications address both individualised healthcare – diagnosis and treatment - and epidemiology with a view to public health. Individualised healthcare is improved by the efficient and secure combination of immediate availability of personal clinical information and widespread availability of advanced services for diagnosis and therapy. Epidemiology healthgrids combine the information from a wide population to extract knowledge that can lead to the discovery of new correlations between symptoms, diseases, genetic features and other clinical data. With this broad range of application in mind, the issues below are identified as key features of our analysis.

- Business case, trust and continuity issues: healthgrids are data- and collaboration grids, but healthcare organisations are required by law to maintain control of their patients' records. Deployment on a scale to make an attractive business opportunity requires a high level security and compliance.
- Biomedical issues: Distributed databases and data mining are important tools for many biomedical applications in fields such epidemiology, drug design and even diagnosis. Expert system services running on the grid must be able to interrogate large distributed databases to explore sources of diseases, risk populations, evolution of diseases or suitable proteins to fight against specific diseases.
- Security issues: These flow naturally from the nature of medical data and from business requirements. Security in current grids is adequate only for research platforms.
- Management issues: The central concept of a 'virtual organisation' (VO) at the heart of eScience, which gave rise to grids, is very apt for healthgrid, but additional flexibility is needed to structure and to control VOs on a broader scale, including, for example, the meta-level of a VO of VOs.

We illustrate the concept of healthgrid with some prototypical examples: GEMSS [3] used a 'high-throughput' numerical simulation of organs obtained from a patients' data and used these to aid understanding or to improve the design of medical devices, with patient-customised approaches at research-level in areas such as radiotherapy, craniofacial surgery and neurosurgery. MammoGrid [4] created a database of standardised mammogram files and associated patient data to enable radiologists in the UK and in Italy to request second opinion and computer-aided detection services. The project also enabled a further study of an epidemiological nature on breast density as a risk factor. In Health-e-Child [5] radiologists are working with oncologists, cardiologists and rheumatologists to identify early imaging signs of conditions that may have a strong genetic component, possibly reducing the need for genetic maps to be obtained on an indiscriminate basis. In Wide In Silico Docking On Malaria (WISDOM) [6] tens of millions of molecular docking experiments have been used to help identify potential antigens for the malaria parasite. The experiment uses large scale virtual screening techniques to select molecular fragments for further investigation in the development of pharmaceuticals for neglected diseases. The economic dynamics in this area are telling: only about 1% of drugs developed in the last quarter century have been aimed at tropical diseases, and yet these are major killers in the third world, with mortality in excess of 14 million per annum. Meanwhile, the results of several major studies of the interface between bioinformatics and medical informatics had been published with a remarkable promise of synergy between the two disciplines, leading to what had already begun to be referred to as 'personalised medicine'. [7,8]

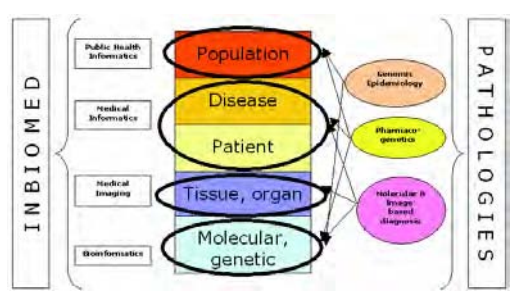


Figure 1 - Disciplines, levels of being and pathology diagnostics (F. Martin-Sánchez)

The SHARE Project: towards a road map

The vision of health that informs the thinking of the White Paper is reflected in European thinking [9] and is depicted in a map of the relationships between the different ontological and epistemological levels and the various modalities of data have been captured by Fernando Martin-Sánchez (cf [1]) in the schematic diagram of Figure 1.

In the White Paper, the HealthGrid community expressed its commitment to engage with and support modern trends in medical practice, especially 'evidence-based medicine' as an integrative principle, to be applied across the dimensions of individual through to public health, diagnosis through treatment to prevention, from molecules through cells, tissues and organs to individuals and populations.

In view of the impact of the White Paper, the EC has funded the project SHARE [10] to explore exactly what it would mean to realise the vision of the White Paper, investigate the issues that arise and define a roadmap for research and technology which would lead to wide deployment and adoption of healthgrids in the next ten years. Thus the project must address the questions, What research and development needs to be done now? and What are the right initiatives in eHealth RTD policy relating to grid deployment?, with all that implies in terms of coordination of strategy, programme funding and support for innovation. Thus the project will define a comprehensive European research and development roadmap, covering both policy and technology, to guide and promote beneficial EU-wide uptake of healthgrid technologies.

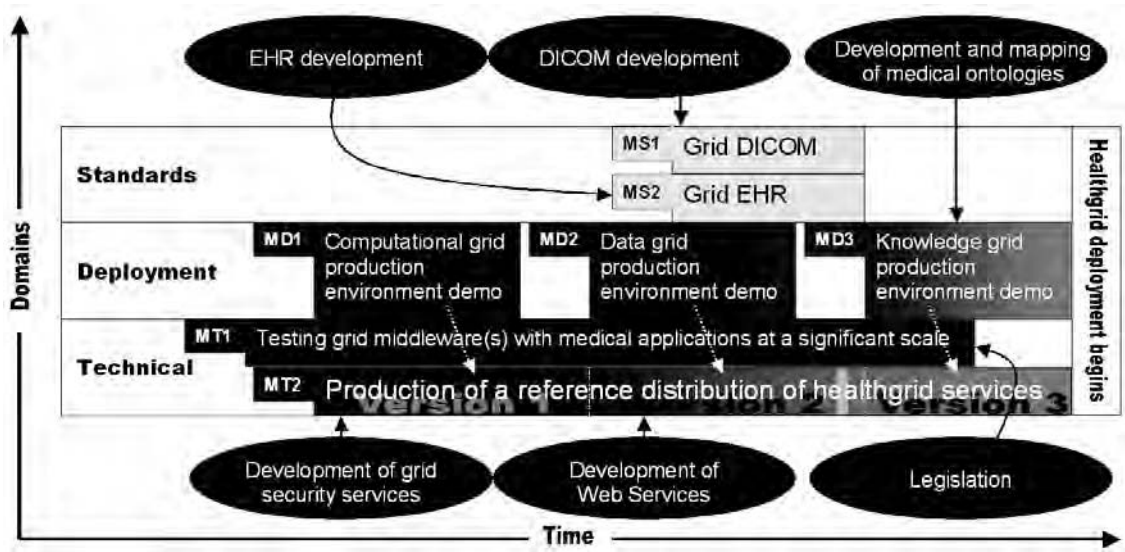


Figure 2 - SHARE technical roadmap diagram, showing milestones and external influences

Technical road map: Step one

SHARE has defined a preliminary technical road map (see Figure 2 above) with two technical milestones for appropriate development of healthgrid services (**MT1,2**), two milestones as examples of grid standards for the medical domain (**MS1,2**), and three deployment milestones of increasing complexity and scope (**MD1,2,3**).

MT1 Before deployment can begin, grid middleware must be tested with medical applications for scalability and robustness. This must begin at an early stage. It is anticipated that this will be an ongoing activity, with different generations of grid operating systems offering newer, faster and more stable capabilities.

A key issue for this milestone will be the robustness of grid solutions based on web services. Scalability, particularly regarding medical applications, is still a concern for grid middleware based on the Open Grid Services Architecture (OGSA) such as GT4 and GRIA. Middleware such as gLite and Unicore on the other hand have been deployed on large scale infrastructures in Europe and have demonstrated their scalability and robustness, but are still awaiting migration to web services.

MD1 The first deployment step will be the rollout of a computational grid production environment demonstrator for medical research. This would seem to be an achievable goal in the reasonably near future, given that there have already been successful deployments of computational grid applications (the WISDOM data challenges, for example) on general purpose grid infrastructures such as DEISA and EGEE. However, convincing healthcare management of the benefits of deploying a computational grid on a hospital or clinic IT infrastructure, which would not be composed of dedicated grid nodes and may already be working near capacity, is a real concern. Many medical centres may simply not have the necessary bandwidth or storage capabilities to make best use of grid technology, or may not have appropriate equipment to capture data in digital form. The management and configuration of a grid is rather complex and may require significant investment in manpower and training.

Installation of grid nodes behind hospital firewalls is incompatible with the present security model, which

required inbound connections. Secure services for data management are under development that could allow firewall rules to be relaxed, enabling connections to grid infrastructures. New architectures and designs should be defined that will minimise the volume of data leaving hospital borders.

Ideally, the grid node would be located outside the firewall with only anonymised or pseudonymised data being stored on the grid. However, even with the most stringent pseudonymisation and de-identification techniques there is still some risk of unauthorised re-identification by a person with sufficient knowledge from other sources. There are therefore legal and ethical implications of storing even anonymised personal data on the grid, and further investigation will be required to determine if this is possible with current national and European policies and legislation.

MT2 starts with **MD1** and ends when production deployment begins. This milestone is the development of a reference distribution of grid services, using standard web service technology and allowing secure manipulation of distributed data. Standards in emerging Web services technologies (WSx and the WSRF specification) will facilitate interoperability between healthgrids built using different underlying tools. A precise, well documented set of requirements is needed to describe the security features and obligation policies at different levels of abstraction in the middleware.

The IBHIS project found that web service description languages and registries are not yet mature enough, particularly WSDL, which describes how to access web services, and UDDI, which provides a registry for service discovery. These technologies are currently not flexible enough, cannot be used for semantic queries or descriptions (e.g. a description of the function a service provides, or the meaning of parameter names) or non-functional descriptions such as quality of service and performance levels. The development of WSDL and UDDI is ongoing, with OWL-S extensions to UDDI to facilitate semantic searching, upcoming versions of UDDI promising to address other limitations, and WSDL 2.0 promising to support semantic descriptions and include non-functional requirements.

Security is not an option but a mandate for healthgrids at all technical levels: networks need to provide protocols for secure data transfer, the grid infrastructure needs to provide secure mechanisms for access, authentication, and authorisation, as well as sites for secure data storage. The grid operating system needs to provide access control to individual files stored on the grid, and high level services need to properly manage the legal issues relating to the protection of medical data. Important progress is being made in terms of fine-grained access control and data encryption.

Revocation of credentials and how to provide temporary access to data is still an open issue, and an important one for healthgrids. There are a number of situations where users would temporarily require access to data that they would not normally have access to, such as a visiting expert being shown an unusual case. Certificate authorisation servers have been developed in both 'Pull' mode, in which sites periodically pull a list of valid members from a central service, and 'Push' mode, in which users obtain a short-lived attribute certificate that they present to sites to prove their membership. However, both of these would leave a window where revoked or expired credentials could be used to gain unauthorised access. Several healthgrid projects have suggested that the data itself should have a 'lifetime' – users with temporary access should not be able to access the data (or a copy of the data) once their credentials have expired.

MD2 Although several prototype data grids for medical research have been demonstrated by healthgrid projects, developing and maintaining a production quality data grid will require a number of issues relating to the distributed storage of medical data to be resolved. In European grid infrastructures, the distributed storage of medical images has been hampered by the limited data management services available, and so the continuation of improvements in this area will be important for the adoption of grids by the medical community. High speed links between data providers and consumers will be a prerequisite, particularly given the high volume of data predicted.

Many legal and ethical issues will need to be resolved, such as the ownership of patient data, ethical control of information, the patient's right to access or be informed about data that concerns them, as well as local, national, and European legislation governing the use of patent data and IT.

Another important concern for this milestone will be the integration of heterogeneous data from multiple sources. While mechanisms for data integration have been demonstrated by previous projects, biomedical data can be exceptionally varied including images with associated metadata and free form text or hand written notes from patient records. There is also the issue of how to deal with missing, inaccurate or obsolete data.

MS1 & **MS2** The use of computer-based tools for clinical research has led to the definition of standards for the exchange of data in many areas but their adoption has not been universal. The exchange of data between bioinformatics and medical informatics is an area where standards are particularly limited. By contrast, in medical imaging the adoption of DICOM for the storage and transmission of medical images has been accepted worldwide. In medical records HL7 is the emerging standard. For both of these standards, there is a question of compatibility with grid technologies.

MD3 After the issues with the distributed storage and querying of medical data have been resolved, the next task will be to deploy services that can build relationships between data items, and will provide appropriate representation to medical researchers. Particularly given that there have been no successful deployments of knowledge grids for medical research to date, this will pose a significant challenge. The data concerned can be extremely varied in nature, structure, format and volume. Depending on the

area of research, the synthesis of knowledge from data could require sophisticated data mining, integrated disease modelling and medical image processing applications, and may also involve the use of techniques from artificial intelligence to derive relationships between data from different sources and in different contexts.

The development of medical ontologies and mapping between ontologies will be particularly important for the successful deployment of knowledge grids. An ontology is the systematic description of a given phenomenon: it often includes a controlled vocabulary and relationships, captures nuances in meaning and enables knowledge sharing and reuse. From an agreed ontology it is possible to define a common data model that describes the format of the data used by all the services. The standardisation of interfaces can dramatically increase interoperability between biomedical resources, and by operating on standardised data formats they can more easily be integrated into complete bioinformatics experiments by eliminating the restructuring of data between each service. The construction of standardised data formats can be improved by defining a domain ontology that covers the concepts used within a given domain. These ontologies will allow relationships between concepts and nuances in meaning to be captured, greatly enhancing the opportunities for communication, knowledge sharing and reuse, and machine reasoning.

Open issues include how to integrate biomedical data using ontologies, how to combine different initiatives and how to employ advanced, semantic reasoning techniques for analysing medical data.

Conclusions and future work

Certain specific features of the community, such as issues of patient ownership of her/his data and the tension between hospitals' IT policies and the requirements of grids, will continue to prove troublesome unless addressed with political will. Another non-functional obstacle is the drag on technology transfer between EC projects. E.g. there is a need for healthgrid projects to begin thinking about data curation and digital libraries, but researchers and providers have not come together to explore this need.

SHARE predicts that it may take ten to fifteen years from a sustainable computing grid to a generalised knowledge grid. However, the transition to data grid may not be as simple as the success of special projects suggests and the transition to knowledge grids will be breaking new ground. It has been suggested that a more realistic timeframe might be twenty to forty years. As a next step, SHARE will focus on the large number of Ethical, Legal and Socio-Economic issues related to healthgrids. These will be integrated with the technical roadmap to recommend both technical and policy actions.

The work reported here has been carried out in collaboration with many colleagues in the HealthGrid community. Thanks are due to these and numerous other colleagues in Europe, the US and Asia for helpful discussions.

References

[1] V. Breton, A. E. Solomonides & R.H. McClatchey. *A Perspective on the Healthgrid Initiative*. Second International Workshop on Biomedical Computations on the Grid, the Fourth IEEE/ACM International Symposium on Cluster Computing and the Grid (CCGrid 2004), Chicago, USA, April 2004

[2] V. Breton, K. Dean & T. Solomonides (Eds.) *The Healthgrid White Paper*. From Grid to Healthgrid: Proceedings of Healthgrid 2005: pp. 249-321, IOS Press, Studies in Health Technology and Informatics, Vol 112, 2005.

[3] GEMSS Project. *Grid Enabled Medical Simulation Services*. Available at: http://www.gemss.de

[4] C. del Frate et al. *Final Results and Exploitation Plans for MammoGrid*. Challenges and Opportunities of Healthgrids, Proceedings of HealthGrid 2006, IOS Press, Studies in Health Technology and Informatics Vol 120, 2006.

[5] J. Freund et al. *Health-e-Child: An Integrated Biomedical Platform for Grid-Based Paediatric Applications*. Challenges and Opportunities of Healthgrids, Proceedings of HealthGrid 2006, IOS Press, Studies in Health Technology and Informatics Vol 120, 2006.

[6] Jacq, N., Salzemann, J., Jacq, F., Legré, Y., Medernach, E., Montagnat, J., Maaß, A., Reichstadt, M., Schwichtenberg, H., Sridhar, M., Kasam, V., Zimmermann, M., Hofmann, M., Breton, V., *Grid-enabled Virtual Screening Against Malaria*, Journal of Grid Computing (to appear 2007).

[7] A. Sousa Pereira, V. Maojo, F. Martin-Sanchez, A. Babic & S. Goes. *The INFOGENMED Project*. ICBME 2002, Singapore, 2002.

[8] BIOINFOMED Project. *Synergy between Medical Informatics and Bioinformatics: Facilitating Genomic Medicine for Future Healthcare*, BIOINFOMED Study Report White Paper, 2003.

[9] European Commission. *e-Health – making healthcare for European citizens: An action plan for a European e-Health Area*. COM(2004) 356 final. A 'Communication from the Commission', available at: http://europa.eu.int/information_society/doc/qualif/health/COM_2004_0356_F_EN_ACTE.pdf

[10] V. Breton, I. Blanquer, V. Hernandez, Y. Legré & A. E. Solomonides. *Proposing a Road Map for Healthgrids*. Challenges and Opportunities of Healthgrids, Proceedings of HealthGrid 2006, IOS Press, Studies in Health Technology and Informatics Vol 120, 2006.

Address for correspondence

All correspondence to Tony Solomonides at Tony.Solomonides@uwe.ac.uk.

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Penetration and Adoption of Health Information Technology (IT) in Thailand's Community Health Centers (CHCs): A National Survey

Boonchai Kijsanayotin[a], Supasit Pannarunothai[b], Stuart Speedie[a]

[a] *Health Informatics Graduate Program, Medical School, University of Minnesota, Minneapolis, Minnesota, USA*
[b] *Faculty of Medicine, Naresuan University, Phitsanulok, Thailand*

Abstract

A universal healthcare coverage program has been implemented in Thailand since 2001 and the Thailand Ministry of Public Health (MOPH) is restructuring its health information systems to support the management of this reform. The MOPH believes that health information technology (IT) is fundamental to the development of an effective health information system, and that users' adoption of health IT is one of the most important factors to the success of health IT implementation projects. However, there is no national data available regarding the penetration and adoption of health IT in Thai community health centers (CHCs). This cross sectional survey was designed to study the penetration and adoption of health IT in the country's community health centers. A random sample of 1,607 regionally stratified CHC's from a total of 9,806 CHCs was selected. With an 82% response rate, the data showed that people who worked in CHCs were currently heavy users of health IT. They exhibited high IT acceptance and positive attitudes toward using health IT. CHCs' staff was less resistant to adopt health IT than previously anticipated. These results are similar in all of the country's geographic regions. Health IT is pervasive in CHCs across the country and penetrates all regions.

Keywords:

community health centers, information systems, computer systems, attitude of health personnel, diffusion of innovation, Thailand

Introduction

The Thailand universal healthcare coverage program (the so-called 30 Baht for all diseases scheme) was instituted in 2001 and continues to evolve[1-3]. The program covers 74% of the 64 million people in Thailand. Thailand's Ministry of Public Health (MOPH) is currently restructuring its national health information system to support that reform. The MOPH is not only responsible for the country's public health system, but also is the country's major healthcare provider. In 2001, it oversaw 868 hospitals (67.1% of hospitals nationwide) and all 9,738 community health centers[4]. A community health center (CHC) is a subdistrict (Tambon) or village-level health service unit. This is a first-line unit, covering a population of about 1,000 - 5,000, with health staff including a health worker, a midwife and a technical nurse. Similar to other developing countries, the Thai government and MOPH administrators have recognized the potential of information technology to improve the quality of healthcare delivery and have embraced the technology[5].

Studies show that approximately 40% of information technology developments in various sectors, including the health sector, have been considered failures or have been abandoned. Moreover, this number has remained approximately the same for the last 25 years[6]. Littlejohns et al (2003) reported that the reasons for the failure of a large computerized health information system project in South Africa were a lack of users' understanding of reasons for new system and the underestimation of the complexity of the healthcare system[7]. Lorenzi and Riley (2003) pointed out that human issues at both individual and organizational levels are the reasons contributing to the information system (IS) implementation failures. They categorized reasons for information system failure as ineffective communication, underestimation of complexity, scope creep, organization problems, technology problems, and leadership issues. They emphasized the importance of organization change management when innovative information systems are being developed and implemented[8]. DeLone and McLean (2002) argued that information system usage is one of the six interdependent dimensions used to measure IS success[9]. The Heidelberg Health Information System (HIS) Working Group stated in the conference of the International Medical Informatics Association (IMIA) in 2003 that ".... people, not technology, will ultimately determine the success of HIS"[10]. Overall, these studies have concluded that the socio-technical aspect of the IT, particularly people and organizations, is essential to the success of the system[11].

With the advent of the universal healthcare coverage program, effective new health center information systems are anticipated in Thailand. The systems will link administrative, healthcare operations and public health information systems together in community health centers, local hospitals, and local and central administrative offices[12, 13]. Knowledge about users and organizational changes is one of the vital components for success of the system design, development and implementation[14,15]. However, national data regarding the current status of IT use in healthcare are not available. We do not know how healthcare providers currently use IT, or their willingness to accept IT in the CHCs. This knowledge is crucial for the

success of health information system development and implementation. Health IT in this study is defined as the information and communication technology that is used for health information systems. The present study had two main objectives. First we aims to describe the penetration of health IT in the country's CHCs and describe how CHCs currently use health IT, their attitude towards health IT and the degree of their acceptance of information technology in their work. Second, we aimed to inform health policy makers and those who are working in health IT projects about the readiness of community health centers to adopt and use health IT.

Methods

A cross sectional national survey was conducted in Thailand during July-October 2005. All 75 provinces except Bangkok, the capital, were stratified by four geographic regions; northern, central, northeastern and southern regions Then we randomly sampled population weighted provinces from each region. This resulted in the selection of twelve provinces with 1,607 CHCs: three from the north with 337 CHCs, four from the central with 376 CHCs, three from the northeast with 708 CHCs, and two from the south with 146 CHCs. At each CHC, we asked an officer who was responsible for the CHC's information management or the CHC's administrative officer (the CHC head officer) to complete the survey. Research collaborators at the provincial health office in each province distributed and collected the self-administered paper-based survey.

The survey instrument was developed by a group of Thai health IT experts. The instrument development was guided by two previous studies in which instruments had been tested and have exhibited acceptable levels of reliability and validity. Computer use and demographic measures were from the first author's previous work[16]. The attitude towards IT and intention to use IT measures were adopted from Venkatesh et al (2003)[17]. They were minimally modified and translated into Thai language so as to be clearly understood by Thai health center personnel. IT use associated with activities in CHCs was measured with ten questions using a four point scale ranging from one (never use) and four (always use). These ten items included four for "use for providing care and routine reporting", three for "use for management and administration" and another three for "use for information searching and collaboration with colleagues". CHCs' attitudes toward health IT were measured with 15 items representing four constructs: 1) performance expectancy (*PE*), 2) effort expectancy (*EE*), 3) social influence (*SI*) and 4) intention to use IT (*IN*), drawn from the Unified Theory of Acceptance and Use of Technology (UTAUT) model[17]. *PE* was defined as the degree to which an individual believes that using the system will help him or her to attain gains in job performance. *EE* was defined as the ease with which the system could be used. *SI* was defined as the degree to which an individual perceives that important others believe he or she should use the system. Intention to use IT was defined as the intention to use a computer system. *PE*, *EE* and *SI* each was measured with four items whereas *IN* was measured with three items. Each item was rated by a seven point scale ranging from one (the most negative perception) to seven (the most positive perception).

The pilot test was conducted with 36 part-time students of Faculty of Public Health at Naresuan University. These students are CHC officers from various parts of Thailand. SPSS V14.0 statistics package was used to perform descriptive statistics and ANOVA analysis.

Results

The response rate was 82% with 1,323 out of 1,607 CHCs responding. Of the 1,323 respondents, male and female were represented in essentially equal proportions (54% and 46%) and average age was 36 years old (SD=7, range 17-59). Three quarters (74%) of the respondents had a bachelor degree and one-third (32%) held a public health administration officer position. Nearly all respondents (99%) were in the middle level of government officer's classification position with 39% in the lower middle and 60% in the upper middle level. This means that they were not new employees but rather had worked in centers for several years, and had worked their way up through the position classification system. They spent aproximately 40% of their work hours providing health services and almost equal proportion of work hours in data management and report production. [Table 1] There was a slight difference in how the respondents were spending their work hours. CHCs in the south exhibited a slightly higher percentage of work hours in data management and reporting activities than the other regions.

Regarding IT resources, on average there were 3.7 persons (median = 3, range = 1-17) working in CHCs and most of these working persons (80%) used computer. [Table 1] Virtually all CHCs have at least one computer system. Only two CHCs (0.02%) reported that they did not have any computer system. On average there were approximately two computer workstations per health center and 36% of the CHCs had a local area network (LAN). Less than half of the CHCs could connect to the Internet and 83% of the connections were through a dial-up modem. Internet connectivity was slightly different in different parts of the country, with the highest proportion in the north and the lowest in the northeast. However, in the central region, the Internet connectivity of CHCs in the province adjacent to the capital, Bangkok was 93.36%. Despite only 47% of CHCs having Internet connectivity, 92% of the respondents reported that they had experience using the Internet. They may have used the Internet at a local Internet café, in educational institutions or at home. Moreover, there was no difference in Internet experience between regions. It is evident that IT is pervasive in CHCs and had diffused to all parts of the country.

Mean scores of the three different types of IT use constructs were 3.6 for "use for providing care and routine reporting", 3.2 for "use for management and administration" and 2.6 for "use for information searching and collaboration with colleagues." Cronbach alpha reliabilities were 0.75, 0.66 and 0.77 respectively. [Table 2] CHCs' personnel frequently used IT for providing care and producing reports, but used IT less in communication and information searching. Moreover, this pattern was similar across regions. This finding is consistent with the result

Table 1 - IT resources in CHCs and work hours spent by personnel in CHCs

Variables	Total	Central	North	Northeast	South	Statistics
1.Number of Community Health Center(CHCs)	1,607	376	377	708	146	
Number of response (% response)	1323(82%)	350(93%)	338(90%)	541(76%)	94(64%)	
2. Number of personnel /CHC - Mean	3.68	3.57	3.86	3.65	3.57	ANOVA *p=0.04* (df=3,1313)
3. Number of personnel using computer -Mean (% of number of personnel)	2.95 (80%)	2.98 (83%)	2.99 (77%)	2.9 (80%)	3.0 (84%)	ANOVA *p=0.54* (df=3,1220)
4. Computer systems in CHC (% of CHCs)	1321 (99.8%)	350 (100%)	338 (100%)	540 (99.8%)	93 (98.9%)	2, *p= 0.10 (df=3)*
Number of computer machines (PC, laptop)	2.10	1.91	2.11	2.16	2.19	ANOVA *p<0.01** (df=3,1305)
LAN (% of CHCs)	478 (36.1%)	113 (32.3%)	160 (47.3%)	175 (32.3%)	30 (31.9%)	2, *p<0.01* (df=3)
5. Internet connectivity in CHCs (% of CHCs)	621 (47.0%)	190 (54.3%)	215 (63.6%)	169 (31.3%)	47 (50%)	2, *p<0.01* (df=3)
6. Reported experience using the Internet	1214 (91.8%)	318 (90.9%)	318 (94.1%)	492 (90.9%)	86 (91.5%)	2, *p=0.35* (df=3)
7. Work hours spent (% of total work hours)						
Providing healthcare	39.1%	40.3%	37.2%	39.5%	39.2%	ANOVA *p=0.17* (df=3,1295)
Data management and reporting activities	39.8%	38.6%	41.4%	38.9%	44.0%	ANOVA *p<0.01** (df=3,1295)
Management & other activities	20.1%	21.0%	21.4%	21.6%	16.8%	ANOVA *p=0.04** (df=3,1295)

* Tukey HDS post hoc test also significant

that respondents spent 80% of their work hours providing health care services and performing data gathering and reporting activities. The use associated with information search and communication was relatively high in the north, which had higher Internet connectivity than other regions. The frequent use of health IT is confirmed by the finding that 92% of respondents reported that they have more than 3 years of experience using computers and two thirds of them use a computer more than one time per day. The results indicate that the computer is an integral part of the CHCs' work.

The mean score of *PE, EE, SI* and *IN* were 5.6, 5.2, 5.1 and 5.6 with the Cronbach alpha reliability 0.96, 0.95, 0.91 and 0.98 respectively. [Table 3] It appears that CHCs' personnel exhibited relatively strong positive attitudes towards health IT. They expressed high IT acceptance. In addition, the findings are similar among regions.

We also provided respondents with the opportunity to express their opinions and suggestions about health IT in CHCs. Half of the respondents (652 out of 1,323 respondents) provided comments. Content analysis of those comments reveals three themes. First, they perceived that IT is vital to their work environment. It helps them do their daily work. Second, they were overloaded with data management and reporting demanded from both local and central administrations. This result is supported by the finding that they spent almost half of their work hours doing the data and reporting activities. Sometimes they have to enter same data into multiple reporting systems. The final theme they perceived was that support from higher levels of the administration was inadequate. They suggested that local and central administrations should provide more technical support, IT training and expand infrastructure such as Internet connectivity and provide more frequent hardware upgrades.

Table 2 - IT use associated with activities in CHCs

Computer use	Total	Central	North	Northeast	South	Statistics
Use computer > 1 time /day : % of respondents	66.4% (866/1305)	55.6% (193/347)	73.1% (245/335)	70.1% (371/529)	60.6% (57/94)	2, *p<0.01* (df=3)
Have used computer ≥ 3 years : % of respondents	92.3% (1025/ 1306)	86.2% (299/347)	96.7% (326/337)	93.0% (491/528)	94.7% (89.94)	2, *p<0.01* (df=3)
Use for providing care and reporting activities: Mean score (SD)	3.6 (0.6)	3.5(0.5)	3.6(0.6)	3.5(0.6)	3.4(0.7)	ANOVA *p=0.06* (df 3,1198)
Use for management and administration activities: Mean score (SD)	3.2(0.7)	3.1(0.7)	3.2(0.6)	3.2(0.6)	3.0(0.6)	ANOVA *p=0.01* (df=3,1055)
Use for information searching and collaboration with colleagues (e.g. email) : Mean score (SD)	2.6(0.8)	2.5(0.7)	2.9(0.7)	2.6(0.8)	2.5(0.8)	ANOVA *p<0.01** (df=3,651)

Measurement scale 1 = Never use, 2 = Sometimes use, 3 = Frequent use and 4 = Always use
* Tukey HDS post hoc test also significant

Table 3 - Attitudes towards Health IT

Attitudes	Total	Central	North	Northeast	South	Statistics
Performance Expectancy (*PE*) Mean (SD)	5.6 (1.8)	5.7 (1.7)	5.6 (1.8)	5.6 (2.0)	5.4 (2.0)	ANOVA *p=0.56 (df=3,1284)*
Effort Expectancy (*EE*) Mean (SD)	5.2 (1.6)	5.2 (1.6)	5.2 (1.6)	5.2 (1.7)	5.0 (1.6)	ANOVA *p=0.62 (df=3,1279)*
Social Influence (*SI*) Mean (SD)	5.1 (1.7)	5.2 (1.6)	5.1 (1.6)	5.1 (1.8)	5.0 (1.5)	ANOVA *p=0.73 (df=3,1283)*
Intention to use IT (*IN*) Mean (SD)	5.6 (1.8)	5.8 (1.7)	5.6 (1.8)	5.6 (1.9)	5.5 (1.9)	ANOVA *p=0.44 (df=3,1292)*

Measurement scale 1 = Strongly Disagree, 2 = Quite Disagree, 3 = Slight Disagree, 4 = Neither Agree nor Disagree, 5 = Slight Agree, 6 = Quite Agree and 7 = Strongly Agree

Discussion

It appears that computer and communication technology is currently in heavy use by CHCs across the country. Information technology utilization penetrates and diffuses to CHCs in every region. CHCs have adopted health IT as part of their work. These results contrast to the previous perception of Ministry of Public Health that CHCs might be slow to adopt health IT because majority of CHCs are in rural areas. People who work in CHCs applied computer technology to conduct and enhance the health services they provide and reporting tasks they perform. This might result from the influence both from within MOPH and from rapid social change toward globalization in Thai society as a whole. Since the universal coverage scheme was implemented, both local and central health administrators have increasingly demanded numerous timely reports of health activities for planning and management. These demands come from reporting systems from various departments and institutions with minimal integration or interaction – typical silos. Utilizing a computer system enables CHCs to accommodate such demands with less effort than they might otherwise need. Relatively high positive attitudes toward health IT among CHCs indicates their high degree of IT acceptance. They demonstrate high receptivity to the use of IT in healthcare. The fact that the majority of the respondents were young with college degrees, and have been exposed to computer technology for years might be a significant contributing factor to the high IT acceptance. However, the study findings illustrate other challenges that might threaten the success of health IT project implementation. The success of the systems

might be threatened by connectivity problems, if the country's goal is to develop integrated health information systems, since over half of the CHCs do not have any ability to connect directly to the Internet. Bandwidth is also a potential problem, since most of those CHCs that are able to connect are using dial-up as the mode of access. The communication infrastructure to all CHCs needs to be expanded. The other challenges are related to the data management and reporting overload including the perception of inadequate IT training and IT resources from the higher level of the organization. These issues should be considered for further studies.

Although this survey was designed to be representative of all CHCs nationwide, our assumption that the respondents represented the CHC (which is the unit of analysis) might be a potential weakness. However, given the consistency of the responses we do not believe the results have been compromised by this assumption or its occasional violation.

Conclusions

Thai health administrators and policy makers anticipate having an effective information system to support the healthcare system similar to administrations in others countries. Lessons learned from previous national health technology implementation projects suggest that user acceptance is the major determinant of project success. We surveyed a representative sample of Thai CHCs nationwide to find out the extent of IT penetration among CHCs and to evaluate their attitude towards and acceptance to IT. The data demonstrate that health IT is pervasive and there is a high degree of IT acceptance and use in CHCs across the country. The study results suggest that many of the necessary socio-technical conditions are already in place, and that the potential for their interference with a successful implementation of the national health information system is smaller than anticipated. However, the inadequate Internet connectivity and the data management and reporting overloads are the challenges that might interfere with the success of health IT projects.

Acknowledgements

The authors gratefully acknowledge Dr.Suvat Siasiriwattana, Mrs Aree Wadwongtham, Dr.Narong Kasitipradith and the staff at the Bureau of Health Policy and Strategy, Ministry of Public Health, Thailand for their great support.

The authors thank Ms.Parunee Yimsabai and staff at the Center of Health Equity Monitoring (CHEM), Faculty of Medicine, Naresuan University, Thailand for helping the survey process. The primary funding source of this project was from WHO Thailand, project activities proposal number 040199.

References

[1] Pannarunothai S, Patmasiriwat D, and Srithamrongsawat S, Universal health coverage in Thailand: ideas for reform and policy struggling, Health Policy, 2004. 68(1): p. 17-30.

[2] Tangcharoensathien V, Pitayarangsarit S, and Vasavid C, Universal health care coverage and medium term financing implications., International Health Policy Program-Thailand: Nonthaburi, 2002

[3] Towse A, Mills A, and Tangcharoensathien V, Learning from Thailand's health reforms. BMJ, 2004. 328(7431): p. 103-105.

[4] Wibulpolprasert S, Siasiriwattana S, Ekachampaka P, Wattanamano S, Taverat R (Editor), Thailand Health Profile 2001-2004, Bureau of Policy and Strategy, Ministry of Public Health.Thailand,2005

[5] Chandrasekhar CP and Ghosh J, Information and communication technologies and health in low income countries: the potential and the constraints, Bull World Health Organ, 2001. 79(9): p. 850-5.

[6] Kaplan B, Shaw NT, Future directions in evaluation research: people, organizational, and social issues, Methods of Information in Medicine, 2004. 43(3): p. 215-31.

[7] Littlejohns P, Wyatt JC, and Garvican L, Evaluating computerised health information systems: hard lessons still to be learnt, BMJ, 2003. 326(7394): p. 860-3.

[8] Lorenzi NM and Riley RT, Organizational issues = change, Inter J Med Inf, 2003. 69(2-3): p. 197-203.

[9] DeLone WH and McLean ER, Information systems success revisited, Proceeding of the 35th Annual Hawaii International Conference on System Sciences (CD ROM) 2002.

[10] Giuse DA and Kuhn KA, Health information systems challenges: the Heidelberg conference and the future, Inter J Med Inf, 2003. 69(2-3): p. 105-114.

[11] Whetton S, Health informatics: a socio-technical perspective, Oxford; New York: Oxford University Press. xxv, 262 p. 2005

[12] Information and Communication Technology Center, Ministry of Public Health, Ministry of Public Health's Information and Communication Technology Master Plan 2004-2006(Thai Language). [Web page] [cited November 29, 2006]; Available from: www.morph.go.th/ict/data/MOPH_ICTPlan.zip.

[13] Bureau of Policy and Strategy, Ministry of Public Health, Health Policy in Thailand 2006, Ministry of Public Health Thailand. [Web page] [cited November 29, 2006]; Available from: http://203.157.19.191/HealthPolicy6.pdf.

[14] Lorenzi NM, Riley RT, Blyth AJC, Southon G, Dixon BJ, Antecedents of the people and organizational of medical informatics: Review of the literature, J Am Med Inform Assoc, 1997. 4(2): p. 79-93.

[15] Dixon DR, *The behavioral side of information technology.* Int J Med Inf, 1999. 56(1-3): p. 117-23.

[16] Kijsanayotin B., Pannarunothai S, and Speedie S, Development of an instrument to measure health enter (HC) personnel's computer use, knowledge and functionality demand for HC computerized information systems in Thailand, Proceeding of the AMIA 2005 Annual Symposium(CD ROM), Washington DC, 2005

[17] Venkatesh V, Morris MG, Davis GB and Davis FD, User acceptance of information technology: Toward a unified view. MIS Quarterly, 2003. 27(3): p. 425-478.

Address for correspondence

Boonchai Kijsanayotin MD., MS.
Health Informatics Graduate Program,
Medical School,
University of Minnesota Twin Cities,
420 Delaware, Minneapolis, MN 55455
Email: kijs0001@umn.edu

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

The Health Informatics Center of Acadiana – Informing Health Policymaking in Post-Katrina/Rita Louisiana

L. Philip Caillouet Ph.D.[a]

[a]*Health Informatics Center of Acadiana, The University of Louisiana at Lafayette, Lafayette, Louisiana, USA*

Abstract

A "healthy communities" initiative in Louisiana led to creation of the Health Informatics Center of Acadiana (HICA) at The University of Louisiana at Lafayette, in the south central United States. Since hurricanes Katrina and Rita devastated the Louisiana coast in 2005, HICA's role has taken on heightened significance. HICA identifies vulnerable populations, documents their risk factors, and evaluates interventions intended to improve community health. HICA collaborates with the Louisiana Department of Health and Hospitals and the Lafayette Community Health Consortium (LCHC), the latter formed for coordination among local healthcare providers and agencies. Both HICA and LCHC were created when "Bonne Santé à Lafayette!" – a locally developed community health improvement plan – was implemented. This paper reports on methods and experiences of HICA and LCHC, offering these as models for addressing community concerns elsewhere. Of special interest is the discussion of Louisiana HABITS, a consumer survey methodology that HICA has developed to measure healthcare access barriers, to provide information that healthcare organizations and governments need to implement workable business strategy and public policy.

Keywords:

public health informatics, health policy, access to health care, health informatics, public health

Problem: Poor health or poor health policy?

Health insurance and healthcare in the USA

In the pluralistic United States healthcare system, healthcare services are offered principally by private-sector professionals and institutions rather than through government agents. These entities are sustained with cash flow margins generated by market-driven compensation for their services. A wide variety of health insurance arrangements reduce the financial risks of prospective patients with varying benefit definitions, limits of coverage, and modes of oversight. Most U. S. residents obtain health insurance coverage for themselves and their dependents through employer-sponsored plans, the premium costs of which are most often shared between the employee and the employer. This somewhat haphazard scheme has been entrenched since the period immediately following World War II, but in recent years a steady increase in the number of uninsured Americans has been of concern to policy makers and ordinary citizens alike. Many federal agencies are involved in the enforcement of public health policy and management of categorical problems related to disease control and prevention, health resource distribution, and health promotion – the latter currently through the U. S. Surgeon General's initiative called Healthy People 2010. [1] However, local management of those programs is largely left to the individual states, as is the dilemma of what to do about the uninsured and under-insured.

Health insurance and healthcare in Louisiana

In the year 2000, Louisiana had a population of 4.5 million. [2] According to data published for that year [3], approximately 19.1% of Louisiana citizens lacked health insurance in 2000. These persons might not have been able to find a private provider willing to treat them unless they are able to pay out-of-pocket for their healthcare costs. An additional 16.2% of the population was insured through the Medicaid program in the same year, meaning that their income and resources were meager. Under Medicaid, state governments administer and share costs with the federal government for health insurance coverage for the qualifying poor. Because Medicaid payments to providers are not always competitive with those from other forms of coverage, providers are often reluctant to see Medicaid patients. Compounding the problem, it is increasingly difficult to recruit and retain private providers willing to locate in both rural and urban areas where pockets of poverty or near poverty exist. Thus the uninsured and under-insured in Louisiana, some 35.3% of its residents, are under-served.

While some federal funding is channeled into qualified community health centers (FQHCs), Rural Health Clinics (RHCs), and Critical Access Hospitals (CAHs) in Louisiana, because of the large number of under-served persons Louisiana has felt it necessary to complement services of private providers by offering both ambulatory and in-patient care directly. Louisiana's governments have had a long history of intervention on behalf of the health of its residents. In the eighteenth and nineteenth centuries, Louisiana's sub-tropical climate was an especially good host to a variety of diseases including yellow fever and malaria, which during periodic outbreaks reached epidemic levels. On March 15, 1855, Louisiana became the first state to enact legislation creating a permanent State Board of

Health. Currently, the Louisiana Department of Health and Hospitals (DHH) is responsible for both the Medicaid program and for the direct care network. Public health services are managed centrally at the state level, with nine regional administrators being responsible for implementation of programs. Region size varies from four to twelve parishes (a "civil parish" is the Louisiana equivalent of a county), with regional populations ranging from 280,000 to 1.2 million. [2] Relative autonomy in public health services administration is exercised locally in the city of New Orleans but not in any other locale. Organized within DHH, the Office of Public Health (OPH) operates at least one clinic or "health unit" in each of Louisiana's sixty-four parishes. These health units provide limited disease prevention and health promotion services as mandated through categorically funded governmental programs. Louisiana is the only one of the fifty United States to have a state-owned, state-operated hospital system. The "Charity System" takes its name from one of the oldest continuously operating hospitals in the United States. Charity Hospital in New Orleans – now part of the Medical Center of Louisiana – was founded in French colonial Louisiana in 1736, sixty-seven years before the United States negotiated the Louisiana Purchase. Until the late summer 2005 flooding caused by Hurricane Katrina rendered its facilities unusable, "Big Charity" was the flagship institution of a geographically dispersed including ten hospitals managed through the Louisiana State University Health Care Services Division.

Despite this long history the state appears to be mired in an extended cycle of poor outcomes, with at least one observer ranking Louisiana as having one of the United States' worst records of relative healthiness. [4] It is widely debated as to whether Louisiana's approach represents a progressive mechanism for dealing with the problem of the insured or is merely a remnant of a failed, two-tiered, paternalistic "plantation system," as some privately describe it. In recent years, a new climate of urgency has arisen in Louisiana that, with the full support of the State Health Officer, seeks to encourage communities and regions within Louisiana to define local needs and seek local solutions – without necessarily relying on state government to take the initiative. Because Louisiana has under-invested in health assessment, a consistent theme in these local initiatives has been the lack of actionable information on which to base effective policy changes.

An informed-leadership approach

Community initiatives in Lafayette, Louisiana

Lafayette – the "capital of Acadiana" – is the hub of south Louisiana's famous Cajun culture, which highly values self-sufficiency and independence, but at the same time embodies an enlightened view of community responsibility among physicians and other healthcare providers. Yet, as the commercial center of the fourth largest metropolitan statistical area in Louisiana, Lafayette shares in Louisiana's poor showing among states. A signal event did occur however in 1994 when two competing not-for-profit community hospitals in Lafayette, Lafayette General Medical Center and Our Lady of Lourdes Regional Medical Center, jointly sponsored and conducted a Community Needs Assessment. Out of that project, the Partnership for a Healthier Lafayette ("the Partnership") was later chartered as a "healthy communities" planning group in 1996. The Partnership was formed as a joint project of the United Way of Acadiana, the Greater Lafayette Chamber of Commerce, the Lafayette Economic Development Authority, and Lafayette Consolidated (City-Parish) Government. OPH Region IV in south central Louisiana was a key player in from the inception of the Partnership. The Partnership's chosen methodology was to engage community stakeholders in envisioning the future and identifying what it would take to realize that vision. "Trend-bender Teams" were formed to propose solutions not only in Healthcare but also in the areas of Education, Youth & Family, Teen Pregnancy, Economic Development, Civic Participation & Volunteerism, Community Infrastructure, and Transportation. The report of the Healthcare Trend-bender Team was completed in December, 1997, and published as part of *A Vision for Our Future* in April, 1998. [5]

Besides the Partnership for a Healthier Lafayette, several other healthy community programs were being launched in Louisiana about the same time. Most notably, the Turning Point Partnership project, sponsored under the auspices of OPH, attracted funding from the Robert Wood Johnson Foundation and the W. K. Kellogg Foundation to support the development of Public Health Improvement Plans. Strategic planning has proceeded at the statewide level and in with the participation of local organizations known as "New Orleans, the City That Cares", "the Southwest Louisiana Partnership" (Region V), and "the Northeast Louisiana Partnership" (Region VIII). [6] Unlike in Turning Point, the planning approach used in Lafayette has focused not so much on improving "public health" (i.e., the services of the state agency) and has instead focused on improving "the health of the public." During the local planning sessions, this subtle distinction evolved into a gradual recognition that the traditional methods of data collection and reporting by public health agencies in Louisiana are inappropriate to reporting on community health status, let alone to enabling decision making for measurable enhancements in community health status. It is to the credit of local public health professionals that they were among the first to agree with this notion when it first surfaced.

Informed leadership via health informatics

In Lafayette, a particular emphasis on "informed leadership" emerged. The rationale for the Health Informatics Center of Acadiana ("the Center" or "HICA") stems directly from the development of *Bonne Santé à Lafayette!* (Good Health to Lafayette!), the community health status enhancement initiative of the PHL Healthcare Trend-bender Team. The *Bonne Santé à Lafayette!* program incorporates both structural and action-oriented elements intended to focus the attention of the healthcare providers on identification and solution of community health problems. The two major critically interdependent structural elements of *Bonne Santé à Lafayette!* are HICA and the

Lafayette Community Health Consortium ("the Consortium" or "LCHC"). While the latter mobilizes healthcare community leadership, the former provides a means of informing the decisions and actions of those leaders. To be truly effective however, the LCHC must exercise *informed* leadership, not just leadership. That is where the Center enters this picture.

Creation of the Health Informatics Center of Acadiana

HICA was established at the University of Louisiana at Lafayette in February 1999, in conjunction with the Department of Health Information Management and the Health Care Administration MBA program. The Center acts as a local clearinghouse for to be used in documenting community health status, in identifying vulnerable populations, and in studying risk factors within these groups.

Mission of the Health Informatics center of Acadiana

The mission of HICA is four-fold:

1. Education of health professionals and healthcare administrators at undergraduate, graduate, and continuing education levels;
2. Research into community health needs and into effectiveness of the healthcare community's response;
3. Health status enhancement in Louisiana, by serving as a resource to healthy communities initiatives; and
4. Health policy enhancement in Louisiana, by serving as a resource to policymakers related to healthcare access, delivery, and financing.

In pursuit of this mission, HICA leverages both long-standing health information management principles and emerging information and communications technologies to collect, aggregate, analyze, and report community health information. The common need for actionable information cries out for the acceptance of consensus standards in the area of computer applications, communications infrastructure, data definitions, and for sharing of best operational practices. HICA serves as a focal point for that development, and for the implementation of appropriate community health information networks, one goal being the release of community health status report cards.

Activities of the Health Informatics Center of Acadiana

HICA gathers, analyzes, and disseminates information essential to identifying and prioritizing community health needs. HICA was created to address many of the concepts of a "management information system for community health" and will ultimately rely on information networking technologies to achieve its mission. Local empowerment and the facilitation of community priority setting should be greatly enhanced in comparison to the current approach of shipping data on locally addressable issues to state and federal agencies only to have it return as out-of-date printed statistics! HICA has even being viewed as "good for business," [7] in terms of its potential for fostering an environment where the community can mobilize to eradicate inequalities in access to healthcare and thereby a healthier labor force. A strong relationship is evolving between HICA and emerging "healthy communities" programs in the area, particularly with the Bayou Teche Community Health Network (ByNet) and the Vermilion Community Health Network (VNET). HICA is working closely in academic-industry partnerships with other groups, including the Emergency Medical Services (EMS) Council of the Lafayette Area, to find funding for projects to obtain, aggregate, analyze, and report on health services utilization data in Region IV.

Related academic programs

The Center also complements the healthcare-related educational and research missions of the University while functioning as a resource to the healthcare community in the immediate area and throughout the state. In Lafayette, both The University of Louisiana at Lafayette (UL Lafayette) and Louisiana State University Health Sciences Center (LSUHSC) each prepare students in the health professions, allied-health fields, and health care administration locally. UL Lafayette currently provides baccalaureate curricula in Nursing and Health Information Management (BS), plus master's degrees Business Administration (MBA) with a Health Care Administration option and Nursing (MSN), with additional preparation for as nurse practitioner (NP) certification. LSUHSC conducts postdoctoral (MD) residency programs for recent graduates. Members of the UL Lafayette faculty have also participated in the continuing education of public health professionals in various settings. [8] HICA serves as a focal point for interdisciplinary curricular and research activities for students and faculty and expects to be an integral aspect of a proposed interdisciplinary graduate program in Health Informatics.

Results – research themes & projects

Research at the Health Informatics Center

Research at HICA has taken several directions. Community health enhancement issues are most clearly associated with what can be called public health informatics. Since Louisiana ranks at or near the bottom of entirely too many categories within which public health is commonly measured, many research questions come to mind immediately. What scales were used? What measurements were taken? Can Louisiana really be that bad off? What are other states doing to better their lot? How could Louisiana get off the bottom of the list? Graduate student research papers have already begun to examine these questions. Louisiana is very fertile ground for advanced research in the public health informatics field, and the existence of HICA will aid in reaping a significant harvest, appealing especially to faculty and students in health care administration, communications, marketing, biostatistics, epidemiology, and public policy.

Louisiana HABITS

HICA is gradually becoming a clearinghouse for data to be used in documenting community health status, identifying vulnerable populations, and studying risk factors in need of remediation within these groups. Reliable parish-by-parish health risk data has not been generally available in

Louisiana. While OPH does publish parish statistics based on results of the Centers for Disease Control (CDC)-sponsored national Behavioral Risk Factor Surveillance System (BRFSS), but the small sample size (limited by data collection budgets) of fewer than 140 random-dialed statewide interviews per month does not lend meaningful statistical significance to findings at the parish level. Given the unique blend of ethnicities and cultures in Acadiana, the applicability of these findings to this area was called into question.

Since March 1999, a HICA research team has worked on the refinement and deployment of HICA's own methodology for assessing healthcare access barriers, real or perceived by healthcare consumers. The tool and approach were dubbed *Louisiana HABITS* (Healthcare Access Barriers In The State). *Louisiana HABITS* has been applied in support of seven distinct health assessment projects supported directly or indirectly by the Rapides Foundation, the Robert Wood Johnson Foundation, and the LSU Foundation/Pfizer Inc.

In its planning to understand consumer perceptions and demand for healthcare services, the *Louisiana HABITS* team defined healthcare access barrier to mean difficulty, delay, or failure in obtaining healthcare or prescribed medications in the past twelve months, or current lack of health insurance, experienced by one or more family members in a household. The team took on the task of answering the question: How should the percentage of all households in the general population that have healthcare access barriers be most easily and accurately determined? Fundamental concerns included statistical validity and protection of respondent confidentiality. The team recognized that random-digit-dialed telephone interviews would be preferred, due to low cost when compared to in-person interviewing and high compliance when compared to mailed surveys.

The *Louisiana HABITS* team reviewed the most current U. S. Census Bureau data, to determine the population and number of households and to set criteria for random sample size sufficient to yield 95% predictive confidence, with a maximum error rate of ±10%. A very costly fifteen-fold increase in interviews would have been necessary to have achieved a ±2.5% interval. The UL Lafayette-developed computer-assisted *Louisiana HABITS* consumer survey was then employed to gather data from a random sample of at least 96 households with telephones, to determine the proportion of the general population of households that report having a healthcare access barrier.

The *Louisiana HABITS* team was concerned, however, that prior surveys using a telephone-only interviews were inaccurate when a certain fraction of all households have no telephone. A review of 1990 U. S. Census data suggested that a substantial fraction of households (e.g., 11.3% of households in St. Mary Parish, Louisiana) were without working telephones at that time. Although Census 2000 data was not available at the time of the first survey, the team felt it reasonable to assume that the proportion of households without telephones had likely fallen due to the increased use of cellular telephones in the decade of the 1990s.

To find a representative sampling of no-phone households, the *Louisiana HABITS* team conducted in-person interviews at locations in each parish where persons whose households might not have telephones could be readily found. In-person interviewing locations included social services offices and parish health units, and occasionally included other public places such as courthouses, public hospital emergency rooms and clinics, and even rural grocery stores and laundromats. As expected, a strong positive correlation was quickly noted between households without telephones and households with healthcare access barriers, with financial reasons as the cause.

In-person interviewing was continued in each parish until data were obtained from at least 96 households in the barrier population in each parish. Demographic stratifiers such as age, gender, marital status, and education level, and general health of the healthcare decision maker were also obtained, as well as household income, household size and composition, and insurance coverage status. Thus in addition to allowing the computation of a no phone adjustment to the findings of the random-digit-dialed telephone survey, the *Louisiana HABITS* team also gained, through in-person interviewing, statistically significant predictive knowledge of the underlying causes of the barriers, including the following:

- Main reason cited by those reporting a problem obtaining healthcare services,
- Main reason cited by those reporting a problem obtaining prescribed medications, and
- Main reason cited by those reporting lack of insurance, and other pertinent statistics.

Knowledge of these underlying causes for healthcare access barriers has already been helpful in the design of interventions in each the twenty-two parishes where *Louisiana HABITS* has been conducted – Acadia, Allen, Avoyelles, Beauregard, Calcasieu, Cameron, Catahoula, Concordia, Evangeline, Grant, Iberia, Jefferson Davis, Lafayette, La Salle, Natchitoches, Rapides, St. Landry, St. Martin, St. Mary, Vermilion, Vernon, and Winn.

Other notable activities

During the several difficult months of population displacement that followed hurricanes Katrina and Rita in the fall of 2005, HICA promoted deployment of an electronic health record for use with evacuees in shelters, but encountered obstacles due to absence of specific authority within the emergency response structure. Since then HICA has assisted in the reshaping of policy at the local, state, and regional levels with a view to making health information technology more responsive to the needs of the victims of such disasters.

HICA has also undertaken partnerships with public and private health-related agencies to act as data collector and analyst on significant federally funding initiatives:

- Process and outcome evaluation for "Healthy Start in Lafayette Parish," working with The Family Tree – funded by the Health Resources and Services Administration (HRSA) of the U.S. Department of Health and Human Services.
- Process and outcome evaluation for "Safe Schools / Healthy Students," working with the Lafayette Parish School System – funded jointly by the U.S. Departments of Education, Justice, and Health and Human Services.
- Process and outcome evaluation for the "Lafayette Jail Diversion Program," working with the Louisiana Office of Mental Health – funded by the Substance Abuse and Mental Health Services Administration (SAMHSA) of the U.S. Department of Health and Human Services.
- Data collection for the Behavioral Risk Factor Surveillance System (BRFSS), in partnership with the Louisiana Office of Public Health – funded by the Centers for Disease Control and Prevention (CDC) of the U.S. Department of Health and Human Services.

At this writing, HICA is in discussions with the Louisiana Department of Health and Hospitals to act as architect for an Aging and Adult Services single-point-of-entry system. Two additional data collection contracts are also pending between HICA and the Louisiana Public Health Institute (LPHI) for the following projects:

- Data collection for the 2007 Louisiana Adult Tobacco Survey, funded by the Centers for Disease Control and Prevention (CDC) of the U.S. Department of Health and Human Services.
- Data collection for the 2007 Steps for a Healthier New Orleans BRFSS, funded by the Health Resources and Services Administration (HRSA) of the U.S. Department of Health and Human Services.

Conclusions

The synergy catalyzed by the presentation of credible community health information in readily usable form to committed, decisive community healthcare leadership is expected energize future efforts to enhancing its health status for years to come! Much remains to be learned as the Health Informatics Center of Acadiana evolves. Perhaps this model of collaboration among academic, community, industry, and government resources will be replicated in other settings where similar concerns exist.

References

[1] U. S. Department of Health and Human Services. Healthy People 2010: Understanding and Improving Health. Washington: U. S. Government Printing Office; 2000.

[2] U. S. Census Bureau. Ranking Tables for Counties: Population in 2000 and Population, Change from 1990 to 2000 (PHC-T-4). Available at: http://blue.census.gov/population/www/cen2000/phc-t4.html. Accessed July 3, 2001.

[3] Louisiana Department of Health and Hospitals, David W. Hood, Secretary. 2002 Louisiana Health Report Card, Submitted to the Governor and the Louisiana Legislature. April 29, 2002.

[4] ReliaStar Financial Corporation. The ReliaStar State Health Rankings – An Analysis of the Relative Healthiness of the Populations in All 50 States, 1997 Edition. 1997.

[5] Partnership for a Healthier Lafayette. A Vision for Our Future. April 1998.

[6] Turning Point -- Louisiana Style: Thinking Differently, Doing Things in a New Way. In: Transformations in Public Health, a publication of Turning Point: Collaborating for a New Century in Public Health. vol 1, issue 2, June, 1998, pp. 5-6, 8.

[7] Health Informatics Center Proposed. In: Lafayette Business Journal, May, 1998. p. 13.

[8] Partnership for a Healthier Lafayette: A Model for Louisiana's Healthy Communities Initiatives, Presented at: Louisiana Public Health Association 1998 Educational Conference, Lafayette LA, April 16, 1998.

Address for correspondence

L. Philip Caillouet Ph.D., Associate Professor & Director
Health Informatics Center of Acadiana
Post Office Box 41007
The University of Louisiana at Lafayette
Lafayette LA USA 70504-1007
E-Mail: caillouet@louisiana.edu

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Informatics Solutions for Emergency Planning and Response

Elizabeth E. Weiner [a], Patricia A. Trangenstein [a]

[a] *Frist Nursing Informatics Center, School of Nursing, Vanderbilt University, Nashville, TN, USA*

Abstract

Early informatics contributions to the emergency planning and response agenda have focused largely on surveillance of threat detection. A broader assessment of possible informatics contributions unveils that informatics can also contribute to increasing the efficiency in disaster response as well as providing a tele-presence for remote medical caregivers. This presentation will explore current and future roles of informatics in emergency preparedness and response.

Special challenges for data management occur with every emergency or disaster. Tracking of victims, electronic health records, and supply inventory are a few of the contributions that informatics can play during disasters. Modeling of response resources can provide the parameters for more effective decision making. Public relations reporting can be made more accurate if given the information in a timely fashion. Databases provide the infrastructure for reporting of data that can be used to manage volunteers or later be mined to determine the effectiveness of planning and response efforts. As informaticists, we have a moral obligation to contribute to the emergency response agenda worldwide.

Keywords:

disaster response, emergency response, emergency planning

Introduction

All disasters begin at a local level, but some situations exacerbate into regional, national, or international events. The complexities of disaster event management are difficult enough, but are magnified when international response is deemed necessary. According to the United Nations International Strategy for Disaster Reduction (UN/ISDR), 2005 saw an 18% rise in natural disasters.[1] An estimated 160 million people – seven million more than in 2004 – were directly affected by natural disasters.

Natural disasters are not the only threats to world health. Currently, there are around 20 significant ongoing armed conflicts throughout the world. [2] Events of terrorism are reported on a daily basis internationally. In addition, there are manmade and technological disasters. The result is that our planning and response efforts need to be improved at all levels, including better preparation and resource management.

One aspect of resource management is that of human resources. Effective planning and response teams need to be reflective of a broad skill set of professionals who can best contribute. One important contribution can be that of the informaticist. Teich, Wagner, Mackenzie, and Schafer noted that with disaster, terrorism, and response to war, "the need for applied informatics expertise may be more pressing, and more in the public eye, than ever before." [3]

The purpose of this paper is to present current informatics contributions and explore possible future solutions to the challenges brought about by emergency planning and response.

Issues and discussion

Biosurveillance and bio-agent detection

The threat of bioterrorism has brought greater attention to the public health infrastructure in all countries. Public health leaders have long complained that there was an inadequate infrastructure to comply with simple public health demands. When that system was further threatened with a biological attack, the breakdowns in the infrastructure became readily apparent. The most noticeable weakness was the relationship between hospitals and public health organizations. Many times there was minimal reporting of possible trends or disease patterns found in hospitals that might possibly have an impact on all of public health. Healthcare workers are now aware of the implications and consequences of their clinical decisions with respect to the entire community. Informatics has played a major role in designing ongoing systematic collection, analysis, and dissemination of data about disease. While some hospitals continue to use manual tracking systems, the best solutions are those that occur in real time and thus immediately identify areas of concern.

Besides the need for additional funding for public health infrastructure, an additional need has been for policy changes in how information is handled. One such example is the Unified Medical Language System from the US National Library of Medicine. This standardized vocabulary tool helps to share descriptions across vocabularies and even link a new bioterrorism-monitoring vocabulary to other terminologies. [4] The synergy between standardized clinical data models and electronic health records have allowed clinicians to use the Internet to rapidly implement large-scale, multi-institutional clinical data integration.

Lober, Karras, Wagner, et al., led a roundtable discussion during AMIA 2001 to discuss and compare six existing bioterrorism detection systems.[5] Although the systems were developed independently, they had striking similarities in the types of data collected and the overall system architectures. All sites indicated concerns with maintaining security and confidentiality. Most used encryption for data transmission. Several systems used clustering of data codes to define disease prodromes of interest in bioterrorism detection. Differences in the systems related to the fact that the projects represented differing relationships with the public health system. As a result, some systems collected multiple levels of data, while others used the visit or case-report level of detail. Continued work in biosurveillance systems has continued since 2001 with funding in the US from the Centers for Disease Control and Prevention and the Agency for Healthcare Research and Quality.

Increasing efficiency in disaster response

Information sharing

Informatics solutions provide the possibility for improving the speed and quality of information that is shared between and among organizations responding to emergency and disaster events. Of particular importance is connecting those in the field who are directly responding to events with those organizations such as hospitals that will be receiving event victims. One example is Maryland's communication network (known as the Trauma Line), which enables prehospital field care providers to communicate directly with physicians in trauma centers and other referral centers. Data on patient vital signs, estimated time of arrival and means of transport, type of injury, level of consciousness, and priority status is put on a fax notepad linked to a cell phone in the ambulance for transmission to the hospital trauma team [3] Another project, know as "MobiDoc," makes use of wireless technology to create an entirely mobile telecommunication system.[3] A field team can perform multiple charting, vital-signs monitoring, image collection, and other data acquisition tasks for multiple patients. The data are sent securely to the hospital's intranet. More patient data and arrival information coupled with algorithms to make use of the data to balance resources will allow for more efficient care, including a need to increase capacity in local areas.

Even the reporting of victims during a mass casualty event has created challenges. After the attacks of September 11, thousands of family members circulated throughout the hospitals in the area in a futile attempt to locate their family members. There was not one central place for them to access the information. Healthcare members in St. Louis wanted to make certain this did not happen to their community. As a result, they developed a bar code system to log and track their victims. [6] In addition, PDAs were used by medics to log patients and belongings as well as notebook computers with wireless technology and networked desktop machines in command centers.

A report of the 2002 Coastal North Carolina Domestic Preparedness Training Exercise described the innovative use of telehealth technologies for terrorism response. [7] During this exercise, East Carolina University tested the in-place telehealth networks as well as deployable communications, networking, and data collection technologies such as satellite communication, local wireless networking, on-scene video, and clinical and environmental data acquisition and telemetry. Specific recommendations were shared based on their experience.

Information sharing during the response to Hurricane Katrina was hampered as it related to supply needs and inventory. For example, a call would go out that bottled water was needed in a certain region. There was no centralized method to determine if those needs were met and by whom, or how long it was going to take to deliver the needed goods. Inventory codes varied by vendors, so that comparing resources across vendor product lines was essentially impossible.

Information sharing during a disaster is essential as it relates to the media. The media provide the interface to the public at large, and it is important that the information they receive is both accurate, timely, and from the designated authority. Secure communication methods need to be established prior to an event. Appropriate informatics tools can help to make that happen.

Preparation (competency-based learning and simulations/exercises)

Provider training and education are also critical elements of a comprehensive plan for bioterrorism and public health preparedness in general. Researchers at Vanderbilt University have developed online modules for nurses in emergency planning and response. In addition to being competency-based from the International Nursing Coalition for Mass Casualty Education, they also used the "How People Learn" (HPL) format. This format was based on a review of the literature for the National Research Council on how people best learn. Current modules can be found at the following address free of charge: www.incmce.org

Other US funding sources have provided for additional learning materials. The CDC supports the linking of a number of academic institutions to state and local health agencies. The result is the Centers of Public Health Preparedness (CPHP) program (http://www.bt.cdc.gov/training/cphp/). The American Medical Association supports its National Disaster Life Support (NDLS) Program, which is also competency based across levels or expertise (http://www.ama-assn.org/ama/pub/category/6206.html)

In Japan, centers of excellence are also funded and recognized for their expertise in emergency planning and response. Hyogo University is such a designated center, focusing on disaster nursing in a ubiquitous society. Central to their efforts is a research focus addressing the development of an information base, the establishment of a nursing support network throughout Asia and the world, and the development of nursing care strategies which includes the Education/Training Program Development Project. [8]

A unique collaboration in emergency preparedness results in a Masters' degree. Called the European Master in Disas-

ter Medicine, The University of Eastern Piedmont School of Medicine, Novara, Italy, and the Free University of Brussels School of Medicine, Brussels, Belgium, together are the founding universities and actually provide the degree. [9] Associate universities include the Centre for Teaching & Research in Disaster Medicine and Traumatology (Linkoping, Sweden), Centre Hospitalier Universitaire Vaudois (Lausanne, Switzerland), the International Emergency Medicine and Disaster Medicine Sections, Harvard Medical School (Boston, Massachusetts, United States of American [USA]), Yale New Haven Center for Emergency Preparedness and Disaster Response (New Haven, Connecticut, USA), and the Vanderbilt University School of Nursing and National Center for Emergency Preparedness (Nashville, Tennessee, USA). Didactic content is presented in an online format for three months prior to the two week live-in course venue. The live-in venue takes place in a selected site in Italy, where interactive exercises culminate in a community-based drill.

Developers at Dartmouth's Interactive Media Laboratory have developed the Virtual Terrorism Response Academy (VTRA)(http://iml.dartmouth.edu/vtra/). The VTRA is a reusable virtual learning environment to prepare emergency responders to handle high-risk, low-frequency events, partiicularly terrorist attacks.[10] Users have the opportunity to sign in according to their profession and get a mentor/host from their designated profession. The most advanced simulations have been funded and used by the US military. The ultimate simulation environment will allow for individuals to participate in teams, preferably team members with whom they will respond to events. Informatics solutions will allow for feedback to be provided to both individuals and teams.

Volunteers

Healthcare volunteers are a necessary component of disaster and emergency response, but they also create challenges. Issues related to registration and credentialing can be solved with informatics participation. The design of the database is only the first challenge, followed by political issues such as who owns the data and how the data is collected, stored, and shared. Scopes of practice are governed by different groups regionally, nationally, and internationally. Liability issues surface with every response.

One example of volunteer registration and credentialing is the Emergency System for Advance Registration of Volunteer Health Professionals (ESAR-VHP). [11] This US based system is a state-based registry of volunteer health professionals. It allows for verification of credentials prior to a disaster, and provides opportunities for education and training in disaster response. In addition, volunteers are urged to register within one and only one registry so that duplicate records are not generated. It is hoped that this practice will discourage the spontaneous volunteer [12]. The success of any response is dependent on all individuals or teams understanding their roles, responsibilities, and the chain of command within the disaster response. Disaster respondents who deploy within a system that is part of the overall disaster planning usually come with the essential equipment that is needed to complete their mission and to keep the responder safe from hazards that might result from the disaster. A spontaneous volunteer who shows up on the disaster scene may not have the equipment, knowledge, skills, or team work skills to be able to be used in a successful response effort.

Electronic health records

Not all victims of disaster and emergency events have electronic records during non-emergency times. However, for those who do, access to such records could streamline the time needed to administer healthcare. Questions also arise during emergency events as to who owns the record? Particularly for those victims that are displaced, should the record follow the person or go to the next healthcare provider? How will either healthcare provider be identified in order to access a centralized database?

Hurricane Katrina provided such a challenge. Dr. David Brailer, (who was then the National Coordinator for Health Information Technology) pulled together informatics specialists from both academic and vendor backgrounds. These experts created www.KatrinaHealth.org, an online service for authorized health professionals. [13] The web site provided access to evacuees' medication information in order to renew prescriptions, prescribe new medications, and coordinate care. This website provided authorized users with access to the medication history of evacuees who lived in the areas affected by Hurricane Katrina, with data or prescription information made available from a variety of government and commercial sources. Sources included electronic databases from community pharmacies, government health insurance programs such as Medicaid, private insurers, and pharmacy benefits managers in the states most affected by the storm. In less than three weeks after Hurricane Katrina hit the Gulf Coast, this information was up and running. Developers agreed that it should not have taken a major disaster to stimulate this sort of collaborative work. As a result of these efforts, KatrinaHealth.org won the 2006 Pinnacle Award from the American Pharmacists Association Foundation.

The Markle Foundation convened a group of industry and government experts following Hurricane Katrina who prepared the summary report which included the need to:

- Foster immediate discussions regionally and nationally among government health leaders, insurers, healthcare providers, and information technology companies to determine what, how, and when patient medical information can be shared securely and quickly in the event of a disaster.
- Create electronic health information systems that are based on simple, open web standards, so that data can be provided in different formats from different users and still be accessible to all.
- Agree upon a method to authenticate the identities of doctors, pharmacists, other health professionals, and patients using the web site, so that they can quickly and securely access private health information needed for their ongoing treatment.

- Make electronic health information records accessible to nurse practitioners, physician assistants, and nurses who will likely be working with physicians and clinics in a disaster's aftermath, rather than just by physicians.
- Examine federal and state public policies governing privacy and medical records-such as the Health Insurance Portability and Accountability Act of 1996 and existing state privacy laws-to be sure they do not hinder the delivery of medical care for displaced persons post-disaster. [13]

Modeling

The establishment of an Emergency Operations Center (EOC) is one of the first steps towards a positive disaster response. Depending on the nature of the disaster situation, a collection of individuals are gathered to respond and plan next steps of the response effort. Based on standardized job action sheets, these leaders each play an assigned role in order to provide a comprehensive based for effective decision making. Recognizing the importance of technology, some EOCs now include a Technology Coordinator at high levels of decision making.

Although the event begins at the local level, regional, national, and even international responders can be called in to respond. However, it may be as long as 48-72 hours before additional help can actually occur on-site. In the interim, members of the EOC need to be able to make decisions regarding priorities, resources, and next steps. Some factors are known prior to an event, such as how long it will take before pharmaceutical stock piles can be brought on-site. Models that include such "known" entities can be used by these decision makers so that more efficient decisions can be made, perhaps saving lives due to the decreased time. Our colleagues in nuclear engineering have taken advantage of such modeling techniques, which can be adapted for use in both public health and healthcare organizations.

Telehealth applications

Expert assistance may not always be available on-site during a disaster. One method to provide this expertise is through telemedicine technologies. This enables an expert in one location to direct field personnel on-site. It is also feasible for that expert to provide assistance to more than one site.

Remote management of trauma patients also allows for remote guidance of procedures. Teich et.al. found that the use of tele-mentors took longer than if the expert were on site, but concluded that the tele-mentor was better than having no trauma surgeon present at all. [3]

Telehealth applications rely on having an intact communications infrastructure. In some situations, the infrastructure is destroyed, leaving clinicians to resort to less technological solutions.

Possible informatics research

The discussion above has generated a number of areas for further informatics research as related to emergency planning and response. These include the following:

- Efficient development and delivery of more advanced biosurveillance systems
- Development of more accurate algorithms for evaluating epidemiological data
- Continued work on vocabulary standardization
- Design of more regional or national databases related to the electronic health record
- Effectiveness of available online learning resources
- Development of enhanced telecommunication/telehealth technologies
- Development of data mining techniques that would assist in determining the effectiveness of response efforts
- Development and evaluation of appropriate modeling techniques to enhance the efficiency of decisions made during response situations

Conclusions and summary

Effective application of informatics to the challenges brought about by disasters can greatly enhance planning and response efforts. Examples of the appropriate use of informatics have been provided in the areas of biosurveillance, efficiency of response, and telehealth. The responsibility for the inclusion of informatics lies with each individual and organization who can lend expertise. Only then can we continue to play a prominent role in the prevention and management of disaster and emergency events.

Acknowledgments

Thanks to the US Department of Health and Human Services for their continued funding and support for the International Nursing Coalition for Mass Casualty Education, which is hosted by the School of Nursing at Vanderbilt.

References

[1] United Nations Fact Sheet. 2006. Retrieved December 3, 2006 from http://www.unisdr.org/eng/public_aware/world_camp/2006-2007/pdf/Fact-sheet-en.pdf

[2] WHO. The first consultation on nursing and midwifery contributions to emergencies. 2006. Available at: www.who.int/hac/events/22_23November2006/en/index.html.

[3] Teich JM, Wagner MM, Mackenzie CF, Schafer KO. The informatics response in disaster, terrorism, and war . J Am Med Inform Assoc. 2002; 9(2):97-104.

[4] Lober WB, Karras BT, Wagner MM, et al. Roundtable on bioterrorism detection: information system-based surveilance. J Am Med Inform Assoc. 2002; 9(2): 105-115.

[5] Kohane IS. The contributions of biomedical informatics to the fight against bioterrorism. J Am Med Inform Assoc. 2002; 9(2):116-119.

[6] Hamilton J. An internet-based bar code tracking system: Coordination of confusion at mass casualty incidents. Disaster Management & Response 2003; 1(1):25-30.

[7] Simmons SC, Murphy TA, Blanarovicii A, Workman F, Rosenthal DA & Carbone M. Telehealth technologies and applications for terrorism response: A report of the 2002 Coastal North Carolina domestic preparedness training exercise. J Am Med Inform Assoc. 2003; 20: 166-176.

[8] Yamamoto A. Disaster nursing in a ubiquitous society. Japan Journal of Nursing Science 2004; 1(1): 57-63.
[9] European Master in Disaster Medicine. Retrieved December 3, 2006 from http://www.dismedmaster.com/
[10] Dartmouth Interactive Media Laboratory. Retrieved December 3, 2006 from http://iml.dartmouth.edu/vtra/
[11] Peterson C. Be safe, be prepared: Emergency system for advance registration of volunteer health professionals in disaster response. OJIN2006: 3(2): Retrieved December 3, 2006 from www.nursingworld.org/ojin/topic31/tpc31_2.htm
[12] American Hospital Association. Emergency system for advanced registration of volunteer health care personnel – Hospital implementation issues and solutions focus group meeting report. 2005. Retrieved December 2, 2006 from http://www.hrsa.gov/esarvhp/FocusGroupReport0705/default.htm
[13] KatrinaHealth. Website for organization 2005. Retrieved December 2, 2006 from http://www.katrinahealth.org/
News release. Website for KatrinaHealth.org. 2006. Retrieved December 2, 2006 from (http://www.markle.org/resources/press_center/press_releases/2006/press_release_06132006.php)

Address for correspondence

Betsy.Weiner@vanderbilt.edu

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

A Multidiscipline Conceptual Framework for Consumer Health Informatics

Robert A. Logan[a], Tony Tse [ab]

[a] *Lister Hill National Center for Biomedical Communications, National Library of Medicine, NIH, Bethesda, MD, USA*
[b] *Clinical Research Policy Analysis and Coordination Program, NIH, Bethesda, MD, USA (current affiliation)*

Abstract

This paper presents an idealized conceptual framework for consumer health informatics research drawing from complementary disciplines: information science and health campaign research. This synthesis is designed to provide researchers with a flexible model to evaluate current research and inform future studies. Following a description of the major components, we describe a recent evaluation of consumer perceptions of a health information system, Genetics Health Reference. This study illustrates how the framework may be applied to provide some direction and insights into ongoing consumer health informatics research. While this model represents a work in progress, we present it in support of efforts to understand the multidimensional impacts of the public's access to health information. We also discuss challenges that remain to develop a better conceptual understanding of how consumers converge on health informatics services.

Keywords:

health behavior; internet; informatics; communication; information science; consumer satisfaction; evaluation studies

Introduction

Napoli [1] suggests that consumer health informatics research underemphasizes the importance of comprehensive theoretical underpinnings in evaluating how consumers interact with health information services. Napoli [1] encourages informatics researchers to:

- Account for an array of institutional, social, professional, individual challenges that impede interactive consumer-based health informatics.
- Propose comprehensive models that underlie and instruct how research might be conceptually conceived and ultimately conducted.
- Borrow theoretical frameworks from health campaign research in conjunction with more traditional informatics research, such as information sciences.

Napoli [1] and Dutta Bergman [2] agree that a challenge in health campaign and consumer health informatics research is to provide a more holistic conceptual foundation, which would help scholars conceive and critique both existing and future research efforts. Health campaigns are non-commercial, research-based efforts that seek to change a target audience's awareness or basic knowledge about a disease, condition, or public health issue [2].

Baker and Pettigrew [3] explain a conceptual framework also serves as a roadmap, which helps a researcher conceive projects in a multidimensional fashion. Reinforcing the idea that there is nothing more practical than good theory, a broad conceptual framework identifies the areas that should be considered in planning an evaluation and clarifies the conceptual omissions that need to be explained.

In this paper, we present an idealized conceptual framework that integrates components from information science and health campaign research. A recent evaluation of a consumer health information Website is used to illustrate how the application of the framework provides some direction and adds insights in ongoing consumer health informatics research.

Our objective is to "sketch" the conceptual landscape representing and integrating existing theories about consumer interaction with health information media. While this model represents a work in progress, we present it in support of efforts to understand the multidimensional impacts of the public's access to health information and account for the related infrastructural, biopsychosocial, and interactional dimensions that have been identified – especially in health campaign research.

Background

In this section, we summarize foundational concepts from two disciplines that have explored information seeking from two different perspectives: (1) information science and (2) health campaign research.

Information science

The well-established information science literature conceptualizes the information-seeking process (ISP) as a dynamic, iterative series of cognitive processes and physical actions required for satisfying information needs (e.g., [4-7]). Several "core" non-linear, dynamic, and iterative information-seeking states are common to these ISP models:

- Need: identifying/expressing information needs;
- Access: finding relevant information; and
- Evaluation: assessing information that is found.

Briefly, information needs represent gaps in knowledge. The process of seeking "missing" knowledge helps to resolve these needs. However, consumers may have trouble identifying missing knowledge (i.e., knowing what is not known). Challenges include ineffective query formulation resulting from a lack of knowledge about medical concepts and terminologies [8] as well as finding credible information resources, constructing well-formed mental representations, and navigating online systems. The retrieved information is evaluated within the context of the original problem, another potential barrier for consumers in the medical domain.

Thus, from a traditional information science perspective, ISP consists of making sense of information within a dynamic environment. While these models typically emphasize decision-making and other cognitive processes, very few conceptual frameworks accommodate well-identified variables within the health campaign literature, including affective dimensions of consumer behavior [6] and the socio-cultural and economic environment that surrounds media use [7,2].

Health campaign research

Although health campaign researchers have embraced an individual-centered model, in recent years they have been prone to discuss the phenomena of how persons interact with health media (e.g., health news and information sources on the Internet), as a process of convergence rather than information seeking [9]. The term *convergence* is used because it is perceived to bridge both consumer information seeking and an array of more affective-oriented rationales for why people use mass media, often collectively referred to as *gratifications* [10].

Overall, health campaign research theory recognizes an array of interactions that conceptually frame the dynamics that occur in consumer health informatics. These dynamic interactions encompass intrapersonal, interpersonal, demographic, and cultural factors, as well as media source credibility, preparation of messages, and channel characteristics [2].

McGuire [9] and Cappella [11] note that the conceptual development of health campaign research has occurred in three phases. The first phase is similar to the ISP model described previously and emphasizes optimizing an individual's exposure to resources with their information needs [9].

The second phase emphasizes the potential communication barriers inherent in the characteristics of how a health media or interpersonal source is perceived. It encompasses potential communication barriers presented by messages, media channel, receiver (individual) and destination characteristics [10]. In other words, health messages may not be optimally understood by consumers if the message is poorly written, if a written rather than a visual media is used to reach some audiences, if a person has little access to mass media, or if messages are poorly timed with an individual's or target audience's needs. The emphasis in the second phase is on both individual perception and media characteristics [9].

The third phase encompasses: a) immediate social influences, such as the influences of peer pressure and commercial advertising on health behaviors, b) cognitive behavioral factors, such as a person's problem and decision making skills, and c) the degree that a commitment to a specific health behavior requires broader skills and individualized training to foster a healthier lifestyle. This phase includes theoretical models as well as approaches to research health message effects [11]. These include behavior change theories, individual information processing modes, message effects research and immediate systemic factors. It also emphasizes a person's individual cognitions and skills as well as his or her behavioral adaptability and milieu, or immediate social surroundings [11].

Dutta Bergman [2] adds that health campaign research has entered a fourth phase where the emphasis is to introduce macro forces, such as a nation's or region's health resources, its global or national context (e.g. developing versus industrialized, cultural and religious heritage) its economic prosperity and geo-political factors. Macro influences are seen as relevant to a conceptual understanding why health campaigns are accepted or rejected by intended audiences.

Dutta Bergman [2] notes a challenge of the fourth phase is to integrate all four dimensions described above. The model proposed below is one approach for integrating the interaction of the phases of health campaign research with the ISP model.

A multidimensional framework

Drawing from a new generation of relevant health campaign research and other ISP models and theories, we propose a conceptual framework (Figure 1) for visualizing high-level health information dimensions that span and bridge two disciplines. Each major dimension is described briefly and encompasses the literature cited above. The model also includes an important added dimension that we term 'outcome.'

Consumer/individual

This dimension covers how a person responds to health communication messages and notes psychological, motivational, immediate social, family factors, and related applied context. This is discussed in the description of phases one, two, and three in health communication research above as well as ISP research. Psychosocial factors may influence the basic information-seeking process (described previously). For instance, an individual's cognitive abilities, affective state, life skills and existing knowledge of the information problem, domain, or information source is likely to interact with the perception of information need, motivation, or effort spent evaluating retrieved information.

Other variables include individual demographic attributes, social influences, personal goals, and how a person perceives a health message. For example, a message that seems incon-sistent with an individual's personal beliefs (i.e., cognitive dissonance) may contribute to ending an

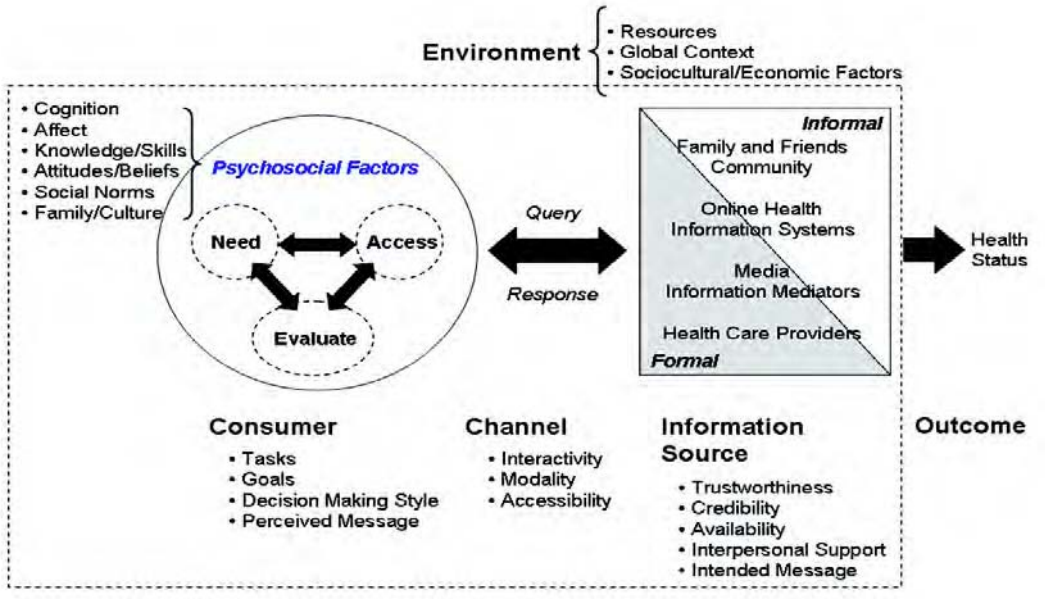


Figure 1 - Conceptual framework for consumer health informatics with five interactive dimensions – Consume (or Individual), Channel, Information Sources, (Macro) Environment, and Outcome

information-seeking session before an information need has been met.

Channel

This dimension describes the influence of the media channel by which health messages are conveyed. This is discussed in the description of phase two of health campaign research above. Channel includes attributes such as the level of interactivity (e.g., unidirectional versus bidirectional), modality (visual, audio, or multimodal), and accessibility (e.g., due to noise or insufficient bandwidth). For instance, some consumers prefer multimedia compared to text (only).

Information sources

This dimension encompasses the perception of the credibility of personal, professional, and media sources and how it influences an individual's reception of health messages. This is discussed in the description of phase two and three of health campaign research above. Many formal and informal sources of health information are available to consumers. Since information sources are typically characterized by perceptions of trustworthiness, credibility, availability, intrapersonal support, and the intended message, health communication from multiple formal (health care providers) versus informal sources (media, health informatics Websites, family and friends) can be a source of confusion for individuals.

Macro environment

This dimension includes the socio-economic, and cultural environment or heritage in which health communication occurs. This is discussed in phase four of health campaign research above. The context, such as available resources due to market and economic forces and socio-cultural trends and traditions, influences all of the other dimensions. For example, the "digital divide" and the impact of religious beliefs impact how individuals perceive health messages.

Outcome

Although health information convergence results in many outcomes (e.g., satisfying an information need), a critical, elusive component is finding therapeutic associations between consumer information access and resulting health behaviors:

> Despite abundant speculation regarding the consequences of consumer participation in interactive health communication, little research has investigated these issues… Ultimately interest and research on effects should focus on quality of health and health care. [12:686-7]

This dimension was added to the above description of ISP and health campaign research.

Overall, each component and its attributes potentially affect health information seeking and convergence on individual behaviors. People do not seek health information in a vacuum—it is integrated into their lives. How and when the need to seek information (perceived or real) overcomes other competing needs, interests, and activities may be based on predictable factors, such as risk-benefit analysis, or simple serendipity. Nevertheless, the design of effective consumer health information systems will likely improve significantly only after researchers better understand the practical role and nature of health information convergence on individual behaviors.

Sample application

While we recognize the complexity of the model and the research challenges posed by multiple interacting components and dimensions, we believe that the framework is useful for presenting a high-level map of areas to be explored. As Greenberg and colleagues state, "the more we know about [online consumer health information seeking] variables, the better we can design educational and technical strategies that help consumers get to the information they seek" [13:1].

Recently, the proposed framework was applied in evaluating the National Library of Medicine (NLM) site, Genetics Home Reference (GHR) [14]. GHR helps consumers understand information about genetic conditions and their related gene or chromosome variations.

A survey was conducted to evaluate perceptions of GHR from consumers' perspectives. Between February and April 2004, 374 members of the Genetic Alliance, an international advocacy group for people with genetic conditions, completed online surveys designed to probe multiple dimensions of consumer perception: content, design, and interface. The survey and data collection were approved by the U.S. Office of Management and Budget.

After consideration, two dimensions (macro environment, channel) were bypassed for investigation because of an anticipated high acceptability of health information seeking on the Internet among the study's participants. Two dimensions (individual and information sources) were pursued. Within the latter dimensions, researchers focused on investigating the perceived uses and gratifications of the Website by exploring participants' thoughts (cognitive) and feelings (affective) about GHR. Researchers also wished to assess participants' assessment of GHR's credibility. The focus on uses and gratifications resulted in questions about the Website's aesthetic appeal and emotions associated with using GHR [15]. Perceived affective dimensions, source credibility, and cognitive uses were derived from semantic differential scales, which are often used in studies of audience media perceptions [9].

By including a range of variables within the individual and information source dimensions, researchers were able to assess whether content quality and aesthetic characteristics in addition to traditional measures such as demographics, online experience, interest and ease of site navigation, predicted user satisfaction.

The study found that content quality and affective dimensions each predicted overall consumer satisfaction [15]. However, age, gender, prior online experience, interest, and education did not predict consumer satisfaction, which is contrary to some previous findings [16].

Further, in a factor analysis that combined all the 13 outcome variables that assessed cognitive, affective feelings and source credibility, the researchers found there were three distinctive perceptual orientations towards GHR [15]:

- Visual design and appeal
- Perceived source credibility/information quality
- Perceived complexity/simplicity and potential bias

In short, participant motivations to use GHR were more holistic than an expected interest to retrieve high quality information about genetics.

The point is that, by using the conceptual framework and integrating variables based on its dimensions, findings about the relative importance of how persons project attitudes onto GHR were better identified. The conceptual framework broadened the research variables that were pursued. In turn, these yielded results that expanded the existing literature about consumer motivations to use a consumer health Website. By initially focusing on a broader conceptual framework, the investigators defined narrow regions to explore and areas to bypass, as well as identified new types of research orientations that yielded surprising results.

Challenges

The health campaign research literature notes an array of methodological challenges, such as understanding audience segmentation and accounting for macroscopic sociopolitical and economic influences on the behavior of both consumers and media organizations [2]. Similarly in information science, there are challenges to capturing the "personal value" consumers place on health information, understanding relevance to particular health needs in the context of situational variables, and tracking the episodic and often serendipitous nature of health information seeking among formal and informal sources. Finally, a dearth of common terms across disciplines and validated instruments for measuring variables hinders consumer health informatics research [17].

Initial methodological challenges we encountered in applying a comprehensive framework include:

- Difficulty in obtaining representative samples online, including underserved populations
- Challenges in identifying and creating operational definitions, and isolating key variables such as "information exposure"
- Lack of standard assessment methods, variables, and operational definitions across research studies

Nevertheless, idealized, comprehensive conceptual frameworks are important in assisting investigators critique the dimensions they have encompassed or eclipsed in modeling consumer informatics research. While the model presented here reinforces this suggestion, it also underscores that even a consumer-centered, psychosocial approach represents only one of the major components of a more expansive, interactive system. It is important to evaluate both motivations for consumer behaviors and operant factors—information source, channel, consumer and environmental. The latter may be useful to explain why intended audiences are drawn or repelled to health Websites and embrace or reject health management and information-seeking concepts in the first place.

Conclusion

Although we may never fully account for the multidimensional spectrum that an idealized model represents, a comprehensive framework adds accountability to social research that fosters consideration of a range of issues and encourages investigator disclosure of the dimensions that are less explored.

An idealized conceptual framework outlines the considerations for which researchers should strive, debate, and defend. A model needs to be comprehensive in order to be ultimately useful and it needs to set a conceptual tradition that is well-grounded in sister disciplines. We hope such a model of consumer information seeking and convergence has been introduced in this conceptually driven manuscript.

Acknowledgments

We thank Dr. Graciela Rosemblat for her assistance and helpful comments. This work was supported by the Intramural Research Program of the NIH, National Library of Medicine.

References

[1] Napoli PM. Consumer use of medical information from electronic to paper media: A literature review. In: Rice RE, Katz JE, (eds). The internet and health communication: experiences and expectations. Thousand Oaks: Sage, 2001:79-98.

[2] Dutta Bergman M. Theory and practice in health communication campaigns. Health Communication 2005 18(2):103-122.

[3] Baker LM, Pettigrew KE. Theories for practitioners: two frameworks for studying consumer health information-seeking behavior. Bull Med Libr Assoc. 1999 Oct;87(4):444-50.

[4] Dervin B. Useful theory for librarianship: Communication, not information. Drexel Library Quarterly 1977;13(3):16-32.

[5] Belkin NJ. Anomalous states of knowledge as a basis for information retrieval. Canadian Journal of Information Science 1980;5:133-143.

[6] Kuhlthau C. Longitudinal case studies of the information search process of users in libraries. Library and Information Science Research 1988;10:257-304.

[7] Marchionini G. Information seeking in electronic environments. Cambridge UK: Cambridge University Press. 1995.

[8] Zeng QT, Crowell J, Plovnick RM et al. Assisting consumer health information retrieval with query recommendations. J Am Med Inform Assoc. 2006 Jan-Feb;13(1):80-90.

[9] McGuire WJ. Input and output variables currently promising for constructing persuasive communications. In: Rice RE, Atkin CK (eds). Public communication campaigns. Thousand Oaks: Sage, 2001: 22-48.

[10] Wimmer RD, Dominick JR. Mass communication research: an introduction. Belmont, CA 2006 Wadsworth, 8th edition.

[11] Capella JN. Integrating message effects and behavioral change theories: organizing comments and unanswered questions. Journal of Communication 2006 56:S265-S279.

[12] Cline RJ, Haynes KM. Consumer health information seeking on the Internet: the state of the art. Health Educ Res. 2001 Dec;16(6):671-92.

[13] Greenberg L, D'Andrea G, Lorence D. Setting the public agenda for online public health: a white paper and action agenda. Washington, DC: URAC [Utilization Review Acc reditation Commission], 2003.

[14] Mitchell JA, Fun J, McCray AT. Design of the Genetics Home Reference: a new NLM consumer health resource. J Am Med Inform Assoc. 2004 Nov-Dec;11(6):439-47.

[15] Logan RA, Fun J, Cheh M. User perceptions of Genetics Home Reference: a study of a health information Web site's image. Intl Comm Assoc Ann Meet. 2006 June.

[16] van der Rijjt GAJ. Determinants of the consumption of health information in the media. Eur J Commun Res. 1998;23(3):255-69.

[17] Bakker TA, Ryce AN, Logan RA, Tse T, Hutcherson L. A consumer health informatics (CHI) toolbox: challenges and implications. Proc AMIA Symp 2005: 21-5.

Address for correspondence

Robert A. Logan, Ph.D.,
Lister Hill National Center for Biomedical Communications,
Building 38A. Bethesda, MD USA. 20894.
Email: logan@nlm.nih.gov

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

A Japanese Model of Disease Management

Naoki Nakashima[a], Kunihisa Kobayashi[b], Toyoshi Inoguchi[b], Daisuke Nishida[c], Naomi Tanaka[c], Hiromi Nakazono[d], Akihiko Hoshino[e], Hidehisa Soejima[e], Ryoichi Takayanagi[b], Hajime Nawata[d]

[a] *Department of Medical Informatics, Kyushu University Hospital, Japan*
[b] *Department of Medical and Bioregulatory Science, Graduate School of Sciences, Kyushu University, Japan*
[c] *Carna Health Support, Japan*
[d] *Graduate School of Medical Sciences, Kyushu University, Japan*
[e] *Saiseikai Kumamoto Hospital, Japan*

Abstract

We started a disease management model, Carna, that includes two programs: one for primary prevention of life-style diseases and one for secondary/tertiary prevention of diabetes mellitus. These programs support the family doctor system and education for participants to allow the concept of disease management to take root in Japan. We developed a critical pathway system that can optimize health care of individual participants by matching individual status. This is the core technology of the project. Under the primary prevention program, we can perform the health check-up/ instruction tasks in the 'Tokutei Kenshin', which will start for all Japanese citizens aged 40 – 74 years in April 2008. In the diabetic program, Carna matches doctors and new patients, prevents patient dropout, supports detection of early-stage complications by distributing questionnaires periodically, and facilitates medical specialists' cooperation with family doctors. Carna promotes periodic medical examinations and quickly provides the result of blood tests to patients. We are conducting a study to assess the medical outcomes and business model. The study will continue until the end of 2007.

Keywords:

disease management, life style disease, diabetes mellitus, critical pathway

Introduction

Disease management is a system of coordinated health care interventions and communications for populations with conditions in which patient self-care efforts are significant [1]. Secondary and tertiary prevention of specific diseases is easier than primary intervention because we cannot predict who will be affected by certain diseases. Therefore, secondary and tertiary prevention is already practiced in the United States [2]. In Japan, however, primary prevention has predominated. This is because the medical institutions and patients have not been adequately motivated to participate in disease management. Medical institutions are independent of insurers, and all citizens of Japan are insured. We have developed a Japanese model of disease management, Carna, which has focused on diabetes mellitus since 2005.

In June 2006, the Japanese Government amended the Medical Care Law to establish a particular health check-up program that includes health instruction (Tokutei-Kenshin). The program will begin in April 2008 [3]. This will be a possible trigger to spread disease management in Japan.

Here, we first introduce the Tokutei-Kenshin system. Then we outline our disease management system and describe the issues involved in system development.

Materials and methods

The Tokutei-Kenshin system

In June 2006, the Japanese Government amended the Medical Care Law to establish an annual health check-up/ management system for all citizens aged 40–74 years. This system will greatly affect insurers because it will cover 57 million citizens (45% of entire Japanese population) and will involve all insurers from April 2008. Insurers will be penalized economically from 2013 when stated goals are not achieved. All Japanese people have basic health care insurance. Therefore, this law will be a very important and have a significant influence on health management in Japan. However, it remains unclear who will provide what kinds of services, even though the government has provided a basic framework for the system.

The government has often told insurers and citizens about the significance of measures against lifestyle disease, and about the Tokutei-Kenshin after the law was passed. Covered citizens can refuse to receive a health-check up. Some insurers are complaining, although the law has been passed, as they are the responsible entity. The health check-up rate will increase by degrees.

The annual health check-up

The annual health check-up involves:

1. Questionnaire (enquiry pertaining to weight change, smoking, exercise).

2. Physical examination (height, weight, body mass index, waist, blood pressure).
3. Blood chemistry tests (triglyceride, HDL-cholesterol, LDL-cholesterol, GOT, GPT, -GTP, Cre, blood glucose (fasting or postprandial), HbA1c, uric acid).

Stratification involves two steps. On the basis of results, participants are assigned to one of three risk groups. The waist and body mass index determine the first assignment. Risk factors identified by blood chemistry tests and smoking history are tallied, and participants are assigned to the 'information provided groug', 'motivation support group', or the 'aggressive support group' (Figure 1).

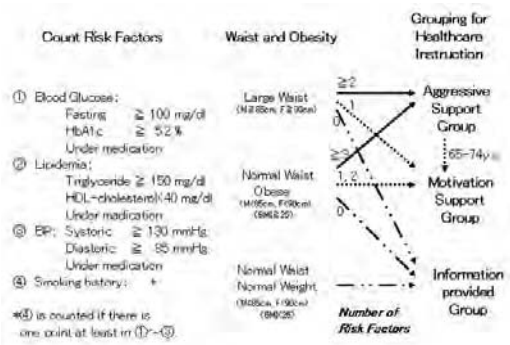


Figure 1 - Stratification into risk groups according to the results of physical examination and blood chemistry tests.

Insurers need to provide health care instruction once to individuals assigned to the motivation support group, and repeatedly to individuals assigned to the aggressive support group. The information provided group is not given health instruction. A physician, a health nurse or a registered dietitian should provide the first health care instruction each year. Non-face-to-face instruction via telephone or information and communication technology (ICT) system is allowed for additional instruction for the aggressive support group. Insurers do not need to provide health care instruction to the person who is under medication for lifestyle disease when they are classed in the motivation support group or the aggressive support group.

If the results of the physical or blood examination are out of the normal clinical range described below, the insurer must encourage the insured person to visit a clinic and confirm the visits that occurred each year.

1) Blood glucose

Fasting	≥	126 mg/dl	or
HbA1c	≥	6.1%	

2) Blood lipids

Triglyceride	≥	300 mg/dl	or
HDL-cholesterol	<	35 mg/dl	

3) Blood pressure

Systolic	≥	140 mmHg	or
Diastolic	≥	90 mg/dl	

4) LDL-cholesterol ≥ 140 mg/dl

All data obtained from the Tokutei-Kenshin will be digitized in a standard protocol as HL7 CDA, and in standard cords as JLAC 10. Insurers will have to maintain records for each insured person as long as they remain insured. The records will contain health check-up and instruction data. If the individual changes insurer as they change occupation, the former insurers will transfer the data to the new insurer so there is no gap in the record. With these systems in place, it is anticipated that the Japanese Government will have access to huge amounts of anonymous data that can be used for statistical purposes. Insurers can keep the data without anonymising it because, for example, they have to analyze it with the data from the medical institutions. Insurers will be able to store data more effectively when the online reimbursement project is fully realized in 2011.

The Japanese Government suggests that the insurer pay an additional 10% contribution to the medical costs of those aged 75 year or over if it does not achieve the stated goals, or be rewarded by a 10% discount if stated goals are achieved from 2013. The Japanese Government is asking for a 25% decrease in the number of diabetes and pre-diabetes patients per insurer, for example.

Insurers will be allowed to outsource the required tasks to health care provider companies. Thus, many tasks will be outsourced to Internet data centers.

Initially, insurers will engage a health care provider company for outsourcing. After registration of the covered citizens, the health care provider will begin service. If providers do not produce adequate outcomes, insurers must pay the penalty as they must take overall responsibility.

Strategy for development of a Japanese model of disease management

We first focused on secondary/tertiary prevention of diabetes mellitus as a Japanese model of disease management, Carna. We then developed a program for the prevention of diabetes complications. After the Japanese Government decided upon the Tokutei-Kenshin system, we developed a program for lifestyle improvements. Our program allows insurers to outsource the Tokutei-Kenshin tasks.

Our disease management program is based on an outbound-call center that manages all information obtained from insurers, medical institutions and individual participants. The call center staff (a nurse or a dietitian) uses the telephone, regular mail and e-mail according to their ICT literacy. The call center always functions as the center of information in data management and communication. We store all information securely at an Internet data center.

We collect information on an individual's initial/improved lifestyle (food, exercise, smoking, alcohol, stress, daily sleep, so on) and knowledge about lifestyle diseases. We also collect the result of blood tests and physical check ups (waist, weight, height, blood pressure, and so on). For primary prevention, the Tokutei-Kenshin includes an initial questionnaire-based interview pertaining to lifestyle. In follow-up health instruction for the aggressive support group, we mainly use e-mail by call center staff. In second-

ary/tertiary prevention, call center staff also communicate with patients by telephone or regular mail.

The main target outcome of the primary prevention program is a decrease in the prevalence of pre-diabetes and diabetes. The main target outcome of the secondary/tertiary prevention program is a decrease in the number/degree of diabetic complications. We collect information about these outcomes on a regular basis.

Insurers will be clients of the primary prevention program after the Tokutei-Kenshin starts in 2008. The family doctors/medial institutions will be clients of the secondary/tertiary prevention program. The Japanese medical insurance pays the lifestyle instruction fee to the medical institutions. The model suggests that the outsourcing be paid for from the fees.

We developed the project with high regard for (1) adaptation to the Japanese medical system, (2) quality control, (3) appropriate matching of services to individuals, (4) adaptation to the Japanese political direction in Japan, (5) efficient and secure data management, and (6) ethical considerations.

(1) Adaptation to the Japanese medical system

Medical institutions and citizens in Japan are less motivated than their counterparts in the United States to participate in disease management. Medical institutions are independent of insurers because all Japanese citizens are covered by public medical insurance and allowed access to any medical institutions. Thus, we have placed importance on providing incentives to medical institutions and to citizens, and we call this a 'Japanese model' of disease management.

(2) Quality control

As disease management will involve call center staff who must deal with enormous numbers of participants and medical institutions, we have given much thought to the quality control of services by compiling tasks and developing ICT systems in the call center. Quality control is particularly challenging because of the scale of the new system.

(3) Appropriate matching of services to individuals

It is difficult to manage and improve the disease status, lifestyle and self-care of individuals by uniform instruction/intervention because individuals vary considerably in basic character, family lifestyle, psychology and health/disease status. Thus, we have attempted to personalize the match-up between services and individuals.

(4) Adaptation to the political direction in Japan

The Japanese Government is focusing on preventing lifestyle diseases. The Tokutei-Kenshin is a product of the new policy. It has been necessary to watch the government and we need to adapt our activities to its policies if we expect the disease management business to take root in Japan.

(5) Efficient and secure data management

For ordinary disease management, data management including collection, storage, analysis and feedback are essential. The scale of the Tokutei-Kenshin system means we expect and are prepared to deal with huge amount of standardized data.

(6) Ethical considerations

There are many legal problems and a new social framework is needed for disease management providers who ensure cooperative, convergent and seamless service to patients by multiple medical institutions and health care service providers. We adhere to the Japanese privacy protection law. In addition, we established a new framework for disease management providers in protecting the privacy of patients' health information, and we propose appropriate handling of patients' health information by multiple service providers working together as a disease management consortium.

Results

(1) Strategy to allow disease management to take root in Japan

To motivate clinics and patients to participate in disease management, we support the family doctor system and education for patients.

In clinics, we match doctors and new patients when they are affected with a lifestyle disease. We also attempt to prevent patient dropout by telephone contact, support patient education and detection of early-stage complications by means of periodic questionnaires, and facilitate medical specialists' cooperation according to the timing described in the critical pathway system.

For patients, we promote medical care described in the critical pathway, report the results of blood tests quickly, and provide 'Carna points' as rewards for the patient's efforts (for instance, regular clinic visits) and for improvement in their diabetic condition (HbA1c). We exchange the points for coupons with which they can obtain certified health-related products such as healthy foods and exercise goods.

(2) Critical pathways for quality control for appropriate matching of services to individuals, and for adaptation to the political direction in Japan

We developed a region-related, outcome-oriented critical pathway as the core competency in the call center. We also standardized workflow in the call center calling 'algorithm'. The critical pathways and the parts of the algorithm are digitized. We prepared an education system with structured questionnaires and comprehensive teaching materials that are closely related to the personalized critical pathway.

We had two kinds of outcome-oriented critical pathways by the end of 2006. One is for the primary prevention program (lifestyle improvement program); the other is for the secondary/tertiary prevention program for diabetes mellitus.

Features of the critical pathway for lifestyle improvement program edge are:

- Using five kinds of critical pathway matching each stage of Prochaska stage model (pre-contemplation

stage, contemplation stage, preparation stage, action stage, maintenance stage) [4].

- Matching the framework of the Tokutei-Kenshin (see Figure 2).

Features of the critical pathway for diabetes mellitus (secondary/tertiary prevention) that we have developed [5] are:

- Scheduling of medical services based on clinical guidelines produced by the Japanese Diabetic Society.
- Supporting general care of diabetic outpatients including timely reminders of the need to visit medical specialists such as an ophthalmologist and a diabetologist.
- Using 'the overlay method' to create an optimal personalized critical pathway for each patient. A personalized critical pathway is created by overlay with a basic sheet for regular examination, and optional sheets matching patient's treatments, the severity of the diabetic complications, and the patient's level of knowledge. We can create 2880 different of critical pathways [5].
- Modifying continuously as the patient's condition

(3) Personalized communication based-on patient characteristics

Patients' responses to interventions vary because the characters of patients vary. For successful intervention, we determine the patient's character type during the registration process, and we depend on this information to personalize our communication with the patient. This approach may also decrease the call center staff's stress.

Figure 2 – Algorithm for telephone call frequency based on the Prochaska stage model and the stratification of the Tokutei-Kenshin. Each risk group has a target line on the staging model

(4) Other algorithms and ICT system for efficient and secure data management

We developed another algorithms ICT system as shown in Figure 3. We used an application service provider system for ICT to input participant records.

We are using an Internet data center in Fukuoka city, Japan for database servers. We send data over the Internet via a virtual private network from the call center.

In the primary prevention program, the initial health instruction requires a face-to-face meeting. We tried a teleconference system using the Internet as a non-face-to-face method. We used VIPS teleconference system developed by Kyuden Infocom Co., and it worked without any problems. In the future, we want to use teleconferencing for the initial health instruction because there are many people living in the countryside in Japan who have access to the Internet. This field is fit to use telemedicine [6]. However, the new law does not allow non-face-to-face methods in the initial instruction.

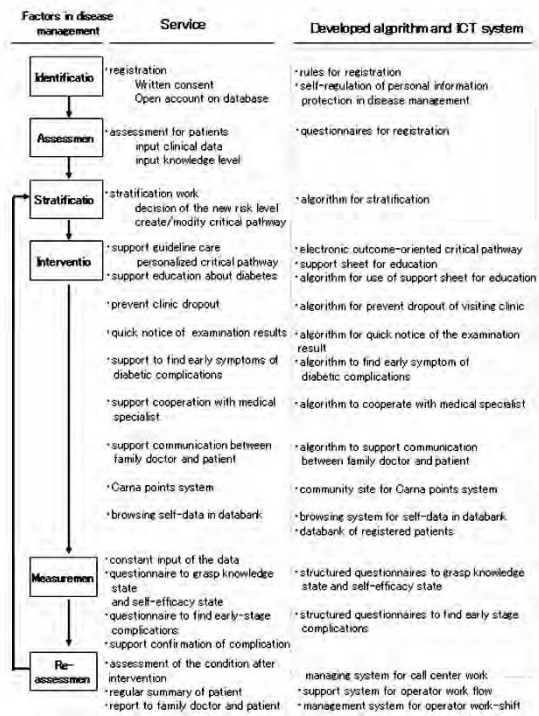


Figure 3 – Disease management services and rules, algorithms, and ICT system in the secondary/tertiary prevention program for diabetes mellitus

(5) Ethical considerations

We quantified the risk of disclosure in terms of information value, the threat arising from inadequate ICT security and areas of vulnerability and showed that the highest risk was posed by databases containing individual patient profiles. Consequently, we need regulations pertaining to the provision of health information and general classifications for various types of patient information that we deal with in disease management work.

We are currently conducting a study to assess the medical outcomes and the business model. The study will be completed at the end of 2007. The ethics committee of the Graduate School of Medical Sciences, Kyushu University approved this study.

Discussion

The Disease Management Association of America defines disease management as a system of coordinated health care interventions and communications for populations with conditions for which patient self-care efforts are significant, and full-service disease management includes the six components: (1) population identification processes, (2) evidence-based practice guidelines, (3) collaborative practice models to include physician and support-service providers, (4) patient self-management education (may include primary prevention, behavior modification programs, and compliance/surveillance), (5) process and outcomes measurement, evaluation and management, and (6) routine reporting/feedback loop (may include communication with patient, physician, health plan and ancillary providers, and practice profiling).

Our disease management 'Carna' meets the definition and includes all six components.

The Tokutei-Kenshin system comes close to meeting the definition of disease management because the Japanese Government focused on metabolic syndrome for the design of the Tokutei-Kenshin system. Thus, we can say that a nationwide disease management project will begin in 2008, although the government has left the details to the insurers who will undertake it. However, typical Japanese insurers are not familiar with health check-ups that are followed by stratification and health care instruction, and they do not have practical knowledge in disease management. Thus, many of them will outsource the task to disease management providers. We fear the possibility that outsourcing will be awarded to inferior providers, and believe the Japanese Government should provide strict recommendations pertaining to the skills of providers who accept outsourcing of the Tokutei-Kenshin tasks.

In the future, we will add a program aimed at chronic disease using the same strategy.

Acknowledgments

This study was supported in 2005 and 2006 by 'a fund to create a new health care service' in the Japanese Ministry of Economy, Trade and Industry. We sincerely appreciate Kyuden Infocom Co., Kyushu Electric Power Co., Tokyo Marine Nichido Co., and Fukuoka Prefecture Medical Association for their support of this study.

References

[1] Disease Management Association of America: Definition of Disease Management, 2003.1.6., available from: http://www.dmaa.org/dm_definition.asp

[2] Rubin RJ, Dietrich KA, Hawk AD. Clinical and economic impact of implementing a comprehensive diabetes management program in managed care. J Clin Endocrinol Metab. 83: 2635-2642, 1998

[3] Japanese Ministry of Health, Labour and Welfare: Particular health check-up/instruction, available from: http://www.mhlw.go.jp/bunya/kenkou/seikatsu/pdf/02.pdf

[4] Ruggiero L, Rossi JS, Prochaska JO, Glasgow RE, de Groot M, Dryfoos JM, Reed GR, Orleans CT, Prokhorov AV, Kelly K. Smoking and diabetes: readiness for change and provider advice. Addict Behav. 24:573-578, 1999

[5] Kobayashi K, Nakashima N, Inoguchi T, Nishida D, Tanaka N, Hoshino A, Hamada M, Kozuma Y, Nakaguma H, Matsushita R, Soejima H, Takayanagi R, Nawata H. Development of a new critical pathway for regional medical network for outpatients with diabetes mellitus. J. Japan Diab. Soc. 49: 817 – 824, 2006

[6] Garcia-Lizana F, Sarria-Santamera A. New technologies for chronic disease management and control: a systematic review. J Telemed Telecare. 13: 62-68, 200.

Address for correspondence:

Naoki Nakashima
Maidashi 3-1-1,
Higashi-ku,
Fukuoka 812-8582,
Japan
E-mail address: nnaoki@info.med.kyushu-u.ac.jp

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Towards Sustainability of Health Information Systems: How Can We Define, Measure and Achieve It?

Sebastian Garde[a, b], Carola M. Hullin[a, b], Rong Chen[c], Thilo Schuler[d], Jana Gränz[a, e], Petra Knaup[f], Evelyn J.S. Hovenga[a]

[a] *Health Informatics Research Group, Central Queensland University, Melbourne VIC & Rockhampton, QLD, Australia*
[b] *Austin Centre for Applied Clinical Informatics, Austin Health, Heidelberg VIC, Australia*
[c] *Department of Biomedical Engineering, Linköping University, Sweden*
[d] *Department of Medical Informatics, University of Freiburg, Germany*
[e] *Faculty of Computer Science, University of Applied Sciences Ulm, Germany*
[f] *Department of Medical Informatics, University of Heidelberg, Germany*

Abstract

Health information systems (HIS) in their current form are rarely sustainable. In order to sustain our health information systems and with it our health systems, we need to focus on defining and maintaining sustainable Health Information System building blocks or components. These components need to be easily updatable when clinical knowledge (or anything else) changes, easily adaptable when business requirements or processes change, and easily exchangeable when technology advances. One major prerequisite for this is that we need to be able to define and measure sustainability, so that it can become one of the major business drivers in HIS development. Therefore, this paper analyses general definitions and indicators for sustainability, and analyses their applicability to HIS. We find that general 'Emergy analysis' is one possibility to measure sustainability for HIS. Based on this, we investigate major enablers and inhibitors to sustainability in a high-level framework consisting of four pillars: clinical, technical, socio-technical, and political/business.

Keywords:

health information systems, electronic health records, sustainability, *open*EHR, computerized medical record systems

Introduction

Information, knowledge management and communication technologies are crucial enablers of system change that can play a vital role in substantially transforming healthcare systems and prevent their failure[1]. From this perspective, we argue further that the sustainability of our health systems depends largely on the sustainability of our Health Information Systems (HIS).

The past few years have seen a myriad of developments and deployments of HIS. Some have very limited focus, other operate on a regional scale, and a few initiatives are underway to establish nationwide HIS, for example in the form of Shared Electronic Health Record (EHR) Systems. The World Health Organisation (WHO) requests international collaboration [1] - good examples include the *open*EHR (http://www.openEHR.org), HL7 (http://www.hl7.org), and the joint Detailed Clinical Models (DCM, http://detailedclinicalmodels.org) initiatives.

With patients being increasingly mobile and treatments and health care providers increasingly specialized, interoperability of HIS has become critical for sustaining current processes. Patient data can be relevant for over 100 years, thus sustainability of patient data is critically important. Not only will sustainability and interoperability save money, we also expect a significant positive clinical impact (cp. e.g. [2]).

While defining HIS failure and success is complex, and current evidence on HIS success and failure rates is insufficient, the "*best current estimate is that HIS failure is an important problem*" [3]. If current customs prevail, very few of these systems will be sustainable, let alone semantically interoperable, costing lives and money. According to Haux, a lot of research and application is necessary to further develop and investigate HIS architectures and infrastructures, in order to identify sustainable approaches [4]. In this context, the aim of this paper is to

- Review general definitions and measurements of sustainability
- Analyse their applicability to HIS
- Investigate major inhibitors and enablers of sustainability in a high-level framework for building sustainable HIS. In this framework, we will relate to the *open*EHR architecture as one candidate for an approach to sustainable development of HIS.

Materials and methods

We conducted a literature review on 'sustainability', a summary of which is presented in this paper. In addition to being heavily involved in many of the following, we

1 Cp. Medinfo 2007 Call for Papers.

reviewed the literature and web resources available on socio-technical issues, Electronic Health Records (EHR), failures of Health Information Systems, recent reports from national EHR initiatives, *open*EHR (http://www.openEHR.org, [5]) as an advanced architecture for EHRs and interoperability, Interoperability Frameworks like the one provided by NEHTA [6], and initiatives like the Detailed Clinical Models initiative.

The enabling/inhibiting pillars for sustainability described in in the second part of this paper are derived from common threads found in the Health Informatics literature on successful and failed systems implementations as well as own experiences. These pillars are related to the structure of Interoperability Frameworks like [6].

Results

Definition and metrics for sustainability

It is believed that the word *sustainability* (German: *Nachhaltigkeit*) was used for the first time in 1712 by the German forester and scientist Hans Carl von Gilinscee in his book *Sylvicultura Oeconomica* [7]. While according to the Canadian *Sustainability Now* initiative more than 300 definitions of sustainability exist (http://www.sustainability.ca), the probably best-known definition stems from the World Council on Environment and Development [8]. It defines sustainable development as that which "*…meets the needs of the present without compromising the ability of future generations to meet their own needs*". From the authors' perspective, the simplest, most generic and compelling definition is found on Wikipedia: "*Sustainability: the ability to continue a defined behavior indefinitely".* However, this simplicity is spoilt, once more specific definitions are required – given that the "the term itself is being applied to so wide a range of issues that it can no longer retain only a single meaning" [9]. If this is the case, how do we know that we have achieved sustainability? In 2003, Maine brought attention to this lack of quantitative indicators of sustainability [9] and a decade earlier the International Institute for Environment and Development ([10], p.2) had already concluded that "*the need for sustainability analysis and particularly for indicators of sustainability is a key requirement to implement and monitor the development of national sustainable development plans […]*".

To quantify sustainability, Maine suggests "*narrowing the use of the term strictly to physical processes, for this appears the only way to achieve the establishment of a secure and robust metric of sustainability […]*" [9]. Consequently, he argues for a "*rigorous metric of sustainability derived from basic scientific principles, and avoiding the application of the term to sociological issues such as the longevity of an organization or society.*" He proposes the energy of reclamation of all outputs of anthropogenic (=derived from human activities) processes as a metric for sustainability. An important attribute any system requires to be sustainable would then be the minimal production of 'waste' or "*the amount of energy that is NOT used to reclaim waste*" [9].

A series of sustainability indicators is based on 'emergy synthesis' as introduced by Odum in [11]. Emergy (with 'm', not 'n') is an abbreviation of the term "embodied energy". Without going into complex mathematical definitions of emergy, it shall be said that emergy expresses the cost of a process or a product in solar energy equivalents, which is regarded as the ultimate energy source. Odum's innovation established a medium for environmental accounting that for the first time made it possible to express economic commodities, services, and environmental work of all kinds on a common basis as emergy [12]. In other words, by expressing the value of products in emergy units (*emjoule*), it becomes possible to compare 'apples and pears' [13].

Once the total number of input flows into a system has been identified and based on this the total emergy driving a process has been evaluated, a set of indicators can be calculated to illuminate different aspects of sustainability as the following important indicator developed by Brown and Ulgiati ([14]):

$$Sustainability\,Index = \frac{Emergy\,Yield\,Ratio}{Environmental\,Loading\,Ratio} = \frac{\frac{Y}{F}}{\frac{N+F}{R}} \quad (1)$$

This index is also called the "Emergy Sustainability Index" (ESI). The *Emergy Yield Ratio* is defined as the ratio of the emergy of the output of the system (Y) and the emergy of purchased services and resources that are input to the system (F). *Emergy Loading Ratio* is defined as the sum of the emergy of local non-renewable sources (N) and purchased resources/services (F) divided by the emergy of the free environmental emergy available from local renewable sources (R).

More recently a joint initiative of Yale and Columbia University, in collaboration with the World Economic Forum and the Joint Research Centre of the European Commission constructed an Environmental Sustainability Index (also called ESI, http://sedac.ciesin.columbia.edu/es/esi/index.html) and compared it to other sustainability indicators such as the Ecological Footprint Index measuring the area of productive land and water appropriated exclusively to produce the resource used and to assimilate the waste generated [15]. Zhao and colleagues further introduce a modified form of ecological footprint calculation by combining emergy analysis with conventional ecological footprint analysis [15].

Sustainability in health and health information systems

Given the long history of sustainability reaching back into the 18th century, it is astonishing that there are no agreed definitions for sustainability of health systems or health information systems – clearly the generic definitions are not sufficient for any measurement and the more specific definitions and indicators of sustainability are not applicable without restrictions to the area of health. Nonetheless we often argue that our health systems are not sustainable as for example Enrico Coiera in a recent paper: "*The health system at present is one that consumes enormous resource, and generates enormous waste, and* ***would not meet any criterion of sustainability****. Injecting new interventions from 'outside' the system, as we currently do in health informatics, is itself not a sustainable approach, as*

the capacity for external designers to meet all the evolving needs of those inside will just never be there" [16]. Thus, intuitively we know that our health information system infrastructure as a whole is not sustainable – however we cannot measure it and thus are not able to take systematic corrective action.

In measuring sustainability of a health information infrastructure, we believe that in analogy to environmental sustainability, it is necessary to 'compare apples and pears' and thus have a common unit like emergy that achieves this for us. In essence, we need to consider what the inputs and outputs and storages of the system under investigation are. We then need to analyse - similar to environmental emergy analysis - which parts are renewable or non-renewable or used non-renewably. Where previously unknown (because no previous studies exist), we need to determine what factors we can use to convert these inputs and outputs into emergy units as detailed by Odum in [11]. Once we have achieved this, all the environmental sustainability indicators that are based on emergy can be applied to health information systems as well. For a given system, we need to identify and analyse the inputs, outputs and stored resources of the system according to Table 1. Other indicators like the ecological footprint seem to be less applicable for HIS.

In contrast to environmental sustainability, the differences between renewable and non-renewable sources however are not always that explicit - e.g. the labour of skilled workers are not always simply renewable – but can be very hard to come by. However, proper training/education programs can make a difference in the long run. For environmental sustainability these resources would be classified as 'used non-renewably' without further distinction [12]. For the analysis of HIS infrastructures, this may not be sufficient and we are currently investigating the use of a 'renewability factor' to rectify this.

Inputs, outputs, storages for health information systems and inhibitors for their sustainability

Table 1 – Analysis of items that are input or output of the system or are stored within the system.

Item	Input/output of the system and any items that stored within it. Some items are the same as typically used for environmental emergy analysis, but others differ.
Data	Raw data measured in joules, grams, dollars or any other appropriate unit.
Solar Emergy per Unit	Factor to transform the data into solar emergy.
Solar Emergy	Calculated: Data x Solar Emergy per Unit

In the following, we investigate these aspects – and especially the enablers and inhibitors for sustainable Health information systems in a high-level framework based on four pillars (each consisting of several high-level building blocks):

- clinical,
- technical,
- socio-technical,
- political & business.

These pillars are related to the structure of Interoperability Frameworks like [6]. Our framework is intended to 'get it right' on a high-level, not about providing all the details – as these details are (with a few exceptions mentioned in the following) relatively well researched.

Some of the inputs and outputs identified are similar to those investigated for environmental sustainability, however some are quite different. This is largely due to the fact that knowledge of various kinds can be seen as one of the most precious resources for us. Table 2 summarises some of the inputs, outputs, and stored items typical for a health information infrastructure, which are untypical from an environmental sustainability point of view and identifies typical inhibitors for sustainability with regard to each item.

Pillar 1: Clinical building blocks

The clinically most important building block for a clinical system is the agreement on clinical content. This fosters semantic interoperability between systems and provides clear meaning – so that we can exchange and migrate data between different systems and support clinical decision making. This clinical domain knowledge needs to be managed and maintained – a complicated task that eventually has become feasible ([5]), although it will always remain difficult on a national or international scale to reach agreement. The separation of technical and clinical concern through *open*EHR's two-level-modelling paradigm seems to be well suited to enable this because it clearly separates clinical content definition from technical concerns.

We suggest the development and international use of a repository of clinical content models that are freely available so that 'flexible standardisation' of this content can occur[2]. For example, *open*EHR archetypes are particularly well suited to serve as a standard form for these clinical content models because archetypes are intuitive to clinicians, but also formal specifications of clinical content technicians can work with.

Pillar 2: Technical building blocks

The technology chosen must be able to cope with the constant changes of health care and health care knowledge without having to change enormous amount of source code (and wait for the vendor to implement it). It must provide the technical basis for semantic interoperability and sustainability. This is not only important because of more and more specialised providers and more and more mobile patients, it also enables the migration of systems without losing considerable amounts of patient data, thus also avoiding vendor lock-in.

2 See http://www.archetypes.com.au for the Archetype Finder, which is designed to support this task as well as the open source Java implementations of *open*EHR at http://www.openehr.org, developed by some of the authors of this paper.

To achieve semantic interoperability, "*[y]ou need the ontology, the information model and services. [..] If you have one and don't have the others, it won't help*" [17]. It is the author's view that ontology and services are relatively well understood – the information model however is largely ignored. This is where an approach like the *open*EHR approach, which is based on a stable and generic information model is of major importance. This works similar to the Java programming language, which decouples itself from the operating system by translating the Java source code to Java bytecode. This code is then run on a native Java virtual machine – thus enabling portability. In a similar way, the *open*EHR two-levelling modelling decouples the technical knowledge (the information model of the software) from the clinical knowledge (expressed in archetypes) to achieve semantic interoperability.

Most importantly, technology needs to be designed to consist of largely independent components, so that replacement can occur without endangering the sustainability of the infrastructure as a whole. Open source implementations will help to validate, improve, evolve their specifications and educate the early implementers [18]. Moreover, if the initial code base is good enough for others to collaborate there is no need to re-invent wheels and thus open source components can serve as building blocks for high level HIS applications. Linux, Apache, OpenBSD, JBoss, Hibernate, Ant and many more are all part of the backbone of our current technical infrastructure and they contribute enormously to sustainability.

Pillar 3: Socio-technical building blocks

While acknowledging the utmost importance of socio-technical issues ranging from comprehensive change management, proper localisation of clinical systems, sufficient training, etc., this has in theory been very well investigated (although not often enough implemented in practice), and "*[s]ocio-technical systems (STS) analysis has provided us with a powerful framework with which to analyse the reasons behind the poor acceptability, uptake and performance […]*" [19]. We therefore refrain here from elaborating on this topic.

Pillar 4: Political/ business building blocks

No matter how much sense this framework makes on a technical, clinical or socio-technical level – sustainability (as well as interoperability) needs to become a major business driver to become reality! We need to ensure that politics is informed and business drivers are 'right'. For this sustainability needs to be measured and a case made for public sector and regional healthcare information systems to be based on open source software to remove the risk associated with any given vendor.

In an era of constant change, political/business decisions do not often hold up long enough to see the rewards of a decision of e.g. implementing a nationwide EHR – as this commonly takes years from planning to roll-out. For this

Table 2 - Examples for inputs, outputs, and stored items in a health information infrastructure – and inhibitors that currently often prevent their sustainability

Sustainability Items	**Description**	**Examples for inhibitors to sustainability**
Health Informatics Knowledge and Skills	Knowledge and Skills of Health Informatitions/ Health IT and IS Professionals. This is vital as we need to work with limited resources to fulfill the great demand.	• Starting similar initiatives from the beginning over and over again, thus losing knowledge inherent in unsustainable systems, which then needs to be recreated • Socio-technical issues, including localisation issues and insufficient change management • Political change and infrastructure not set up appropriately for political change. Independent umbrella organisations may be one solution. • Wrong infrastructure to cope with major changes (e.g. disaster management)
Clinical/ Patient Information	The clinical information on a patient stored in a clinical system. If this information is lost, it has major consequence for patient life and money. We need to sustain clinical data for 100years and more.	• Migration of systems without loss of patient data not feasible (e.g. information model not sufficiently clear) • Vendor lock-in • Insufficient system ability to match clinical practice / workflow • Wrong business drivers • No open source or otherwise available and agreed (e.g. standardised) specifications for data that needs to be shared
Clinical knowledge	The clinical knowledge of Health Professionals – maintained, structured and evidence-based.	• Hard-coded clinical knowledge • Evolvement of clinical practice and general processes which causes systems to slowly become obsolete

reason, we propose (independent) umbrella organisations on a national and eventually internationally level that render political decisions more predictable and sustainable. These organizations can provide leadership and stability for specific purposes.

Finally, while obviously appropriate funding is essential, it does not seem to be wise to provide more and more funding to a system that struggles - without identifying and addressing the fundamental problems within the 4 pillars.

Discussion and conclusion

As shown in this paper, there are many problems to solve in achieving any state, which could be called sustainable. Some are technological, some are socio-political and some organizational. However, most importantly, the lack of a definition for sustainability and agreed and standardised metrics to measure sustainability in health is problematic because without we cannot quantify the status quo of sustainability at any given point of time - and as a consequence sustainability will be largely ignored by decision makers as an important business driver. Indicators generally simplify in order to make complex phenomena quantifiable so that information can be communicated efficiently to decision makers. We thus have to ensure that suitable indicators are applied to make the complex phenomenon sustainability quantifiable for decision makers. In this paper, we showed one possible way of developing such an indicator.

Shabo's Model for the Sustainability of Longitudinal EHRs [20] is one that tackles the concrete problem of sustaining EHRs. As such Shabo's model is more specific to this concrete problem, but also limited to it; his suggestions are in harmony with our results.

While Return on Investment (ROI) is an inherent part of this this paper, its quantification is not part of this paper. For a first estimate, a similar model as the model used by Walker and colleagues to quantify the value of health care information exchange and interoperability [2] could be employed. Also, the World Business Council for Sustainable Development, has formulated the business case for sustainable development and argues that "*sustainable development is good for business and business is good for sustainable development*". The same we believe is true is for the sustainable development of HIS. We need to develop such a business case for sustainability for our health information infrastructure and change business practices on every level. One important step for this is to develop, evaluate and use indicators for sustainability that can be used by decision makers to quantify sustainability to justify expenditures on fighting barriers to sustainability and take a systematic approach towards sustainability of HIS. This can be based on the analysis presented in this paper.

References

[1] WHO. WHA58.28 - Resolution on eHealth. http://www.who.int/gb/ebwha/pdf_files/WHA58/WHA58_28-en.pdf. 2005 (Last access: 08.03.2007).

[2] Walker J, Pan E, Johnston D, Adler-Milstein J, Bates DW, Middleton B. The value of health care information exchange and interoperability. Health Aff (Millwood) 2005: Suppl Web Exclusives: W5-10-W15-18.

[3] Heeks R. Health information systems: failure, success and improvisation. Int J Med Inform 2006: 75 (2): 125-37.

[4] Haux R. Individualization, globalization and health - about sustainable information technologies and the aim of medical informatics. Int J Med Inform 2006: 75 (12): 795-808.

[5] Garde S, Knaup P, Hovenga E, Heard S. Towards Semantic Interoperability for Electronic Health Records: Domain Knowledge Governance for *open*EHR Archetypes. Accepted for Methods of Information in Medicine 2007.

[6] National E-Health Transition Authority (NEHTA). Interoperability Framework. 2006.

[7] von Carlowitz HC. Sylvicultura oeconomica. Anweisung zur wilden Baum-Zucht. Freiberg: TU Bergakademie Freiberg und Akademische Buchhandlung, 1713. Reprint edited by Irmer K and Kießling A, 2000.

[8] UN World Commission on Environment and Development. Our Common Future (Brundtland Report). Oxford: Oxford University Press, 1987.

[9] Maine T. Towards a Metric of Sustainability, in: B. Douglas, B. Furnass and J. Goldie, (Eds.), ISOS online conference - In Search Of Sustainability, 2003.

[10] Dalal-Clayton B. Modified EIA And Indicators Of Sustainability: First Steps Towards Sustainability Analysis. Environmental Planning Issues 1993: 1: International Institute For Environment And Development.

[11] Odum HT. Environmental Accounting: Emergy and Environmental Decision Making. New York: Wiley, 1996.

[12] Campbell DE, Brandt-Williams SL, Meisch MEA. Environmental Accounting Using Emergy: Evaluation of the State of West Virginia. United States Environmental Protection Agency. 2005.

[13] Jorgensen SE, Nielsen SN, Mejer H. Emergy, environ, exergy and ecological modelling. Ecological Modelling 1995: 77: 99-109.

[14] Brown MT, Ulgiati S. Emergy Evaluation of the Biosphere and Natural Capital. Ambio 1999: 28 (6).

[15] Zhao S, Li Z, Li W. A modified method of ecological footprint calculation and its application. Ecological Modelling 2005: 185 (1): 65-75.

[16] Coiera E. Putting the technical back into socio-technical systems research. Int J Med Inform 2006: to appear.

[17] Moore J. A new scheme for data sharing - Semantic interoperability aims to ease data sharing among disparate health systems. Government Health IT 2006: 13. Nov. 2006.

[18] Chen R, Klein G. The openEHR Java Reference Implementation Project, accepted for Medinfo 2007.

[19] Iversen KR, Grotan TO. Socio-technical aspects of the use of health related personal information for management and research. Int J Biomed Comput 1996: 43 (1-2): 83-91.

[20] Shabo A. A global socio-economic-medico-legal model for the sustainability of longitudinal electronic health records - Part 2. Methods Inf Med 2006: 45 (5): 498-505.

Address for Correspondence

Dr. Sebastian Garde
Health Informatics Research Group, Central Queensland University, Austin Health, Melbourne, VIC, Australia
Phone: +61 (0)3 9496 4040
s.garde@cqu.edu.au
http://healthinformatics.cqu.edu.au

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Bermuda Triangle or Three to Tango: Generation Y, e-Health and Knowledge Management

Kwang Chien, Yee

Research Fellow, eHealth Services & research Group, School of Information Systems, University of Tasmania, Australia.

Abstract

Generation Y workers are slowly gathering critical mass in the healthcare sector. The sustainability of future healthcare is highly dependent on this group of workers. This generation of workers loves technology and thrives in stimulating environments. They have great thirst for life-experience and therefore they move from one working environment to the other. The healthcare system has a hierarchical operational, information and knowledge structure, which unfortunately might not be the ideal ground to integrate with generation Y. The challenges ahead present a fantastic opportunity for electronic health implementation and knowledge management to flourish. Generation Y workers, however, have very different expectation of technology utilisation, technology design and knowledge presentation. This paper will argue that a clear understanding of this group of workers is essential for researchers in health informatics and knowledge management in order to provide socio-technical integrated solution for this group of future workers. The sustainability of a quality healthcare system will depend upon the integration of generation Y, health informatics and knowledge management strategies in a re-invented healthcare system.

Keywords:

Generation Y, health informatics, knowledge management, human resource development, technology

Introduction

The healthcare system is at a cross-road. In the next few decades, this industry is undergoing the most extensive transformation ever seen in any industry in history [1]. There are many internal and external factors that drive changes to the healthcare system; some of these factors are very relevant to health informatics and knowledge management:

1. The population that the current healthcare structure serves is rapidly ageing [2]. Patients are getting older and they are more likely to have multiple medical problems with increasing number of prescribed medications. A recent study from Australia has shown that the average age of patients admitted to medical wards is 74 and the average number of prescribed medication is 9.6 [3]. The problem of ageing, increasing complexity and polypharmacy is only going to worsen with time.
2. Secondly, the healthcare system is experiencing rapid subspecialisation, with increasing utilisation of technology. This creates the need for increase communications and information exchanges among various healthcare professionals.
3. Thirdly, advances in technology have become so rapid that it is difficult for healthcare professionals to keep up to date with all areas [4]. Therefore, point-of-care decision support and evidence based guidelines will become increasingly important.
4. The proliferation of knowledge management models and health informatics innovation aims to provide the right information at the right place at the right time. The delivery of information to clinicians at the bedside, however, is often not filtered and therefore an increasing amount of information reaches the desks of clinicians everyday.
5. We have an ageing workforce that needs replacing. This ageing workforce is now being slowly replaced with generation Y workers. Generation Y workers have very different expectations and work practices that a generational gap is widening within the healthcare workforce.

This paper examines the characteristics of generation Y workers applicable to the healthcare system, drawing on the experience from other industries. It then presents a discussion about the potential benefits and pitfalls of working with generation Y in the areas of health informatics and knowledge management. The paper argues a likely adverse outcome if the healthcare re-invention process fails to take these social-cultural issues into account. Finally, the paper proposes some simple rules to deal with generation Y in the health informatics and knowledge management areas and argues that a successful re-invention with generation Y will produce a sustainable healthcare system.

Generation Y and healthcare workplace

Generation Y is commonly defined as those born after 1978 [5], although the exact cut off year is arguable [6]. This generation of young workers has very different understanding and expectation of the world. They are creating a big impact in other industries [7]. As they are now getting

into their 20's, their impact on the healthcare system is slowly being acknowledged. Many characteristics of generation Y have been described and reviewed elsewhere [5-6]. This section will focus on the characteristics which are relevant to electronic healthcare implementations and knowledge management

Technology savvy

Generation Y grows up, surrounded by technology [5]. They see the diffusion of computer from academic research centres into everyday life. They experience the impact of technology not only on their activities of daily living but also career creation and financial stability. Technology has not only transformed their lives, but also provided limitless opportunities for generation Y from "Yahoo" to "Youtube". The healthcare system is unfortunately an industry that is slow to take up technology. When it does utilise technology, however, the story of failure is common. In fact, 75% of big IT projects in healthcare fail [8].

Stimulation and challenging tasks

Generation Y hates routine tasks [9]. They want to have fun during their routine work. They thrive in stimulating and challenging conditions [5]. The industry revolution and now the information age revolution have seen the automation of most manufacturing jobs. Unfortunately, most of the routine day-to-day jobs in healthcare system is still carried out by human beings. Generation Y loves multi-tasking, especially using multiple technological devices. This is a challenging issue in healthcare, especially when quality and safety is taken into consideration.

Information gathering and presentation

Generation Y gather information with lightning speed. They are street smart and they gather information by the fastest means, including the utilisation of technology, but most likely, they obtain essential information through social networking and mobile messaging from their friends [5-6]. Generation Y wants simple information presented in simple language. In generation Y terms: "U get info U wnt 2 r mates only. :)". Unfortunately, this culture does not fit into the formality and hierarchy of healthcare system [4] and generation Yers are perceived to often bypass the usual protocols.

Education and training

Education and training is part of generation Y's life [5]. They are the most educated generation of all time. They, however, learn in a different way. Generation Yers want on-demand, ubiquitous and relevant education and training. These education and training needs are often supported by technology. The healthcare system is however, based heavily on an apprenticeship model [10]. Education and training provided are often irrelevant to their perceived immediate needs.

Fluidity of workforce

Generation Y workers like variety. They are likely to change jobs or institutions frequently [5]. The work model of students working their way through to consultants at the same hospital is being challenged. Generation Yers feel bored when they need to remain in one job for a long time. This creates significant problems in the medical workplace as every hospital is slightly different. While explicit knowledge representation has been emphasised, rising on the backdrop of evidence-based medicine, representation and delivery of tacit knowledge which until now has largely been ignored, becomes difficult because of the fluidity of workforce.

Potentials for health informatics and knowledge management with generation Y

There is significant potential for health informatics and knowledge management to flourish with the increasing critical mass of generation Y in healthcare workforce. Most generation Y healthcare professionals are computer literate and love technology. This section will explore the opportunity for health informatics and knowledge management, working with generation Yers.

Health informatics

Generation Yers swim in technology since birth. They believe in technology! Generation Y finds it easy to adapt to changes and utilise new technology. Given the ageing population, the increased complexity of medical disease management and increased number of prescription medications, implementation of an electronic health record is the only sustainable future. The implementation of an electronic healthcare system will not require significant upskilling of generation Y. In effect, generation Y will probably suggest the use of technology for most of their routine work. They want technology to fit into their needs and their world. This creates more needs for information systems researchers, not only to search for techno-social integrated solution, but also to provide a direct bridge between technology evolution and medical workforce evolution.

The speed of technology roll out offers the necessary stimulation for generation Y to remain interested in the healthcare system and to participate in the redesigning process of the healthcare system. While this might sound like a daunting task to many, generation Yers will find it challenging and they will want to be part of this evolution.

Their familiarity of search engines, multi-tasking and electronic storage system from young age means that generation Y will adapt to electronic information much easier. They will be able to find information quickly through the maze of digital coding. As generation Yers like to improve the efficiency of performing common tasks, they will provide useful suggestions for future healthcare revolution. Their suggestions will likely to be not only useful, but also practical!

Knowledge management

The traditional knowledge management model discusses the stages of elicitation, representation, sharing, evolution and delivery of knowledge [11]. While the theory of knowledge management seems to suggest a distinction between tacit and explicit knowledge [12], in the medical field, tacit and explicit knowledge seem to be inter-con-

vertable, especially with clinical disease management [11].

The request for on-demand knowledge representation by generation Y and the increasing fluidity of the workforce with generation Y are likely to rapidly increase the need for better knowledge management within the healthcare system. The distinction between tacit and explicit knowledge management will become more evident with generation Y.

Firstly, generation Y wants rapid and on-demand access to education, training and clinical decision support systems. The proliferation of electronic resources makes the integration and development of up-to-date guidelines difficult. The rapid expansion of subspecialisation and research means that guidelines are almost out of date the moment consensus has been reached. There needs to be a better way to elicit knowledge electronically. Third generation search machines and artificial intelligence will be essential if we are to gain the confidence of generation Y.

Secondly, generation Yers will not be interested in learning knowledge irrelevant to their current needs. Therefore, they rely heavily on technology to deliver up-to-date, on-demand knowledge representation. As knowledge representation and sharing become more rapid and cross-institutional, the difficulties in version management will increase. The problem not only relates to updating the current knowledge representation, but also ensuring the removal of older versions from searchable knowledge data-base [11].

Thirdly, the delivery of knowledge and information retrieval systems for generation Y will have to be intuitive. Ubiquitous feedback algorithms that provide "related" information for generation Y will need to be relevant, individualised and user-centered.

Finally, the distinction between explicit and tacit knowledge will become obvious as generation Y moves from one institution to the other. The medical workplace thrives on individuality. There are not only issues with different guidelines for clinical disease management, more importantly, significant differences exist in leadership, management skills, radiology ordering, pathology ordering, pharmacy dispensing as well as interpersonal communication. The tacit knowledge which enables this has previously been transferred through the process of socialization. With the increasing number of short-term employment, part-time and locum workforce [13], the representation of tacit knowledge will have to be delivered to the workforce rapidly, allowing another dimension of knowledge management to flourish.

Pitfalls for health informatics and knowledge management with generation Y

While there are potentials for health informatics and knowledge management to expand, there are also pitfalls working with generation Y.

Problems with technology

Generation Y has very high expectations of what technology should deliver. While the technology might be available, there are multiple issues to be considered, such as ethics, security, data integrity, cultural factors, environmental factors and social factors [14]. It will be difficult to communicate the relevance of these factors to generation Y.

The healthcare budget is limited. It is not possible to fund cutting edge technology all the time. Low end technology will probably be as good and as efficient as cutting edge technology in performing some of the routine clinical jobs. While from a health economics point of view, it is imperative that these are taken into account, it will not fit into generation Y's culture of "being cool!"

Generation Y is already creating a generation gap in leadership, clinical governance and work-life balance [15]. Their familiarity with technology and their desire to utilise technology, if not managed appropriately, will lead to widening generation gap and disharmony in the healthcare workplace.

The implementation of new technology in a complex system often leads to multiple unforeseeable problems. The healthcare system is arguably the most complex of all systems. While there are socio-technical integrated solutions, there are still more failures that successes [14]. While the experience of generation Y in technology evolution might assist them in their adaptation to technology, it might also foster the misunderstanding that technology will fix all problems. The complexity of the healthcare system and therefore the difficulties in Information Technology (IT) implementation might not be easily comprehensible to generation Y.

Generation Y might want to be involved and lead in early stages of IT implementation. They, however, lack the necessary experience of project management. Their nature of experimentation and overestimation of their capacity are significant problems with IT implementation in healthcare as safety is often not the priority for generation Y.

Problems with knowledge management

While generation Yers are very good at information gathering, their information presentation is difficult to be shared with outsiders. The SMS messaging forms a unique language structure. This language is spreading rapidly, including its utilisation in healthcare messages for generation Y [16]. While the SMS language is gaining acceptance, including formal examination in New Zealand, many healthcare workers might struggle to understand the language. Should we use SMS language to transmit medical data? This is a great dilemma that we will need to resolve soon.

While it is often possible to retrieve on-demand information and to receive on-demand education, one needs to have a certain level of knowledge and competency in the medical world. Furthermore, the hierarchical structure assumes certain functions for each individual within that organisation. The on-the-spot clinical decision support and

knowledge representation for generation Y will challenge the basic fundamental assumption of the medical world. This has the potential to create significant conflict within the medical profession.

While knowledge elicitation and representation might be achieved regarding disease treatment, the tacit knowledge of "know-how" might realistically be very difficult to be elicited and represented. This is especially so for interpersonal interaction, leaderships and management styles. It might cause disastrous consequences if attempts were made to elicit, present and share this sensitive tacit knowledge.

Three to tango: Integration

Is the combination of generation Y, health informatics and knowledge management a Bermuda Triangle that should never be entered? Is it possible to integrate generation Y, health informatics and knowledge management together to form an irresistible force to transform the healthcare system? This paper argues that it is possible to integrate the three together to form a tangible connection to build a sustainable future healthcare system. This section will provide health informatics and knowledge management researchers and practitioners some simple guidance to work with generation Y.

A recent paper recommends ten commandments for IT designs in healthcare[17]. These ten commandments are:

1. Speed is everything;
2. Anticipate needs and deliver in real time;
3. Fit into the user's workflow;
4. Little things can make a big difference;
5. Recognise that physicians will strongly resist stopping;
6. Changing direction is easier than stopping;
7. Simple interventions work best;
8. Ask for additional information only when needed;
9. Monitor impact, get feedback and respond;
10. Manage and maintain your knowledge based systems.

These ten commandments are proposed in order to achieve socio-technical integration for IT implementation, especially decision support system in the current healthcare system. The future, however belongs to generation Y! We therefore need to take into account the characteristics of generation Y when building the future healthcare system. This paper suggests 10 different commandments for IT implementation in the healthcare system with generation Y. This is based on the synthesis of the understanding of generation Y in the literature as well as the author's experience of working with generation Y.

1. **Everything is speed**

   Generation Y wants outcomes and wants them fast. Therefore, from design and communication to feedback and improvement, speed is the key. Otherwise, generation Y will get bored and move on.

2. **Find out their needs and deliver wirelessly.**

   It is very important to find out the needs of generation Y and to involve them early in the design phase. While some of their suggestions might be difficult to achieve, their early involvement motivates them to help you.

3. **Empower juniors to change workflow through health informatics.**

   While IT implementation that fits into the workflow is desirable, IT implementation that drives the change of workflow is the future! It is a golden opportunity to empower generation Y to suggest current work flow changes, and deliver these changes through IT implementation.

4. **What are little things and big things?**

   There is nothing that is big or little in the minds of generation Y, as they think globally but act locally. Therefore, it is essential to ensure that all the important issues are addressed despite the perceived triviality by the IT team.

5. **Recognise generation Y wants progress all the time!**

   Generation Yers get bored easily. It is therefore of upmost importance to have a clear version management strategy and communicate the plan to generation Y. If improvements are made based on feedback from generation Y, communicate that clearly.

6. **Don't stop them doing the wrong thing; ask them to do the right thing more often.**

   The implementation of IT project, especially clinical decision support and guidelines will inevitably change the underlying process. The challenge to do the right thing more often is much more powerful than a condemnation to stop doing the wrong thing.

7. **Simple intervention that multi-tasks, works best**

   Generation Y loves multi-tasking and they are good at it. While intervention needs to be simple, it also needs to provide generation Y with the potential to multitask.

8. **Don't ask for information, "google" for information.**

   Generation Y is good at searching for information, especially electronic information. It is essential that IT programs are intuitive and manage to "google" information that is already available in database rather than asking generation Y repeatedly for the same information.

9. **Check, evaluate, change or move on!**

   Generation Y learns from experience and errors and they expect others to do the same. It is therefore important to share the experience and errors of IT implementation with them. In the event of unexpected outcomes, evaluation and changes need to be acted upon quickly. In the event that improvement is impossible, a direct response to generation Y is essential and then move on!

10. **On-demand knowledge representation in their language.**

Knowledge representation, either tacit or explicit knowledge should be ubiquitous and on-demand. While the SMS language might not be the official language and might be difficult to understand, it is the language that will communicate the best with generation Y!

Conclusion

In the next few decades, the healthcare system is facing the most massive transformation of any industry. Among all the challenges, the replacement of the ageing workforce by generation Y workers will create a massive impact. Generation Y workers have very different characteristics. Their inclination to utilise technology and to create life experience generates an opportunity for health informatics and knowledge management to flourish. There are, however, pitfalls working with generation Y. This paper proposed ten recommendations to integrate generation Y with health informatics and knowledge management for a sustainable future healthcare system:

1. Everything is speed;
2. Find out their needs and deliver wirelessly;
3. Empower juniors to change workflow through health informatics;
4. What are little things and big things?
5. Recognise generation Y wants progress all the time!
6. Don't stop them doing the wrong thing; ask them to do the right thing more often;
7. Simple intervention that multi-tasks, works best;
8. Don't ask for information, "google" for information;
9. Check, evaluate, change or move on!
10. On-demand knowledge representation in their language.

Acknowledgments

Thanks to all generation Yers who worked with me, inform me, educate me and enlighten me on their journey towards excellence. "U R the BEST!"

References

[1] Fitz-Enz J. Industry change = internal change. HRD/OD, Am. Management Association International. 1998: 1: 5.
[2] Lozada M. The health occupation boom. Vocational Education Journal 1995: 70: 34-7.
[3] Yee KC, Barbour L. Improving medication safety at the interface: The role of pharmacist versus information technology in the real world. Program and abstracts of Improving Medication Safety Conference: Sharing the lessons learned , Sydney, New South Wales, Australia. 2006.
[4] Benson S and Dundis SP. Understanding and motivating health care employees: integrating Maslow's hierarchy of needs, training and technology. J Nurs Manag 2003: 11: 315-20.
[5] Sheahan P. Generation Y: Thriving and Surviving with Generation Y at Work. Victoria: Hardie Grant Books, 2005.
[6] Huntley R. The World According to Y: inside the new adult generation. New South Wales: Allen & Unwin, 2006.
[7] Salt B. Y the younger generation is so different. The Australian. 16th November 2006. p. 26.
[8] Littlejohns P, Wyatt JN and Garvican L. Evaluating computerised health information systems: hard lessons still to be learnt. BMJ 2003: 326: 860-3.
[9] Workout what Gen Y wants. The Australian, Weekend Career One. 25th-26th November 2006. p. 1.
[10] Gorman D, Sir J Scott. Time for a medical educational change in time. Intern Med J 2006: 36: 687-9.
[11] Peleg M and Tu S. Decision support, knowledge representation and management in medicine. Methods Inf Med 2006: 45 (s1): S72-S80.
[12] Nonaka I, Takeuchi H. The Knowledge-Creating Company: How Japanese Companies Create the Dynamics of Innovation. London: Oxford Unviersity Press, 1995.
[13] Skinner CA, Riordan RL, Fraser KL, Buchanan JD and Goulston KJ. The challenge of locum working arrangements in New South Wales public hospitals. Med J Aust 2006: 185: 276-278.
[14] Wears RL and Berg M. Computer technology and clinical work: Still waiting for godot. JAMA 2005: 293: 1261-3.
[15] Yee KC, Tsunami to SurfWave: Toward harmony in generation Y gap challenge. Program and abstracts of Towards Harmony in the Health Workplace Conference: Practical approaches to managing difficult interactions and preventing the escalation of conflict and grievances , Sydney, New South Wales Australia. 2006
[16] Bisset K. Get set for 24/7 help with sextxt. The Australian, Weekend Career One 2nd-3rd December 2006. pp. 19.
[17] Bates DW, Kuperman GJ, Wang S, Gandhi T, Kittler A, Volk L, Spurr C, Khorasani R, Tanasijevic M and Middleton B. Ten commandments for effective clinical decision support: making the practice of evidence-based medicine a reality. J Am Med Inform Assoc 2003: 10: 523-30.

Address for correspondence

Dr. Kwang Chien Yee
eHealth Services & Research Group
School of Information Systems
University of Tasmania
Hobart, Tasmania, Australia.
Email: kwang.yee@dhhs.tas.gov.au

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Health Service Organisations and Professionals: An Information Systems Model for Transforming the Nexus between Accreditation and Practice

Jo-Anne Kelder

School of Information Systems, University of Tasmania, Australia

Abstract

This paper presents a qualitative research approach used to generate data and theoretical insights for information systems design in a highly regulated health service organisation. Ethnographical and sociological analytic techniques were used for sensemaking in the domain, and then identifying key structures affecting the conduct of the health service. A key aspect was taking a futures perspective of possibilities for evolution of these structures. From this, a continuum model representing the nexus between accreditation and practice was developed, varying the degree of integration of information systems for organisation- and individual-level accreditation. The paper uses the model to discuss the implications of possibilities for design in organisations where accreditation requirements have implications for work practice and information systems design.

Keywords:

Accreditation, health informatics, organizations.

Introduction

National health care systems are currently subject to transformational change drivers that threaten sustainability; health informatics is preoccupied with the question of how to model and design sustainable technical systems in complex, dynamic socio-technical health settings. This research assumes that a 'whole of system' design approach is needed to avoid failure in health service ICT design and implementation [1].

Health service organisations and health professionals who work for them are increasingly subject to quality assurance processes and accreditation requirements. But recent research raises questions on whether the abundance of very costly accreditation processes delivers better health systems [2].

Quality frameworks for organisations consist of standards for operation and metrics for measuring performance as the basis for determining accreditation status. ICTs have an integral role in quality assurance of health services: they enable collecting and managing vast quantities of data onto digital information systems and interrogation of the system to evaluate an organisation's compliance to the standards [3].

This paper presents the outcome of a research approach used to generate data and theoretical insights for information systems design. The research setting was a highly regulated health service organisation, BreastScreen Tasmania (BST). BST is a member of BreastScreen Australia (BSA), the accreditation organisation for the Australian breast screening Program. Each member organisation is accredited on the basis of compliance to 176 standards, comprising the 'best practice' according to experts from professions including radiography, radiology, pathology and surgery as well as best practice in data management, management of client psycho-social needs and other contributing professions.

The research method has three phases and draws on established ethnographic and sociological techniques and frameworks for generating and structuring data from the setting. The criterion for selection is their capacity to accommodate and deliver insight into the interactions and relations between the people, the place and the things (PPT) in the setting. Many such PPT frameworks have been fruitfully used in medical/health informatics research projects [4, 5].

PPT frameworks used for applying initial structure to the data were Distributed Cognition theory (Dcog) [6] which focuses on people-artefact interactions developing over time as a cognitive system, and Communities of Practice theory (COP) [7] which directs attention to the existence of mechanisms for sharing and distributing knowledge and information (people-brokers, artefacts-boundary objects). Activity Theory (AT) [8] was applied for its capacity to engender thinking about PPT interactions as affected by tools, rules and division of labour for a community activity. It facilitates representing multiple perspectives (different subjects with different objectives) and helps identify misalignment between the possible objectives of an activity and elements used in the conduct of the activity

Method

This method was designed for research in complex socio-technical settings. It varied the use of ethnographic and sociological analytical techniques over three phases of data collection and analysis. The first two phases of the method can be applied to any complex socio-technical domain for the purposes of making sense of the domain and creating structures understanding it. The outcome of third phase is specific to the research setting, but the principles of the method can be applied more generally.

Phase one: Sense-making

The first phase was a sense-making exercise for the researcher. The initial research questions were: 1) *How to approach IS research in complex and sensitive (highly social) domains?* and 2) *What are the relationships between the situation of the work place, digital information technology/ other artefacts in the domain and the ways people perceive and do their work?*

This phase was deliberately exploratory and involved several months of immersion in the research setting. Ethnographical techniques were used (field observations, semi-structured interviews, document collection and iterative analysis) to explore the setting; PPT techniques (from Dcog, COP and Activity Theory) were used to initiate understanding the setting by applying some structure to the raw data, and directing the researcher's attention to specific types of inter-relationships between the people, place and things in the domain.

Immersion involved observing people at work and conducting semi-structured interviews for insights into what was observed. Informal and observation-associated comments and conversations with staff members were recorded as field notes. The researcher reviewed over 200 organisation documents and scanned the wider environment using automatic alerts of electronic content relevant to the health service context. BST community education and cancer policy development staff members also made available their information resources.

A technique used for exploring the setting was to follow multiple client trajectories through the series of interactions with the organisation to identify *client* perspectives and assumptions about why and how they engage with the health service [9]. Field notes were taken for 24 clients having a routine mammogram, and seven clients attending an assessment clinic.

Client trajectories for personal interactions with BST were then extended by tracing the trajectory of artefacts and people connected with client interactions beyond the immediate context. For example, the client record (CR) was moved by trolley between the clinic area and the data management area. Clinic data entered onto the physical CR had to be entered onto the Client Information System (CIS). The trajectory of CR-CIS interactions gave a rich data set of coordination and breakdowns of activities, problem situations and multiple interpretations involving individuals, work teams and the organisation.

PPT approaches share techniques from ethnography (field notes, interviews, document collection) but collect data from their own perspective. Collecting data without particular reference to a PPT theory, and experimentally applying the differing theoretical frameworks enabled exploring a variety of possible structures for framing understanding the data.

Phase two: Identifying problems

This phase of the research focused on identifying problem themes. Theme discovery was facilitated by paying attention to aspects of the setting connected to problems in conducting the organisation enterprise (screening women for breast cancer). Themes were identified by iteratively applying PPT techniques to gain theoretical insight into the setting and by using the trajectories technique to uncover perspectives and structural relations not identified by specific PPT frameworks.

A theme was selected for further investigation according to its capacity to express a significant problem-situation, affecting at a structural level the information systems and work practice systems design in the research setting. It was the basis for new research questions in Phase 3. For BST, the theme was *the relation between accreditation and practice.* And the research questions: 3) *What are the consequences of the way accreditation is framed for people and the artefacts they use in their work situation?* and 4) *What relationships are included in the nexus between accreditation and practice and what are the implications for IS design?*

Phase phree: Researching an emergent theme at the level of structure

In phase three, the theme was researched iterating a sequence of techniques designed to conceptualise and model the structures of social organisation expressed in the theme. Ethnographic observations were extended by following trajectories of interactions beyond the immediate observation context followed by conceptual modeling of the data at the level of structure and testing the structural models via further ethnographic observations.

Ethnographic observations focussed on identifying the elements and relations in PPT interactions connected to the problem-theme; trajectories of interactions beyond the organisation boundary were followed to identify the contexts and drivers for different perspectives found within the organisation.

Conceptual modelling of the data was used to identify key structures shaping and constraining the attributes of the organisation. Some of these structures could be derived from the PPT theories used in the previous phases; some were created using data that did not fit the theories. COP theory and AT provided constructs for identifying key elements in the data and modelling their structural relations. Trajectories pointed to data sets and relations that required creating additional constructs to theorise them adequately.

This sequence of observations modelling testing the model was iterated to produce models of the elements and relations structuring the relationship between being accrediting for breast screening practice and how the enterprise was conducted within BST.

The nexus between accreditation and practice was conceptualised in four related models: 1) organisation-level accreditation; 2) individual-level accreditation; 3) the nexus between accreditation and practice for organisation-level and individual-level accreditation, mediated by boundary maintenance (Figure 1) and 4) a context diagram setting out the artefact and membership (information) connections between individuals, networks of practice who accredit them, an organisation conducting a health enter-

prise and an organisation whose role is to accredit that organisation (Figure 2).

Application of the models - Futures perspective

The third and fourth models were used as the basis for imagining alternative structures for the future, occurring either by natural evolution or by design. The technique of trajectories was used to investigate the changes over time of visions, negotiations, decisions and activities that led to the current partially integrated structures for legitimating individuals and organisations for conduct in health-related activities. An integrated structure for accreditation was based on current change drivers and desirable attributes for a future health system .

A continuum model of the *accreditation – practice nexus* (Figure 3), varied according to the degree of integration of individual- and organisation-level accreditation systems provided the focus for envisaging structures for accreditation. The properties of the IS and work practice systems were derived for the integrated end of the continuum by considering possibilities for data sharing between the different entities constituting an.

Results

This section outlines the constructs and the structural models developed in Phase 3.

Constructs

COP analysis provided models of systems of work practice within the clinic at individual-level that highlighted the role of *membership* in a competent *community of practice* for establishing competence at the level of work practice. Trajectories of professions within the multi-disciplinary team at BST established that health professionals are accredited in a social system of *multi-membership*: entry into a specific work domain (and COP) is dependent on prior accreditation as a continuing member of a professional association, a *network of practice* (NOP) [10], with membership obligations such as demonstrating continuing professional development (CPD). PPT interactions observed around the nexus of accreditation and practice were 'collegial' that is, operating at a local level and based on personal relations and trust; artefact use was limited; individuals were responsible for collecting and maintaining the evidence of *competency* for membership in the NOP, and for their reputation in the COP.

AT was applied to the data at the level of activity because it addresses 'rules' and 'tools' operating in a given activity and the research had documented organisation effort in ensuring it complied with the national accreditation standards (NAS). The data also evidenced influence of the accreditation requirements originating from BreastScreen Australia (BSA) on practice within the breast screening clinic context. The *accrediting organisation* designed artefacts to monitor and evaluate the competence of BST at organisation-level and required that *accreditation artefacts are naturalised into the practice* of the breast screening COP. All policy and procedure manuals and the client record in particular, were designed and written to comply with the NAS for organisations accredited for membership in the national Program.

New constructs were created to describe and model the organisation-level response: a set of activities designated *boundary maintenance*, to enable coordination and compliance from the breast screening *enterprise organisation* (BST) to the *accrediting organisation* (BSA). The concept of *trajectory into a community of practice* was also constructed to capture the two aspects of membership linked to individual membership of a *network of practice*. That is: entry into the breast screening COP is predicated on NOP membership, but membership of a COP is negotiated and established over time as an individual demonstrates competence in applying NOP skills and knowledge in the joint conduct of the enterprise. This is a social construction of accreditation.

Conceptual models

Figure 1 represents the nexus between accreditation and practice in situations of different systems for organisation-level and individual-level accreditation, mediated by boundary maintenance. The boundary maintenance required to bridge the disjunctions between the individual-level social construction of accreditation (mediated by an individual's membership of both the community of professionals with whom they work and the professional community of their vocation) and the organisation-level artefact construction of accreditation (mediated by information contained on client records (CR) converted to information on digital client information systems (CIS)).

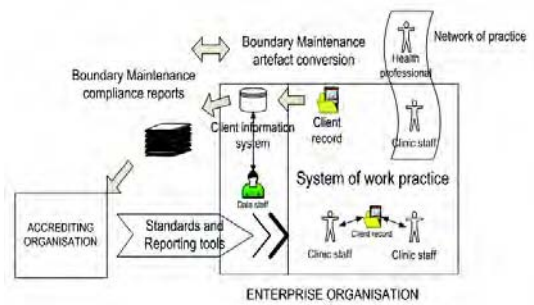


Figure 1 - Accreditation mediated by boundary maintenance

Boundary maintenance is the activity set required within the organisation to convert the information artefacts used for practice (the CR for each woman registered with the breast screening program) into information artefacts for organisation accreditation (the CIS which aggregates the data of all clients into statistics such as screening rate and cancer detection rate for the target population which are used to generate reports used to determine BST's compliance to the NAS).

Accreditation of an organisation is conducted largely via digital information artefacts processing data and generating reports. Membership in the BSA Program is contingent on successful implementation of the NAS into all aspects

of an organisation's practice. This naturalisation of standards is supported by naturalising BSA's artefacts for use in the organisations engaged in the breast screening enterprise. This takes the form of templates, the standards and accompanying manuals; it includes the effect of requiring that work artefacts developed within the organisation (CR, policy and procedure manuals) comply with the NAS. It is an artefact construction of accreditation.

Accreditation for individual health professionals is separated from organisation accreditation systems. Individuals maintain membership with their professional organisation by participating in approved continuing professional development activities and being able to provide documentary evidence.

Figure 2 represents the connections (evidenced by information transfer processes) between four major s in the social organisation of accreditation in the health context: individuals, networks of practice, health enterprise organisations and the accrediting organisation.

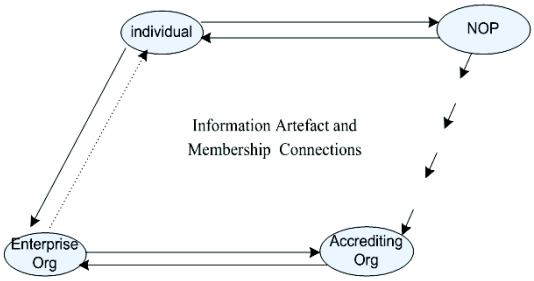


Figure 2 - Accreditation infrastructure entities

The accrediting organisation, BSA, authorises the National Accreditation Standards as current best practice for the professions engaged in its Program. Networks of Practice (NOP) connect with the accrediting organisation by their representatives on committees determining standards, and when individuals are invited to participate in a site accreditation visit to audit the data and practice of an enterprise organisation. BSA and NOP information artefacts and systems are separate.

The standards and measures expressed in the NAS act recursively to educate and support their adoption as standards of best practice generally within the NOP and across the health domain beyond the immediate context of breast screening.

BSA primarily interacts with member breast screening organisations via the output of digital artefacts. Individual professionals working for BST create data which is aggregated for demonstrating the organisation is a competent member of the Program. But they have limited access to data relating to their individual work practice. Radiologists are given feedback on key indicators such as the cancer detection rate and discuss this data privately or in the context of multi-disciplinary meetings with surgeons and pathologists to improve the knowledge and skills of everyone involved. Other members of the clinic community of practice rely on information generated and maintained within their local system of work practice, including feedback from colleagues, comments by clients and the data on the client record.

Future thinking for IS design - A continuum model

Figure 3 represents a continuous spectrum of possible accreditation structures for establishing the legitimacy of health service organisations and individual professionals working within them: each possible structure associated with a system of work practice and information systems. Work practice systems (WPS) and information systems (IS) for any point on this continuum will have certain properties which can be identified by systems analysis techniques such as business process analysis or PPT analyses. Transformation of the structures and associated systems requires taking into account those properties and identifying the issues and drivers for transition from one nexus to another.

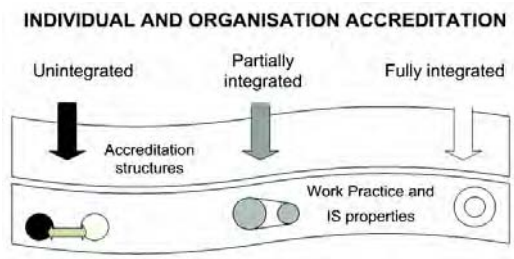


Figure 3 - Continuum model accreditation-practice nexus

Unintegrated accreditation structures are characterised by separate artefacts and information systems: the information system for collecting data for organisation level accreditation is not used by individuals in their work or for their accreditation. Separate WPS and IS are used in the conduct of the enterprise and individual accreditation activities are also separate from their work practice. Data from practice artefacts (e.g. the CR) must be added onto accreditation artefacts (e.g. the CIS) by additional processes. This boundary maintenance activity requires additional staff members and organisation resources and is a locus of breakdowns in organisation activities.

Partial integration occurs via the naturalisation of accrediting organisation standards and other artefacts into the enterprise organisation's WPS. Competent use of these artefacts and compliance with standards becomes part of the individual's identification of what it means to be an accredited professional in that environment. Partial integration at BST is evidenced by use of data on cancer detection rates that is measured for organisation-level accreditation and also provides feedback for individual radiologists. However, boundary maintenance activity is still required to utilise data for all levels of accreditation. For example, a BSA standard for accreditation data at organisation-level includes evidence of individuals attending multi-disciplinary meetings, data which can also be used as CPD evidence for continuing eligibility for NOP membership. Here, the systems are separate and require human work to collect and format into reports.

Where accreditation systems are fully integrated, a single artefact acts as a boundary object [11] for all entities in the accreditation infrastructure (Figure 2), providing information sufficient and appropriate for the needs of each entity. If the IS used within the organisation in the course of conducting the enterprise also takes the work practice data to measure competency for individuals and the organisation, *and* that IS artefact and its measurement is accepted by the NOP, this displaces the individual's work of demonstrating meeting NOP membership requirements onto the IS that records information about their work. Boundary maintenance activity is eliminated.

Discussion

Unintegrated, or partially integrated systems for organisation and individual level accreditation are problematic for health services such as BST at the level of WPS and IS design. Boundary maintenance (BM) activity at the intersection between practice artefact (CR) and accreditation artefact (CIS) constantly breaks down, with implications for the organisation's accreditation and resource requirements. Figure 2 can be used to consider alternative information flows and artefacts to support accreditation and practice.

Designing a single artefact that meets both practice and organisation accreditation requirements removes the need for resources for boundary maintenance. This is predicated on the organisation standards being incorporated into the relevant standards and requirements for professional bodies, thus removing the individual's burden of fulfilling distinct sets of requirements for multiple memberships. It is also predicated on appropriate permissions for access to data being agreed.

Australia's demographic profile points towards a continuing shortage of qualified health professionals paralleled by an increasing client base. Digital mammography is quicker than film and provides data that can be linked to an electronic client record and performance record of the radiographer and radiologist. Providing feedback to practitioners on individual skills and knowledge competence enables communication for improvement. This happens successfully with radiologists in the national Program. Information of individual radiologists' performance in their work (which contributes to the organisation accreditation) is provided and used as an opportunity for the team to discuss and take steps to improve knowledge and skills. Individual performance linked to organisation-level competence status has contributed to a COP-based drive for excellence and improving the quality of client care.

Integrated accreditation systems which provide real-time performance feedback to professionals and their NOP may provide a less time-consuming and costly process of accrediting professionals for work – relevant to the debate of importing health professionals trained overseas to work in Australia, or of exporting digital images to radiologists overseas and wishing to be assured of the quality of their work.

Conclusion

This method produced a set of conceptual models for thinking about the nexus between accreditation and practice in the health care setting. It enables consideration of the impact of different structures for accreditation on the design of information systems and work practice systems for a given health service enterprise, and a vision for integration that can contribute to the transformation of the wider national health system. In particular, identifying the location of an organisation on the continuum model is critical for understanding the constraints on ICT design for a given nexus: design in the context of unintegrated accreditation systems will be constrained to provide for boundary maintenance between the two.

Acknowledgments

The researcher thanks her supervisors, A/Prof. Paul Turner and Dr John O'Neill, the SIT CRC for their support. She especially thanks the staff members at BreastScreen Tasmania. This paper presents her doctoral research and is her own work.

References

[1] anon. Key Issues and Future Directions in e-Health-perspectives on 2010 and beyond: User Environment Program, School of Information Systems, University of Tasmania; 2004.

[2] Braithwaite J, Westbrook J, Greenfield D, Pawsey M, Naylor J, Iedema R, et al. A prospective, multi-method, multi-disciplinary, multi-level, collaborative, social-organisational design for researching health sector accreditation. BMC Health Services Research 2006;6:113-123.

[3] Clarke AE. Situational Analyses: Grounded Theory Mapping After the Postmodern Turn. Symbolic Interaction 2003;26(4):553-76.

[4] Nemeth C, O'Connor M, Klock PA, Cook R. Discovering Healthcare Cognition: The Use of Cognitive Artifacts to Reveal Cognitive Work. Organization Studies 2006;27(7):1011-1035.

[5] Bate SP, Robert G. Knowledge management and communities of practice in the private sector: Lessons for modernizing the national health service in England and Wales. Public Administration 2002;80(4):643-663.

[6] Hutchins E. Cognition in the Wild. Cambridge, Mass: The MIT Press; 1995.

[7] Wenger E. Communities of Practice: Learning, Meaning and Identity. Cambridge: Cambridge University Press; 1998.

[8] Engeström Y, Middleton D, editors. Cognition and Communication at Work. Cambridge: Cambridge University Press; 1996.

[9] Wales R, O'Neill J, Mirmalek Z. Ethnography, Customers, and Negotiated Interactions at the Airport. IEEE Distributed Systems Online 2002(September/October):15-23.

[1] [10]Seely Brown J, Duguid P. The Social Life of Information. Boston, Mass.: Harvard Business School Press; 2000.

[2] [11]Bowker GC, Star SL. Sorting Things Out: Classification and Its Consequences. Cambridge, Mass.: MIT Press; 2000.

Address for correspondence

School of Information Systems, GPO Box 252, Hobart, TASMANIA, 7000; Phone +613 6226 6200; Fax: +613 6226 6211; Email: Jo.Kelder@utas.edu.au

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Health Informatics: An Intercultural Perspective

Quynh Lê

Department of Rural Health, University of Tasmania, Australia

Abstract

Health informatics is a significant contribution to health care. It provides health professionals with powerful technologies to enhance their performance in caring for patients. The introduction of health informatics has added a new dimension in the health discourse. However, there are also issues and problems which are associated with health informatics, particularly in relation to privacy, confidentiality and data security, which are deeply embedded in culture. As privacy and confidentiality are linguistically and culturally constructed, health workers, patients and the public may have different views and exhibit different behaviours towards health informatics. The discussion of these issues is situated in an intercultural discourse.

Keywords:

electronic health record, health care systems, data integrity, electronic medical record, health informatics.

Introduction

Computer has permeated many aspects of society. It is hard to imagine any social activities which are immune directly or indirectly from the influence of digital technology. Computer technology is one of the fastest changing technologies. Technologies which were developed several years ago can easily become out-of-date today. In health science and health care, the impact of computer technology is widespread [1-3]. The development of health informatics is indicative of the impact of computer technology in health science and health care.

Health informatics is a timely contribution to health science and health care in Australia. It is an indication of the growing power of computer technology in health science and health care. While health informatics has made many contributions [4-6], there are also problems which have been identified [7-10]. This paper focuses on the cultural dimension of health informatics.

Health informatics: a new paradigm

The impacts of computer technology in society are marked with the prefix 'e' in many areas of computer development and implementation such as e-learning, e-communication, and e-commerce. In health, the broad term 'e-health' covers a wide domain including electronic health records, health information networks, telemedicine services, health portals, etc. It is an umbrella term covering two areas: health informatics (collection, analysis and movement of health information and data to support health care), and telehealth (videoconferencing and website delivery of health information or health care to a recipient). Health informatics is the appropriate and innovative application of the concepts and technologies of the information age to improve health care and health [11].

Health informatics has much to offer in community health care. Computer networks and telecommunications provide particular support that can enhance the collaboration among clinicians, care providers and patients. Special-purpose computer tools referred to as Consumer Health Informatics (CHI) represent the application of computer and information technologies specifically to support the health information and communication needs of patients and lay persons [12]. Health informatics plays an important role in the management of health information, particularly information of patients. It enables healthcare workers and policy makers at different management levels to plan and manage services. For example, health screening planning does not function well if there are no well-kept records of individuals who have undertaken certain kinds of tests or missed them due to personal or service problems. Health informatics may also record information about patients' health care experience, treatment and financial costs.

The Danish health information network MedCom [13] is a good illustration of health informatics implementation. It handles over 80,000 messages daily. All hospitals, pharmacies and emergency doctors, 90% of general practitioners, 98% of laboratories, 55% of specialists, and 20% of municipalities are connected to it. MedCom enables hospitals to use electronic referrals, and avoid data re-entry. The professional quality of referrals has risen, and discharge letters are stored directly [14].

It is worth pointing out that while e-health is becoming powerful tool and can make a huge contribution to health care, it is still at an early developmental stage in many countries [15].

Cultural factors in health informatics

Culture has been defined as the shared products of the society, including the ideas, norms, and material objects that describe how people handle daily tasks and make

sense of their experiences. Culture is also dynamic and adaptive [16].

> The culture of an individual has a profound effect on the perspective from which they deal with health and illness. Culture has influenced peoples' convictions, attitudes, types of knowledge, and values; modes of behaviour, habits and customs; language and tradition.[17]

Acculturation is a process in which people of a different cultural and social discourse have adapted to accommodate a new discourse. It can be a process filled with confusion, resistance, and reluctance and sometimes sufferings. Health informatics is not just a technology or a simple approach which can be introduced to a human discourse without any problems. There are cultural and social issues associated with health informatics.

> We also learnt that when things go wrong – as they seem to do in more than half the cases – people tend to blame 'the technology' whereas social, behavioural, psychological, and cultural factors are the most likely culprits [18].

The development and implementation of health informatics in health care can go through an acculturation process, which may include negative and positive experiences.

The first problem can be expert-orientated. Health informatics experts and enthusiasts can contribute to the formation of negative attitudes among prospective health informatics users. The worst case is when such experts hold the assumptions that health informatics is the magic solution to heath care and do not take the social and cultural factors seriously. Secondly, the introduction of health informatics can be seen as a paradigm shift in certain discourses. According to Roberts [19] this is the rejection of one set of values and ideas and the adoption of a new set with regards to what constitutes effective implementation. This paradigm shift is occurring worldwide but faster in some parts than the others depending on the availability of resources, existing infrastructure and the stage of development reached. If health informatics is viewed as a new paradigm, strategies have to be planned carefully to facilitate acculturation of current and prospective users to a new health care discourse. Otherwise the acculturation experiences can be painful and sometimes destructive. It is important to involve users (e.g. doctors, nurses, patients, administrative staff, etc) in the decision making process in their acculturation into an unknown or less familiar territory.

Health informatics operates under key principles covering confidentiality, privacy and security. These three concepts are inter-related and are important in evaluating the success or otherwise of health informatics implementation. However, concepts and principles such as privacy, confidentiality and security which govern health informatics have different cultural meanings and values and they are perceived differently by users of different cultural backgrounds. Thus, these three fundamental concepts and principles in health informatics should be examined in terms of cultural discourse.

The cultural discourse of privacy

Humans are social beings. Individuals live together in a community. They belong to a community but this does not mean that their community owns them. They have the right to be left alone. Individuals are entitled to personal privacy which covers three domains:

- *Physical privacy*: such as bag searching, use of our DNA
- *Information privacy*: the way in which governments or organisations handle our personal information such as our age, address, sexual preference and so on.
- *Freedom from excessive surveillance*: our right to go about our daily lives without being monitored or have our actions caught on camera. [14]

Health informatics should adhere to the privacy principle to ensure that individuals' privacy is respected. We tend to take information privacy for granted or do not seriously appreciate it unless it is threatened or lost. Individuals' health information is their personal privacy which should not be 'violated' by government agencies. In special cases when individuals' health condition is a serious threat to the community, their right to privacy may be exercised differently. For example travellers contracted a highly contagious life-threatening disease are expected to reveal fully their conditions to health authorities.

According to Le [20], privacy is something which is personal, belonging to an individual and is not in the public domain. It normally refers to an individual's private life. Thus, according to this definition, an individual's life consists of private and public domains. The private domain includes his/her personal belongings such as home, relationship, thoughts and feelings. The public domain includes social belongings such as professional life, policy, social activities. The following example illustrates what information is private and what is public.

> Mr. Green is working for a company in Tasmania. He joined the Liberal party when he was a student and now he is an independent. His mother is very poor and old but Mr. Green seldom visits his mother even though they are living in the same suburb. They argue a lot when he visits her.

The text given above consists of two kinds of information: private and public. The problem is that the text does not linguistically mark the information in such a dichotomy. To a great extent, privacy is culturally determined. What is private to an Australian may not be so to a Vietnamese.

Not all cultures view privacy in the same way. In Western cultures, individuality is very important. Each person is entitled to their own privacy. Children are introduced to the concept of privacy at an early stage in their childhood. They are taught to respect other people's privacy and they also expect others to respect theirs. In Asian cultures, the division between the public domain and the privacy of individuals is not always clearly prescribed.

In a report about an intercultural experience of a group of Australian students in Australia, Harbon [21] described an

instance in which an Australian student was very upset when she discovered that her host family had searched her suitcase while she was billeted by them. To her it was a serious violation of privacy. Whereas, the host family felt it was interesting to know more about their guest, whom they treasured and cared for tremendously.

Collectivism is very strong in Asian cultures. In an Asian family, privacy is not greatly valued. Parents have 'the right' and 'the duty' to know the private life of their children. It is not a matter of privacy intrusion but a responsibility of the parents to know their children' private domain well so that they can adequately and meaningfully protect their children and ensure their wellbeing. In a Confucian society, interpersonal relationship is the foundation of social coherence. This relationship is characterised by the social roles assigned to each member in a family and in a community. While it is a social violation to ask personal questions in Western cultures, it is a common speech subject in many Asian countries to inquire about someone's age, health conditions, and personal life.

Privacy is an important factor in health informatics. However, users of health informatics may interpret this concept differently due to their social and cultural backgrounds. It is possible that migrants in Australia may violate the principle of privacy in health informatics without being aware of the seriousness.

The cultural discourse of confidentiality

Confidentiality refers to the treatment of information disclosed or provided by individuals on the basis of trust that it will not be made available or disclosed to unauthorised people or services. In health, generally the patient's consent must be sought before his information can be used for a specific purpose.

According to the Australian National Privacy Principles [22, 23], an organisation must take reasonable steps to protect the personal information it holds from misuse and loss and from unauthorised access, modification or disclosure. It must take reasonable steps to destroy or permanently de-identify personal information if it is no longer needed for any purpose for which the information may be used or disclosed.

A study conducted by Lindenthal, Thomas, and Ghali [24] compares the handling of confidentiality among American, Egyptian, and Israeli psychiatrists, and American and Israeli psychologists and internists. The study supports the view that no significant differences exist between practitioners of the same professional groups practicing in different countries while also showing significant and parallel between-group differences. According to Akhter [25], in some cultures, on one hand, sharing personal information among family members indicates a strong bond of coexistence and on the other hand the desire to keep any weaknesses, medical or otherwise, from the extended family is not uncommon. For a societal structure in which the family plays a central role, both allegiance to the family and a desire to keep its reputation strong is an understandable concept. The bond in an extended family provides solace and support in times of need. However, it can also become oppressive and limiting individual freedom.

From the professional duty perspective, confidentiality is based on the trust between patients and health professionals. McClelland and Thomas [26] suggest that confidentiality is grounded in the principle of respect for autonomy – health professionals explicitly or implicitly indicate to their patients that they will keep confidential the information provided to them. Patients are reluctant to share their private and sensitive information if this trust is lost. McClelland and Thomas point out that the duty of confidentiality exists within a wider social context in which other moral obligations may compete. These competing appeals set limits to medical confidentiality and arise from two principal sources: the patient's best interests and public interest. Problems arise when the patients' best interests vary according to their cultural and religious backgrounds, which may not be easily detected or decided by those involved.

Tai and Lin [27] give an interesting example about the cultural concept and practice of confidentiality in a Confucian society. When a patient has been diagnosed with terminal cancer, the first person to be notified is often not the patient himself, but the head of the family, such as the father or the husband. He then will confer with other family members to see what course must be taken. After the decision is made, the patient may be advised in a disguised way, to ease his anxiety. Furthermore, when considering different treatment options, the family members, especially husband or father, are again consulted first rather than the patient himself/herself. When the patient is a father or husband, the family member who becomes the spokesperson for the family, with whom physicians consult, is usually the eldest son.

Gossiping is a good example of cultural variation in dealing with personal privacy and confidentiality. Quite contrary to the principle of confidentiality, gossiping is a sociolinguistic activity which is widespread among cultures. A gossip is a casual conversation between at least two participants about the private life of someone. Morally it is an offence to participate in gossiping. However the seriousness of this moral offence is perceived differently in various cultures. In Western societies, gossiping is condemned and it could be treated as a criminal act if it is proved to cause damage and harm to the victim. In Asian cultures, gossiping is generally discouraged but it is not treated seriously. Gossips are often mentioned in folktales and historical events. The acceptable attitude towards gossiping is a big concern to health informatics as it violates the principle of confidentiality as health workers are expected by the health authorities, patients and the public to strictly adhere to this principle and they should incorporate the spirit of Hippocratic Oath into the social contract.

The cultural discourse of security

To protect individuals' privacy and confidentiality, it is important to ensure that security measures are taken so that health data is kept safely. In health informatics, computer technology provides a range of approaches and strategies

to improve security of health data. Two main approaches include restriction of access and anonymisation of records. Security protection of data requires sound physical as well as logical access controls. Encryption is a method for anonymising electronically held patient information. It is the process by which data are converted into a sequence of alternative characters, by applying a set of rules (or keys) that both generates the encrypted material and is capable of recreating the original information. Another method for anonymising patient information is the use of separate databases in which clinical information is separated from patient-identifier information. The secondary database retains the non-identifiable patient information, which may be used for a range of purposes[28].

Security is an important factor in health informatics. The loss of or unauthorised access to personal and sensitive data can result in financial and legal costs and personal trauma. From an intercultural perspective, there are two issues involved. Firstly, health workers and patients of different cultural backgrounds may treat data security differently. Data security in health informatics needs absolute commitment from those who are privileged to have authorised access. However, such commitment can vary due to different cultural attitudes towards data security and the cultural discourse in which security is reinforced.

One of the most common computer security problems is the management of passwords. A password is a key to access a computer system or a computer file. Though technical security can be very effective, it is the user whose handling of passwords can make computer security vulnerable. It first appears that culture has nothing to do with password security. However, human errors reflect cultural influence on users' attitudes and behaviours in dealing with computer security. In a culture which emphasises collectivism, sharing is a common feature in human interaction, particularly among family members and close friends. Ownership does not belong to individuals but it can be extended to close others. Friendship and kinship are based on mutual trust. In this cultural context, sharing security passwords can occur. Health informatics should take into account this cultural phenomenon.

Implications for policy planning

According to McClelland and Thomas [26] there is a need to establish a new culture for handling health care information – a culture that recognises, understands and responds to the changing structure of health care and health care delivery systems, which depends increasingly on the ready sharing and manipulation of patient information. The digitised health communication and interaction has not only provided an innovative approach to health care but also created a new discourse of health care which requires adjustment and adaptation. Policy makers and health authorities need to introduce programs and strategies for health workers to facilitate their acculturation into the new digitised health discourse.

Australia is a land of cultural diversity. Health workers and patients come from different cultural backgrounds, which may affect their behaviours and attitudes towards health issues and health care, particularly in relation to privacy, confidentiality and data security. As privacy and confidentiality are linguistically and culturally constructed, one would expect different views and behaviours of health workers, patients and the public in response to health policy.

The Linguistic Relativity Hypothesis [29] states that language is so intricately linked to its own culture that it is impossible to fully understand the message through a different language.

> The 'real world' is to a large extent unconsciously built on the language habits of the group. We see and hear and otherwise experience very largely as we do because the language habits of our community predispose certain choices of interpretation (p.177).

The implication for policy planning is that we should not assume that lucid translation of written and spoken health information from English to other languages or vice versa automatically leads to perfect understanding and interpretation.

Cultural diversity should be taken into account when developing and implementing health informatics programs that reflect culturally and linguistically diverse population [30]. Miscommunication or communication failure in the health discourse tends to happen to migrants whose knowledge of English is very limited or whose cultural metaphors and stereotypes influence their health behaviours and attitudes.

Conclusion

In summary, this paper has discussed some cultural issues associated with health informatics. The focus is on issues relating to privacy, confidentiality and security which are fundamental in the implementation of health informatics, particularly from an intercultural perspective.

Health informatics is a significant contribution of computer technology to health care. Metaphorically it is like a superhighway which traverses various roads and alleys of the health discourse, locally, nationally and globally. It has enabled health professionals and health services to improve their effectiveness. However, it is not all smooth. In a culturally diversified discourse, the implementation of technology in dealing with people needs to take into account the social and cultural aspects of human behaviours and attitudes. It is no exception with health informatics, particularly in Australia, which is a land of cultural diversity.

Acknowledgments

I would like to thank Cecilia Chiu for her valuable proofreading and comments.

References

[1] Ammenwerth E, Iller C, Mahler C. IT-adoption and the interaction of task, technology and individuals: a fit framework and a case study. BMC Med Inform Decis Mak 2006;6:3.

[2] Curry A, Knowles G. Strategic information management in health care - myth or reality? Health Services Management Research [NLM - MEDLINE] 2005;18(1):53.
[3] Dixon DR. The behavioral side of information technology. Int J Med Inform 1999;56(1-3):117-23.
[4] France E. Data privacy in medicine. A perspective offered by the Data Protection Registrar. British Journal of Healthcare, Computing and Information Management 1997;14:20-2.
[5] Lewis A. Health informatics: information and communication. Advances in Psychiatric Treatment 2002;8:165–71.
[6] Linda K. An electronic emergency. Modern Healthcare 2003;33(19):21.
[7] Beauchamp T, Childress J. Principles of Biomedical Ethics. Oxford: Oxford University Press; 1994.
[8] Binetti P. Biotechnology and the birth of a third culture. Journal of Biological Regulators and Homeostatic Agents 2004;18(3-4):255-60.
[9] Brooks F, Scott P. Knowledge work in nursing and midwifery: an evaluation through computer-mediated communication. Int J Nurs Stud 2006;43(1):83-97.
[10] MacPherson S. Health informatics and the cultural divide: a UK perspective. Health Informatics Journal 2001;7(2):62-5.
[11] Health Informatics Society of Australia (HISA). What is health informatics? [cited 2006 18 October]; Available from: http://www.hisa.org.au/100127.php.
[12] Brennan P. Health informatics and community health: support for patients as collaborators in care [cited 2006 8 October]; Available from: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=10805012&dopt=Citation.
[13] MedCom. MedCom projects. MedCom; 1999-2006 [cited 2006 18 october]; Available from: http://www.medcom.dk/wm109991.
[14] Lawlink NSW. What is privacy? The NSW Privacy Commissioner; 2004 [cited 2006 18 October]; Available from: http://europa.eu.int/eur-lex/pri/en/dpi/cnc/doc/2004/com2004_0356en01.doc.
[15] Commission of the Europe Communities. e-Health - making healthcare better for European citizens: An action plan for a European e-Health Area; 2004 [cited 2006 18 October]; Available from: http://europa.eu.int/eur-lex/pri/en/dpi/cnc/doc/2004/com2004_0356en01.doc.
[16] Carolyn P. The impact of culture on rural women's descriptions of health. Journal of Multicultural Nursing & Health 2001;7(1):50.
[17] Janelle T, Celeste Mulry B. Palliative Care and Culture: An Optimistic View. Journal of Multicultural Nursing & Health 2006;12(2):28.
[18] Jadad A, Delamothe T. What next for electronic communication and health care? BMJ 2004;328:1143-4.
[19] Roberts D. The educational paradigm shift: possible implications for higher education and flexible learning. In: Osborne J, Roberts D, Walker J, editors. Open, flexible and Distance Learning: Education and Training in the 21st Century. Launceston: University of Tasmania; 1997.
[20] Le M. Privacy: an intercultural perspective. In: Lakoff R, Ide S, editors. Broadening the Horizon of Linguistics Politeness. Amsterdam: Johnh Benjamins; 2005.
[21] Harbon L. Interpreting culture: preservice teachers taking control and making meaning across cultures. Language, Society and Culture [serial online] 1999 [cited 10 October 2006](4):5-12. Available from: http://www.educ.utas.edu.au/users/tle/JOURNAL/.
[22] The Office of the Privacy Commissioner - Australia. National Privacy Principles (Extracted from the Privacy Amendment (Private Sector) Act 2000). Privacy Commissioner, 2000 [cited 2006 10 October]; Available from: http://www.privacy.gov.au/publications/npps01.html.
[23] Inan H. Australian National Privacy Principles by Hurol Inan 2001-2006 [cited 2006 10 October]; Available from: http://www.hurolinan.com/book/br_detail.asp?LocatorCode=201.
[24] Lindenthal J,Thomas C, Ghali A. A cross-cultural study of confidentiality. Social Psychiatry and Psychiatric Epidemiology 1985;20(3):107-51.
[25] Akhter J. Case Study: Confidentiality in an Extended Family; 2003 [cited 2006 12 October]; Available from: http://www.parkridgecenter.org/Page1885.html.
[26] McClelland R, Thomas V. Confidentiality and security of clinical information in mental health practice. Advances in Psychiatric Treatment 2002;8:291-6.
[27] Tai M, Lin C. Developing a culturally relevant bioethics for Asian people. Journal of Medical Ethics 2001;27:51-4.
[28] NHS Connecting for Health. Better information for health, where and when it's needed; 2005 [cited 2006 18 October]; Available from: http://www.connectingforhealth.nhs.uk/.
[29] Littlejohn S. Theories of Human Communication. New Mexico: Wadsworth; 2002.
[30] Kreuter MW, McClure SM. The role of culture in health communication. Annual Review of Public Health [NLM - MEDLINE] 2004;25:439.

Address for correspondence

Dr Quynh Lê
Department of Rural Health, University of Tasmania
Locked Bag 1372, Launceston, Tasmania Australia
Tel: +61 (0)3 6324 4053; Fax: +61 (0)3 6324 4040;
Email: Quynh.Le@utas.edu.au

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Successful Systems Sustaining Change

Sheila Bullas[a], John Bryant[b]

[a] *Associate Director, Princess Alexandra Hospital NHS Trust, UK*
[b] *Fellow, Centre for Health Informatics Research and Development (CHIRAD), UK*

Abstract

Much has been published on the success and particularly the failure of IT projects; still failures are commonplace. This prospective study focused from the outset on assessing risk of failure and addressing critical success factors. The aim was to apply existing methods in a challenging acute care hospital where success demanded rapid achievement of sustainable improvements in clinical and administrative processes. The implementations were part of the English National Programme for IT. The desired outcomes required the integration of accepted tools and techniques to provide a pragmatic approach to systems implementation: Lean, Six Sigma, PRINCE2 and Benefits Management. The outcome and further insights into success and failure of IT projects in healthcare are described. In particular lessons are identified related to the business need for the project and the successful achievement of the required benefits and business change.

Keywords:

information systems success and failure; implementation strategies; organizational change

Introduction

The organisation

The Princess Alexandra NHS Hospital Trust (PAH) is a 550 bed general acute hospital. This previously failing hospital had made great improvements over the previous two years; a new senior management team was in place. The organisation required further improvements to reach the standards defined by the Department of Health and other external stakeholders. Information systems were poor, basic information governance was lacking and there were few informatics procedures in place. As in most NHS hospitals there was immense unrealised talent amongst the workforce. The strategy and systems offered by the national programme suited this hospital well. This project started in October 2004.

In common with many NHS organisations, the Trust was under considerable financial pressure, being required to pay back debts and make efficiency improvements.

The English National Programme for IT (NPfIT)

The programme originated from the 1998 Department of Health Strategy entitled 'Information for Health' [1]. A supporting document [2] outlined the information and IT system required for delivery of the NHS Plan [3] and to support patient centered services.

'Securing our Future Health' [4] concluded that to meet people's expectations and deliver high quality the UK needed to devote more resources to healthcare matched with reform. Information and communications technologies (ICT) were recognised as a major driver of this reform. NPfIT was formally established in October 2002 to procure, develop and implement modern, integrated IT infrastructure and systems for all NHS organisations in England by 2010.

NPfIT is a wide ranging programme covering national infrastructure and applications as well as applications to support local organisations. This project focuses on applications to support clinical care and administration of the acute hospital: the incremental implementation of an electronic patient record. The systems and services of a Local Service Provider (LSP) were procured nationally.

NHS Connecting for Health is an agency of the Department of Health whose purpose is to deliver the National Programme for IT.

Local implementation at the Princess Alexandra Hospital

The local implementation comprised the replacement of the existing patient administration system (PAS) with a strategic PAS based on which clinical functionality would be built.

In 2004, the Trust initiated the first project of the programme, namely the replacement patient administration system, the implementation of order communications and the implementation of a data warehouse. In 2005, a second project commenced to replace the radiology information system (RIS) in preparation for the implementation of the picture archiving and communication system (PACS).

Accenture was the Trust's Local Service Provider (LSP). The systems being implemented were iSoft (originally intended to be Lorenzo and later changed to the more established ipm and icm products), HSS (RIS) and AGFA (PACS).

Approaches and methods

Lessons on success and failure

A great deal has been written on the success and failure of information systems. The definition by which success of this project would be measured was adapted from the work of Robert Block in 1983 [5]. Success of the project was defined as: the implementation of the systems on time and within budget, meeting their goals and specified requirements and satisfying the users.

The risks of failure for this project were distilled from sources [6-9] that addressed failure related to information systems implementations. These were used to identify the critical success factors for the project:

CSF 1. Implementation of systems that were tried and tested.

CSF 2. Securing and retaining the commitment of the Trust Board and Executive.

CSF 3. Senior management having limited patience when waiting for tangible results or when delays occur.

CSF 4. Describing, through the project objectives, how the project contributes to the main aims and objectives of the organisation.

CSF 5. Providing sufficient and appropriate resources (people and money)

CSF 6. Recognising that the project is principally a business change project with IT as a major enabler of that change. Linked to this the acknowledgement that change must come from within the organisation and may be facilitated from outside.

CSF 7. Ensuring robust management of the project, including project governance.

CSF 8. Understanding, influencing and managing the expectations of users.

CSF 9. Communication. Linked to this, recognising the fear of retribution when conveying bad news to the Executive and Board.

In addition, the design took into account the view that most failures are the victims of organisational and people related issues [10], even those which initially appear as technical failures. The point at which the failure becomes inevitable may be many months before failure becomes apparent. This suggests that avoiding failure should be part of project design.

There were a number of constraints outside the control of the Trust within which it had to operate:

- Governance arrangements from the Trust, through the local health community, the Strategic Health Authority, to Connecting for Health. This defined the structure of Project and Programme Boards, structure and content of key documentation and project dates.
- The systems and the suppliers were determined through the national procurement. This gave a good fit with the local strategy for improvement.

Approaches, tools and techniques

A range of tried and tested approaches, tools and techniques were adopted for the project: some at the outset and others emerged and were integrated during the project as a result of changes in the internal and external environment. The choice was not an entirely free one: it included best practice standards, approaches required by NPfIT, and those adopted for the Clinical Service Improvement programme [11].

Project management

PRINCE 2 is the UK Office of Government Commerce (OGC) recommended approach to project management [12]. The approach was tailored to suit the nature and size of the project and to fit with requirements of the LSP and reporting bodies. Once the approach was agreed, this was strictly adhered to. Within the Trust, this standard was poorly implemented before starting this project, even for IT projects where it is a requirement. PRINCE2 was used for both technical implementation and business change. A qualified and experienced PRINCE 2 practitioner was an essential requirement of the project manager.

Change management

Change management was entirely the responsibility of the Trust. The LSP introduced a Business Change Workstream and supported the Trust in undertaking Business Change Workshops to a specific format. However LSP business change covered only the simplest level of change where the new system necessitated the redefinition of processes at the systems use level.

It was necessary to adopt additional methods if business change was to be created and the objectives of the project achieved. The following model shown in Figure 1 was designed, describing the different levels of change that were required if the objectives of the system were to be met. This was developed from various concepts of change management [13,14].

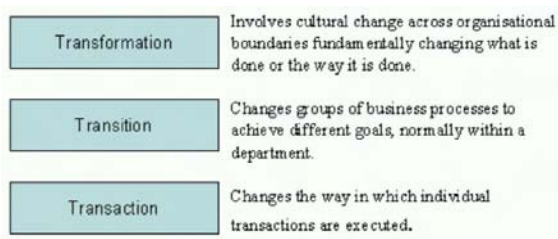


Figure 1 - Levels of change

These different levels of change required different approaches.

At the **transaction level**, workshops introduced by the LSP were used. In discussion with the users, the current 'as is' processes for all transactions involving the old and/or new information systems were documented. Users worked with the new systems in order that they could envisage how transactions would change ('to be' processes). These workshops also stimulated ideas for transitional change by encouraging discussion on the effi-

ciency and effectiveness of current processes. At the transaction level existing transactions or processes (series of transactions) are replicated using the new system. These new processes were documented and used to tailor the training which focused on business processes and not just what buttons to press. While it is essential that all current processes are comprehensively translated to the new system, to only do this gives rise to limited benefits and would not have met the objectives of the project.

During the transaction level workshops, areas were explored where transition level change might be identified. This included identifying bottlenecks and constraints in the processes, elements of waste and inefficiency, and duplication. These ideas input to the **transition level.** Further changes were identified at this level by conducting benefits workshops using the Cranfield Benefits Management method (see below). Processes were re-designed in workshops involving all stakeholders in those processes. Re-designed processes, including changes to who did what, when and where, were documented and tested before individual transactions were defined. The aim was to optimise whole processes including both computer based and manual steps.

The Trust was subject to more fundamental change as a result of external requirements, for example the introduction of the 18 week referral to treatment target [15], and internal pressures to improve performance, eliminate waste and achieve financial balance. In line with healthcare organisations around the world, NHS organisations were beginning to adopt industrial processes to improve healthcare [16,17]. The Trust adopted a tailored Kaizen approach as the means by which it would achieve **transformational change.** This approach is based on Lean Thinking, started by Toyota in the 1950s and developed by Womack and Jones in the 1990s [18]. The approach focused on the patient journey. It aimed to ensure that processes flow efficiently, value to the patient is increased, and steps that fail to add value are minimised. This approach provided a single transformational change programme within the hospital. The information systems implementations and improvements needed to integrate with this programme if maximum success were to be derived from both the transformation and the systems implementation programmes.

Benefits management

The Cranfield Benefits Management method [19] was adopted by the Trust as its standard to identify the benefits, particularly transitional benefits that would arise from the implementation of the information systems. This approach was adopted by much of the NHS for use as part of the NPfIT. The approach provides a structure to identify benefits that is linked firmly to the business drivers for the project. These drivers are derived from the business strategy of the Trust. In summary, workshops involving all key stakeholder groups systematically identify and measure benefits plus business and IT enablers to achieve the benefits. Benefits are measured and ownership for realisation established at this early stage.

This method was adapted locally to ensure that benefits were linked into the main strategic objectives of the organisation at the outset and further refined at stages throughout the project. Early workshops identified the processes to be re-designed as part of the Transitional Change Plan and provided clear guidance for configuration and training work packages. Output from this stage provided the means of benefits monitoring and the basis for benefits realisation.

Stakeholder analysis, as included in the Cranfield Benefits Management approach was also the basis for the development of the communications plan. This identified the stakeholders, their current commitment, and that required for success.

Organisation-wide Lean events identified transformational change that existing and new information systems were required to support. The proposed changes were further subjected to benefits workshops to ensure that information flows supported both new and continuing processes.

Resources: people and money

When an organisation is under financial pressure, as this one was, it is tempting to under-estimate resource requirement in order to get the project agreed. This limits the success that is achievable if the project is attempted with inadequate resources. It also undermines the credibility (and hence challenges commitment) if additional resources, that might have been foreseen at the outset, are requested at a later stage. For these reasons, great attention was given to estimating the required resources as soon as possible, ensuring (through project governance structures and processes) that these were clearly understood and then robustly defended. This particularly applied to the people and their skills.

Resource requirements, identified during plan development, led to the development of roles, job descriptions, person specifications and recruitment of the project team.

Results

The first system to go-live (RIS) met the definition for success: on time, within budget and meeting the user's expectations. PACS goes live in May 2007 and has met all early milestones successfully on or before planned dates.

The PAS/Order Communications package has been deferred for two to three years in a complex scenario including supplier delay and local financial pressures. While an option was presented to the Executive that was highly likely to have brought the project in on time and within budget, the delays had compromised confidence and commitment with not entirely unexpected consequences.

Discussion

This section discusses the progress in the light of each of the critical success factors.

The programme being implemented ranged from the tried and tested to the new and not fully developed (CSF1). This

was inextricably linked to CSF3: Executive patience is limited when results are not quickly coming or delays are encountered. Furthermore, different systems were subject to different contractual and financial conditions. This proved important in the progress and outcome of the various elements in the project. Commitment for the PAS/ Order Communications package was lost when delays resulted in the supplier being unable to deliver on time. This was due to delays in the development of the Lorenzo application. Although substituted with tried and tested applications, confidence was already compromised. Commitment might have been sustained had the delay not coincided with financial pressures which led the Trust to withdraw support. On the other hand, the tried and tested RIS and PACS systems, which came with a far higher price tag but with financial penalties on the Trust if milestones were missed, sustained commitment (CSF2). Although it was made clear that support for this could also have been withdrawn had financial arrangements been different. Financial balance in the current year took precedence over all other objectives.

The formal governance arrangements were one essential element in gaining and sustaining the commitment of the Executive and Trust Board: providing formal communications on progress against plans and budgets (CSF9). At least as important were the close working relationships without which continuing support cannot be expected at times of pressure. This was a particularly important aspect of this project: severe financial pressure combined with delay in delivering the technology put strains on the commitment.

How the project objectives contributed to the aims and objectives of the organisation was clearly identified through the business change workstream and communicated to the Executive and Board through project governance arrangements established through the formal project management arrangements (CSF4). Board and Executive understanding of this was maintained throughout.

Provision of sufficient and appropriate resources (people and money) was supported by the Executive (CSF5). In the financial climate, there were, unsurprisingly, attempts to compromise on project resources in order to reduce costs. However the risk that this posed was understood and appropriate resource levels maintained. The relationships developed with executives and the open and robust approach to project governance proved pivotal in maintaining resources.

The programme was managed very much as a business change project but not to the exclusion of excellent technical implementation of the systems: both are essential requirements for success (CSF6). The adoption of the 3-tier model (Figure 1) and associated methods, the inclusion of a business change workstream within the project structures and processes, and the full integration of information change into the transformation programme for the Trust as a whole were crucial to success. The latter proved to be the most difficult to achieve. The Trust-wide transformation programme was defined as separate from informatics and promoted as such by people both within and outside the organisation. As a result, project processes were needed to make up for shortcomings in the level of integration. There can only be one major change programme in an organisation, to which IT projects contribute. In the NHS NPfIT, conflicts could arise as a result of different external organisations imposing different approaches.

Robust management of the project, including project governance (CSF7) was ensured through the adoption and implementation of PRINCE 2 appropriately tailored to nature of the project. In practice this included defining governance structures, regular meetings, appropriate agenda, maintained risks and issues logs, comprehensive status reports, and response to exceptions agreed at the right level. This gave confidence that the project was progressing to plan and any variations were dealt with before critical milestones were in danger of being missed.

Understanding, influencing and managing the expectations of users is at the heart of the very definition of success (CSF8). The business change programme was designed to incorporate this requirement. It is inevitable at the start of a project that the expectations of the users are diverse and contain some elements that the project cannot meet. The workshops brought as many of the users into the process as was practically possible. Developing user understanding of the project and the products brought expectations into line with what could reasonably be delivered and the project was adapted to more closely meet expectations where possible. Meeting those expectations was then dependent on the successful technical and organisational implementation of the project as agreed with the users. One aspect of user expectation management that had not been adequately addressed was at the immediate go-live period. The implementation of the RIS system went smoothly with few teething problems that were promptly and efficiently addressed. Users had not been prepared to expect these and, since all other aspects of the project (including time and budget) were as expected, this caused some temporary dissatisfaction.

The final critical success factor (CSF9) addressed in this study was the need to communicate. The communication plan was an important part of the project but just what constituted appropriate and adequate communications? Stakeholder analysis, as included in the Cranfield Benefits Management approach, was the basis for the development of the communications plan. This determined the minimum requirements for communications. In practice, communications played a greater role than such analysis suggests. The balance between listening and telling, the credibility of the person or medium providing the communication, and the need to continually assess and repeat or reinforce should be emphasized.

Fear of retribution when conveying bad news to the Board, or in this case Executive, was part of the culture and conveying news contrary to the pertaining view of the Executive was not welcomed however diplomatically expressed.

Conclusions

This study set out to determine whether one could take lessons about success and failure from the past, systematically design a project to take account of these lessons, and as a result improve the likelihood of success. From such a case study approach, it is difficult to know conclusively. It is uncertain whether progress or outcomes would have been worse had this not been the case.

This case study has shown that attention to all the critical success factors chosen proved important to the success or failure of the projects. One factor that became apparent during the study was the subjective nature of the critical success factors and how they were interpreted in the design of the project. Just how did you judge whether the commitment of the Executive was sufficient to ensure that it would not disappear under difficult circumstances? How could you define the required relationships required to ensure the engagement of clinicians and other stakeholders?

The critical success factors and how they should be addressed should be further defined for the benefit of those implementing projects. Small organisations cannot expect to develop and retain the expertise necessary to direct the complex scenarios that lead to failure of large IT projects. Expert external resources may provide support but cannot replace a strong internal lead. Given the wide ranging nature of the factors and cultural impact, the informatics professionalism of the local programme director and their direct relationship with the Executive is proposed as an additional critical success factor.

Hospital IT projects aim to create change and achieve benefit within a complex environment. It is necessary to work within the broader constraints and not against them, and to turn these constraints to advantage. This requires adopting the change approaches that the organisation is comfortable with rather than those the IT profession might wish to adopt.

Acknowledgments

Thanks to all the staff at the Princess Alexandra Hospitals NHS Trust and all others involved in the projects including local health community, Strategic Health Authority, Connecting for Health and the suppliers.

References

[1] Department of Health, NHS Executive. Information for Health an Information Strategy for the Modern NHS. Dept of Health, 1998

[2] Department of Health. Building the Information Core: Implementing the NHS Plan. Dept of Health, 2001.

[3] Department of Health. The NHS Plan A plan for investment a plan for reform. Cm 4818-1. Stationery Office, 2000.

[4] Wanless D. Securing our Future Health: Taking a Long Term View. HM Treasury, 2002.

[5] Block R. The Politics of Projects, Yourdon Press, New York, 1983.

[6] Sauer C. Why Information Systems Fail: A Case Study Approach. Alfred Waller, 1993.

[7] Collins T. Crash: ten easy ways to avoid a computer disaster. Simon & Schuster, 1997.

[8] Fortune J, Peters G. Information Systems: Achieving Success by Avoiding Failure. Wiley, 2005.

[9] Lorenzi NNM, Riley RT. Organisational Aspects of Health Informatics. Springer-Verlag, 2004.

[10] Lorenzi NM, Smith JB, Conner, SR, Campion TR. The Success Factor Profile for Clinical Computer Innovation. In: Fieschi M et al (Eds), Medinfo 2004. Amsterdam. IOS Press, 2004.

[11] Rogers H, Silvester K, Copeland J. NHS Modernisation Agency's way to improve health care. BMJ 2004 328: 463 (letter).

[12] Office of Government Commerce. Managing Successful Projects with PRINCE2 Manual. Stationery Office, 2005.

[13] Noehr C and Bernstein K (1989). Can HIS be developed without Organisational Change? In Barber B et al (Eds), Medinfo 89. Amsterdam. North Holland.

[14] ICL. A window on the Future. An ICL Briefing for Management on the Findings of the Management in the 1990s Research Programme, conducted by the MIT Sloane School of Management. ICL, 1990.

[15] Department of Health. Tackling hospital waiting: the 18 week patient pathway. Dept of Health, 2006.

[16] Young T, Brailsford S, Connell C, Davies R, Harper P, Klein JH. Using industrial processes to improve patient care. BMJ 2004; 328: 162-164

[17] Babensky JA, Roe J, Bolton R. Lean Sigma: will it work in healthcare? Journal of Health Information Management, 2005, vol 19, no. 1.

[18] Womack JP, Jones DT. Lean Thinking. London: Simon and Schuster, 1996.

[19] Ward J, Daniel E. Benefits Management: Delivering value from IS and IT investments. John Wiley & Sons Ltd, 2005.

Address for correspondence

Sheila Bullas, BSc MBA, MBCS, CITP,
iBECK, Willsmead, Abington Pigotts, Royston, Hertfordshire, UK, SG8 0SD.

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

MUST - A Participatory Method for Designing Sustainable Health IT

Finn Kensing, Hrönn Sigurdardottir, and Arjen Stoop

IT University of Copenhagen, Denmark

Abstract

Several important issues in designing sustainable health IT, such as coherent visions for change and genuine user involvement, are too often neglected or not paid enough attention to in practice. The MUST method addresses the early stages in the design of sustainable IT applications. The method highlights how those issues can be dealt with as it provides practical recommendations in terms of principles, tools and techniques. The method has proven helpful in assisting project participants to focus on and combine issues that are also crucial when designing health IT. MUST has been developed and tested in commercial settings. Here however, we illustrate the method's potential for health IT as it was recently used for the evaluation of a faulty health IT project intended to support shared care in relation to pregnancy.

Keywords:

health informatics, action research, methodology, systems design, shared care, sustainability

Introduction

Several important issues in designing health IT, such as a coherent visions for change, real user involvement, understanding the work that is supposed to be supported by IT, and the way proposed changes have to be incorporated in daily work, are in practice too often neglected or not dealt with properly [1-4]. As a consequence, sustainability of the IT system lacks. By sustainability we refer to the need for a balance between development, use and protection of an organizations' resources (financial foundation, personnel and IT systems). The MUST method we present here has proven to be able to deal with these issues. The method was developed and tested through a longitudinal action research program involving 14 IT projects in public and private companies [5]. Here the method is used for evaluation rather than for design, therefore the paper is of an exploratory character.

The MUST method

MUST [6] consists of four resources to assist project participants in planning, conducting, and evaluating their projects. By resources we stress that we are not dealing with a cook book type of method, instead project participants each time have to decide which method elements to include and how. The resources are: 1) A range of concepts and conceptual frameworks, 2) Four guiding principle, 3) Four phases and 4) Sixteen tools and techniques. We restrict ourselves to presenting the principles in some detail and to simply list relevant tools and techniques for each of them. Further we give the rationale behind the four phases.

The four phases

The method suggests an IT project to be divided into two separate projects [6], a design project and a realization project. In between the two we find the call for tender and contractual negotiations. MUST deals only with the design project. A design project results in one or more coherent visions for change on the basis of which the customer decides which vision, if any, should be pursued through a realization project. This division is especially needed when IT is designed to support communication and cooperation between actors from different professions and organizations. The MUST method suggests an IT design project to be divided into four phases as illustrated in Table 1. The table shows the focus and the result of each phase as well as the decisions supported by each phase.

Table 1 - The four phases of MUST, their focus, the subsequent results, and the decisions they support

Phases	Focus	Results - decisions
The initiation phase; establishing the project.	The scope of the IT design project: time, costs, contents, participants.	Project charter & a plan - deciding the scope and basis of the IT design project.
The alignment phase; understanding management strategies.	Relations between the goals of the IT design project and the organization's business and IT strategy.	Strategic analysis report - selecting work domains for further analysis.
The in-depth analysis phase; understanding current work practices.	The work practices of the selected work domains.	Analysis report & descriptions of work practices - prioritizing goals, problems and needs, and ideas for IT.
The innovation phase; developing visions.	Visions of IT systems and their relations to the organization of work and the users' qualifications. The realization project.	IT design report & mock-ups and prototypes - deciding which visions should be realized and how should this take place.

The four principles

The principles express an overall perspective built into the method while the tools and techniques provide more concrete recommendations on how to conduct various activities. These may be substituted by other tools and techniques, which serve the same needs. Instead, structuring the design process according to the four phases, and striving to live up to the four principles is considered important – as we shall see in the evaluation below.

Principle 1. Coherent visions for change

The result of an IT design project is one or more coherent visions for change in the organization in question and in relations to its environment. The proposed change should meet the organization's revealed goals, needs, and opportunities within its business and IT strategy - which may itself need revisions as part of the project. By a coherent vision we mean that the following three elements of an IT application are designed to support each other: IT systems, work organization, and the qualifications users need to perform their job with the help of the proposed IT systems in the proposed work organization. If management and clinicians don't share the same vision of how to support the strive for better economy, quality, and service for the patients, the intended changes will not occur [7]. Relevant tools and techniques that support this principle include workshops, prototyping, and scenarios.

Principle 2. Genuine user participation

If users contribute solely as informants, it is not considered genuine user participation. Instead this principle prescribes the active participation of end-user representatives influencing the process of design as well as the visions it results in. There are two rationales for this, a pragmatic and a political. The pragmatic argument is that mutual learning between users and IT designers is needed, while the political argument shows a concern for the employees' rights to influence their own working conditions. The political argument may also be justified by the simple fact that e.g. health care professionals have demonstrated that they have the power to block an IT application that they find irrelevant or too cumbersome to use [8, 9]. Relevant tools and techniques that support this principle include user representatives in project groups and steering committees, prototyping, workshops, and hearings.

Principle 3. IT designers' need to experience the users' work practices

Basically, there are three different ways of obtaining new knowledge relevant for an IT design project: Reading about the subject matter, asking knowledgeable people to tell you about it, or allowing yourself to experience the subject matter first hand. This principle argues for the need to include also the latter to really understand the work practices for which you are developing IT support. In recent years a substantial critique has been put forward to the development of health care IT not taking account of the social aspects of health care work [10]. By observing the users while performing their regular business or while trying out prototypes in situations as realistic as possible, IT designers acquire knowledge and understanding of the work and communication processes that need to be supported by the IT system. Relevant tools and techniques that support this principle include observations, in-situ interviews, and thinking aloud experiments.

Principle 4. Anchoring visions for change

This principle focuses on three groups, whose members usually cannot all participate directly in an IT design project: 1) Management, who have the power to decide whether or not the proposed visions will be implemented, 2) Employees and other interested parties, who will either use the IT systems in question, or who will be affected by them, 3) Internal and external people taking care of the technical and organizational realization activities that are required to implement the proposed visions. The principle prescribes that these groups must be informed and involved in various ways to be able to evaluate the consequences of the proposed changes, as seen from each of their perspectives. This needs to occur in time for the project group to incorporate their reactions into the final design proposals. If visions are not anchored with these groups the expected changes are unlikely to take place. Relevant tools and techniques that support this principle include prototyping, scenarios, reviews, and hearings.

The case: Sundhed.dk

To illustrate the MUST method's relevance within Health Informatics we have used it for the evaluation of a web based pregnancy application. The intent of the project was to support communication and collaboration among midwifes, obstetricians, general practitioners, and the pregnant women. Especially this type of health IT has been proven difficult to manage [11].

The pregnancy project was launched as a part of a large Danish public e-health portal initiative in the year 2004. The portal (in Danish Sundhed.dk) provides a framework for electronic communication between the parties involved in the Danish National Health Service and communications with the patients. Furthermore the portal provides information and services for the citizens and for health care professionals. According to the contract, the vendor who won the tender for the portal should also design web applications to support shared care for two widespread diseases. These web applications should be designed in a way so that they would be able to function as a generic model for future web based services. Even though pregnancy is not a disease it was chosen since pregnant women are known to be competent users of the Internet and web based services. In addition at the outset, pregnancy was considered relatively simple to support. The portal has received several national and international rewards for providing excellent service to citizens and health care professionals and outstanding user interface.

Research methods

For a detailed description and evaluation of the research method used to design the MUST method readers are referred to [5]. Instead here we concentrate on the method-

ological considerations as to how the evaluation of the pregnancy project was carried out. The evaluation is here used as the basis for proposing the MUST method as a resource for designing health IT. The main data collection tools were 1) participant observation at project group meeting over a period of eighteen months (documented in notes); 2) eight qualitative interviews (documented in notes and on tape) with project participants (IT specialists, user representatives, and management form both the customer and the vendor side); 3) a full day workshop (documented in a report approved by the participants), and document analysis.

Results

This section draws on a recent evaluation of the pregnancy project, focusing on the process rather than on the resulting product. The evaluation report was approved with minor corrections by all involved in the evaluation and later used as the basis for a reformulation of sundhed.dk's business strategy and for developing future services. In the evaluation below we end each section by listing proposals for improvements. We use the four principles of the MUST method as a conceptual framework for the evaluation of the pregnancy project, thereby also demonstrating the design potential of the method. The argument is that if the project had lived up to these principles it would have had a greater chance to succeed. Unfortunately, in contrast to clinical trials, we are not in a position to isolate a set of factors and conduct an experiment based on which we may claim evidence as to a design method. This is due to the complexity of IT projects and their highly contextual dependencies.

Coherent visions for change

For the main course of the project the participants did not share a common understanding of what they were designing. This applies to the functionality of the pregnancy application - the first element in MUST's understanding of a coherent vision for change. It also applies to the second element – work organization – since the web application's integration to other IT systems remained a battleground for most of the project's lifetime. Therefore it remained unsettled for a very long time how doctors should organize their record taking in relation to their current EPR's and the new web application. Instead the third element – users' need for new qualification – was by all parties considered not to be a problem. Despite the above problems the web application in itself was considered simple. The vendor produced a preliminary draft for a general description of the expected course of events and of the associated systems integration. However, this should be perceived only as a first general project draft to be developed further in close cooperation with the partners involved in the pilot project. Therefore the participants from the two counties that signed up for a pilot project expected to be able to obtain great influence on the design of the web application. This turned out to be a problem since few resources were set aside for experiments. Thus constant discussions took place on whether ideas for improvements were inside or outside the scope of the contract. Like for the web portal, the idea of the pregnancy application is "simply" to provide access to data already available. The participating counties differed on this issue, as they had different prerequisites for participating in the pilot project. The consequence of this was never analyzed, or at the least no consequences were drawn on the basis of this. The one county did not have an IT system that could be used for generating data for the web application, while the other had such systems. Therefore the participants from the former kept arguing that the web application should support doctors in entering and retrieving health data about their patients, while participants from the latter preferred to use their existing IT systems. The solution became to support both. This landslide in the conceptualization of the product, developed slowly. Some key actors let it happen – they wanted to keep decisions open as long as possible, thus maintaining the reasons for the counties to be part of the project. Others pursued openings to have the application go in the directions they wanted.

The character of the contract between the customer and the supplier was the main reason for the participants not being able to agree on the product's functionality since it consisted of an incomplete requirement specification. Both the customer and the supplier used the contract against the counties in the latter's attempt to get what they wanted. There are both negative and positive consequences of this course of events. The pilot project did clarify to some degree the concept of 'shared care' and how IT may support this. This in turn helped clarifying what 'a generic model for future web based services' might mean. This is crucial for future sundhed.dk projects. On the other hand 'generic' also turned out to be a fuzzy concept. Concerns were raised about the possibility to rely on a generic model in future shared care projects. Finally, the constant questioning of the scope of the project resulted in the inclusion and exclusion of processes and work practices in ways that did not contribute to the progression of the project. The following ideas were suggested based on the evaluation:

- Sundhed.dk needs to take on the job of explaining to its funding partners that its different projects move in unreclaimed territory, which means that room for experiments and systematic evaluations are needed.
- Before commitment to a fixed time, fixed functionality, fixed prize type of contract, the project should have developed a coherent vision for change through an experimental approach using MUST.
- Sundhed.dk has to take the role of coordinating the counties needs for IT support for shared care, or it must see to that another agent plays that role.
- The funding partners of sundhed.dk need to make an informed choice - as to various types of projects - about the role(s) of sundhed.dk at a national level.
- The partners need to understand that different roles come with different prerequisites in terms of project funding capabilities and different qualifications and authority to run a project.

Genuine user participation

Three central groups of users were involved too late and not in adequate ways for them to influence the process of design as well as the product it resulted in. These groups

are: Pregnant women, visiting nurses, and general practitioners and their IT vendors. They are a mix of individual and organizations that are not part of the same line of command. Here there is only room for including an evaluation in relation to the pregnant women.

Pregnant women were involved too late and mainly for testing purposes. The main reason for this was that the project group decided that user needs had been collected once and for all three years before the project really started. However, not even these needs were taken care of. The consequence was that the wishes, needs, and requirements of perhaps the primary user group were never really dealt with. The following ideas were suggested based on the evaluation:

- A proper stakeholder analysis should be conducted before requirements are developed and decided upon.
- Users should be involved in order to develop their wishes, needs, and requirements as well as in an evaluation of the degree to which their interests are met by an application under development.
- An iterative process is recommended for projects where wishes, needs, and requirements are not clearly stated at the outset or where conflicts of interests are to be expected.

Experience the users' work practice

The project group totally ignored this principle. The reason was that they were either afraid to - or not qualified to - expose themselves to the situations for which their IT application was intended. Thus, they managed to keep a distance to the real needs of the various types of intended users. The consequence was devastating for the quality of the pregnancy record, as its functionality and modes of interactions did not meet the needs of its intended users.

The supplier's suggestions for how to meet the requests of the customer, was based on a too narrow understanding of the work processes that the project group finally tried to support. This flaw was never faced. Sundhed.dk took on the task of a usability study, but pregnant women were not included even though an important part of the rhetoric around sundhed.dk was that the portal was about supporting patients/citizens. Further, the clinical work practices were not analyzed.

The following ideas were suggested based on the evaluation:

- Members of the project group should experience at the least one visit of a pregnant woman to her general practitioner's consultation.
- Members of the project group should observe a pregnant woman in her home for one day in the beginning, half way trough, at the end of her pregnancy, and during the first weeks after the delivery.

Anchoring

The project was organized so it should have been possible to live up to this principle. At the national level both management and the vendor, who took care of the technical implementation, were fully able to evaluate the consequences of the proposed changes, as seen from each of their perspectives. Further, both counties had representatives in the national project group and they each sat up county-based project groups. They arranged meetings to inform the clinicians, who were also involved in pilot testing and teaching workshops for them to familiarize themselves with the pregnancy application and provide their feedback. However, one county withdraw from the project after waiting many months for its vendor to estimate the costs of integrating the web application with its local system – the costs were considered too high. The other county also withdrew from the project since the application disappointed the clinicians.

The anchoring problem was not so much due to bad project organization and management at the national or at the county level. Rather, the problems were caused by the lack of a coherent vision for change as already accounted for and by the fixed time, fixed functionality, fixed prize type of contract that did not allow for experimentation and learning.

The following ideas were suggested based on the evaluation:

- Distinguish between at least two types of projects that correspond to different types of contracts. For the one type it is possible to freeze a requirement specification before the start of the technical and organizational realization project. For the other type, user representatives need to take part in iterative explorations of problems and possibilities along side the development of solutions. For the first type of project, a fixed time, fixed functionality, fixed prize type of contract may work. Instead for the other type of project a development contract is needed, that allows for experimentation and learning.
- Establish mechanisms to detect if the wrong project model has been selected, and adjust accordingly.

Discussion

Many participatory design *tools* and *techniques* have been developed [12]. Developing a coherent participatory design *method* has not been the aim of the PD community. However, a few groups have systematically organized their design practices into a coherent method. Most of these are only documented in one ore more papers see e.g. Grønbæk et al [13] and Blomberg et al. [14]. In addition to the MUST method, the only other coherent method published as a usable guideline is Contextual Design (CD). Readers are referred to [15] for a proper comparison of MUST and CD, here we just list that 1) CD and MUST have the same *application area*: a participatory design approach that focuses on early design activities. 2) CD blurs the distinction between managers and users referring simply to 'customer' where as to MUST 'users' and 'managers' are different categories. 3) MUST acknowledges that there may be different interests at stake and it provides ways to deal with them without assuming that conflicts may be resolve in a harmonious way, instead CD is not explicit about conflicts and thus provides no means for dealing with them.

4) MUST provides tools and techniques for project management, CD does not. Other traditional software development methods also have addressed the issues and concerns that are dealt with by the MUST method. There are thought significant differences between the MUST and the traditional methods that we in the following will address shortly. Compared to traditional waterfall models and evolutionary models, MUST is designed to incorporate the best – and avoid the worst – aspects of these approaches. We recommend iteration between analysis and design activities within and among the project phases. Other contemporary methods, like Rapid Application Development (RAD) and Extreme Programming (EP) also have abandoned the traditional phase model, but differ from the MUST method particularly in relation to how and when the decision for what kind of a system is to be built, is acquired. Both RAD and EP have incorporated the assumption that a decision to build a system of a particular kind already has been made. RAD, EP and and Rational Unified Process (RUP) focus on building systmes from scrach. In turn, RUP does incorporate early design activities but has strong focus on modeling, specifications and implementation. Our last comparison is to the Business Process Reengineering method. Beeing simliar in scope, MUST and BPR both aim at formulating visions for the future use of IT and leave out the organizational implementation but consider the relationship between a design project and an organizations business and IT strategies. BPR does not deal with ethical or practical isssues relating to the users and therefore does not provide any help in understanding, developing or presenting relations between IT and user's work practices.

Here we have used elements of the MUST method for the evaluation of a recent failing health IT project. We demonstrated that if two of the method's four resources had been applied, the project would have had a better chance of success. Due to space limitations we did not find room for neither the introduction nor the application of MUST's other two resources. Anyhow, we have demonstrated that methods like MUST are relevant for health informatics especially when it comes to the type of IT applications that have the ambition to support shared care. This type of application is characterized by the fact that neither health professionals nor IT specialist have good models for what the application should be like. Further, such projects need to be conducted in a context in which there is no unity of command. This implies that experiments are called for, and that all parties are allowed to systematically evaluate the degree upon which their interest are taken care of. MUST is such a method.

Acknowledgements

We want to thank Keld Bødker and Jesper Simonsen, who together with the first author of this paper developed the MUST method. Further, the project group of the pregnancy project and sundhed.dk as an organization participated openly and honestly in the evaluation of a rather troubled project for the purpose of improving their performance in future projects.

References

[1] Atkinson CJ, Peel VJ. Transforming a hospital through growing, not building, an electronic patient record system. Methods Inf Med 1998; 37(3): 285-93.

[2] Ash JS, Berg M, Coiera E. Some Unintended Consequences of Information Technology in Health Care: The Nature of Patient Care Information System-related Errors. J Am Med Inform Assoc 2004; 11(2): 104-12.

[3] Berg M. The search for synergy: interrelating medical work and patient care information systems. Methods Inf Med 2003; 42(4): 337-44.

[4] Forsythe DE, Buchanan BG. Broadening our approach to evaluating medical information systems. Proc Annu Symp Comput Appl Med Care 1991: 8-12.

[5] Kensing F. Methods and Practices in Participatory Design [Dr. Scient dissertation]. Copenhagen: ITU Press; 2003.

[6] Bødker K, Kensing, Finn., Simonsen, J. Participatory IT Design. Designing for Business and Workplace Realities: MIT Press; 2004.

[7] Nikula R. Why Implementing EPR's Does Not Bring about Organizational Changes - A Qualitative Approach. Studies in Health Technology and Informatics 2001; 84:666-9.

[8] Levitt JI. Why physicians continue to reject the computerized medical record. Minn Med 1994; 77(8): 17-21.

[9] Scott JT, Rundall TG, Vogt TM, Hsu J. Kaiser Permanente's experience of implementing an electronic medical record: a qualitative study. BMJ 2005; 331(7528): 1313-6.

[10] Coiera EW, Jayasuriya RA, Hardy J, Bannan A, Thorpe ME. Communication loads on clinical staff in the emergency department. Med J Aust 2002; 176(9): 415-8.

[11] Berg M. Health Information Management. Integrating Information and Communication Technology in Health Care Work: Routledge; 2004.

[12] Kensing F, Blomberg J. Participatory Design: Issues and Concerns. Computer Supported Cooperative Work (CSCW) 1998; 7(3): 167-85.

[13] Grønbæk K, Kyng M, Mogensen P. Toward a cooperative experimental system development approach. In: Computers and design in context: MIT Press, 1997; pp. 201-38.

[14] Blomberg J, Suchman L, Trigg RH. Reflections on a Work-Oriented Design Project. Human-Computer Interaction 1996; 11(3): 237-65.

[15] Beyer H, Holtzblatt K. Contextual Design: Defining Customer-centered Systems: Kaufmann; 1998.

Address for correspondences

Finn Kensing,
IT University of Copenhagen, Rued Langgaards Vej 7
DK-2300 Copenhagen S, Denmark, e-mail: kensing@itu.dk

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Change Management and the Sustainability of Health ICT Projects

Karen Day[a] and Tony Norris[b]

[a] *School of Population Health, University of Auckland, New Zealand*
[b] *Institute of Information and Mathematics Sciences, Massey University, Auckland, New Zealand*

Abstract

The development of the electronic health record (EHR) is a strategic and important enabler of the delivery of integrated healthcare. As each innovative aspect of the EHR is implemented in New Zealand, long-term success is essential for its overall sustainability on the national scale. How we achieve this success is dependent upon how people adapt to the changes brought about by the implementation of these innovations. The transition period of the change process we follow during this adaptation is characterized by a capability crisis, in which we tend to predict failure in our attempts to make the changes to which we are committed. This could be a signal of the first step toward sustainable change as people adapt to changed processes, technology and relationships. Once we have mastered the incremental changes brought about by health ICT projects for the implementation of the EHR, we are able to connect health services by means of the same EHR and provide enabled, sustainable integrated healthcare.

Keywords:

change, sustainability, failure, electronic health record, integrated healthcare

Introduction

Health ICT projects are notorious for their failures, despite the efforts of project managers to embed as many critical success factors into the project design as possible. Despite these failures we successfully implement innovations such as the electronic health record (EHR), build on our successes, and sustain their effect.

This paper takes a look at the sustainability of the electronic health record in New Zealand, which is the result of a series of incremental innovative implementations of aspects of the EHR, within the national framework of providing an infrastructure and capacity to build capability for a distributed EHR. The concept sustainability is defined and is linked to change in the complex adaptive system of healthcare. A description is given of the capability crisis that marks the transition period in the change process, and a discussion follows on the relevance of predictions of failure to ultimate sustainable success of an innovation. It is concluded that although we tend to predict failure in the short term, especially during the transition to change, we do not enact our predictions. On the contrary, we adapt to and adopt new processes, technology and relationships that support the growing EHR in order to predispose our healthcare providers to deliver integrated care.

Sustainability defined

There is a general impression that ICT projects tend to fail, with failure rates ranging between 50% and 80%. (1-3) This failure rate appears to apply to health ICT projects as well. There are different descriptions of such failure. A project may be partially successful where it achieves some but not all of its goals or it could simply be an outright failure or be abandoned for any number of reasons. It may be considered a failure because there is no change in an organization after completion of a project, or there may be unpredicted undesirable project outcomes. Some projects are successful in one setting but when implemented in a different setting they fail. Lastly, there are sustainability failures. (3-5)

The term sustainability is used in several contexts and with different meanings. Conversations about sustainability usually include words and phrases such as institutionalization, realization of long-term benefits, long-term continuation of an innovation, financial self-sufficiency, efficiency, and durability. (6) Thus one can assume that an ICT innovation that has been introduced has been sustained if it has been diffused into an organization (7): it has become institutionalized, long-term benefits are occurring, it is financially self-supporting (especially after the funding for the initial project is no longer there) and it is proving to be efficient. It has become part of business as usual, the way we have always done things. It is also essential that the innovation is scaled up to the whole organization and in the case of the electronic health record, the region and whole country wherever appropriate. (8)

Institutionalizing a health ICT innovation is fraught with sustainability issues, especially when funding ends with the completion of an initial project. (8) Political support is essential in the introduction of innovations, particularly long-term support. Once there is executive and leadership support, it is essential that capacity is built and resources are provided for ongoing support and maintenance of the innovation. (6, 8) New technology, processes and relationships need to be institutionalized in order for the innovation to become business as usual.

Sustainability and change

Sustainability is therefore a part of the change process, that is, the change cycle involved in the introduction of an innovation does not end when its implementation project ends. (6) When we change it appears that we follow a process that approximates Lewin's (9) unfreeze, change, refreeze cycle. Most change theories indicate that we follow a process that marks our acknowledgement of the change confronting us and our transition through to final adaptation to the new way of working or living, aiming to achieve a better state than when we started out. (10, 11) The trough in this process, as depicted in Figure 1, can be destructively deep if we attempt to change too much or make a fundamental change without preparatory changes. (4) In order to reduce the impact of such an effect, and to improve the probability of sustained change, we could fill the trough with small accumulative changes so that the larger, more radical changes are laid over, using the preceding ones as their foundation. (4, 12)

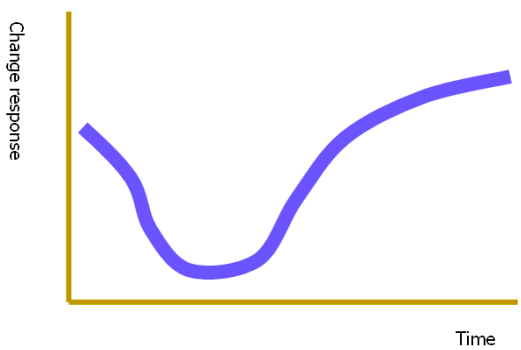


Figure 1 - The change process(10)

In the development of the EHR in New Zealand, the trough of change was filled initially with infrastructure projects that would support future implementation of more components of the EHR. (12) This means that once the ICT infrastructure and foundation administrative IT (information technology) were in place, other more complex clinical applications could be implemented. Incremental changes could be represented in Figure 1 as multiple smaller change curves within the larger curve of the complete EHR. (13) This means that the overall change is cumulative and therefore possibly more likely to be sustained. In this way we accommodate the large change in small increments, thus allowing for the institutionalization of the EHR as we adapt from one change to another, interdependent change. The accumulation of changes results in an accumulation of sustained innovations as each innovation is absorbed as the starting place for the next innovation.

In this way long-term benefits start to appear as we adapt to the changes, as the innovations are absorbed into an organization. (4) The proof of the pudding of sustainability lies in the long-term benefits for the stakeholders and the extent to which they perceive the added value of the innovation being implemented. The challenge for the success of health ICT projects lies more in its sustainability than in the success of the implementation project itself. It lies in the long-term adaptation of the system in which it is introduced rather than in the short-term success of the project. (8) This means that a partially successful project can become fully successful in the long-term, assuming that it achieves the final planned outcomes and sustainability. The accumulative development of the EHR in New Zealand requires a sustained capacity for adaptation to change. In the complex adaptive system of healthcare there is a constant stream of changes that we experience every day. This is compounded by the innovative EHR applications that are being implemented with a view to sustained change in the way we deliver healthcare in a paperless environment. (14)

Complexity and sustainable change

Health can be viewed as a complex adaptive system (15), in which many parts of the system interact interdependently in varying and unpredictable degrees with one another and their environment (16-18). The continuum of complexity ranges from simple and unambiguous with high degrees of perceived certainty, to chaos which extends beyond complexity, uncertainty and ambiguity. Within this context capability is potentially at its best when there is a moderate degree of complexity. (19) It is when things are reasonably complex that change is most stimulating and best received. We usually function well in the position where most of our world is reasonably certain and predictable, fairly unambiguous, familiar, mostly known and knowable, and where interdependencies and relationships are fairly simple. (14) Paradox and tension co-exist in this environment where fuzzy boundaries and sensitivity to initial conditions make it hard to predict the future impact of the implementation of innovations. Emergence of unexpected and unpredictable phenomena is common in health ICT projects, as is characteristic of complex adaptive systems. Changes occur within changes as a result of the butterfly effect (20), which results in disproportionate consequences to even the best laid plans.

The diffusion of a health ICT innovation is not a simple, linear process of plan, implement, review, and reap the benefits. The diffusion itself is fraught with unpredictable consequences, the difficulties of long-term benefit management, and paradoxes that are difficult to assess and exploit. The Health Information Strategy (HIS) for New Zealand (21) attempts to take advantage of these characteristics of healthcare as a complex adaptive system by focusing on the development of a technological infrastructure to support the EHR, parallel to the development and nurturing of technological capability. In this way a distributed EHR is developed that is predisposed to national dispersion and thus the scale required for sustainability. (8) This allows for emergence in such a way that incremental innovations are supportable and scalable according to the capacity of the services in which they are implemented. Once the ICT that is needed for in-house EHR management for one health service is on its way to being institutionalized, an organization is able to make forays

into connecting up with other services, such as primary care organizations. In this way linkages are made between services that allow for more innovations to be implemented. On the surface this approach may appear useful. However, even when attempts are made to maintain manageable projects for implementing these innovations, there are times when the scale of an innovation itself is large, or is perceived to be overwhelming by those involved.

The enormity of change implications

Although we appear to follow a process (depicted in Figure 1) when adapting to our changing and complex environment, the implications of planned change appear to overwhelm most of us during the transition phase of the process. This transition is marked by commitment to planned change but an overwhelming sense that it is all is too much too soon. This transition period usually lasts for a short while (hours or days), but for some people it continues for months. In combination these features appear to comprise a capability crisis. (22)

This capability crisis was researched during the implementation of a large infrastructure project that aimed at establishing a single Information Services platform for an emerging shared services organization for two New Zealand district health boards. (22) Action research was used as a way of managing change in this project in addition to providing a platform for the research to explore the role of the transition aspect of the change process.

The capability crisis described by the project participants is characterized by a heightened sense of ambiguity and complexity where nothing seems certain and everything seems more complicated than ever before – it occurs at the transition period of the change process when people are moving out of their usual way of working and into the complexity of a world in which everything seems to be changing, fluid, unpredictable and difficult. Sensitivity to initial conditions is exhibited in the disproportionate increase in workload, where people are working on multiple levels of the old ways, learning and adopting new ways, and integrating the change into their operational activities. One participant in the project described her experience in an interview towards the end of the project as follows.

It was a terrifying experience because ...it was all very well for other people to be quite glib about that but the scope was creeping. The scope of my task was expanding by the minute.

Paradox and tension are evident in the disproportion between the need and availability of resources. Also, communication appears to be an issue in that people need and demand information, yet they seem to be unable to use it. There is a demand for leadership and yet people appear to be unable to relate to the leadership that is available. People resonate with leaders of their choice rather than taking leadership from appropriate sources as described by another project participant.

He knew he was going to deliver, and he knew it would work and he knew a lot of people thought it wouldn't work. In terms of his leadership there was no question – they could say whatever they like, it was going to work. In terms of leadership and taking a stand that was great.

In addition, predictions of failure mark this crisis: although people are committed to a project's goals and ultimate success, there is a period in which they are uncertain of that success and cannot see how it could be achieved. This tendency to predict failure appears to contradict commitment to project success. It seems that although people appear to be resisting the changes brought about by health ICT innovations, the capability crisis marks their initial attempt at adapting to the change. It is a sign of commitment to the project rather than to the past, the status quo, as described in an interview with a team leader.

You're identifying problems but you don't know there are solutions in the situation. Since you don't know, you feel it's a losing battle.

Predictions of failure

It is during the capability crisis that we realize more fully the implications of our commitment to the project at hand. We suddenly feel inadequate for the task and frustrated that we are unable to learn and develop skills fast enough to be able to make the transition to the new ways of working, using new technology, processes and in new relationships with our colleagues and the project stakeholders. There seems to be a background chatter predicting failure, as described by the training team leader:

There was a lot of talk around me about the project ... "not going to work" ...people around you creating an uncertain environment by the things they say ...making it hard to deal with the situation ....saying things like "how's this going to work? We're never going to get it off the ground."

At the same time as the background chatter of predictions of failure, we hear in counterpoint the management mantra that we must make the project work. Some participants found it hard to relate to what they called "Hitlerism" when the managers did not appear to take their concerns seriously and merely repeated that the project would go ahead and that it would be successful. This was described by an interviewee quoting snatches of conversations she witnessed:

... "but you can't do things like that" ... "but we will do it this way" ... "but you don't understand you can't do it like that because..." "we have to do it, we have a time frame and we have to go ahead" ... "we're rolling..."

Although predictions of failure appear to occur in the early stages of the change process and seem to be resistance to change, we cannot assume that they will disrupt a project or lead to its actual failure. These predictions of failure could be a sign that people are progressing through the early stages of the change process. (10) This means that the commitment to the change brought about by a project remains intact but the predictions of failure may be signaling attempts to accommodate the finer aspects of change in

order to contribute to project success. This is supported by the appearance of frustration and heightened awareness of issues and the inability to perform effectively during a period of change.

Despite protestations that the project would be a failure and that it could not progress to its predicted outcome, people forged on with their project activities. Although there is a temptation to ignore these protestations and predictions of failure, it may be more productive to respond with a mantra that continues to support continued efforts aiming at project success. The capability crisis marks commitment to the project's goals, and therefore its success and eventual sustainability. It is important for managers and leaders to acknowledge the predictions of failure, and to deliberately use these predictions as the background to their forward-looking mantra that focuses project participants on future success. Towards the end of a project the predictions of failure and mantra of success fade as we focus on winding up project activities.

Conclusion

We adopt innovations in many ways, levels and degrees of complexity. The important thing is that the changes brought about by the implementation of the EHR are sustained, especially in a complex system such as healthcare. Sustainability is characterized by the long-term realization of benefits, the adoption of an innovation as part of our daily routine: the institutionalization of the implemented innovation. The adoption of an innovation requires changes to the way in which we work. People appear to follow a change process is which we commit to the change, transition from old ways of working and assimilate the new processes, technology and relationships into our daily activities: the changes become part of how we have always worked.

The transition period in the change process is marked by a capability crisis in which we are overwhelmed by the implications of the project in which we are implementing an innovation. The project marks a foray into complexity that results in a temporary capability crisis. An aspect of the crisis is predictions of failure which is more an expression of adaptation to the innovation than resistance to change. Although there is fear of failure in the early stages of implementation of an innovation, these predictions of failure and the capability crisis are not heralds of project failure – they are more likely a sign that people are making efforts to adapt to the innovation's associated changes. It is up to managers and leaders to keep a look-out for the capability crisis and use it (complete with its predictions of failure) to support efforts to adapt to the changes brought about by such a project. A mantra that encapsulates the project vision should form the counterpoint for these predictions.

What this means for the implementation of the EHR is that large changes could be disruptive and destructive in that they create capability crisis that is too big to handle for individuals as well as for organizations. The incremental approach used in New Zealand has predisposed people working in healthcare to prepare for and adopt innovations in the EHR in an interdependent manner. In this way, successes can be built upon one another, resulting in the building of capability and IT in a sustained manner. As we master one innovation we are able to prepare for and master the next one, consequently absorbing them into our daily work activities. In this way we can exploit paradoxes, tensions and unplanned consequences that emerge as a result of small successes.

As saturation of EHR innovations is achieved within organizations such as district health boards, each service should become capable of making links with other organizations in the interests of integrated healthcare. Small successes accumulate until the distributed EHR develops the capacity to become a nationally connected EHR, thus achieving the scale required for the success of a sustained innovation, the electronic health record.

Acknowledgments

The authors thank project participants at healthAlliance for their contributions to this research.

References

[1] Shore B. Failure rates in global IS projects and the leadership challenge. Journal of Global Information Management. 2005;8(3):1-5.
[2] Glaser J. Why does an IT investment fail to deliver returns? Healthcare Financial Management. 2003;57(4):114 - 21.
[3] Cozijnsen AJ, Vrakking WJ, van Ijzerloo M. Success and failure of 50 innovation projects in Dutch companies. European Journal of Innovation Management. 2000;3(3):150-9.
[4] Heeks R, Mundy D, Salazar A. Why health care information systems succeed or fail. Information systems for public sector management. Working paper series 9. Manchester: Institute for Development Policy Management; 1999.
[5] Keeling R. Project management: An international perspective. London: MacMillan Press Ltd; 2000.
[6] Johnson K, Hays C, Center H, Daley C. Building capacity and sustainable prevention innovations: a sustainability planning model. Evaluation and Program Planning. 2004;27:135 - 49.
[7] Rogers EM. New product adoption and diffusion. Journal of Consumer Research. 1976;2(March):290-301.
[8] Braa J, Monteiro E, Sahay S. Networks of action: sustainable health information systems across developing countries. MIS Quarterly. 2004;28(3):337-62.
[9] Lewin K. Field theory in social science. New York: Harper and Row; 1951.
[10] Elrod II PD, Tippett DD. The "death valley" of change. Journal of Organizational Change Management. 2002;15(3):273-91.
[11] Moore GA. Crossing the chasm: marketing and selling high-tech products to mainstream customers. New York, N.Y.: Harper Business; 1991.
[12] Orr M. Evolution of New Zealand's health knowledge management system. British Journal of Healthcare Computing and Information Management. 2004;21(10): 28-30.
[13] Orr M, Sankaran S. Mutual empathy, ambiguity and the implementation of electronic knowledge management within the complex health system. Systems Thinking and Complexity Science: Insights for Action 11th Annual

ANZSYS Conference: Managing the Complex 2005 5-7 December Christchurch, New Zealand. ; 2005.
[14] Plesk PE, Greenhalgh T. Complexity science. The challenge of complexity in health care. British Medical Journal. 2001;323:625-8.
[15] Dooley K. A complex adaptive systems model of organization change. Nonlinear Dynamics, Psychology, and Life Sciences. 1997;1(1).
[16] Tan J, Wen HJ, Awad N. Health care and services delivery systems as complex adaptive systems. Examining chaos theory in action. Communications of the ACM. 2005;48(5):37-44.
[17] Plesk PE, Wilson T. Complexity science: complexity, leadership, and management in healthcare organisations. British Medical Journal. 2001;323:746-9.
[18] Plesk PE, Greenhalgh T. The challenge of complexity in health care. British Medical Journal. 2001;323:625-8.
[19] Fraser S, Greenhalgh T. Coping with complexity: Educating for capability. British Medical Journal. 2001;323(7216):799-803.
[20] Gleick J. Chaos: making a new science. New York: Penguin Books; 1987.
[21] Health Information Steering Committee (NZ). Health Information Strategy for New Zealand. Wellington: New Zealand Ministry of Health; 2005.
[22] Day K, Norris AC. Supporting information technology across health boards in New Zealand: themes emerging from the development of a shared services organization. Health Informatics Journal. 2006;12(1):13-25.

Address for correspondence

Karen Day,
School of Population Health
University of Auckland
Private Bag 92019
Auckland Mail Centre
Auckland, New Zealand
k.day@auckland.ac.nz

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

A Sustainability View on the EPR System of N. N. Burdenko Neurosurgical Institute

Michael A. Shifrin, Elena E. Kalinina, Eugene D. Kalinin

N. N. Burdenko Neurosurgical Institute, Moscow, Russia

Abstract

This paper aims at the analysis of the "sustainability status" of an electronic patient record system developed at the Medical Informatics Laboratory of the N. N. Burdenko Neurosurgical Institute (EPR/NSI). It includes some of the principles that allowed a small team of developers to create a sustainable EPR system for a large medical institution with complicated diagnostic and treatment processes.

Keywords:

electronic patient record, business processes, database, formalization, sustainability.

Introduction

This paper is concerned with the analysis of factors of sustainability of medical information systems on the example of a single system – the EPR for the N. N. Burdenko Neurosurgical Institute. It takes a shape similar to a clinical case report in which the analysis of a single complicated case helps to understand general rules.

The N. N. Burdenko Neurosurgical Institute is a hospital specializing in the neurosurgical treatment of patients presenting all types of neurosurgical pathology. It has 300 beds in the clinical departments, 40 beds in ICU, 14 operating rooms, diagnostic, radiological and rehabilitation departments, an Outpatient Department, and a full set of diagnostic and laboratory equipment. The Institute carries out a number of important scientific programs in medicine and neuroscience.

The EPR system has been running since 2000. It has about 600 users; it supports the documentation of almost all events, investigations and manipulations taking place during the diagnostic and treatment process (D&TP). The vast majority of paper documents forming the case reports for patients (required under Russian legislation) are hard copies of electronic documents generated in EPR/NSI. There are three main categories of users of the EPR: medical staff (both nurses and doctors), medical administration and scientific researchers. The system is running in 7/24 mode, all the data are inputted at the working places of the users.

Technically the system is based on the Microsoft technologies. It uses MS SQL Server as DBMS, MS IIS as application server, VBScript as programming language and MS Internet Explorer as client. A special piece of technology MEDSET (Modelling, Engineering, Development, Support and Evolution Technology) supporting all stages of the information system lifecycle was developed to guarantee the evolutionary development of the EPR/NSI (see [1 - 3]).

Sustainability features of the EPR/NSI

Sustainability of medical information systems is a complex phenomenon characterized by many features. Below we discuss some features of the EPR/NSI that make this system sustainable.

Lifetime

The EPR/NSI is running for almost 7 years in non-stop mode. It was stopped only once for the weekend for migration to the new system core, but even so all crucial functions of the system were running. Although this timespan is not sufficient to arrive at the ultimate conclusion about the future of the system, it is long enough to confirm the vitality of the principles it was based on.

Evolutionary development

As mentioned above, special efforts were made in order to allow for the evolutionary development of the system. The system was, and is, evolving in three "dimensions":

- "the spatial dimension" : the number of working places increased from 1 at the start (at the Admissions) to more than 400 nowadays, the number of users increased from 4 nurses working in shifts in the Admissions to more than 600;
- "the functional dimension" : the number of functions supporting special kinds of personnel's activity increased from 1 (admission/discharge) to more than 100;
- "the cognitive dimension" – the formalisation of the inputted data changes all the time in order to better espouse the requirements of the different categories of users; it was necessary to achieve the most comfortable proportion of free text fields and fields with a fixed set of values.

Usability of the system (see [4])

The system is actively used by the users: as just mentioned, almost all documents forming the paper case reports are printed copies of electronic documents gener-

ated in the EPR. Moreover, medical administration needs to spend special efforts to have doctors print out all electronic documents because in many cases they are not necessary for everyday activity. The system is actively used by the heads of the departments and the head physician to control the working processes of departments and of the hospital as a whole. The medical statisticians use the option of getting current statistical reports "by click" so that the shifted their attention to more intentional domains (e.g., medical classifications).

The user interface is simple, all Internet users are accustomed to it hence in most cases the teaching efforts are focused on specific features of the workflow organisation. The point is that the migration from conventional paper records to EPR requires more strict data handling discipline.

In fact, seven years of developing the EPR/NSI system in actual practice demonstrated that this system is stable and can resist challenges of reality and can accommodate to permanent changes. All this counts in favour of the sustainability of the system.

The size of the developer team

One more fundamental evidence of the system sustainability is that the team supporting the system has - already for 4 years – consisted of only 4 people: the head of the group (analytical tasks, new users' relations, the testing of new options), a programmer (analytical and programming tasks), and 2 system administrators (users' support, database administrating, testing, as well as teaching at times). This team supports both existing functions of the system and ensures the development and implementation of new ones.

The methodological basis of sustainability

All facts about the EPR/NSI discussed above would be unimaginable without a strong methodological basis. The principal issues of our MEDSET approach are described in this section.

Formalisation

The first principle is that any information system is just a formal representation of a certain part of the real world and human activity in it. Implementation of an information system consists in the immersion of staff's activity in formalized environment. So developers of the system intended to support activity in some subject domain, have to formalize:

- the users' knowledge of processes in the subject domain
- their own knowledge of processes in this subject domain

And, as it always happens in poorly formalised domains like medicine, this knowledge is fuzzy and implicit.

Adequacy of formalisation

Formalisation may be effective only if it is goal-driven. Any complex subject knowledge may be formalised in many ways, accordingly to the formulated goals. Formalisation has to be adequate to the goals.In the case of EPR/NSI, the goal was formulated as follows:

Delivering informational support to everyday doctors' and scientific researchers' activities.

Note that the administrative needs were not included in the list of goals. Our experience demonstrated that requirements of the medical administration may be satisfied almost freely if clinical data are organized and stored in adequate manner.

Ways of knowledge representation

There are only three ways of representing knowledge in an information system: as data structures, as programme code, as stored data. For every information system the proportion of these methods needs to be determined adequately to the goals of the system. The most stable parts of the knowledge may be represented in the database schema, the relatively stable ones in the program code while the most rapidly changing parts may be stored in the database.

Business processes and activity processes

The basic objective for the lot of the EPR/NSI decision was to treat separately business processes (BP) and activity processes (AP). The structure of the first ones reflects the organization of business, the structure of the second ones – the organization of personnel's activities. The BP-structure is relatively stable in a mature institution, and building a model of the data, it is worth being to follow it. The AP-structure is, on the contrary, rather flexible, and it is reasonable to reflect it in the interface model.

In other words, it may be said that the body of BPs reflects the view at the institution from the outside whereas the AP represents the view from the inside. The set and structure of the BP processes is relatively independent on the medical institution: patient registration, multiple laboratory investigations, surgical treatment and so on are typical for a great number of institutions. The APs, by contrast, may be organised in many different ways.

The main results of this separation are listed below:

- the stability of the data base structure and "local" nature of changes required for the implantation of new processes;
- the possibility of the evolutionary development of the system;
- the possibility of a fine tuning of the system interface to the needs of the staff.

The “Drama Principle” for user interface organization

We use the drama principle of "the union of place, time and action" to develop the users' interface, especially its functional part. This means that users' functions are organized in such a way that the screen forms may be filled at one working place in one action. This principle allows us to design simple screen forms containing only those data which are necessary for the user to input the data generated in his activity process.

Conclusion

All of the issues considered above have been essential to the development and implementation EPR system using an extremely small number of developers. We hope that the EPR/NSI will be sustainable until new achievements both in medicine and techniques ultimately force the developers to change their paradigm and to migrate to a new hard-, middle and software.

Now it is necessary to add just one more principle which is the foundation to the success of any information system:

Love the user.

The information system developer should always keep in mind that he works for the sake of the user and not the user works for the sake of his system. It is only under this condition that it is possible to develop a really sustainable system.

Acknowledgments

The authors are grateful to their colleagues, technicians and physicians, whose dedication, patience and good will were crucial for the development and the implementation of the system.

We also thank those who demonstrated criticism to our efforts, for they made us to find better ways of the system development.

Special thanks to Alexandre Vaxman, MS, for help in editing.

References

[1] Shifrin M, Kalinina E, Kalinin E. MEDSET – an integrated technology for modeling, engineering, deployment, support and evolution of information systems. Proceedings of the 15th IEEE Symposium on Computer-Based Medical Systems, Maribor 2002

[2] Shifrin M, Kalinina E, Kalinin E. Electronic patient record for N.N.Burdenko Neurosurgical Institute: on the road to paperless patient record. Proceedings of Medical Informatics Europe – 2002, Budapest, 2003.

[3] Shifrin M, Kalinina E, Kalinin E. EPR project for the NN Burdenko Neurosurgical Institute: goals, technology, results, future trends. British Journal of Healthcare Computing & Information Management 2003; 20(7): 18-21.

[4] Pantazi SV, Kushniruk A, Moehr JR.. The usability axiom of medical information systems. Int J Med Inform 2006; 75(12):829-39.

Address for correspondence

Michael Shifrin
Phone: +7 499 972 8525
Mobile: +7 916 692 6937
Fax: +7 495 250 9350
E-mail: shifrin@nsi.ru

Chapter 9.

Genomics

This page intentionally left blank

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Predicting Coronary Artery Disease with Medical Profile and Gene Polymorphisms Data

Qiongyu Chen*, Guoliang Li*, Tze-Yun Leong*, Chew-Kiat Heng[+]

**Medical Computing Laboratory, School of Computing, National University of Singapore*
3 Science Drive 2, Singapore 117543
[+] Department of Paediatrics, National University of Singapore
5 Lower Kent Ridge Road, Singapore, 119074

Abstract

Coronary artery disease (CAD) is a main cause of death in the world. Finding cost-effective methods to predict CAD is a major challenge in public health. In this paper, we investigate the combined effects of genetic polymorphisms and non-genetic factors on predicting the risk of CAD by applying well known classification methods, such as Bayesian networks, naïve Bayes, support vector machine, k-nearest neighbor, neural networks and decision trees. Our experiments show that all these classifiers are comparable in terms of accuracy, while Bayesian networks have the additional advantage of being able to provide insights into the relationships among the variables. We observe that the learned Bayesian Networks identify many important dependency relationships among genetic variables, which can be verified with domain knowledge. Conforming to current domain understanding, our results indicate that related diseases (e.g., diabetes and hypertension), age and smoking status are the most important factors for CAD prediction, while the genetic polymorphisms entail more complicated influences.

Keywords:

Coronary artery disease, single nucleotide polymorphisms, data mining, machine learning, Bayesian networks.

Introduction

Coronary artery disease (CAD) is the most common form of heart disease in America and Europe [1, 2]. It occurs when the arteries that supply blood to the heart muscles (coronary arteries) become hardened and narrowed. The major risk factors documented for CAD include related diseases (e.g., diabetes and hypertension), family history, gender, plasma lipid, etc [3].

Research in predicting the risk of CAD usually relies on medical profile and family history information. For example, Lapuerta et al [4] used seven different mean lipid values with the neural networks method to predict the occurrence of a complication of CAD. Wilson et al [5] predicted the risk of CAD by identifying risk categories and statistical tests, e.g., linear regression and logistic regression.

With the rapid advancement of biomedical technologies, now we can combine genetic information, such as microarray-based genotyping, to predict the disease risk. This integrative approach gives us better understanding of the fundamentals of the disease. Tham et al [6] combined medical profile, family history and microarray-based genotyping information in a neural network committee approach to predict the risk of CAD, and achieved reasonably good results.

In this work, we studied a set of profiles with comprehensive genetic polymorphism data and non-genetic data, collected with respect to CAD risk analysis. The objective is not only to predict the risk of CAD, but also to infer the dependency relationships among the relevant domain variables. We compared the prediction of several well known classifiers, including Bayesian networks, naïve Bayes, support vector machine (SVM), k-nearest neighbor, neural networks and decision trees. We experimented with predictions on the non-genetic and genetic factors separately, as well as collectively. Two feature selection methods based on Chi-squared test and Gain Ratio show that the non-genetic information, such as diabetes, hypertension, smoking status, and age, are more important for CAD prediction.

To study the dependency relationships among variables, we examined the structure of the learned Bayesian Network. We observed that the learned network identifies important dependency relationships among variables. Most of these relationships can be verified from the information on gene polymorphisms (*e.g.* polymorphisms at different locations in a single gene). Some other dependency relationships deserve further investigation. To our best knowledge, this is the first work to examine the relationships among both genetic and non-genetic variables in the CAD domain with a probabilistic graphical model (Bayesian networks in particular).

Background

Bayesian networks are graphical tools for modeling uncertainty and inferring causal relationships among the variables. The nodes in a Bayesian network represent the

variables in the target domain. The links between nodes represent probabilistic dependencies among different variables, and often imply causal relationships. Together, the graphical representation models qualitative information, while the conditional probability table in each node models quantitative information in the domain.

Bayesian networks can be constructed entirely from domain knowledge, but the process can be very time-consuming. To overcome this problem, there are many research efforts to build Bayesian networks from data [7-9]. For example, to learn the structure of a Bayesian Network, a family of score-based algorithms search through potential network spaces for the network with the best score using some pre-defined scoring function. The scoring function usually reflects how well a network topology fits the data set, e.g., Bayesian Information Criterion (BIC) [10]. Search methods may range from greedy search to exhaustive search, although the latter is not always feasible for Bayesian networks with moderately large number of variables. Other BN structure learning methods include constraint-based learning algorithms that infer network topology based on dependencies among variables. These dependencies are measured by pre-defined statistical tests.

In this study we also considered other major classification algorithms. Naïve Bayes is a probability-based classification method that assumes all the features are independent of each other given the class. A support vector machine (SVM) finds the optimal decision plane by selecting the fewest instances as the support vectors with the largest margin in the feature space. It is one of the most effective classification methods to date, although it sometimes suffers in the presence of noisy data. Neural networks are a type of black-box methods which can approximate any continuous function to arbitrary accuracy provided that the model has sufficiently large number of nodes and the parameters of the model are chosen properly. K-nearest neighbor classifiers are instance-based methods. They make predictions based on the distances between the test data and the training data. The decision tree method is a *de facto* classification method to evaluate other classification methods. It is recursively built based on the information gain of the features.

CAD prediction

The heart disease data set

The data set used in the study is an expanded version of what was originally collected from a Singaporean hospital for predicting CAD using neural networks [6]. It contains information from 2,949 medical profiles. There are 41 variables for each profile, including 10 non-genetic risk factors and 30 candidate gene polymorphisms. The term "non-genetic factors" is loosely used here to include all risk factors that are not directly collected from genotyping. Some non-genetic factors are clearly non-genetic in nature, such as smoking; but some are implicitly related to traits that may be genetic in nature, such as race and family history of disease. Eight of these non-genetic factors are discrete variables and two others are continuous variables. The distributions of continuous variables Age and body mass index (BMI) roughly follow the normal distribution. Both of them are discretized into 9 equal-width categories separately. The non-genetic risk factors are summarized in Table 1.

Table 1- The summary of the non-genetic variables

Code	# of states	Remarks
CAD	2	Healthy or diseased
SEX	2	Male or female
RACE	3	Chinese, Indian and Malay
DM	2	Diabetes, healthy or diseased
HY	2	Hypertension
SM	2	Smoker and non-smoker
FCAD	2	Family history of CAD
FDM	2	Family history of diabetes
FHY	2	Family history of hypertension
AGE	continuous	
CBMI	continuous	body-mass index

Each medical profile also contains 30 candidate gene polymorphisms. Studies have indicated that these gene polymorphisms may affect the patient's chance of developing CAD in Caucasian populations [3]. However, the influences of these genes on Asians (Chinese, Indian and Malay in this study) are unclear. Most of these variations are single nucleotide polymorphisms (SNPs), but there are also a few indels of longer base pairs. In this data set, each gene marker has 3 possible polymorphisms. Some of these markers are different polymorphisms located in various regions of a single gene. We list them in Table 2. Obviously, the included genetic attributes that may affect the chance of suffering from CAD are non-exhaustive.

The class label is CAD, with 0 denoting healthy subjects and 1 denoting patients suffering from CAD. Classification of CAD is based on the presence of at least 50% narrowing in at least one of the major coronary arteries by angiography. Of the 2,949 subjects, 1,462 or 49.6%, are constituted by patients diagnosed with Coronary Artery Disease at the time of data collection; the rest of the subjects were considered healthy at the time of recruitment.

Prediction results

We applied the classification algorithms implemented in the WEKA [11] software package. The entire data set was randomly split into 10 folds for cross-validation. The following results are based on the same 10 folds of data.

Sensitivity is defined as TP/(TP + FN), specificity as TN/(TN + FP), and prediction accuracy as (TP + TN)/(TP + FP + TN + FN), where TP, TN, FP and FN denote the numbers of the true positives, true negatives, false positives, and false negatives, respectively.

Table 2 - Candidate genetic polymorphisms

Code	Gene	Description
G1	Angiotensin Converting Enzyme	Enzyme involved in metabolism of angiotensin
G2	Angiotensinogen receptor	
G3	Angiotensinogen	Precursor of the hormone angiotensin
G4 – G8	Apolipoprotein B (ApoB)	Protein associated with cholesterol
G9 – G10	Lipoprotein Lipase	Enzyme that breaks down fat
G11 – G12	Antithrombin III	Anticoagulating factor
G13 – G18	Fibrinogen	Molecule that forms the blood clot
G19 – G20	Factor VII	Precursor of blood clot formation
G21 – G22	Apolipoprotein A1	Protein associated with cholesterol
G23	Glycoprotein 3A	
G24	5,10-methylenetetrahydrofolate reductase	
G25	Connexin	Protein that composes vertebrate gap junctions
G26	Cholesteryl ester transfer protein	Protein involved in cholesterol homeostasis
G27 – G30	ATP-binding cassette A1 (ABCA1)	Protein involves in cellular lipid removal

Table 3 - Prediction results (Non-genetic variables only)

Algorithm	Sensitivity	Specificity	Accuracy
Bayesian Network	0.886	0.887	0.887
Decision Tree (J48)	0.884	0.884	0.884
K nearest neighbors	0.854	0.896	0.875
Naïve Bayes	0.883	0.886	0.884
SVM	0.896	0.879	0.887
NN (MLP)	0.871	0.853	0.862

To study the effects of non-genetic and genetic factors, we performed three sets of experiments: prediction with non-genetic factors only, with genetic polymorphisms only, and with all the variables. The results are shown in Table 3 ~ 5.

Table 4 - Prediction results (Gene polymorphism only)

Algorithm	Sensitivity	Specificity	Accuracy
Bayesian Network	0.660	0.619	0.640
Decision Tree (J48)	0.625	0.571	0.598
K nearest neighbors	0.603	0.566	0.584
Naïve Bayes	0.672	0.608	0.640
SVM	0.713	0.634	0.653
NN (MLP)	0.611	0.599	0.605

Table 5 - Prediction results (All variables)

Algorithm	Sensitivity	Specificity	Accuracy
Bayesian Network	0.897	0.882	0.889
Decision Tree (J48)	0.884	0.884	0.884
K nearest neighbors	0.772	0.865	0.819
Naïve Bayes	0.889	0.879	0.884
SVM	0.895	0.889	0.892
NN (MLP)	0.877	0.874	0.875

With the same set of variables, the classifiers are generally comparable to one another in terms of prediction accuracy, with SVM showing a slight advantage, and K nearest neighbors being the worst. Further study showed that the resulting receiver operating characteristics (ROC) curves of these classifiers are similar as well. The improvement of these results over those reported by Tham et al [6] may be attributed to the different classification algorithms, and that more data was collected in this study, and thus the patterns in the data are more representative.

The prediction results based on only non-genetic variables are comparable to those based on the entire data set, while the results based on only genetic polymorphisms are considerably worse than the other two. This indicates that in our setting, non-genetic factors have more significant effects on the risk of CAD. There are two possible reasons for this phenomenon. One is that the set of genetic polymorphisms in our data may be a poor representation of the overall genetic factors for CAD, for it does not cover all possible relevant genes. Another reason is that some of the genetic effects were already expressed through the "non-genetic" variables such as race, family history and related diseases. The interactions of genes with other factors, such as smoking and diet habit, are also important. But these interactions are too complex to be analyzed by classification algorithms alone.

Feature selection

To further improve the prediction results, and to identify statistically significant variables, we applied two feature selection algorithms – Chi Squared and Gain Ratio – on the data set. For the most part these two algorithms agree with each other on the ranking of the variables in terms of significance.

Most of the non-genetic variables are highly ranked, except for the family history variables, with FHY (family history of hypertension) being the lowest ranked variable with both algorithms. Domain experts suggested that it may be due to data collection limitation, as the values of these variables come from the subjects' recalled history, instead of medical records. Most of the genetic polymorphism variables are ranked lower than non-genetic variables. Our final selected variable set consists of all non-genetic variables except for FHY, as well as G4, G17, G23, G26 and G30 from genetic polymorphisms variables.

As shown in Table 6, we see a general though slight improvement of prediction accuracy in the new round of classification experiments with feature selection.

Table 6 - Prediction results (with selected features)

Algorithm	Sensitivity	Specificity	Accuracy
Bayesian Network	0.902	0.883	0.892
Decision Tree (J48)	0.884	0.888	0.886
K nearest neighbors	0.847	0.898	0.873
Naïve Bayes	0.901	0.881	0.891
SVM	0.889	0.889	0.889
NN (MLP)	0.869	0.856	0.863

Dependency relationships identified from Bayesian network

Many classification algorithms used in our study are "black box" algorithms. Although they gave reasonably good results, it is difficult to interpret the models built from data. The Bayesian network, on the other hand, includes a graphical representation of the dependency relationships among variables in the problem domain. Figure 1 shows the Bayesian network learned from all the variables using score-based algorithm with hill climbing, without any input of domain knowledge.

The model identifies many important dependency relationships among genetic variables in the study. Some of the polymorphism variables in the group show high dependencies among one another. For example, G13 – G18 are highly correlated to one another and they are closely-connected in the learned Bayesian network. Domain knowledge confirms that G13-G18 are six SNPs in different locations of the same gene Fibrinogen. Another identified group is G27 – G30, which are SNPs in the ABCA1 gene. This group is associated with both "race" and related diseases, which agrees with a prior study on ABCA1 polymorphisms in local population [12].

Our network also identified other patterns, one of which is that some of the gene polymorphisms are irrelevant to CAD when the race of the subject is known in the studied population. For example, G1 (polymorphism in Angiotensin Converting Enzyme) is shown to be blocked from CAD by race. This finding is reasonable since G1 is a silent indel (287 bp, at Intron 16). Therefore, while it is

very likely to be correlated with race, it is unlikely to have an impact on the risk of CAD. Another similar example is the pair G11 and G12, which are both polymorphisms at ATIII. G11 is a 76 bp polymorphism at the 5' un-translated region, while G12 is a silent indel at intron 5.

In addition to the known correlations, the learned Bayesian Network also shows some interesting unknown patterns. For example, there is a strong dependency between G3 (SNP in Angiotensinogen) and G23 (SNP in Glycoprotein 3A). Whether this is a mere artifact or an indication of actual correlation requires further study.

Sensitivity analysis on the final network shows that the top four risk factors for CAD are diabetes, hypertension, smoking status and age, which conform to domain knowledge. No gene polymorphisms are highly ranked in the analysis, which indicates that no single gene makes a major contribution to the risk of CAD. This is certainly a valid finding since CAD is a complex disease with multifactorial etiology and polygenic. No single gene is therefore expected to contribute substantially to the disease risk except in the case of familial hypercholesterolemia.

4. Discussion and future work

The combined effects of genetic and non-genetic factors on the risk of CAD have not been extensively studied in the medical domain. We have applied six well-known classification algorithms for CAD risk prediction in Singapore. Our experiments showed that the prediction results are better than those in a previous study. We have also identified some interesting patterns in the variables. The Bayesian network constructed from structure learning identified some correlated groups among the genetic factors. Most of them can be explained by relevant domain information. In addition, some gene polymorphisms are shown not to have significant correlation with CAD prediction in the studied population based on the data set used. The identified dependency relationships among the genetic factors from the learned Bayesian network call for further studies.

Our experiments showed that non-genetic variables have a more direct influence on CAD, while the effects of genetic factors are more complicated and subtle. This agrees with the general belief that no single gene is a major determinant factor for CAD. The interaction of genetic factors with non-genetic factors, such as smoking, diet and physical activity habit, are important for the disease risk prediction.

The statistical analysis in our experiments has successfully identified several genetic and non-genetic factors that have significant influence respectively on CAD. However, the interplay of these two groups of factors and their combined effects are still unclear. Our study showed that classification algorithms alone may prove to be inadequate in this regard. Bayesian networks, on the other hand, have shown some promising results as effective exploratory data analysis tools. We are also looking at other probabilistic graphical models with richer representation power, e.g., Probabilistic Relational Models [13], that may give us further insights into this problem.

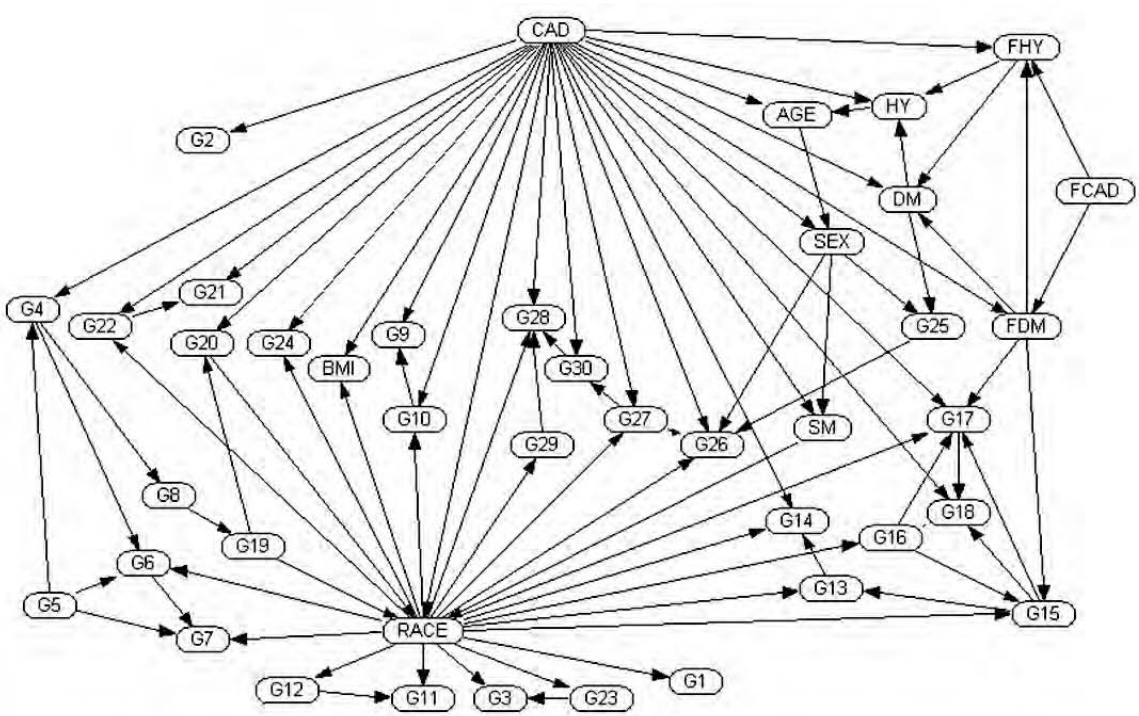


Figure 1 - The learned Bayesian Network

Acknowledgments

We thank Mr. Rohit Joshi for helpful discussions. This research was partially supported by Research Grant No. BM/00/007 from the Biomedical Research Council (BMRC) of the Agency for Science, Technology, and Research (A*Star) and the Ministry of Education in Singapore.

References

[1] Coronary Artery Disease Overview, http://imaginis.com/heart-disease/cad_ov.asp, accessed on February 19, 2005.

[2] Coronary Artery Disease, http://www.nhlbi.nih.gov/health/dci/Diseases/Cad/CAD_All.html, accessed on November 25, 2006.

[3] Heng CK. Candidate Genes for Coronary Artery Diseases, PhD Thesis: National University of Singapore; 1996.

[4] Lapuerta P, Azen SP, LaBree L. Use of neural networks in predicting the risk of coronary artery disease. Computers and Biomedical Research 1995;28(1):38-52.

[5] Wilson PWF, D'Agostino RB, Levy D, Belanger AM, Silbershatz H, Kannel WB. Prediction of Coronary Heart Disease Using Risk Factor Categories. Circulation 1998(97):1837-47.

[6] Tham CK, Heng CK, Chin WC. Predicting Risk of Coronary Artery Disease from DNA Microarray-Based Genotyping Using Neural Networks and Other Statistical Analysis Tool. Journal of Bioinformatics and Computational Biology 2003;1(3):521-39.

[7] Cooper GF, Herskovits E. A Bayesian method for the induction of probabilistic networks from data. Machine Learning 1992;9:309-47.

[8] Heckerman D. A Tutorial on Learning with Bayesian Networks. In: Jordan M, editor. Learning in Graphical Models. Cambridge, MA: MIT Press; 1998. pp. 301-354.

[9] Heckerman D, Geiger D, Chickering DM. Learning Bayesian networks: The combination of knowledge and statistical data. Machine Learning 1995;20:197-243.

[10] Schwarz G. Estimating the dimension of a model. The Annals of Statistics 1978;6:461-64.

[11] Witten IH, Frank E. Data mining: practical machine learning tools and techniques with Java implementations. San Francisco: Morgan Kaufmann; 1999.

[12] Tan JH-H, Low P-S, Tan Y-S, Tong M-C, Saha N, Yang H, et al. ABCA1 gene polymorphisms and their associations with coronary artery disease and plasma lipids in males from three ethnic populations in Singapore. Human Genetics 2003;113(2):106-17.

[13] Friedman N, Getoor L, Koller D, Pfeffer A. Learning Probabilistic Relational Models. In: Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence (IJCAI '99): Morgan Kaufmann Publishers; 1999. pp. 1300-1309.

Address for correspondence

Qiongyu Chen
School of Computing, National University of Singapore
3 Science Drive 2, Singapore, 117543
Email: chenqy@comp.nus.edu.sg

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Towards a Top-Domain Ontology for Linking Biomedical Ontologies

Holger Stenzhorn[a,b], Elena Beißwanger[c], Stefan Schulz[a]

[a] *Department of Medical Informatics, Freiburg University Hospital, Germany*
[b] *Institute for Formal Ontology and Medical Information Science (IFOMIS), Saarbrücken, Germany*
[c] *Jena University Language and Information Engineering (JULIE) Lab, Germany*

Abstract

In this paper we present the ongoing development and extension work on BioTop – a top-domain ontology for linking biomedical domain ontologies. We start by making the case for the application of a common ontology to interface independent biomedical domain ontologies by introducing a set of more general classes. Then we briefly depict the relation of BioTop to the GENIA ontology as starting point of its initial developement. Afterwards we propose our distinction of ontologies into top, top-domain and domain ones and describe our approach to the integration of the top ontology BFO into BioTop. Then we present our plans to join the OBO and OBO Foundry repository of ontologies and list its admission principles in relation to our ontology. Some actual BioTop interface classes are shown subsequently. We conclude by detailing on some planned BioTop usages in the area of BioNLP and cancer research and show some further intended improvements.

Keywords:

biomedical ontologies.

Introduction

The last couple of years have seen a tremendous increase in the amount of data collected within the life sciences, especially in biomedicine and its subfield of genomics research. This in turn has spurred many scientific efforts to analyze and structure the newly gained data and to extract further knowledge from it. In the following we are now focusing on the application of ontologies for this particular task and specifically examine one currently existing drawback: Most existing biomedical ontologies – even when having overlapping content – are developed mostly independently from each other. Also, each ontology embraces only some distinct scenario with a mere partial view of the overall scientific field. What has therefore been missing so far is an overarching resource to help with linking and interfacing those independent ontologies. With such a facility in place, new methodologies could be conceived to employ ontologies more efficiently in concert and to create synergetic effects. To this end, we have developed the top-domain ontology BioTop, to be presented in the following.

We concentrated our work on interfacing a smaller selection of about 60 ontologies in the Open Biomedical Ontologies (OBO) framework [1]: the Gene Ontology (GO) [2], the Sequence Ontology (SO), the Cell Ontology (CO), the Chemical Entities of Biological Interest (ChEBI) and the Foundational Model of Anatomy (FMA) [3]. At this point we want to stress that the methodology of our work can be easily applied to create additional interfaces to other ontologies in both this particular topic area as well as similar ones. We are highly interested to further investigate the latter in our future research.

Project background

Relation to GENIA

The initial version of BioTop rested upon the idea to create a comprehensive, formally-based redesign and expansion of the original GENIA ontology [4]. The basic development policy was to follow the fundamental principles of formal rigor, explicitness and precision of ontological axioms and to maintain the clear overall scope of creating a biomedical upper ontology. The implementation was to be based on the Description Logic subtype of the Web Ontology Language (OWL-DL) [5].

The GENIA ontology had originally been developed for and within the biological natural language processing (BioNLP) community and had quickly become a de-facto standard in this field. Its authors claim the ontology to be a formal model of cell signalling reactions in humans and regard its main application to serve as basis for creating thesauri and semantic dictionaries in BioNLP applications (e.g. the semantic annotation of named entities in biological literature abstracts). The GENIA authors also consider it as providing a mutual basis for an integrated view over multiple biological databases.

The GENIA ontology itself is very small, containing only 45 distinct terms, arranged in a simple taxonomy with a maximal depth of six levels. It also limits itself to a set of highly general upper-level classes centred on the notions of biochemical substances and their corresponding locations in the organisms.

During our work, we found some non-trivial shortcomings within the GENIA ontology: Firstly, for most classes a proper documentation and/or textual definition was missing or lacked clarity. Secondly, some class names or their

particular position in the taxonomy contradicted common biological or ontological intuitions of consistency. Both issues can obviously lead to conflicting interpretations and incorrect applications of the classes. A complete analysis of the deficiencies found in GENIA and our proposed solutions can be found in [6].

Top, top-domain and domain ontologies

We propose to distinguish ontologies into three basic types (with their approximate size proportions shown in Figure 1):

- A top ontology (also called top-level or upper ontology) contains only a very small and restricted set of the most high-level, general classes such as "Continuant", "Occurrent", "Function" or "Object" – together with some accompanying relations. Examples for this kind of ontologies are BFO [7] and DOLCE [8].
- A top-domain ontology (also called upper-domain (level) ontology) holds the essential core domain classes to interface to both upper and domain ontologies, like "Organism", "Tissue" or "Cell" in the case of biology. A top-domain ontology can also include more specific relations and further expand or restrict the applicability of relations introduced by the top ontology. An example for this kind of ontologies is BioTop.
- A domain ontology has as its members a multitude of low-level, domain-specific classes to comprehensively describe a certain (aspect of a) domain of interest, e.g. "Antisense RNA Transcription" or "DNA Replication" from the Gene Ontology.

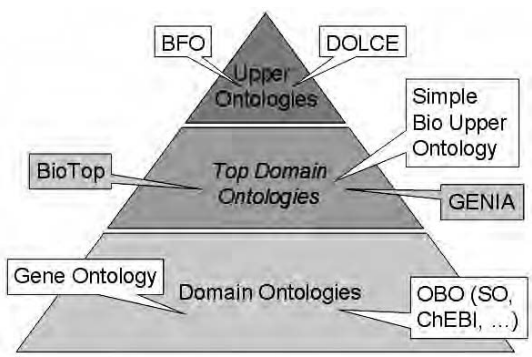


Figure 1 – Ontology layer pyramid (adapted from Alan Rector)

Issues and goals for the second development phase

When we published the initial version of BioTop we received valuable comments that both encouraged us to continue our work on BioTop but also pointed out some existing shortcomings from the standpoint of biomedical and ontology experts. We also received an open invitation to apply for a membership in the OBO Consortium and (subsequently) in the OBO Foundry. Therefore we set up the following list of basic goals we wanted to achieve in the further development:

- Full inclusion and adoption of the Basic Formal Ontology (BFO) and the OBO Relation Ontology (RO) [9].
- Joining the OBO Consortium as well as the OBO Foundry by fulfilling its given set of principles.
- Improving and expanding the existing interface classes to further link the OBO ontologies.

Methods

Integration of top ontologies

Importing BFO and the OBO Relation Ontology (RO)

In the initial version of BioTop we created a set of top-level classes which was based on the terms and definitions found in the publications on both BFO as well as DOLCE (i.e. we initially had a mixture of both top-level ontologies). We also created some additional relations that we based on the ones published in the Relation Ontology (RO).

When the first official version of BFO and RO became available in OWL-DL we decided to modify our ontology by removing the mentioned mixture of BFO and DOLCE top-level classes and to employ the OWL import mechanism to include the now available BFO and RO versions in our ontology.

This integration task was obviously straightforward where actual BFO classes were used before: Here we only had to replace our self-defined class as parent of a given child class with the corresponding class from the BFO import. More problematic were the cases in which DOLCE-inspired classes had been used. Here we had to find either a class in BFO that matched or closely resembled the respective DOLCE class or we had to remodel the class by introducing one or more mediator classes that were deemed to be ontologically sound.

The inclusion of the RO was also straightforward. We had basically only used the relations already defined by the RO and hence simply had to change the references in our ontology from our relation definitions to the imported ones. For the two additional relations we had defined ourselves, we only needed to change the reference to their parent property since they were directly based (i.e. subproperties) on ones from the RO.

Because of ongoing discussions to consider the addition of new relations to the RO, the need for keeping our self-defined relations might disappear and hence we could remove them. In case the new relations in the RO do not meet our needs we plan to suggest the addition of our relations to the original RO.

Joining the OBO Consortium and the OBO Foundry

After receiving the above mentioned open invitation to join the OBO Consortium we decided to complete the following steps:

- Select the most important published ontologies contained within the OBO (in regard to their user base size and direct relevance to biomedical research), analyze and compare their current top-level classes with the existing interface classes of the initial BioTop version.

Then detect the potential overlap in their respective scope with the scope of BioTop.

- Actively contact the curators and developers of each ontology for which we had developed interface classes and inform them of our plans to join the OBO Consortium and the OBO Foundry and also introduce BioTop as a possible additional layer ontology on top of their ontologies. Additionally invite them to report further ideas on how they think their ontologies and/or BioTop needs to be changed to meet their respective needs.

BioTop and the OBO Consortium and Foundry Principles

In order to apply for a membership in the OBO Consortium, an applicant must take several measures to ascertain that the prospective member ontology fulfils a predetermined set of principles [1]. The following list identifies those principles in relation to the application of BioTop to the OBO Consortium:

1. The BioTop ontology is completely open-source and therefore directly accessible and available to everybody. Links to the latest published version – as well as to older versions – can be found on its website. To adopt BioTop in other projects the respective developer or user must only acknowledge its original source and agree not to alter and distribute the modified ontology under its original name and with the same identifiers.
2. The implementation of BioTop is based on OWL-DL which is now accepted as a common and formally defined language by the OBO Foundry and is established as an official standard for building ontologies for the Semantic Web published by the World Wide Web Consortium (W3C). This in turn entails the availability of a wealth of documentation and supporting tools (for editing as well as classification) and thus allows for a straightforward implementation and adaptation process.
3. BioTop is a top-domain ontology for biomedical research and hence exhibits a clearly defined and delineated subject matter that is distinct from the existing consortium ontologies. Hence it contains only classes and relations necessary to define the higher and (more) general level of this subject field and allows further to link lower level domain ontologies with each other.
4. The string "biotop" is utilized as the unique identifier space for our ontology and serves as the namespace prefix of its OWL-DL implementation. This facilitates avoiding possible naming conflicts as a result of identical class names in BioTop and other ontologies. Also each single class within the ontology holds a unique identifying name to prevent inner-ontology confusion.
5. We include in our ontology precise, plain textual definitions for all classes and relations to resolve the ambiguity that many terms possess in the biomedical sciences. By doing so BioTop cannot only be processed by computer systems but is also understandable for humans and applicable it in their regular work.

In addition to these principles the OBO Foundry requires some additional principles to be followed in order to join:

6. We employ a common version control system to make possible the easy identification and retrieval of all available and different ontology versions. This mechanism simplifies greatly the joint collaboration effort by helping to keep track of all changes happing during the everyday development circle, such as the renaming or the deletion of classes or relations for example.
7. The BioTop ontology uses the relations defined by the OBO Relation Ontology through importing its official, published OWL-DL representation. It additionally introduces two new relations that have been formally defined and strictly follow the prescribed pattern of definitions of the original relations.
8. From the beginning we tried to not solely develop the OWL-DL implementation of the ontology but we were also careful not to overlook the need for a comprehensive and comprehensible documentation. We achieved this in two ways, namely by introducing extensive comments and remarks to each class and relation – directly into the implementation – and also by creating descriptive publications targeted at domain experts with no or little background in ontologies.
9. The interest in our work expressed by various researchers after the first release of BioTop showed us that there exists a multitude of potential users for our ontology. By reaching across several independent domain ontologies all users interested in the combination and interoperability of those ontologies can be regarded as possible users of BioTop also.
10. The development started off as a collaborative effort between researchers from two different institutions. Through face-to-face discussions at conferences, postings on mailing lists and personal mail communication, many more people specialized in ontology and biology are currently getting involved and provide input and comments for the continuous BioTop refinement.

Interfaces to other ontologies

We tried to achieve a comprehensive coverage of the interface classes in BioTop to link together as many biomedical domain ontologies as possible. But we also tried not to lose track of our original goal: We wanted our ontology to be focused on biomedical matters on a more generic level without any limitations to specific subdomains or some particular species.

Thus we wanted to avoid as much overlap as possible with the given domain ontologies wherever possible and sensible from an ontological point of view. When we found a place of considerable overlap we tried to ascertain whether this overlap problem should be solved on the side of BioTop or rather on the side of the domain ontology. So far this problem has not been tackled in a satisfactory fashion but we plan to contact the responsible curators to further discuss matters and to come up with a principled way of handling such cases.

BioTop sometimes does not provide a direct, logical link to the upper-level classes of the domain ontology. In such cases we obviously see the need to talk to the respective ontology curators to find out whether they should intro-

duce some additional top-level classes to their ontology or whether BioTop is missing any important classes that could be general enough to be relevant to other ontologies as well.

Table 1 lists a small sample of the links we have created so far from BioTop to other domain ontologies and which have been accepted as being valid by the respective curators. A more comprehensive treatment of those links can be found in [6].

Table 1 – Example links from BioTop to OBO ontologies

BioTop	**OBO Ontologies**
Biological Process	Biological Process (GO)
Protein Function	Molecular Function (GO)
Cell Component	Cell Component (GO)
Cell	Cell (CO), Cell (FMA)
Atom	Atoms (ChEBI)
Subatomic Particle	Elementary Particles (ChEBI)
Organic Compound	Organic Molecular Entities (ChEBI)
Tissue	Tissue (FMA)
DNA, RNA	DNA, RNA (SO)
Protein	Protein (SO)

Figure 2 depicts some small, restricted portions of the BFO, BioTop and GO ontologies to demonstrate the layering and interfacing in between them. In this particular example the domain-specific GO class "Transcription" is linked to the BioTop class "Biological Function" which in turn is linked to the generic, domain independent class "Function" in BFO.

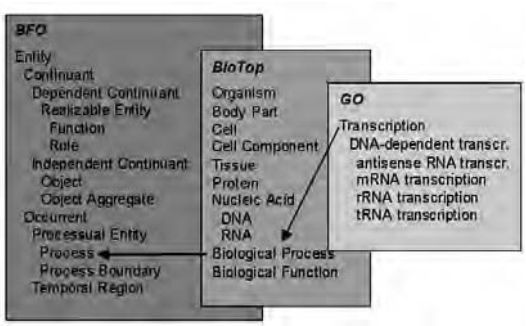


Figure 2 – Classes showing the ontology interfacing

Outlook and Future Plans

Usage Scenarios

For the further evaluation and application of BioTop we are currently pursuing the following two main scenarios.

Natural language processing and named entity annotation

The European project BOOTStrep [11] will apply the BioTop ontology for various natural language processing purposes in the biology domain (BioNLP). Amongst them is its application to improve the quality of semantic annotation of biological text corpora, i.e. the class names from the BioTop ontology are used as the vocabulary to semantically annotate named entities automatically identified in literature abstracts. The resulting annotated corpora are subsequently employed as a training material for statistical methods which in turn build the basis for more complex BioNLP applications such as meaning disambiguation, relation extraction or anaphora resolution.

Involvement in the development of other ontologies

After successfully joining the OBO Consortium and the OBO Foundry we hope that BioTop will become even more publicly visible and subsequently adapted in the development processes of new and existing ontologies in the context of OBO and elsewhere. By doing so, the developers would acknowledge their (and also our) belief in the necessity for interoperability among individual ontologies, as we have stressed above. An application of BioTop by a larger user base will furthermore facilitate the discovery and resolution of residual problems, bugs and shortcomings in the current BioTop implementation.

Further improvements

Amelioration and expansion of interface classes

Future versions of BioTop will incorporate continuously improved interface classes. We are discussing the suitability and applicability of the current interface classes with the curators of several OBO domain ontologies. This will lead to the addition of new or the removal of existing BioTop classes, as well as to the clarification of class definitions or documentations.

One specific interest lies in the present development of an ontology for clinical trials within the OBO Consortium. We believe that BioTop could be applied in this context as a link between biological ontologies on one side and medical ones on the other ameliorating the creation of classes that are both of biological and medical interest. To this end, the first author participates in the creation of an ontology for clinico-genomic trials on nephroblastoma (a specific type of kidney cancer found in children) as part a European research project [12]. This ontology will serve after its completion as major input and the basis for the mentioned OBO clinical trial ontology.

Re-integration of DOLCE

As mentioned above, the initial BioTop version contained a mixture of classes from both BFO and DOLCE at its top level. We now believe that it would be an interesting

experiment to reintroduce DOLCE through including its official OWL-DL version as a second top-layer (in addition to keeping the BFO classes). Firstly, this could perhaps allow us to elicit whether the two ontologies are indeed equivalently applicable as the top-level of our ontology. Doing so could also identify places in BioTop where and why one top ontology might excel the other. Secondly, through the addition of DOLCE we might attract users who are accustomed to this top ontology and are therefore reluctant to use an ontology (solely) based on BFO.

Related work

We are aware of two other projects that are currently engaged in setting up a top-domain ontology for biology. These are the Simple Bio Upper Ontology [13] and GFO-Bio [14]. It seems that at the time of this writing both projects exist in an experimental implementation stadium only and have not produced any publications. Nevertheless we intend to contact both groups for discussions and a possible cooperation.

Our initial intention for BioTop was to improve the interoperability between different biomedical domain ontologies by having a common top-domain ontology. The creation of several top-domain ontologies in this field would obviously be counterproductive and hence some cooperation is essential to achieve a unified solution. Then it would be ideal to have a dedicated workshop to gather the views from more experts in the field to reach a consensus about a single top-domain ontology (which could be based on BioTop or have it as a source).

Conclusion

In this paper we described our current efforts to further develop and extend the biomedical top-domain ontology BioTop. We made the case why an overarching ontology with general classes is important and needed to link independent domain ontologies. We described our integration of BFO into BioTop, discussed our intention to join the OBO Consortium and the OBO Foundry and listed their principles in relation to BioTop. Then we showed some actual interface classes and concluded by detailing on planned BioTop usages and related projects.

Availability

All BioTop material (including its OWL-DL implementation) is available from its website http://www.ifomis.org/biotop.

Acknowledgments

This work was supported by the European Union Network of Excellence "Semantic Interoperability and Data Mining in Biomedicine" (NoE 507505) (http://www.semanticmining.org) as well as the European Commission STREP project "BOOTStrep" (FP6 - 028099) (http://www.bootstrep.eu).

References

[1] OBO Consortium. Open Biological Ontologies (OBO), 2004. [http://obo.sourceforge.net] – Last accessed: March 25, 2007.

[2] Gene Ontology Consortium. Creating the Gene Ontology Resource: Design and Implementation. *Genome Research*, 11(8):1425–1433, 2001.

[3] Rosse C, and Mejino JVL. A Reference Ontology for Biomedical Informatics: the Foundational Model of Anatomy. J Biomed Inform. 36:478-500.

[4] Ohta T, Tateisi Y, and Kim JD. The Genia Corpus: An Annotated Research Abstract Corpus in the Molecular Biology Domain. In Proc. of the 2nd International Conference on Human Language Technology Research (HLT 2002), San Francisco: Morgan Kaufmann, 2002; pp. 82–86.

[5] Horrocks I, Patel-Schneider PF, and van Harmelen F. From SHIQ and RDF to OWL: The Making of a Web Ontology Language. Journal of Web Semantics, 1(1):7–26, 2003.

[6] Schulz S, Beißwanger E, Wermter J, and Hahn U. Towards an Upper-Level Ontology for Molecular Biology. In Proc. of the American Medical Informatics Assn. Annual Conference (AMIA 2006), Washington, 2006; pp. 694–698.

[7] Grenon P, Smith B, and Goldberg L. Biodynamic Ontology: Applying BFO in the Biomedical Domain. In Proceedings of the Workshop on Medical Ontologies, Number 102 in Studies in Health Technology and Informatics, Amsterdam: IOS Press, 2004; pp. 20–38.

[8] Gangemi A, Guarino N, Masolo C, Oltramari A, and Schneider L. Sweetening Ontologies with DOLCE. In Proceedings of the 11th International Conference on Knowledge Engineering and Knowledge Management (EKAW 2002), Siguenza, Spain, 2002; pp.166–181.

[9] Smith B, Ceusters W, Klagges B, Köhler J, Kumar A, Lomax J, Mungall C, Neuhaus F, Rector AL, and Rosse C. Relations in Biomedical Ontologies. *Genome Biology*, 6(5):R46 (1:15), 2005.

[10] OBO Foundry Consortium. Open Biological Ontologies Foundry (OBO Foundry), 2006. [http://www.obofoundry.org] – Last accessed: March 25, 2007.

[11] BOOTStrep Consortium. Bootstrapping Of Ontologies and Terminologies Strategic Research Project (BOOTStrep), 2006. [http://www.bootstrep.eu] – Last accessed: March 25, 2007.

[12] ACGT Consortium. Advancing Clinico-Genomic Trials on Cancer (ACGT), 2006. [http://www.eu-acgt.org] – Last accessed: March 25, 2007.

[13] Rector AL, Stevens A, and Rogers J. Simple Bio Upper Ontology, 2006. [http://www.cs.man.ac.uk/~rector/ontologies/simple-top-bio] – Last accessed: March 25, 2007.

[14] Höhndorf R. GFO-Bio: A Biomedical Core Ontology, 2006. [http://onto.eva.mpg.de/gfo-bio.html] – Last accessed: March 25, 2007.

Address for correspondence

Holger Stenzhorn, Department of Medical Informatics,
Freiburg University Hospital,
Stefan-Meier-Str. 26, 79104 Freiburg,
Germany

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

The Molecular Medicine Informatics Model (MMIM)

Marienne Hibbert [a], Peter Gibbs [b], Terence O'Brien [c], Peter Colman [d], Robert Merriel [e], Naomi Rafael [f], Michael Georgeff [g]

[a] *MMIM Project Director, Melbourne, Australia*
[b] *MMIM Chief Scientist , Melbourne Health and Ludwig Institute, Melbourne, Australia*
[c] *MMIM Neurosciences leader, Melbourne Health and University of Melbourne., Melbourne, Australia*
[d] *MMIM Diabetes leader, Melbourne Health and Walter and Eliza Institute, Melbourne, Australia*
[e] *MMIM Steering Committee Chairman, Melbourne Health, Melbourne, Australia*
[f] *MMIM Senior Database Administrator, Melbourne Health, Melbourne, Australia*
[g] *MMIM Solution Manager, Monash University, Melbourne, Australia*

Abstract

In 2005 a major collaboration in Melbourne, Australia successfully implemented a major medical informatics infrastructure .The convergence of life sciences, healthcare, and information technology is now driving research into the fundamentals of disease causation and toward tailoring individualized treatment. The Molecular Medicine Informatics Model (MMIM) is a 'virtual' research repository of clinical, laboratory and genetic data sets. Integrated data, physically located within independent hospital and research organisations can be searched and queried seamlessly via a federated data integrator. Researchers must gain authorisation to access data, and obtain permission from the data owners before the data can be accessed. The legal and ethical issues surrounding the use of this health data have been addressed so data complies with privacy requirements. The MMIM platform also record links individual cases across multiple institutions and multiple clinical specialties. Significant research outcomes in epilepsy and colorectal cancer have already been enabled by the MMIM research platform. The infrastructure of MMIM enables discovery research to be accessible via the Web with security, intellectual property and privacy addressed.

Keywords[1]:

Medical Informatics, Systems integration, genomics

Introduction

The convergence of life sciences, healthcare, and information technology is revolutionizing the discovery and developmentof new treatments and the optimal use of available therapies [1-2]. Researchers now have the capability to analyse intricate details of human biology through linkage of genetic data with clinical outcomes data giving them the potential to understand the fundamental causation of human disease and predict outcomes. This is driving the development of new drugs, new diagnostics, and lead us to the era of personalised medicine [3]. Key to making the required associations between genotype and phenotype is access to detailed clinical data in sufficient numbers of patients to ensure studies are adequately powered. Few institutions alone have sufficient numbers to perform meaningful analyses, particularly where stratification is performed to look at specific disease attributes. Further, , to utilize all of the available information, clinicians need to look beyond their own specialty into datasets of other disease groupings, analyzing the impact of co-morbidity. To achieve the research objectives promised by this new era in science, sharing of research and clinical data between disparate research groups and institutions becomes critical.

The impetus for the development of MMIM came from a recognition of the need to maximise collaborative research across Australia and internationally. Europe's INFOMED[2], the CaBig in USA [6] and The UK Cancer Grid [7] also recognized the need for collaborative integrated data, data standards and tools. A cohesive approach between disciplines was needed so that research data collection becomes a one time only exercise with the data stored in such a way tha it remained readily accessible, and in a format that facilitated rapid interrogation, permitting multiple research questions across various clinical disciplines and jurisdictions to be addressed. In addition to the value in new and emerging data sets such as genomic data being linked to clinical and outcome data, the is great value to be obtained from existing data. Thus researchers will be able to examine genetic, genomic and proteomic profiles, all factors that may influence treatment outcome, toxicity and potential benefit.

The MMIM Project can enable research from multiple perspectives, including:

- Co morbidities;
- Treatment strategies;
- Genetic predisposition;
- Health Screening activity;
- Genomics, proteomics & epigenetics;
- Outcomes.

1

The objective of the project is to maximise collaborative research efforts, both in Australia and internationally through the development of a federated data integration infrastructure.

Materials and methods

MMIM background

Phase 1 of the MMIM project was successfully completed in 2005. This phase delivered the integration of data across five hospital sites and two medical research institutes involving colorectal cancer, epilepsy, diabetes and the tissue banks.

Phase 2 of the MMIM Project is integrating data across additional Victorian and interstate hospitals including the additional disease types - multiple sclerosis, stroke, cystic fibrosis, asthma, and brain tumours.

Phase 3 of the project is funded over a three year period until June 2009 to provide support for the creation of an Australian Cancer Grid across all tumour types and:

- The Infrastructure expansion – the data grid;
- The research activity and outcomes;

Overview of the MMIM federated mode - technologies

The MMIM project is a federated model where each participating site retains ownership and control over their own data sources and data collection systems. The architecture can be seen diagrammatically in Figure 1.

The data sources used for integration were established as clinical research databases written and maintained by specialist clinicians in their own healthcare facilities. These included datasets that address single questions up to comprehensive datasets capturing information across the full range of a disease, from data regarding genetics and screening through to care of end stage disease. Data was uploaded from these databases nightly, or manually loaded (for static datasets) into a 'cache' database, a local DB2 UDB database termed a Local Research Repository (LRR) located at each site. The distributed LRR databases were federated using IBM Websphere Information Integrator running on a single server termed the Federated Data Integrator (FDI). Information Integrator makes remote databases appear as local DB2 table views, allowing single SQL queries to be executed against all federated data.

Public domain databases were also federated into the system including a local XML flat file and resources from the National Library of Medicine (Genbank, Medline and Uniprot) via an Internet web service. Both data sources appeared relational to the end user even though they were not.

A unique number was assigned to each patient termed a Unique Subject Index (USI) by transferring certain identifying information to Sun (SeeBeyond) e-Ways and replicated to the LRRs via the FDI.

The security system included a number of features. Each LRR was connected to the FDI via Virtual Private Network (VPN) connections, which ensure data privacy and encryption. Views block all identifying information, allowing end users to see only the clinical data in conjunction with the USI. User access to these views on the FDI is controlled by assigning DB2 database roles which define privileges to the table/view level. DB2 Query Patroller is used on the FDI to track all queries for audit purposes. Access is controlled by assigning permission at the table level.

SAS™ Enterprise Guide was used as the interface for researchers to perform queries, statistical analysis and construct reports.

Connectivity:

Each participating research institution has a local data storage facility (the LRR) which are connected via a secure technology involving double encryption, Virtual Private Network and DB2-DB2 encryption. This is the technology commonly used in industries to link various sites of operations for major corporations, such as financial institutions,

Data loading:

On a nightly or ad-hoc basis, the clinical research data is loaded into the LRR at each individual site. This loading process uses an extraction, transform and load software feature that is installed on each LRR.

Record linkage - The Unique Subject Index

The Unique Subject Index (USI) is the key element in linking patient records across disparate data sources within and across institutions. It ensures compliance with privacy. Linking patient/subject records and assigning USI identifiers to data allows patients to be linked across multiple institutions and databases while also observing legal, ethical, privacy and data ownership constraints

The USI was developed based on matching of six key demographic data items (Surname, Given name, Middle name or Initial, Date of Birth, Gender, Digits 5 to 9 of the Medicare Number). The software checks new records for a match against existing subjects, using probabilistic matching and a score is assigned on the basis of match / non-match for each data item. "Fuzzy logic" is used for transpositions, soundex matches, and common "dummy" names (e.g. Babe of, Twin 1). Manual checking of subjects in the "grey area" between thresholds can be undertaken by the data owners.

The USI is a 10 digit unique number assigned by the system to each patient. The process involves sending selected identifying data on a nightly basis from each LRR to a Unique Subject Index (USI) program, where the unique number is generated and stored in an Oracle database. This number is then stored the LRR in encrypted form.

End-user queries:

Researcher access is provided to specified tables in MMIM only on application with the research/purpose fully described and only after approval of the application by the MMIM steering committee and data is only available in de-identified form from the FDI.

The identifying data remains stored on the USI data store with extremely restricted and controlled access. No Health Data is stored externally and only authorized researchers can log in to MMIM via the SAS™ terminal server and send queries to the FDI which 'interrogates' the LRRSs. The results of the query is then put into the researcher's secure folder for statistical analysis (the SAS™ server). All queries are tracked and logged for security audit purposes.

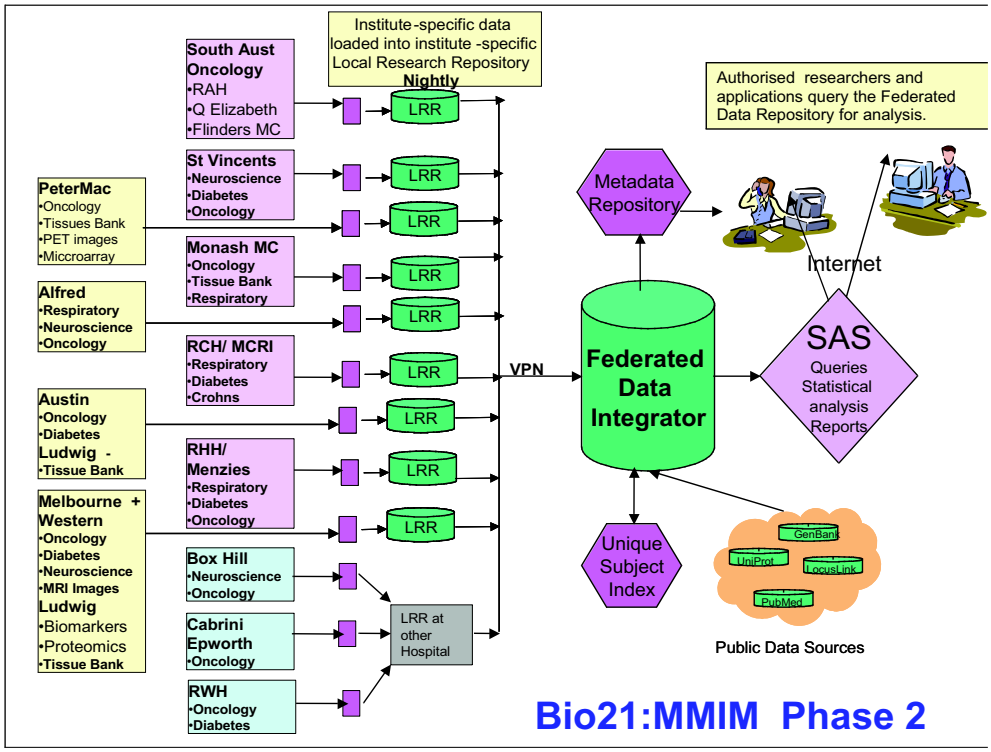


Figure 1. Schematic Architecture of the Bio21:MMIM data grid showing the secure data flow and technologies. Authorized researchers can access the integrated and de-identified data via the Internet.

The MMIM architecture

Diagrammatically the phase 2 architecture is shown in the diagram above.

Security levels:

There are two main levels of security for access to MMIM. Firstly, the users: authorization and usernames and password must be obtained, so only authorised researchers have access to the FDI. Secondly, the MMIM System Administrators who are responsible for building and maintaining the system who have access to all parts of the system (currently there are only two MMIM project authorised personnel with this level of access).

Improving the quality of source data

In MMIM source data was accepted 'as-is' - as provided by participant sites. Standards were used where appropriate (eg NCCI), however provided researchers agreed to the same data definitions, so data was comparable, then specific health data standards were not implemented. However, the databases and systems used for clinical data collection vary at local sites compromising data consistency, but the data quality has improved as the data was available to the clinical researchers. Indeed the project disease/tumour leaders have worked with colleagues in other hospitals and disease areas to standardise data fields and collection processes as far as possible. Further MMIM is working with Australian groups trialing standardised cancer data capture models. These initiatives will contribute to standardised and better quality data over time.

The metadata management – business glossary

The MMIM system provides the ability to search for the range and type of information being captured by MMIM for a particular disease and to discover whether the required data is available - the metadata as opposed to accessing the data itself – the technical data. MMIM users can search and discover the

- clinical areas covered (diseases types& database purpose)
- sites contributing
- types of data collected (pathology, procedures, genetic)
- detail of data elements

and have enough information to request access

MMIM chose IBM's Websphere Business Glossary (WBG) to provide terminology management capabilities. Definitions of standard terms, attaching standard terms to items (databases, tables, or columns), defining the hierarchies of terms, specifying the preferred terms, synonyms and having categories of terms with hierarchies have been implemented. This functionality is important in search and discovery of metadata as users may use non standard or non-familiar terminology (or may misspell words) when describing items. For example, "date of birth" may be described as "dob", "gender" may be described as "sex", etc.

The Business Glossary has open access from the Internet and searching for information can occur by browsing, by drilling down the 'trees' or searching using keywords. In addition the technical - authorized data users have sufficient information about the data items to interpret the data fields, to run queries and to understand the data models so they have the knowledge to join data across databases.

Privacy

The project has continued to seek independent legal advice from lawyers and privacy experts at all stages to ensure that measures taken privacy issues continue to be addressed.

Site and project Human Research Ethics Committee approval has been obtained for all participating sites, and is a prerequisite to proceeding with implementation of MMIM at participating sites..

Intellectual property

A Collaboration agreement that all participating sites must sign to join MMIM explicitly provides for recognition of both Background and Project Intellectual Property.

The project has a set of standard IP management and commercialization processes. Default IP positions are agreed. However, individual research projects are free to negotiate appropriate terms on a case by case basis.

Research results

Examples of research outcomes to date include the areas of epilepsy and neuropsychiatry[8-10], evaluating the sensitivity and specificity of FOBT compared to colonoscopy over a 25 year period [11], and the evaluation of colorectal cancer patients in the areas of biomarkers and therapy [12-15].

Discussion

Building the federated data integration infrastructure

MMIM has successfully built the system infrastructure and federated database integration capability outlined in the methodology section above. This technology allowed the issue of patient privacy, patient record-linkage to be covered as well as ensuring researcher intellectual property was protected. This stage of the project involved data sets (clinical, genomic, tissue bank & biomarkers) for three disease types, namely, colorectal cancer, epilepsy and diabetes.

MMIM is building on this to include a further ten public health sites in Victoria and three states and territories ultimately linking more than 35 disease databases. It is further expanding the technology across the Regional Cancer Services within Victoria and the Metropolitan Melbourne to create the South Eastern Australia part of an Australian Cancer Grid

Powering future research

The MMIM Project has transformed the way that research can be undertaken giving approved and authorised researchers access via the internet to a virtual repository of privacy–protected data not previously available.

From their own work stations researchers can now:

- Link genomic data to clinical / outcome data in Colorectal Cancer and Epilepsy;
- Test multiple hypotheses without collecting / recollecting their own data (with data owner approval);
- Research suitable pre-symptomatic testing and early intervention based on genotype data;
- Research genetic, genomic and proteomic profiles, factors that may influence treatment outcome
- Find out if tissue samples are available for patients with selected clinical profiles
- Analyse summary/statistical information across institutions and from diverse databases.

The following table summarises the traditional approach to research data collection and assembling of databases with that offered by the MMIM Project

Table 1 - Comparative Advantages of Using MMIM

Using Traditional Standalone Research Databases	**Using the MMIM Data Grid**
Static Data at one point in time Often one-off 'dump' Linked once	**Dynamic** Data refreshed & updated Live link to clinical research data Links made on-demand
Often anonymised data	Codified–ethically re-identifiable in exceptional circumstances and privacy protected
Data leaves 'owners' control	Data owners control access
Minimum data sets	Minimum + legacy data
Must specify exact data/ query up-front - can only answer one off specific research question	**Discovery** Research analysis / explorative 'Quality' type clinical reports Clinical Data collected at healthcare site Discovery tools and potential for iterative and exploratory research on a theme (for approved data)
Usually population based studies	Clinical and Biomedical Data collected at healthcare sites and population data

The MMIM virtual repository has enabled collaboration between multiple institutions, both within and across disease specialties, and between clinical researchers, bio-

informaticians and Information Technology specialists. This in turn has expanded research capacity and productivity as follows:

- Within and across disease groups;
- Between data owners;
- Between data owners and researchers across academic institutions in Australia;
- Between data owners and researchers overseas;
- New research data types –e.g. imaging (MRI)

At the same time MMIM can enable expanded teaching and training resources (data sets and tools) in the health research field (medical informatics, genomics, proteomics) which is being developed.

Conclusion

MMIM provides a privacy protected data grid for connecting heterogeneous and dispersed data for medical researchers. The platform is operational and is developing to become scalable and sustainable. It continues to incorporate new data and provide tools for researchers.

Acknowledgments

Phase 1 of t his project was funded by the Victorian Government Department of Innovation, Industry and Regional Development via Bio21 Australia Limited. Phase 2 was funded by the Commonwealth Government of Australian, Department of Education, Science and Training, via their Systemic Infrastructure Initiative. Phase 3 is funded by the Victorian Government Department of Innovation, Industry and Regional Development via the University of Melbourne.

References

[1] Martin-Sanchez F. Iakovidis I. Norager S. Maojo V. de Groen P. Van der Lei J. Jones T. Abraham-Fuchs K. Apweiler R. Babic A. Baud R. Breton V. Cinquin P. Doupi P. Dugas M. Eils R. Engelbrecht R. Ghazal P. Jehenson P. Kulikowski C. Lampe K. De Moor G. Orphanoudakis S. Rossing N. Sarachan B. Sousa A. Spekowius G. Thireos G. Zahlmann G. Zvarova J. Hermosilla I. Vicente FJ. Synergy between Medical Informatics and bioinformatics:Facilitating Genomic medicine for Future Healthcare. Journal of Biomedical Informatics. 37(1):30-42, 2004 Feb.

[2] INFOMED; A virtual laboratory for accessing and integrating genetic and medical information for health applications. EC Project IST-2001-39013. 2002-04 European Commission, Brussels.

[3] Sanz, F, Diaz C, Martin-Sanchez F ,Maojo V. Structuring Eurropean biomedical Informatics to Support individualized Healthcare: current issues and future trends. MedInfo 2004 ;11(pt 2);8030-7.

[4] Maojo V. Kulikowski CA. Reflections on biomedical informatics: from cybernetics to genomic medicine and nanomedicine. Studies in Health Technology & Informatics. 124:19-24, 2006

[5] Maojo V, Garcia-Remesal M, Billhardt H, Alonso-Calvo R, Perez-Rey D, Martin-Sanchez F. Designing New methodologies for Integrating Biomedical Information in Clinical Trials Methods Inf Med 2006;45:180-5.

[6] Von Eschenbach, Buetow K. Cancer Informatics Vision: caBIG ™ Cancer Informatics 2006:2;22-24.

[7] CancerGrid http://www.cancergrid.org/

[8] Yerra R, Kilpatrick C, Matkovic Z, King B, Rafael N, Hibbert M, Brand C, and O'Brien, First Seizure Clinic Experience: Heterogeneity of patient population and prognosis, *Epilepsia* **46** suppl 8 (2005), 360.

[9] Yerra R, Kilpatrick C, Matkovic Z, Belbin S, Rafael N, Hibbert M, Brand C, and O'Brien T, Syndromic diagnosis of epilepsy in the First Seizure Clinic population, Journal of Neurosciences, 238 (2005), S150

[10] Mark Walterfang, Ronald Siu, Dennis Velakoulis. The NUCOG: validity and reliability of a brief cognitive screening tool in neuropsychiatry patients. Australian and New Zealand Journal of Psychiatry 40 (2006), 995-1002

[11] Macrae FA, Slattery MA, Brown GJ, O'Dwyer MA, Murphy C, Hibbert ME, St John DJB. Sensitivity of faecal occult blood testing (FOBT) for asymptomatic colorectal cancer and advanced adenomas over a 25 year experience in colorectal cancer screening. Gastroenterological Society of Australia Australian Gastroenterology Week, Brisbane 2005, Journal of Gastroenterology and Hepatology 20 (2005), A34.

[12] Lim E, Jones IT, Gibbs P, McLaughlan S, Faragher I, Skinner I, Chao MW, and Johns J. Subsite-specific colorectal cancer in diabetic and non-diabetic patients . Cancer Epidemiology, Biomarkers & Prevention 14 (2005), 1579–1582.

[13] Rodrigues J, Lim E, McLaughlin S, Faragher I, Skinner I, Chao M, Croxford M, Chapman M, Johns J, Gibbs P The Influence of Language Spoken on Colorectal Cancer Diagnosis and Management'. ANZ Journal of Surgery 76 (2006), 671.

[14] Gibbs P, McLaughlin S, Skinner I, Jones I, Hayes I, Chapman M, Johns J, Lim L, Faragher Re: Completion of therapy by Medicare patients with stage III colon cancer. . Journal of the National Cancer Institute 98(21) (2006), 1582.

[15] Reiger NA, Barnett FS, Moore JWE, Neo E, Badahdah F, Ryan AJ, Ananda SS, Croxford M, Johns J, Gibbs P The quality of pathology reporting impacts on lymph node yield in colon cancer, Journal Clinical Oncology 25 (2007), 463.

Address for correspondence

Dr Marienne Hibbert, Ph D
MMIM Project Director
6 North, Main Building
Melbourne Health
Grattan Street, Parkville
VIC 3050

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Cancer Genomics Object Model: An Object Model for Multiple Functional Genomics Data for Cancer Research

Yu Rang Park[1], Hye Won Lee[1], Sung Bum Cho[1], Ju Han Kim[1,2*]

[1] *Seoul National University Biomedical Informatics (SNUBI), Seoul National University College of Medicine, Korea*
[2] *Human Genome Research Institute, Seoul University Medicine College of Medicine, Seoul 110-799, Korea*

Abstract

The development of functional genomics including transcriptomics, proteomics and metabolomics allow us to monitor a large number of key cellular pathways simultaneously. Several technology-specific data models have been introduced for the representation of functional genomics experimental data, including the MicroArray Gene Expression-Object Model (MAGE-OM), the Proteomics Experiment Data Repository (PEDRo), and the Tissue MicroArray-Object Model (TMA-OM). Despite the increasing number of cancer studies using multiple functional genomics technologies, there is still no integrated data model for multiple functional genomics experimental and clinical data. We propose an object-oriented data model for cancer genomics research, Cancer Genomics Object Model (CaGe-OM). We reference four data models: Functional Genomic-Object Model, MAGE-OM, TMA-OM and PEDRo. The clinical and histopathological information models are created by analyzing cancer management workflow and referencing the College of American Pathology Cancer Protocols and National Cancer Institute Common Data Elements. The CaGe-OM provides a comprehensive data model for integrated storage and analysis of clinical and multiple functional genomics data.

Keywords:

cancer, genomics, data model, standards

Introduction

Functional genomics includes studies of the abundance of gene transcripts by microarrays (transcriptomics), the abundance, localization and interactions of the translated proteins (proteomics), the flux in related metabolites (metabolomics), and various others [1]. For managing and representing of these functional genomics data, several technology-specific data models have been proposed, including MAGE-OM for microarray [2], PEDRo for proteomics [3], SMAR [4], ArMet [5] and MIAMET [6] for metabolomics, and TMA-OM [7] for tissue microarray.

Current researches emphasize the need to integrate data from multiple functional genomics [8, 9]. Following these trends, many cancer researches have been conducted using multiple functional genomics technologies including DNA microarray, 2DE/MS and Tissue Microarray for the understanding of global biological characteristics [3-5].

As the number of cancer studies using multiple functional genomics technologies increases, there are increasing demands for flexible solutions for systematic management of these data. Several databases have been implemented for specific functional technologies or specific purposes [12]. Despite the necessity, there is a few number of integrated data models (Table 1). It only supports a few genomics technologies or document models for genomics and clinical data. Furthermore, most approaches are designed without consideration of extendibility for integration with new biological data models.

In the present study, we designed a new data model for cancer genomics research using multiple functional genomics data, Cancer Genomics Object Model (CaGe-OM).

Table 1 - Previous approaches for integrated model

	Method	Target data	Reference model	Implementation
FGE-OM (Jones A et al., 2004)	Integrated object model	Transcriptomics, and proteomics (2DE and MS)	MAGE-OM, PEDro, Gla-PSI	RAPAD (microarray, 2DE and MS data)
SysBio-OM (Xirasagar S, et al., 2004)	Integrated object model	Transcriptomics, proteomics and metabolomics	MAGE-OM, PEDro, and a model for protein-protein interaction and metabolite	CEBS (only for microarray data)
Genotype Shared Model (HL7 clinical genomics SIG)	Document (XML)	Transcriptiomic, proteomics, sequence and clinical data	HL7 CDA	Genetic testing : BRCA Tissue typing: BMT
IBM GMS (Robson B, et al., 2004)	Document (XML)	Clinical and genomics (protein structure) data	HL7 CDA	Genomic Messing System Language (GMSL)
caCORE (Covitz PA, et al., 2003)	Object oriented API (caBIO)	Clinical and genomics data	Object Model	caBIG, CGAP, MMHCC, caArray etc..
XDesc (Shifman MA, et al., 2004)	EAV and Relational model	Clinical and genomics (Transcriptomis) data	TrialDB	YMD

Materials and methods

We used class diagram of Unified Modeling Language (UML) to represent the concepts, objects and relationships in multiple functional genomics data for cancer research. We reference four experimental data models (FuGE-OM, MAGE-OM, PEDRo and TMA-OM) and two clinical and histopathological data models (College of American Pathologist Cancer Protocols and National Cancer Institute Common Data Element) to design a data model for cancer genomics research.

Functional genomics experiment data modeling

For designing a framework to represent results from multiple functional genomics investigation, we reference four data models; the FuGE-OM for common aspects of experiments, the MAGE-OM for microarray, the PEDRo for proteomics, and the TMA-OM for tissue microarray. The FuGE-OM focused on modeling the common artifacts of functional genomics, such as sample preparation, protocols, instruments, and contact details [1]. Following the wisdom of the FuGE-OM, we reference packages associated with common aspects of functional genomics in three data models (MAGE-OM, PEDRo and TMA-OM) and modify the existing packages and classes within FuGE-OM for describing common biological experimental data which belongs to CommonBioData namespace. We extract technology-specific packages for each three data model. These extracted packages comprised in TechnologySpecific namespace.

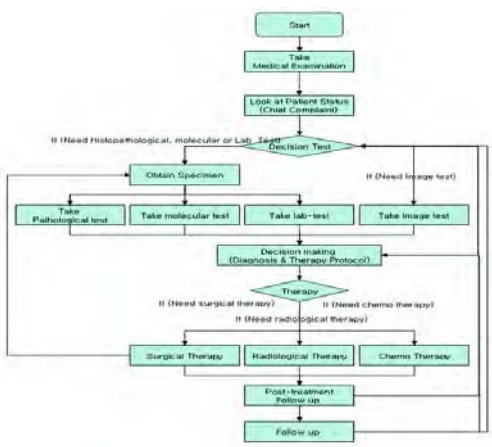


Figure 1 - Workflow diagram of clinical management of cancer. Diamonds indicate events and rectangles are physical entities.

Clinical and histopathological data modeling

For designing the clinical and histopathological data modeling, we analyzed cancer management and referenced document models of clinical and histopathological information, such as College of American Pathologists (CAP) Cancer Protocols (CPs) [13] and National Cancer Institute (NCI) Common Data Element (CDE)[14]. Figure 1 shows the workflow diagram of clinical management for cancer. To obtain comprehensive and extensible data models, we have created 6 packages (i.e., MedicalExamination, Histopathol, Specimen, DecisionTest, Therapy, and FollowUp) by systematically capturing the event and process from the workflow diagram (Figure 1) and category and value from the 43 CAP CPs and NCI CDE.

Results

Workflow of clinical management of cancer

For structured modeling of clinical data for cancer, it is required to analyze workflow of clinical management for cancer like diagnosis and therapy (Figure 1). When a patient arrives at a hospital, she/he takes a medical examination (captured by MedicalExamination class). Medical examination is an event to look at a patients status by a doctor based on physical examination (i.e. inspection, auscultation, and palpation) (PhysicalExam). Then the doctor writes down chief complaints of the patient. Then the patient takes a decision test such as clinical laboratory, images, histopathological and molecular tests (DecisionTest). The doctor makes a decision about the diagnosis and therapy protocols based on the results from various decision tests (Diagnosis & Plan). There are three types of therapy: surgical, radiological and chemotherapy (Therapy). In solid tumor, tumor specimen is obtained after surgical therapy. And then, histopathological test is taken on specimen (HistoPathol). After the therapy, post-treatment follow up is taken on the patient (FollowUp). After primary treatment, the patient is observed according to the follow up schedule (FollowUp).

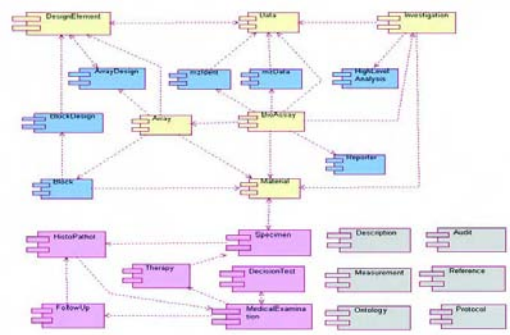


Figure 2 The relationships of 26 packages in Cancer Geno-mics Object Model (CaGe-OM). Most packages in this model are catego-rized into three namespaces; the CommonBioData (yellow-colored), ClinicalData (pink-colored) and Technolo-gySpecificData (blue-colored) namespae. Six packages (gray-colored) are remaining for general purpose.

Overview of cancer genomics object model

CaGe-OM is a data model that contains 183 classes grouped into 25 packages. Figure 2 shows the relationship of CaGe-OM packages, which grouped in three namespaces: the CommonBioData, TechnologySpecific, and ClinicalData models. The CommonBioData contains a

set of packages that describe common aspects of functional genomics from microarrays, proteomics, tissue microarray or potentially other functional genomics techniques. The packages belong to ClinicalData namespace represents clinical and histopathology data of cancer. The TechnologySpecific namespace contains the packages for describing technology-specific data. The remaining 6 packages for common annotation: Description, Measurement, Ontology, Audit, Reference, and Protocol. This model has three abstract classes at the top-level, Extendable, Describable and Identifiable that are unchanged from MAGE-OM. Most classes inherit their attributes. CommonBioData and TechnologySpecificData can refer to ClinicalData through Material packages in CommonBioData namespace.

CommonBioData namespace

The CommonBioData namespace represents common aspects of functional genomics experiments. Experimental design (captured by Investigation package), biological sample preparation (Material) and biological molecules such as DNA or protein sequences (ConceptualMolecule) are common components of all functional genomics investigation. The CommonBioData namespace consists of six packages; Investigation, Material, Array, DesignElement, Data and BioAssay.

To represent data from experiments using any type of technology, packages contained in this namespace have a generic structure. In Data packages, for instance, the Data object represents a container for a set of multidimensional data matrices and the coordinate set found in each of the dimensions.

The Material is a package for all biological and physical materials involved in an experimental workflow. For integrating genomics and clinical data, this package has a relationship with Specimen package that belongs to ClinicalData namespace. As a result, CommonBioData namespace has a reference to ClinicalData via Material and Specimen classes, representing the clinical and histopathological information of the specimen used in functional genomics experiments.

The Array, ArrayDesign, and DesignElement in the MAGE-OM contain information regarding the design, manufacturer and contents of microarrays (http://www.omg.org/docs/formal/ 03-02-03.pdf). In these packages, the ArrayDesign package is a microarray specific package. However, Array and DesignElement is commonly used in the functional genomics experiments such as microarray, tissue microarray and proteomics. Thus we are adding these two packages into CommonBioData namespace.

Clinical Data namespace

The Clinical Data namespace includes package with classes covering clinical and histopathological data 43 cancer types considered by CAP CPs, and clinical contexts from the workflow analysis of cancer management and NCI CDEs. The ClinicalData namespace is composed of six packages; MedicalExamination, Histopatho, Specimen, DecisonTest, Therapy and FollowUp.

The MedicalExamination package, the core package in ClinicalData namespace, contains classes for Demography, PhysicalExam, History, Diagnosis and Plan. Through MedicalExamination classe, all the other packages contained in ClinicalData namespace have associations with Medical-Examination package (Figure 3(a)).

The HistoPathol package provides classes describing histopathological information of specimens (Figure 3(b)). The BasicHistoPathol class stores elements that should be included regardless of the organ and tissue. The OrganSpecific class store elements for specific organs. The BasicHistoPathol class is an abstract class, the subclasses of which are the TumorInfo and Histology classes.

Classes in DecisonTest package are describing several medical tests such as image, laboratory, molecular and histopathological test. Therapy package contain classes to store data from medical therapy; surgical, radiological, and chemotherapy. Specimen package provide classes describing information of tissue obtained by surgical therapy or biopsy. The FollowUp package defines the classes for follow up data like recurrence and vital sign of patient.

TechnologySpecificData namespace

For storing technology-specific data of the experiment, the TechnologySpecificData namespace contains the packages derived from MAGE-OM, PEDRo, and TMA-OM. The TechnologySpecificData namespace is composed of eight packages; ArrayDesign, HighLevelAnalysis, Assay, mzData, mzIdent, Block, BlockDesign and Reporter.

The ArrayDesign and HighLevelAnalysis packages are microarray-specific packages. These packages are reused from corresponding MAGE-OM packages. ArrayDesign includes the manufacturing protocols, contacts, and details of the exact materials used for each feature in Array. The HighLevel-Analysis is a package for the analysis results.

The Assay, mzData and mzIdent packages, which come from the PEDRo, are proteomics-specific packages. The Assay package provides classes and attributes that contain information and annotation on the event of proteomics experiment using 2D or MS and the acquisition of images. The mzData package stores the output from mass spectrometry (MS). The mzIdent package contains the output (and input parameters) from database searches with mass spectrometry data to identify proteins or to quantify protein abundance.

The Block, BlockDesign and Reporter are tissue microarray-specific packages. These packages are identical to packages of the same name in TMA-OM. The BlockDesign package stores the intended pattern of individual block elements. The block with large number of tissues is constructed according to the BlockDesign and the block is sliced to arrays. The Block package records information on the actual events manufacturing blocks. The Reporter package contains classes for reporters used in TMA experiments. The reporter represents materials to identify a particular molecule like gene, protein, or DNA sequence.

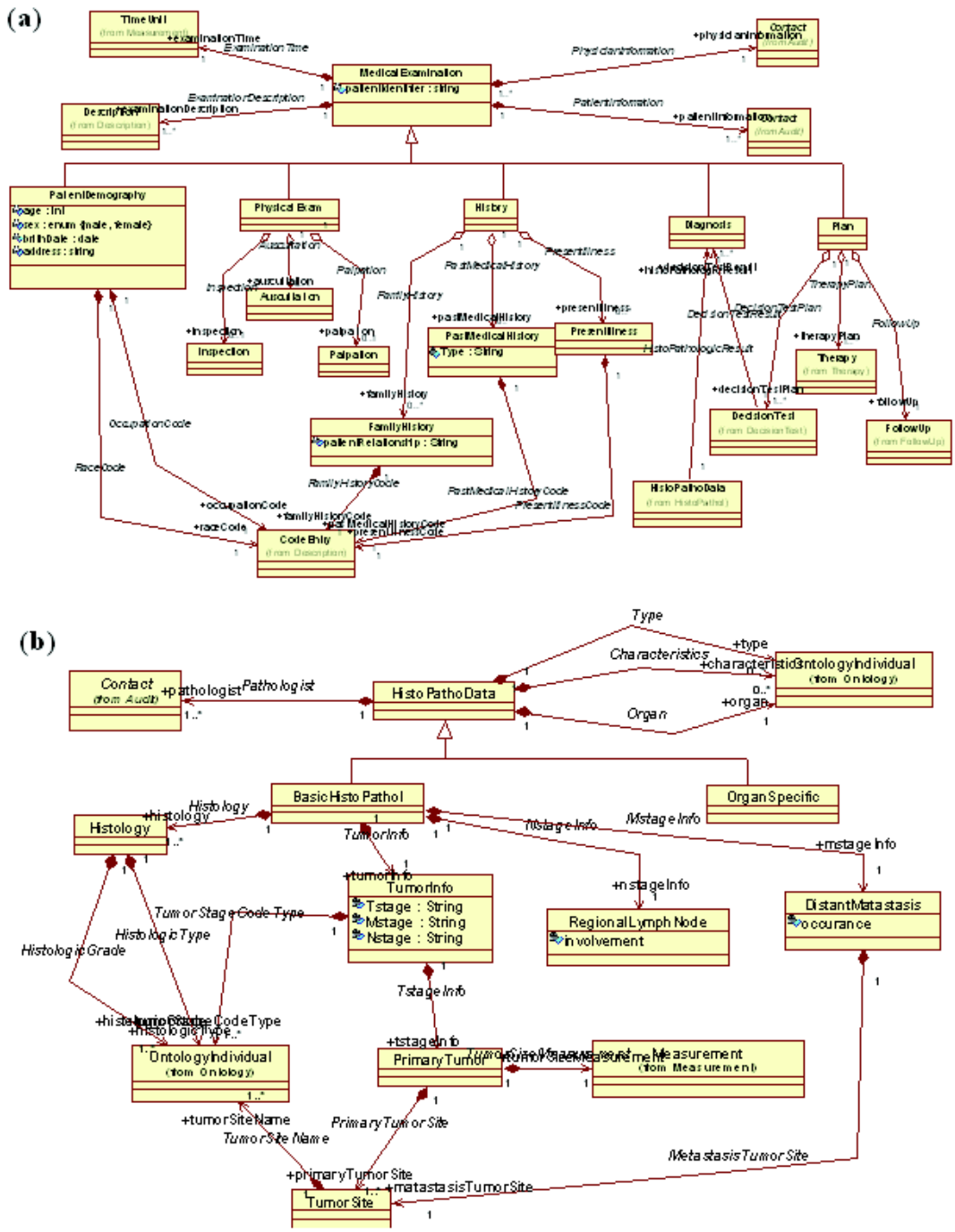


Figure 3 - Class diagram of (a) MedicalExamination and (b) Histopathol packages in ClinicalData namespace.

Discussion

We developed a data model, CaGe-OM, to store and integrate data generated from microarray, proteomics and tissue microarray experiments performed on the same biological samples. The CaGe-OM can represent clinical and histopathological information as multiple functional genomics data for any type of cancer. This integrated data model allows the combined analysis of multiple functional genomics data for understanding of the underlying biological nature in a systematic fashion.

The CaGe-OM can integrate easily a new biological data model without significant difficulty by representing common aspects of the new models as CommonBioData and technology-specific parts as TechnologySpecificData separately, while it is hard to modify the models in previous studies to condider and integrate a new model (Table 1). Because the CaGe-OM is independent of implementation, several applications based on this model such as relational database schema, web application and XML document can be constructed.

The development of an integrated data model for cancer genomics researches may facilitate tight integration of technology-specific data models and clinical data models. As functional genomics are increasingly used in cancer research, the CaGe-OM will be useful for the structured data management of clinical data and for the analysis of functional genomics data combined with clinical data.

Acknowledgments

This study was supported by a grant from Korea Health 21 R&D Project, Ministry of Health & Welfare, Republic of Korea (0412-MI01-0416-0002).

References

[1] Jones AR, Pizarro A, Spellman P, et al. FuGE: Functional Genomics Experiment Object Model. OMICS. Jun 2006; 10(2): 179-84

[2] Spellman PT, Miller M, Stewart J, et al. Design and implementation of microarray gene expression markup language (MAGE-ML). Genome Biol. Aug 2002; 3(9):Research0046.1-9

[3] Garwood K, McLaughlin T, Garwood C, et al. PEDRo: a database for storing, searching and disseminating experimental proteomics data. BMC Genomics. Sep 17 2004;5(1):68.

[4] Castle AL, Fiehn O, Kaddurah-Daouk R, et al. Metabolomics Standard Workshop and the development of international standards for reporting metabolomics experimental results. Brief Bioinform. Jun 2006;7(2):159-65.

[5] Jenkins H, Hardy N, Beckmann M, et al. A proposed framework for the description of plant metabolomics experiments and their results. Nat Biotechnol. Dec 2006;22(12):1601-6.

[6] Bino RJ, Hall RD, Fiehn O, et al. Potential of metabolomics as a functional genomics tool. Trends Plat Sci. Sep 2004;9(9):418-25.

[7] Lee HW, Park YR, Sim J, et al. The tissue microarray object model: a data model for storage, analysis and exchange of tissue microarray experimental data. Arch Pathol Lab Med. Jan 25 2006; 130(7):1004-13

[8] Ideker T, Galitski T, Hood L. A new approach to decoding life: systems biology. Annu Rev Genomics Hum Genet. 2001;2:343-72.

[9] Waters M, Boorman G, Bushel P, et al. Systems toxicology and the Chemical Effects in Biological Systems (CEBS) knowledge base. EHP Toxicogenomics. Jan 2003;111(1T):15-28.

[10] Callagy G, Pharoah P, Chin SF, et al. Identification and validation of prognostic markers in breast cancer with the complementary use of array-CGH and tissue microarrays. J Pathol. Feb 2005;205(3):388-396.

[11] Ippolito JE, Xu J, Jain S, et al. An integrated functional genomics and metabolomics approach for defining poor prognosis in human neuroendocrine cancers. Proc Natl Acad Sci U S A. Jul 12 2005;102(28):9901-6.

[12] Elfilali A, Lair S, Verbeke C, et al. ITTACA : a new database for integrated tumor transcriptome array and clinical data analysis. Nucleic Acids Res. Jan 1 2006;34(Database issue):D613-6.

[13] Warzel DB, Andonaydis C, McCurry B, Chilukuri R, Ishmukhamedov S, Covitz P. Common data element (CDE) management and deployment in clinical trials. AMIA Annu Symp Proc. 2003:1048.

[14] The College of American Pathologists. Cancer protocols in January 2006 revision version. Available at: http://www.cap.org/apps/docs/cancerprotocols/protocolsindex.html. Accessed November 8, 2006.

Address for correspondence

Ju Han Kim, MD, PhD,
Seoul National University College of Medicine,
28 Yongon-dong, Chongno-gu, Seoul 110-799, Korea.
Tel: +82-2-740-8320; Fax:+82-2-742-5947
e-mail: juhan@snu.ac.kr

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Automatic Pedigree Reconstruction for Genetic Studies in Isolated Populations

C. Larizza[a], I. Buetti[b], G. Milani[a], A. Nuzzo[a], C. Sala[b], D. Toniolo[b], R. Bellazzi[a]

[a] *Department of Computer Science and Systems, University of Pavia, Pavia, Italy*
[b] *DIBIT - San Raffaele Scientific Institute, Milan, Italy*

Abstract

This paper describes a tool implemented to automatically reconstruct the pedigree of an isolated population of Northern Italy with the aim of supporting genetic studies. The goal of such studies is to analyze genealogic, clinical and genetic data for genetic dissection of complex diseases. In this context the reconstruction of the population pedigree is fundamental to verify that such population is a genetic isolate and obtain the parental relationships among the individuals participating to the study. The algorithm presented in the paper, from heterogeneous data sources (demographic municipal and parish archives and other data sources), derives the pedigree applying several heuristic rules in a predefined order. One of the main difficulties in performing such task stands in the "record linkage" process that requires the definition of a sufficiently general strategy for managing the ambiguities caused by missing or imprecise/erroneous input data. The paper, finally, presents and discusses the preliminary results obtained by reconstructing the pedigree of four villages from the data collected during the first eighteen months of project.

Keywords:

pedigree, record linkage, isolated population.

Introduction

A specific interest of the post-genomic era is the correlation of genotypical and phenotypical information with the aim of investigating the genetic components of complex disorders. This research activity requires a strong involvement of Biomedical Informatics and, more in general, of Informatics to provide knowledge and tools for dealing with such an ambitious goal [1].

The term complex disorder refers to the phenotypes resulting from abnormalities that cannot be considered as dependent from a single gene, as in Mendelian diseases [2]. In this case, the problem of understanding their nature is typically very complex, since the risk associated with a mutation may depend for a large part on interaction with other genetic or environmental risk factors. As a consequence, it is usually impossible to find a genetic marker which turns out to be a perfect predictor of the complex trait. Different approaches can be applied to cope with this problem, ranging from linkage analysis to association studies [3]. Those studies, however, may be hampered by several factors, in particular by the small effects of the genetic variants on complex disorders and by the genetic heterogeneity of such variants [4].

In recent years there has been growing interest in mapping disease genes in genetically isolated populations, in which, due the small number of founders, the genetic heterogeneity is reduced, the environmental noise is usually minimized, and a wide set of genealogical data may be available [5]. For this reasons, many concurrent studies on isolated populations have nowadays started in Europe [6-7], Canada [8] and United States [9]. Within these studies, a cohort of citizens (or all citizens) belonging to the isolated population is visited, screened and genotyped. The information about their family trees is derived from public registries government and from other sources, such as the parishes' archives. The final goal of such studies is the joint analysis of genealogical, clinical and genetic data with the aim of obtaining long term research outcomes and, at the same time, short term public health outcomes for the population under screening. In such studies it is evident that the correct reconstruction of the population genealogic tree is a fundamental step for two different reasons: first, because it allows verifying whether the population has few founders and, therefore, can be considered genetically homogeneous, as expected; second, because it allows finding parental linking among individuals belonging to the same phenotype and therefore identifying their common ancestors.

Given the great number of individuals usually involved in such studies (usually the number ranges from 2.000 to 3.000 living people) and the need of reconstructing their pedigree at least from 1600 (the total number of individuals to be considered can increase up to 100.000), it is very important to be able to perform such task automatically or by minimizing, as much as possible, the manual intervention. However, the intrinsic characteristics of such kind of populations (e.g. a little set of frequently recurrent names and surnames) and of the available data sources (e.g. paper registries containing often imprecise or illegible data, very frequent name/surname variations due to transcription errors, language evolution or poor quality of the archives) require the adoption of specific strategies. In particular, it is necessary to choose a correct and robust approach for performing "record linkage" [10], basic processing step of the pedigree reconstruction. Record linkage is the task of quickly and accurately identifying records corresponding to the same entity (in our case, the entities are individuals) from one or more data sources, like municipal and parishes archives.

In this paper we describe the algorithm for the pedigree reconstruction, defined for an Italian project started in 2005. It involves the DIBIT-Hospital San Raffaele (HSR) of Milan and the Laboratory for BioMedical Informatics of the University of Pavia in studying the genetic component of complex disorders in the population of the Val Borbera, an isolated valley in the North of Italy. The aim of the paper is to present the different problems encountered in deriving the pedigree of an isolated Italian population and to discuss the computational solution adopted within the project. Moreover, the first results obtained by reconstructing the pedigree of the population since 1838 and an evaluation of the performance of the algorithm are presented. Finally, the open problems and the future developments aimed at improving the algorithm are discussed.

Materials and methods

The Val Borbera project

The Val Borbera is an isolated valley in the Appenine Region in Northern Italy that has been geographically isolated from the surrounding area until recent years. The population was mainly farming until 50 years ago and seems to have had a constant, but limited, growth over the centuries until a large portion immigrated to the Americas at the beginning of the last century. The part of the population that did not emigrate is still living in the valley or in the surrounding and corresponds to about 2000 people. We started in June 2005 collecting the acts of the municipal archives of four of the eight little towns and the birth and marriage certificates of the several little parishes of the valley, by transferring them from paper to electronic form. Moreover, we exploited the electronic demographic archives available since 1985 to manage the visits booking of the people participating to the study. In the future we will record also the death certificates of the population, which data will be exploited to refine and improve the reconstruction algorithm.

Computational methods

This section describes the different data sources available for the pedigree reconstruction, the architecture of the database designed to support the different phases of the algorithm, the problems encountered during the analysis of the data and the computational solution adopted.

The data sources

The study exploits the demographic information coming from municipal and church archives, quite easily traced in the few towns and parishes of the valley, stored in an MS-Access database. Such data represent the input to the pedigree reconstruction algorithm. The overall data processing covers the following sequence of steps, reported also in Figure 1:

1. *data import* from the heterogeneous databases (visits booking and municipal and parish archives) into a unified format;
2. *data cleaning and merge*, for data standardization and duplicates elimination with data merge;
3. *pedigree reconstruction*.

Step 1 is a sort of data structure standardization process, so that after the data import into a Unified Data Structure, each individual record has the same format, as shown in Figure 1. The unified structure is filled in by extracting and preprocessing the data coming from the historical archives, in order to reduce its imprecision and complete the partial information. This phase entails the definition of some rules for deriving new information from the available data. As an example, it is possible to reduce or completely define an individual birth date range, starting from the son's birth date, when the exact one is not available or it is completely omitted.

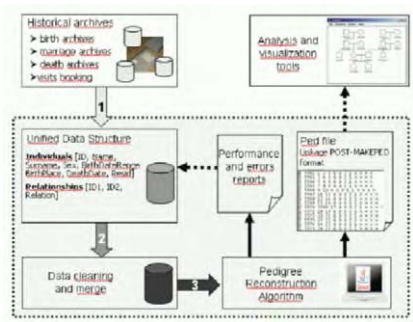


Figure 1- The overall strategy for data processing. The grey box highlights the procedures described in this paper

For this purpose, we distinguish the individuals reported in the archives into three categories: *Registered Persons, Relatives, and Spouses. Registered Persons* are the individuals identified trough their birth certificate or reported in the electronic municipal demographic archives (used also for visits booking). We assume that their personal identifying data are known without ambiguities (complete names, surnames and the precise birth date are always reported). *Relatives* are the persons reported in the different types of certificates (birth, marriage or death) and in the visits booking archive as parents or grandfathers of registered persons or spouses. Usually, such persons are identified only with their name and surname. Sometimes their age is reported, but frequently it is imprecise, having an error of ±3 years, while the birth date is omitted. *Spouses* are the individuals registered in the marriage certificates, whose demographic data (especially name and surname) are frequently different from those reported in birth certificates. This very frequent situation hampers the possibility of a strict match of a spouse with the corresponding registered person and forces us to resort to some additional strategies for record linkage performed in steps 2 and 3. The unified data structure allows the data cleaning and the reconstruction algorithm steps to be independent from the source databases structure, so that, if any modification of the source occurs or new information become available, after a suitable adjustment of the import procedure, it is necessary only to run again the algorithm without any other modification. On the other hand, the limit of this solution is that the reconstruction is not incremental with respect to new information inserted into the database.

Step 2 is fundamental to simplify the mechanisms of record linkage applied at Step 3. Its final goal is to eliminate possible duplicated individuals, by preserving and merging into a single record all the information available. If any information is reported in the duplicated records with different precision, it is necessary to resort to some a priori assump-

tion on the source archives reliability, in order to reduce the imprecision and the ambiguities during the merge. Before duplicates elimination, it is sometime necessary to perform names standardization for solving the problems in record linkage due to their different kind of variations.

Finally, in step 3 the reconstruction algorithm is performed in order to identify the parents of each registered person by exploiting all the available information derived at step 2. This phase requires to perform multiple record linkage, since the same person can be referred in several records (for example his birth certificate, all his sons birth certificates and his marriage certificate) with slightly different data. The strategy we adopt here is to perform different levels of checks between any couple of records, in order to detect all the relatives and, therefore, to correctly reconstruct the families composition. The final output is a *LINKAGE POST-MAKEPED* formatted file [11] containing the list of all the trios[1] of the population. This format has become a standard for most genetic analysis and visualization tools, so that the results need no more process to be evaluated and used. Moreover, some further output reports are provided about possible or proved data inconsistencies, which may be used to revise ambiguous, corrupted or missing information.

The following paragraphs describe in detail how steps 2 and 3 are performed. It is important to notice that such steps are completely independent from the original data sources and, after appropriate setup, can be applied also to different data sets.

Record linkage methods

In the different phases of the algorithm (data cleaning and pedigree reconstruction) we resort to record linkage procedures based on the match of the name, surname and birth date/age attributes. From our own experience, the most common problems in record linkage for pedigree reconstruction are generated by errors or imprecision in the attributes caused by *name variations*, *lack of precision in dates* (birth, marriage or death date), *incomplete information*, like partial name and/or surname, or availability of the individual age instead of his/her birth date. The more critical problem is represented by name variations, which cause failure to bring linkable records pairs together, due to different types of modifications on the name. The most frequent variations are: *spelling variations, phonetic variations, multiple and/or alternate names*.

- *Spelling variations*, usually due to data entry errors or ambiguity in demographic data, create a real problem which needs its own solution.

Example of name/surname variations are:
Bellazzi/Bellazi, Catterina/Caterina, Aluisa/Luisa.

- *Double or multiple names* which are very frequent in our data base. Typical situation is the one where the same person is referred in three archives with different names (which we call tokens):
  - birth certificate: *Anna Maria*
  - marriage certificate: *Anna*
  - son's birth certificate: *Maria*

If we should adopt an exact match of the names to link the records, a great number of relationships should be lost. To cope with this situation, we resort to a names comparison function based on the Levenshtein distance [12] applied to each token. Such function calculates a similarity score between two names taking into account: the number of tokens composing the name, the frequency of each token in the population (e.g. the name Maria is very frequent, so its contribution to the similarity score is lower than other less recurrent names) and the edit distance between each token. A threshold score is defined to accept or reject the match of two records based on the names.

- *Phonetics variations* result in significant modification of the name/surname usually due to the natural Italian language evolution and affect surnames coming from certifications registered before 1838. A dedicated strategy is needed for managing this kind of problem. At this time the idea is to identify (manually or automatically, using the edit distance) all the possible variations of every single name/surname, organize them into categories and associate each category to a "standardized" one (which typically is the one used since 1838). Once such categories have been defined, a preprocessing phase aimed to standardize all the names/surnames will be performed, before starting the actual reconstruction algorithm. Such strategy has not yet applied to real data, since currently only demographic data since 1838 are available in electronic form.
- Another frequent source of error is represented by imprecision in dates. This seems not so critical like the name's variation's problem. Our strategy here is to define some heuristic rules in order to derive a plausible range for the birth date of a person, by resorting to the fertile age concept. We assume that at the birth of a child the mother is never less than 14 or more than 60 years old, and the father is not less than 15 or more than 75 years old. This rule helps us to establish, with some other minor hints, a range for birth, death and marriage dates.

As the overall data processing cannot be completely automated, due to errors or imprecision caused by manual entries in the original archives, some reports of the ambiguous cases are provided to the archivists to manually check and correct the data in the source archives, as described in the following section.

An important step of the data cleaning process takes care of the elimination of duplicates and implements a strategy of data merge, which completes the information of the records with the data contained in the duplicates discarded.

The duplicates problem

Due to the source data organization, some individuals could have double or even three records (e.g. municipal and parish birth certificates and visits booking records), containing slightly different data (name, surname, birth date/age of an individual and of their parents). The issue at this step is to define a strategy for recognizing the duplicates and a series of rules to reduce the imprecision of the data during the merge. Such criteria establish, for example, which kind of record is considered the master and which are considered duplicates. This choice is driven by a priori

1 A *trio* is the numeric triplet that identifies an individual (called *proband*) together with his/her parents: *IDproband IDfather IDmother*

knowledge about the reliability and completeness of the different source archives. As an example, when we find duplicates in birth and municipal electronic archives, we maintain the second one, because its data are checked by the individual at the time of his/her visit and, therefore, is considered more reliable. In any case, we complete the information with the data coming from the birth certificate and, if the dates are reported with different precisions, we maintain the ones more accurate. A simple example of duplicates data is shown in Table 1, while Table 2 reports the resulting merged record.

Table 1 - The two original records

Archive type	Name	Surname	Date-Of-Birth-From	Date-Of-Birth-To
Municipal	Cristiana Francesca	Larizza	10-11-1956	10-11-1962
Birth	Cristiana Grazia	Larizza	10-11-1956	10-11-1957

Table 2 - The resulting record after duplicate elimination and data merge

Archive type	Name	Surname	Date-of-birth-from	Date-of-birth-to
Municipal	Cristiana Francesca Grazie	Larizza	10-11-1956	10-11-1957

The same kind of merge is performed also on data related to parents/relatives of each individual. In this case, it is very frequent having in one record the age and in another a birth date range (derived by exploiting the fertile age concept). In this case, for example, the mother/father age can be used to reduce his/her birth date range.

The reconstruction algorithm

The final goal of the pedigree reconstruction process is to create a trio for each registered person, by performing a series of cross checks between the information reported within the individual record (demographic data of her/his parents) and the marriage certificates of the potential parents. Such procedure can take to one of the following results: a) identification among the registered persons of the actual individual father/mother; b) creation of a fictitious father/mother that will be reused like any other registered person as possible parent of the remaining individuals.

In any case at the end of the procedure the relationship father-child/mother-child will be inserted into the corresponding table. More precisely, the algorithm covers the following steps, as shown in Figure 2.

1. Extraction from the data base of the list of registered persons potentially matching with the individual's father/mother. This phase is based on a blocking strategy which reduces sensitively the computational time in the comparison phase. The selected blocking attributes are: parent's sex, surname and birth date/age. Depending on the precision of the available data and on the extraction results, the blocking phase can be reiterated by relaxing some constraints in the birth date in order to try to obtain a not empty records list.
2. For every possible parent included in the list, a cross check, based on his/her marriage certificate, is performed in order to confirm the true match. It consists in verifying whether the potential father/mother is married with the mother/father declared in the son's birth certificate and in checking the compatibility of their ages.

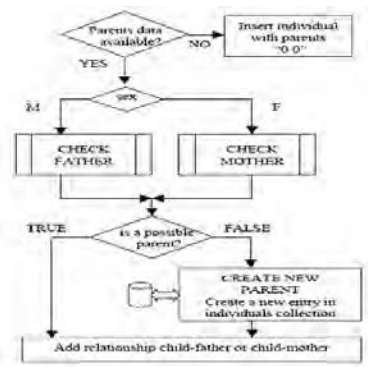


Figure 2 - Flow chart of the reconstruction algorithm.

3. If the marriage data confirm such parental relationship, the relation parent-child is created and stored in the corresponding table and the search stops.

If it is not found any possible parent among the registered persons of the list, including also fictitious individuals, (since the sequence of cross checks stops, due to failure of one of the matching conditions), or if the generated list is empty, the algorithm creates a further fictitious individual, based upon the data declared in the son's birth certificate and creates and stores the relation parent-child.

Results

After one a half year of project we collected and analyzed the data coming from municipal birth and marriage certificates registered since 1838 and municipal electronic archives of four towns available since 1985.

Table 3 - The data analyzed

Archive type	records #
Births	17.351
Municipal demography	3.480
Marriages	12.915
Individuals category	
Registered Persons (excluding duplicates)	19.139
Spouses	25.829
Relatives	75.529
Fictitious individuals	12.111

Table 3 reports the archive type and the number of records filled in into the Unified Data Structure during the import procedure from the historical archives. The total number of individuals whose pedigree has been derived is 31.164. The remaining records are used for cross checking between birth and marriage certificates. The data cleaning step detected 1692 records referring to duplicated individuals. Such records have been discarded by the reconstruction process, after addition or merge of the contained information into the remaining records.

The results of the pedigree reconstruction provided encouraging results being able to reconstruct a very big family of 10.634 individuals, corresponding to the 33.6% of the population analyzed. The total number of fictitious individuals created to complete the broken families is 12.111, but only 2.454 have been included into the big family. The remaining families are very small: we have 4.219 families whose dimension ranges from 3 to 117 individuals. Moreover, for some individuals it is possible to reconstruct the family up to seven generations. Also twins are in general correctly identified. The quality of the reconstruction algorithm was measured in terms of number of true positives n_m (number of relations correctly detected), number of false positives n_{fp} (number of multiple relations n_{mul} occurring when individuals are related to more than one father and/or mother) and number of false negatives n_{fn} (considered equal to the number of fictitious individuals $n_{fict,}$). Given N_m the number of expected relations (which is twice the number of individuals of the population) we can assume $n_m = N_m - n_{mul} - n_{fict}$. Given the previous assumptions, the *Recall* and *Precision* of the algorithm [10] were respectively estimated to be about 67.7% and 97.1%. At a first evaluation of the algorithm performance, it seems that a problem, although not very frequent, is the separation of a unique family into many family groups, whereas the aggregation of individuals belonging to different families into a single one never occurs. The debugging reports detected some conflicts in about 290 families' composition (spouses have a different number of sons), so our next goal is to identify and correct manually the errors causing this problem (at a first investigation they seem mainly due to transcription errors in the paper archives). On the basis of the debug performed until now, the number of records that will require a clerical manual review is about 3400. After the manual correction of the source archives, we expect to reduce the number of fictitious individuals and to unify some families into a single one. However, the results obtained until now are considered satisfactory by both the archivists and by the biologists that are able, on the basis of the pedigree obtained, to schedule the future visits and the genotyping of the most interesting individuals.

Conclusion

The automatic reconstruction of a population pedigree is a difficult problem that requires specific record linkage strategies. Various groups in the world are working on such techniques for various purposes (administrative, clinical, epidemiological studies, etc.), many of them being based on probabilistic approaches, but each methodology needs a customization based on the application context and the country (in particular, when names and surnames are chosen as linkage attributes) which heavily affects the performance (in terms of rate of errors) of the algorithm. In this paper we propose an approach that takes into account several context dependent ties on the linkage attributes in order to assess a true or non-true match between each record pair. The pedigree obtained automatically with the current algorithm seems satisfactory. However, to improve the results it will be necessary the manual intervention for detecting the causes of the errors in the tree and the correction of the data in the source archives. In the future we will make a more systematic evaluation of the matching rules, in order to tune the configuration parameters (for example, the threshold for the edit distance) and obtain an improvement in the algorithm performance. Moreover, we will evaluate the adequacy of the solution defined for managing the phonetic variations in names and surnames on real data.

References

[1] Martin-Sanchez F, Iakovidis I, Nørager S, Maojo V, De Groen P, Van der Lei J et al. Synergy between medical informatics and bioinformatics: facilitating genomic medicine for future health care. JBI, 2004 Feb; 37(1):30-42.

[2] Lander ES, Schork NJ. Genetic dissection of complex traits. Science. 1994 Sep 30; 265(5181):2037-48.

[3] Botstein D, Risch N. Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. Nature Genetics, 2003 Mar;33:228-237.

[4] Cardon LR, Bell JI. Association study designs for complex diseases. Nature Genetics, 2001 Feb;2:91-99.

[5] Chapman NH, Thompson EA. Linkage disequilibrium mapping: the role of population history, size, and structure. Adv Genet. 2001;42:413-37.

[6] Mancosu G, Cosso M, Marras F, Borlino CC, Ledda G, Manias T et al. Browsing isolated population data. BMC Bioinformatics. 2005 Dec 1;6 Suppl 4:S17.

[7] Izzi C, Sanna-Cherchi S, Prati E, Belleri R, Remedio A, Tardanico R e al. Familial aggregation of primary glomerulonephritis in an Italian population isolate: Valtrompia study. Kidney Int. 2006 Mar;69(6):1033-40.

[8] Hamet P, Merlo E, Seda O, Broeckel U, Tremblay J, Kaldunski M et al. Quantitative founder-effect analysis of French Canadian families identifies specific loci contributing to metabolic phenotypes of hypertension. Am J Hum Genet. 2005 May;76(5):815-32. Epub 2005.

[9] Lin JP, Hirsch R, Jacobsson LT, Scott WW, Ma LD, Pillemer SR et al. Genealogy construction in a historically isolated population: application to genetic studies of rheumatoid arthritis in the Pima Indian. Genet Med 1999; 1(5): 187-93.

[10] Gu L, Baxter R, Vickers D, Rainsford C Record linkage: Current practice and future directions. Technical Report 03/83, April 2003, CSIRO Mathematical and Information Sciences, GPO Box 664, Canberra 2601, Australia.

[11] G. M. Lathrop, J. M. Lalouel, C. Julier, and J. Ott Strategies for Multilocus Linkage Analysis in Humans PNAS 1984;81:3443-6.

[12] V.I. Levenshtein, Insertions and Reversals, Cybernetics and control Theory 1966; 10(8):707-10.

Address for correspondence

cristiana.larizza@unipv.it

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Enhancing the Quality of Phylogenetic Analysis using Fuzzy Hidden Markov Model Alignments

Chrysa Collyda[a], Sotiris Diplaris[b], Pericles Mitkas[b], Nicos Maglaveras[a], Costas Pappas[a]

[a]*Lab of Medical Informatics, Faculty of Medicine, Aristotle University of Thessaloniki, Greece*
[b] *Department of Electrical and Computer Engineering, Aristotle University of Thessaloniki, Greece*

Abstract

Any effective phylogeny inference based on molecular data begins by performing efficient multiple sequence alignments. So far, the Hidden Markov Model (HMM) method for multiple sequence alignment has been proved competitive to the classical deterministic algorithms with respect to phylogenetic analysis; nevertheless, its stochastic nature does not help it cope with the existing dependence among the sequence elements. This paper deals with phylogenetic analysis of protein and gene data using multiple sequence alignments produced by fuzzy profile Hidden Markov Models. Fuzzy profile HMMs are a novel type of profile HMMs based on fuzzy sets and fuzzy integrals, which generalize the classical stochastic HMM by relaxing its independence assumptions. In this paper, alignments produced by the fuzzy HMM model are used in phylogenetic analysis of protein data, enhancing the quality of phylogenetic trees. The new methodology is implemented in HPV virus phylogenetic inference. The results of the analysis are compared against those obtained by the classical profile HMM model and depict the superiority of the fuzzy profile HMM in this field.

Keywords:

bioinformatics, phylogenetic analysis, multiple sequence alignment, fuzzy integrals, Hidden Markov Models.

Introduction

The use of molecular data for inferring phylogenetic trees has gained considerable interest among biomedical researchers. Organisms, such as viruses do not leave fossil records, thus the only way to study their past is through the phylogenetic relationships of existing viruses. Phylogenetic analysis of protein and gene data can be accomplished by analyzing the genomic and proteomic sequence of the species. Many automatic methods have been developed for inferring phylogenetic trees, such as maximum parsimony[1], maximum likelihood[2] and distance methods[3]. The common feature of all these methods is that rates or patterns of change in sequences cannot be analyzed unless the sequences can be aligned [4], thus a robust multiple sequence alignment (MSA) is required as an input.

Bioinformatics offer a series of methods addressing this problem, such as CLUSTAL-W [5], PSI-BLAST [6] and HMMER [7] that can overpower classic methods of pairwise sequence alignment [8].

The well-known and widely used statistical method of characterizing the spectral properties of the residues of a genomic or proteomic pattern is the HMM approach. Profile HMMs have proved to offer a robust solution for MSA. Their wide use in bioinformatics has led to the creation of large profile databases [9],[10] that can offer biological knowledge (alignments, phylogenetic distribution, domain organization) for solving various problems, such as protein classification [11], building of phylogenetic trees [12], or gene function prediction.

However, an issue with the use of HMMs in MSA is its simplifying assumption on stochastic independence. This property, though, is not at all obvious when examining genomic or proteomic sequences; an underlying dependence may exist between current and previous states. Fuzzy HMMs have been introduced in speech recognition [13], in order to relax this assumption and resolve similar model definition issues, while in a previous paper we have introduced them in biological sequence analysis [14], by mathematically formulating the fuzzy profile Hidden Markov Model. Relative work has also been done in [15], where alignments from a fuzzy profile model were used for the description of the protein domain of kinases.

A characteristic of profile-HMMs, which have been used for MSA so far, is that these are finite models for the probability distribution over an infinite number of possible sequences. Profile-HMMs have the great benefit on generalized profiles that they are formally built on the probability theory. The disadvantage is that this theory restricts the flexibility of the models because the sum of the probability distribution over all modeled sequences must equal to one. In consequence, the probability of one sequence cannot be increased without decreasing the probability of another sequence in the profile-HMM. Fuzzy profile HMMs lack this restriction, thus they can be effectively used in order to better represent the sequences common residues and ultimately construct better phylogenetic trees.

In this paper, a new methodology of phylogenetic tree inference that makes use of the fuzzy profile HMM for multiple sequence aligning is presented. The fuzzy profile HMM representation, as we introduced it in [14], is defined by using fuzzy integrals and fuzzy operators in HMMs instead of probability theory. The classical HMM probabilities are replaced by fuzzy possibilities. The Choquet integral [16] is used for the integration over the HMM states and a new fuzzy measure is used for its application. Multiple sequence alignments are then obtained from the model and are used for the phylogenetic analysis of viruses coming for the HPV family. Phylogenetic analysis is finally performed with different algorithms and is evaluated using a bootstrapping schema.

The rest of the paper is structured as follows. First, the description of the experiment data is presented, while the following section describes in brief the methodology of constructing, training and acquiring alignments from the fuzzy HMM. In consecutive, we present the phylogenetic analysis schema, as well as the experimental results. Finally, we discuss the potential impact of the proposed methodology and end up with our conclusions.

Materials and methods

Data description

In the conducted experiments we used the E6 protein of several types, subtypes and variants of the Human Papillomavirus. Specifically, we acquired 78 different variants of the E6 protein using PSI-BLAST, coming from all known HPV types, as well as 30 protein sequences with homology to the HPV E6 protein, coming from various organisms. The latter were incorporated in the dataset in order to identify the discrimination capability of the different methodologies.

MSA with fuzzy profile Hidden Markov Models

Profile Hidden Markov Models

HMMs are statistical models used for MSA that allow the comparison of one gene or protein with a group of others, therefore facilitating the production of distinct differences between itself and the others. HMMs are a generalization of the profile in terms of statistical weights, rather than scores. At each position, the profile HMM gives the probability of finding a particular residue, an insertion, or a deletion. A profile HMM is composed by a number of interconnected states, each of which is able to emit observable output symbols, i.e. residues or gaps. Each state contains symbol emission probabilities and state transition probabilities. The symbol emission probabilities b_{jk} represent the probability of emitting each possible symbol k from a state j, whereas state transition probabilities a_{ij} are the probabilities of moving from the current state i to a next one j. An observation sequence $O=O_1O_2\ldots O_T$ can be generated by starting at an initial state and continuously changing of states, by also emitting symbols, until a specific end-state is reached at time T. The only visible outputs in this procedure are the emitted symbols, while the actual transition between states remains "hidden". Figure 1 depicts the structure of a profile HMM used for MSA, as introduced in [17]. Multiple alignments are used as a training set to build the model. One match state is assigned for each alignment column, insert states serve to insert extra symbols relative to the match states, while delete states allow for skipping positions in the training set aligned sequences. There are totally N=3*m+3 states in a profile-HMM, where m is the number of its match states.

The utilization of profile HMMs for MSA can be divided in three problems [18]:

- Problem 1: Computation of an observation probability according to the model: $P(O|\lambda)$. This is the problem of evaluation (HMM scoring), and it is usually solved using the forward-backward procedure.
- Problem 2: Computation of the state-sequence which fits the best to an observed sequence. This is the alignment problem. The Viterbi algorithm is usually used solve this problem and recover the hidden part of the model.
- Problem 3: Computation of the model parameters a_{ij}, b_{jk} and π to maximize the probability of one observation. This is the training problem, and the EM algorithm is usually exploited to this end.

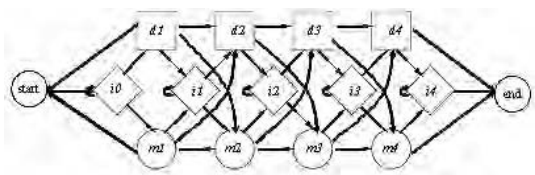


Figure 1- An example profile HMM with four match states.

MSA with fuzzy profile Hidden Markov Models

In a fuzzy profile Hidden Markov Model the classical HMM probabilities are replaced by fuzzy possibilities. Fuzzy integrals are used for aggregation over the states, while fuzzy operators are used instead of the algebraic ones. Though, the profile HMM structure is kept intact in terms of states and observations. The advantage of using fuzzy operators is that they are less constrained than classical integrals and probabilities, thus relaxing the independence assumptions that are necessary with probability functions in classical HMMs. This transformation also reduces the space of computations, thus yielding better response times.

In order to relax the additivity property of the classical HMMs constraint and take the relations between subsets into consideration, a generalization in terms of fuzzy measures has been introduced [19]. For the implementation of fuzzy HMMs, a possibility measure can be defined, such that

$$P(A \cup B) = \max(P(A), P(B)) \quad (1)$$

The max operator in the above equation is the fuzzy intersection operator. For a finite fuzzy set $X = \{x_1, x_2, \ldots x_n\}$, the density μ^i of a fuzzy measure μ can also be defined [20] as $\mu^i = \mu(\{x_i\})$, where $\mu(\{x_i\})$ represents a fuzzy measure, such

as the max operator, that is computed over the variables of the fuzzy set *X*.

This fuzzy measure can then be used with fuzzy integrals to compute integrations over fuzzy sets.

For the case of profile HMMs, the integration is done over the states that are a discrete set. In such cases, the discrete Choquet integral [21] can be used:

$$I = \sum_{i=1}^{n} \left[h(x_i) - h(x_{i-1}) \right] \mu_i^n \quad , \tag{2}$$

with defined $x_i : 0 = h(x_0) \le h(x_1) \le \ldots \le h(x_n)$, and the discrete step μ_i^j:

$$\mu_i^j = \begin{cases} \mu\left(\{x_i, x_{i+1}, x_j\}\right) & \text{if } i \le j \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

The fuzzy profile HMM $\bar{\lambda} = (\bar{A}, \bar{B}, \bar{\pi})$ with *N* states $S=\{S_1,S_2,\ldots,S_N\}$ that can be observed through a space of observations Ω with observations $O=O_1O_2\ldots O_T$ corresponding to unknown state sequences $Q=q_1,q_2,\ldots, q_T$ can be fully defined by the matrices $\bar{A}$, $\bar{B}$ and $\bar{\pi}$, where $\bar{A}$ is the fuzzy state transition matrix, $\bar{B}$ is the fuzzy observation matrix and $\bar{\pi}$ is initial state fuzzy density. Two fuzzy variables $x \in X=\{x_1,x_2,\ldots, x_N\}$ and $y \in Y=\{y_1,y_2,\ldots, y_N\}$ are used to represent the state at time *t* and *t+1* [17].

In these terms, $\bar{\pi}_S(A)$ is the grade of certainty that the initial state is in A. Respectively, for $X_0 \subset X$ and $Y_0 \subset Y$, $\bar{\alpha}_Y(X_0 \mid Y_0)$ is the grade of certainty that the state at time *t+1* is in Y_0, given that the previous state was X_0. Concerning the observation space $\Omega_0 \subset \Omega$, $\bar{b}_i(\Omega_0)$ is the grade of certainty that the current observation is in Ω_0, given a current state S_i.

After defining the model we are able to address the three HMM problems. Specifically, the problem of HMM evaluation can be solved using the fuzzy forward-backward algorithm. $\bar{\alpha}_i(t)$ is the grade of certainty of $O=O_1O_2\ldots O_T$ and x_i at time *t*. The initialization step is $\bar{\alpha}_1(i) = \bar{\pi}_i \wedge \bar{b}_i(O_1)$, while the induction step becomes:

$$\bar{\alpha}_{t+1}(i) = \sum_{i=1}^{N} \bar{\alpha}_{ij} \left[\mu_i^n(t,j) - \mu_{i+1}^n(t,j) \right] \wedge \bar{b}_i(O_{t+1}) \tag{4}$$

where the sum is the discrete Choquet integral, the $\wedge$ operator stands for the fuzzy intersection operator, and μ_i^j is defined in Equation 3. From the above equation, it is possible to observe that the assumption of independence of the observation until time *t* is not necessary anymore neither is necessary the knowledge of the next state. The answer to the evaluation problem for the forward and backward variables respectively is:

$$P(O \mid \lambda) = \sum_{i=1}^{N} \bar{\alpha}_T(i), \; P(O \mid \lambda) = \sum_{i=1}^{N} \bar{\beta}_1(i) * \bar{b}_i(O_1) \tag{5}$$

In the fuzzy case, the grade of certainty for a sequence is used to score the model. The Choquet integral is computed over the states at each time *t*, where the integration step $(\mu_i^n(t,j) - \mu_{i+1}^n(t,j))$ becomes a value *j* at time *t+1*.

Respectively, the fuzzy Viterbi algorithm, which is used for the alignment of new sequences to the model, uses the Choquet integral and multiplication for the fuzzy intersection operator in order to define the variable δ for the fuzzy case:

$$\bar{\delta}_t(i) = \max_{q_1,q_2,\ldots,q_{t-1}} \left\{ \bar{\pi}_{q_1} b_{q_1} \prod_{\tau=2}^{t} \left[\bar{\alpha}_{q_{\tau-1}q_\tau} \rho_\tau(q_{\tau-1}, q_\tau) \right] \bar{b}_{q_\tau}(O_\tau) \right\} \quad , \tag{6}$$

where $\rho_t(i,j) = [\mu_i^n(t,j) - \mu_{i+1}^n(t,j)] / \bar{\alpha}_t(i)$. $\bar{\delta}_t(i)$ is the degree of certainty for a single state sequence finishing at time *t* in a state S_i.

Similarly, for training the fuzzy HMM model, the fuzzy version of the EM algorithm can be derived, again by using the fuzzy coefficient that multiplies the state transition coefficients and summing up using the Choquet integral.

Using fuzzy profile HMMs in phylogenetic analysis

After obtaining a multiple sequence alignment using the fuzzy Viterbi algorithm, the methodology of performing and evaluating phylogenetic analysis is depicted in Figure 2.

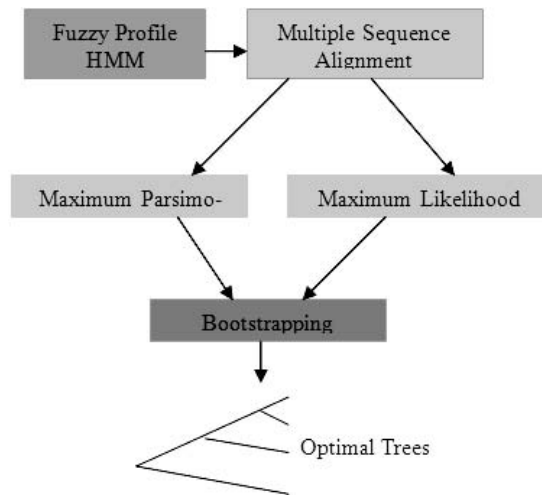


Figure 2 - Phylogenetic analysis using fuzzy HMM alignments

First, alignments are used to compute phylogenetic trees with the maximum parsimony and maximum likelihood methods.

The maximum parsimony method constructs trees on the basis of the minimum number of amino acid changes required to account for the data. This often results in the generation of hundreds of equally most parsimonious trees, making it difficult to justify the choice of a particular tree.

Similarly, for the maximum likelihood method, testing all possible trees is impossible, and it is also not computationally feasible to estimate the model for each tree. Therefore, the accepted strategy is to infer a "reasonable" tree topology with faster – although less reliable – reconstruction methods and use that tree to estimate the parameters.

By implementing a bootstrap analysis [22] of the produced trees though, it is possible to acquire a measure for the reliability of the alignments. Bootstrapping can approximate

the underline distribution by resampling from the original dataset and constructing a majority-rule consensus tree. Bootstrapping values can then be used as a confidence measure for the quality assessment of the alignments.

In the conducted experiments the abovementioned phylogenetic analysis schema was used for both classical profile HMM and fuzzy profile HMM alignments, and the confidence levels of the bootstrapping procedure were measured.

Results

In order to evaluate the performance of the fuzzy profile HMM alignments we have performed several comparative tests against the classical approach. The dataset was randomly divided in a 90%-10% manner into training and test set for the HMM models. Specifically, 97 sequences were used for training the model, while 11 sequences were used to create the multiple sequence alignment and consequently construct their phylogenetic tree. The test sequences came equally from the E6 protein HPV family and the homologous proteins obtained from PSI-BLAST. The classical profile HMM was then trained in 20 cycles, while the fuzzy profile HMM required 12 cycles for training. Alignments were obtained using the fuzzy Viterbi algorithm. Part of the resulting alignments yielding from the fuzzy profile HMM is depicted in Figure 3.

```
M-------L-M--L-A----M-M----E---LH-H----G----------|
D-VlIRC-Y-L-CH-KPL-C-EV----E--KVK-HIL-T-KAR-------
V-L----P-K---M-K---C-P------------HRL-E-----------
A-IsIDQ-KTT-SR-NPR-ST-VgtvtE-----iH--D-YLR--------
D-LsIRC-M-C-CL-KPL-S-PA----E--KLR-H-LN-SKRR-------
--------M-E-KL-QPK-T-I--------KN----C-S-QD--------
ElIfVGE-I-S-SS-KPA-I-SN------------IL--M-RCspgsssk
D-VlIRC-Y-L-CH-KPL-C-HV----E--KVR-HILD--KAR-------
Q-VrIRC-C-K-CH-KPL-S-PV----E--KTN-HIV--KKTQ-------
Q-LfMRC-Y-I-CH-KPL-S-WE----E--KEA-LLV-GNKR--------
--------YGD-YQ-SEYyG-P------gg-------N-YDFTpqgppss
```

Figure 3 - Part of alignment using fuzzy Profile HMM

In consequence, phylogenetic trees were inferred using the maximum likelihood (ML) and the maximum parsimony (MP) methods. Bootstrapping was also applied for each case. Table 1 illustrates log likelihood scores for the two cases of ML applications, while also contains the MP scores for each case. Optimal trees for the bootstrapped cases can be observed in Figures 4 and 5.

The fuzzy profile HMM was implemented in Java, partially using the BioJava API, while phylogenetic analysis was performed with the PHYLIP package [23].

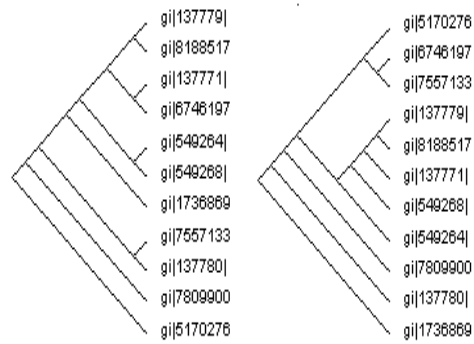


Figure 4 - Best trees for fuzzy and simple HMM by maximum likelihood method as calculated by the Phylip Package

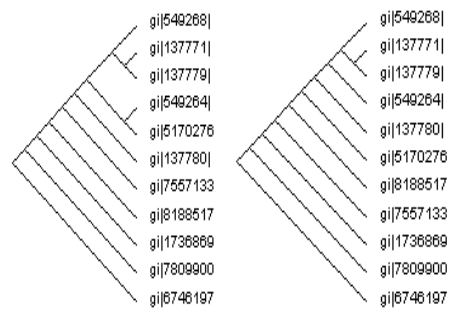


Figure 5 - Best trees for fuzzy and simple HMM by Maximum Parsimony method as calculated by the Phylip Package

Table 1 - Phylogenetic evaluation scores for different models and algorithms

Methods	HMM	FuzzyHMM
ML Simple	-13298.89	-13317.902
MP Simple	8507	8067
ML with Bootstrap	-13671.514	-13633.146
MP with Bootstrap	8609	8347

Discussion

By inspecting the inferred trees it is possible to observe that both models can discriminate between the E6 proteins and the rest of the homologous sequences. Concerning the evaluation scores, in the maximum-likelihood case, the tree topology, as well as the evaluation scores is similar for the two alignments in both simple and consensus trees. The maximum parsimony method for the fuzzy case yields models with lower values for both simple and bootstrapping methods, compared to the simple HMM alignment. This means that the fuzzy HMM can produce more parsimonious phylogenetic trees, a property that seems to come naturally from the non-independence assumption of the fuzzy profile HMM. Another issue worth pointing is the ability of the fuzzy HMM trainer to converge in less

cycles, thus building the model faster. In this sense, the computational cost for performing the training stage in the fuzzy case is reduced without sacrificing the quality of the alignments. Instead, the relaxation of the statistical independence assumption provides enhanced biological meaning to the alignments.

Conclusions

In this paper we presented a new methodology of phylogenetic tree inference that makes use of the fuzzy profile HMM for multiple sequence aligning. The fuzzy approach relaxes the independence restriction implied in classical profile HMMs, thus providing more biologically meaningful alignments. In terms of phylogenetic analysis this implies the constructions of more parsimonious trees in comparison with the classic HMM approach. Finally, we have shown this property by experimenting with HPV virus protein data. Future work involves the application of fuzzy HMM alignments in the creation of a whole new series of profiles that can then be used in protein classification.

Acknowledgments

This work was supported by the EPEAEK II project: "IRAKLEITOS: Scholarships targeted in basic research. Phylogenetic analysis of biological data".

References

[1] Fitch WM. Toward defining the course of evolution: Minimum change for a specific tree topology. Systematic Zoology 1971; 20:406-416.

[2] Felsenstein J. Evolutionary trees from DNA sequences: A maximum-likelihood approach. Journal of Molecular Evolution 1981; 17:368-376.

[3] Page RDM, Holmes EC. Molecular Evolution: A Phylogenetic Approach. Oxford: Blackwell Science, 1998.

[4] Salemi M, Vandamme AM, editors. The Phylogenetic Handbook. Cambridge University Press; 2003.

[5] Thompson D, Higgins DG, Gibson TJ. Clustal W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 1994; 22:4673-80.

[6] Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Oxford d University Press Nucleic Acids Research 1997; 25(17):3389–3402.

[7] Eddy SR. Profile Hidden Markov Models. Bioinformatics 1998; 14:755-763.

[8] Pearson WR. Empirical statistical estimates for sequence similarity searches. In J. Mol. Biol. 1998; 276:71-84.

[9] Bateman A, Coin L, Durbin R, Finn RD, Hollich V, Griffiths-Jones S, Khanna A, Marshall M, Moxon S, Sonnhammer ELL, Studholme DJ, Yeats C, Eddy SR. The Pfam protein families database. Oxford University Press, Nucleic Acids research 2002; 30(1):276-280.

[10] Mulder NJ. InterPro. In: Encyclopedia of Genetics, Genomics, Proteomics and Bioinformatics, Jorde LB, Little PFR, Dunn MJ, Subramanian S, editors, John Wiley & Sons, 2435-2439; 2005.

[11] Diplaris S, Tsoumakas G, Mitkas PA, Vlahavas I. Protein classification with multiple algorithms. In Proceedings of the Tenth Panhellenic Conference on Informatics 2005, Volos, Greece. Bozanis P, Houstis EN (Eds), Springer, Lecture Notes in Computer Science; 3746:448-456.

[12] Collyda C, Maglaveras N, Pappas C. A Multiple Sequence Alignment Tool Using Hidden Markov Models For Phylogenetic Analysis. In Proceedings of the 4th European Symposium on Biomedical Engineering 2004; Patra, Greece.

[13] Tran D, Wagner M. Generalised Fuzzy Hidden Markov Models for Speech Recognition. In Proceedings of the International Conference on Fuzzy Systems 2002; Calcutta, India.

[14] Collyda C, Diplaris S, Mitkas PA, Maglaveras N, Pappas C. Fuzzy Hidden Markov Models: A New Approach In Multiple Sequence Alignment. 20th International Congress of the European Federation for Medical Informatics (MIE 2006), Maastricht, Netherlands. A. Hasman et al. (Eds.), IOS Press, Studies and Health Technology in Informatics; 124:99-104.

[15] Bidargaddi NP, Chetty M, Kamruzzaman J. Fuzzy Profile Hidden Markov Models for Protein Sequence Analysis. Proceedings of the IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology 2005; San Diego, CA, USA.

[16] Narukawa Y, Murofushi T, Sugeno T. Conditions for choquet integral representations. Proceedings of the Eighth International Fuzzy Systems Association World Congress 1999; 911–925.

[17] Krogh A, Brown M, Mian IS, Sjolander K, Haussler D. Hidden Markov models in computational biology: Applications to protein modeling. Journal of Molecular Biology 1994; 235:1501-1531.

[18] Rabiner A. A tutorial on Hidden Markov Models and selected applications in speech recognition. Proceedings of the IEEE 1989; 77(2):257-285.

[19] Wang Z, Klir G. Fuzzy measure theory. Plenum Press; 1992.

[20] Grabisch M, Murofushi T, Sugeno M, editors. Fuzzy measures and integrals : theory and applications. New York: Springer; 2000.

[21] Mohamed M, Gader P. Generalized hidden markov models. i. theoretical frameworks. IEEE Transactions on Fuzzy Systems 2000; 8(1):67–81.

[22] Felsenstein J. Numerical methods for inferring evolutionary trees. Systematic Zoology 1982; 27:783-791.

[23] Felsenstein J. PHYLIP: Phylogenetic Inference Package, Version 3.65. Seattle: Department of Genetics, University of Washington, 1993.

Address for correspondence

Lab of Medical Informatics, Faculty of Medicine, Aristotle University, GR54124, Thessaloniki, Greece

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Determining Transcription Factor Activity from Microarray Data using Bayesian Markov Chain Monte Carlo Sampling

Andrew V. Kossenkov[a], Aidan J. Peterson[b], Michael F. Ochs[c]

[a] *The Wistar Institute, Philadelphia, PA, USA*
[b] *Howard Hughes Medical Institute at the University of Minnesota, Minneapolis, MN, USA*
[c] *The Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins University, Baltimore, MD, USA*

Abstract

Many biological processes rely on remodeling of the transcriptional response of cells through activation of transcription factors. Although determination of the activity level of transcription factors from microarray data can provide insight into developmental and disease processes, it requires careful analysis because of the multiple regulation of genes. We present a novel approach that handles both the assignment of genes to multiple patterns, as required by multiple regulation, and the linking of genes in prior probability distributions according to their known transcriptional regulators. We demonstrate the power of this approach in simulations and by application to yeast cell cycle and deletion mutant data. The results of simulations in the presence of increasing noise showed improved recovery of patterns in terms of χ^2 fit. Analysis of the yeast data led to improved inference of biologically meaningful groups in comparison to other techniques, as demonstrated with ROC analysis. The new algorithm provides an approach for estimating the levels of transcription factor activity from microarray data, and therefore provides insights into biological response.

Keywords:

Microarray analysis, Bayesian analysis, transcription factors

Introduction

The regulation of gene expression is a primary form of response in all cellular systems. This response is typically mediated by activation of transcription factors or complexes (both referred to here as TFs) that can induce or repress transcription of sets of genes by binding to upstream elements known as promoters. Identification of the activity level of individual TFs provides insight into biological processes activated or deactivated in specific samples. For instance, identification of upregulation of the ELK–1 TF can indicate activation of the oncogenic RAS-RAF pathway in a tumor cell, which is difficult to measure directly.

Microarrays provide insight into the global transcriptional response of cells, which should be useful for identification of TF activity. Although early studies using microarrays focused on "guilt-by-association" identification of genes that may function similarly to known genes [1] or biomarkers of disease [2], efforts have been made to use microarrays to link to transcription factor promoter sites [3].

These methods overlook a key aspect of transcriptional regulation, as they rely on clustering of genes into groups with each gene belonging to a single group. However, most, if not all, genes are likely to be multiply regulated, as evolution has been very effective in borrowing function by using existing genes in new roles. Even among genes regulated in the yeast cell cycle, only roughly 10% are associated with a single cell cycle phase [4]. This results in the identification of a large number of clusters with genes improperly grouped away from those involved in the same biological processes (e.g., in analyses of the yeast cell cycle data, a typical cluster analysis shows more than 20 clusters). This results in a significant loss of power for identification of TF activity.

Although our knowledge of transcriptional regulation is growing rapidly, in a recent study using Agilent human arrays, only ~1000 genes out of 20,000 were annotated with high reliability as to their TFs using TRANSFAC professional [5]. In order to recover the signal of TF activity, it is therefore highly desirable to maximize the signal by correctly grouping genes into multiple overlapping groups.

The problem of multiple regulation was identified reasonably early in the development of microarray technology. The application of singular value decomposition (SVD) to microarray data [6] addressed multiple regulation, however the orthogonality constraints led to less than ideal results, since biological processes are not independent. We took a different approach, applying our Bayesian Decomposition algorithm to microarray analysis [7]. This algorithm, described below, uses a series of constraints and a structure minimization argument to identify overlapping sets of genes.

A significant advantage to our approach is the ability to encode biological knowledge through prior probability distributions. In this work, we demonstrate how knowledge of coregulation through TFs can be encoded into the algorithm, leading to improved statistical power for the determination of the activity levels of biological processes.

Methods

In order to recover signatures of transcription factor activity, the analysis of microarray data will need to isolate patterns related to a biological process governed by a transcription factor, identify the genes associated with this pattern in the background of multiple regulation, and link these genes to transcriptional regulators. We will describe the Bayesian Decomposition (BD) algorithm in general, then show specifically the modifications that allow direct inference of transcriptional regulation to improve statistical power.

Bayesian decomposition

The fundamental factorization needed to identify overlapping groups of coexpressed genes is the recovery of a distribution matrix (**A**) and a pattern matrix (**P**) that multiply together to form a mock ("fitted") data matrix (**M**), which reproduces the data matrix (**D**) within the noise (). This relationship can be written as

$$\mathbf{D} = \mathbf{M} + \varepsilon = \mathbf{AP} + \varepsilon \,. \qquad (1)$$

For microarray data, **D** is generated from replicated experiments and therefore represents the best estimate for the expression of each gene in each condition. The computed matrices then provide the assignment of genes to patterns, **A**, and the assignment of conditions to patterns, **P**, as shown in Figure 1 for a hypothetical analysis of the cell cycle. In this example, the data is approximated by the multiplication of **A** and **P**, so that a gene (*N*) with complex behavior (transcribed strongly in G1 and weakly in G2), can have that behavior explained as a mixture of simpler behaviors (G1 and G2).

The factorization of **D** into **A** and **P** is generic, and as noted above, approaches using orthogonality criteria have been used. However, biological patterns will not be orthogonal, as this would imply independence. In fact, SVD applied to cell cycle data does not even readily recover phase signatures [6], while BD recovers signatures for the cell cycle phases as well as a signature for the entrained metabolic oscillator [7].

Probability distributions

BD implements a Markov chain Monte Carlo (MCMC) approach in order to solve Equation 1. The Markov chain uses a Gibbs sampler requiring relative probability measures between points in the distribution of possible solutions. These are provided according to Bayes' Equation,

$$p(\mathbf{A}, \mathbf{P}|\mathbf{D}) \mu\, p(\mathbf{D}|\mathbf{A}, \mathbf{P})\, p(\mathbf{A}, \mathbf{P}). \qquad (2)$$

The posterior probability, $p(\mathbf{A},\mathbf{P}|\mathbf{D})$, describes the probability of a model (**A** and **P**) given the data, and it is the distribution sampled by MCMC. The prior, $p(\mathbf{A},\mathbf{P})$, provides the probability of the model independent of the data. A simple example is that a model with negative copies of mRNA can be ruled out *a priori* and has zero prior probability. The likelihood, $p(\mathbf{D}|\mathbf{A},\mathbf{P})$, gives the probability that the data comes from the model, and this probability is related to the χ^2 distribution.

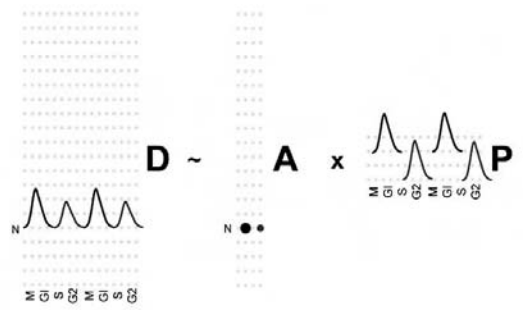


Figure 1 - The decomposition performed by BD

The prior encoded by the algorithm has three important features: positivity, correlated structure, and minimization. Positivity is incorporated by the inclusion of a one-dimensional atomic domain for each matrix in the model (**A** and **P**). In this domain, point masses (i.e., atoms) are created according to a prior distribution that is exponential in amplitude and uniform in location. The creation or destruction of atoms does not modify the prior distribution, which allows an approach to minimization of structure, as discussed below.

Correlations are introduced between points in the model (**A** and **P** matrices) by the mapping of atoms from the atomic domain to the matrices [8]. The mappings use kernel functions to spread the amplitude of each atom to one or more matrix elements, allowing linking of samples through correlations in **P** and linking of genes through correlations in **A**. We have used correlated structure in **P** to perform supervised learning [9], and here use correlations in **A** to provide prior information on TF regulation.

We use "birth and death" MCMC techniques for creating and destroying atoms. Since the prior distribution is unaffected by these actions, atoms can be eliminated readily, as long as the fit to the data is not adversely affected. This, coupled to internal mechanisms of amplitude exchange, leads to a minimization of structure (i.e., number of atoms). In the simplest application, this matches sparse matrix approaches.

Markov chain Monte Carlo sampling

The Markov chain begins with empty atomic domains, and thus empty **A** and **P** matrices. The algorithm attempts to birth atoms (created *ex vacuo*), move or exchange amplitude between atoms, and remove atoms in separate MCMC steps. Atoms are created according to the prior distribution and mapped through the kernel functions to the **A** and **P** matrices. The log likelihood is

$$\log L = \sum_i \sum_j \left\{ \frac{1}{2\sigma_{ij}} (D_{ij} - \sum A_{ip} P_{pj}) \right\}, \qquad (3)$$

so that changes in the likelihood can be easily calculated for any change in the matrices. The algorithm calculates this change in such a way as to allow resampling of amplitude to increase the speed of exploration of the posterior distribution.

After an equilibration period determined by the user, sampling of the distribution is done by recording atoms and mapping them to the **A** and **P** matrices. Statistical measures, such as the mean and standard deviation for each matrix element, can be calculated. The rows of the inferred **P** matrix (patterns) link conditions, while the columns of the inferred **A** matrix assign genes to each pattern. Convergence is checked by insuring that the χ^2 fit to the data is stable, and that multiple chains starting at random points reach the same solution.

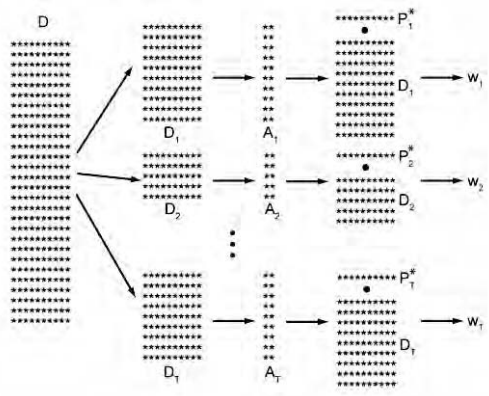


Figure 2 – Calculating weights for TF analysis

Estimating transcription factor activity

Previously we estimated TF activity without using prior information successfully for *S. cerevisiae*. We applied the original BD and subsequently linked genes associated with patterns to TFs [10]. However, for mammalian systems, a more statistically powerful approach will be required. The approach we propose here is to encode the knowledge of transcriptional regulation *a priori* during Markov chain sampling, thus borrowing power across genes.

Normalizing expression for each transcription factor

In order to map an atom to multiple matrix elements, it is necessary to determine the amount of the amplitude that should be devoted to each gene. This mapping cannot be made uniform across the matrix elements, as the overall copy number of mRNA produced will vary by gene and by biological process.

We address this issue by implementing a preprocessing step to determine weight vectors for each gene linked through a TF. The overall approach is shown in Figure 2. The data, **D**, are divided into T overlapping subsets, with each subset containing all genes regulated *a priori* by a given TF. Each subset, $\mathbf{D_t}$, is analyzed using the original BD without *a priori* correlations and positing $P+1$ patterns, where P is the total number of groups in T that contain any gene in $\mathbf{D_t}$. This provides for a pattern for each TF and a pattern for routine metabolic function, which BD typically isolates separately. The rows of the $\mathbf{A_t}$ matrices are normalized to unit amplitude, and the column from each with the lowest variance is taken to represent the pattern related to that TF. The dot product of the corresponding pattern, $\mathbf{P_t}^*$, with each row of $\mathbf{D_t}$ provides the weights, $\mathbf{w_t}$, for spreading an atom linked to this TF into the **A** matrix.

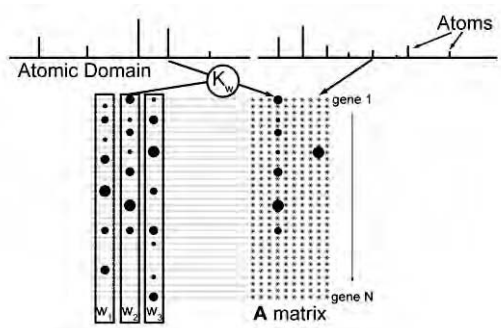


Figure 3 - Atomic domains and prior information

Sampling with prior information

We encoded TF information within BD by dividing the atomic domain related to the **A** matrix into two subdomains (see Figure 3). Atoms in the left half of the atomic domain spread to multiple elements in **A** through kernel functions, K_w, using the weight vectors, $\mathbf{w_t}$. The location of the atom determines the column of **A** and the weight vector $\mathbf{w_t}$ that are used. An atom in the right half maps to a single element, allowing prior information not supported by the data to be ignored.

As with BD, the new algorithm, BD-TF, starts with empty **A** and **P** matrices. As atoms are created, moved, or destroyed, changes are mapped to the matrices according to the scheme shown in Figure 3. If there is correlated structure in the data due to TF activity, a single atom will successfully recover this, and statistical power will be gained by use of K_w. Since the prior is defined on the atomic domain, the form is unchanged from the original BD algorithm. The calculated change in the log likelihood is affected by the correlation functions, K_w, but the likelihood function in Equation 3 includes summation across all matrix elements, so no change of form is needed. The equilibration and sampling proceed as in the original BD.

Analysis

We have analyzed three separate sets of data and compared our BD-TF results to BD and to some standard techniques. We first used simulations of the yeast cell cycle, which allowed us to increase noise levels to determine the behavior of the algorithm across many levels of noise. Second, we analyzed the widely studied yeast cell cycle data set [1], permitting comparison with other methods. Third, we analyzed the Rosetta compendium of yeast deletion mutants, which included error estimates for all data points [11]. An open problem in the field is the correct dimensionality for analysis (i.e., number of patterns). We have estimated this previously by multiple methods for the data used here, while, for the simulated data, we use the simulated dimensionality.

For the simulated yeast cell cycle data, we created expression levels for 288 genes at 48 time points over two cycles. The **P** matrix comprised 5 overlapping patterns - four that reflect cell cycle phases over two periods and one representing a metabolic oscillator with an amplitude 5% as large as the cell cycle pattern amplitudes and with twice the frequency [7]. The **A** matrix assigned genes to patterns in the expression profile, with most genes assigned to 2 – 4 patterns, reflecting the fact that in yeast only ~10% of cell cycle genes belong to a single phase [4]. Noise was added to the data matrix, including different levels of additive and multiplicative noise, using the widely accepted noise model [12]

$$D = N\left(0, \sigma_a\right) + e^{N(0,\sigma_m)} A_S P_S , \qquad (4)$$

where $\mathbf{A}_S$ and $\mathbf{P}_S$ are simulated amplitude and pattern matrices, and σ_a and σ_m are additive and multiplicative levels of noise respectively. Simulated data matrices with 154 different noise levels were created, varying σ_a from 0 to 6.5 and σ_m from 0 to 3. The data matrix without noise had maximum amplitude 3.15 and mean amplitude 0.65. We simulated 4 replicate arrays and calculated mean and standard deviations for each simulated element in each **D**.

The Cho data was analyzed using BD, as described previously [7], and using BD-TF. Groups of coregulated yeast genes were assembled based on literature reports of regulatory relationships between TFs and target genes, and only coregulation groups with at least five members were retained. In addition, we sought to enrich for true TF-target relationships by including only genes supported by evidence both of TF binding and of alteration in transcription when the TF was absent or overexpressed. For the cell cycle analysis, the regulators Mot3, Ndt80, Ste12, Swi5, Cbf1, Fkh1, Fkh2, Swi6, Mcm1, Swi4, and Rlm1 provided prior information with 5 – 16 target genes each.

In our previous work [7], we established that the best factorization used 6 dimensions (5 cell cycle phases, due to early and late G1 signatures as seen previously [4], and a metabolic oscillator signature). An ROC analysis was performed for BD by increasing the stringency of assignment of a gene to a pattern. Essentially, each gene had a mean value of its strength within a pattern and an uncertainty on that assignment from **A** based on MCMC sampling. By increasing the number of standard deviations away from zero required to assign a gene to a group, multiple estimates of the assignment of the genes to the patterns were made, allowing the ROC curve to be constructed. The gold standard for the analysis was based on the known molecular biology of gene coregulation independent of microarray studies, and this comprised 9 groups with 43 genes total [13]. The results were compared between hierarchical clustering, the original BD, and BD-TF.

The Rosetta compendium data was previously analyzed using BD at multiple dimensions, and consistency analysis determined that 15 dimensions were optimal [10]. The data was reanalyzed at 15 dimensions using BD-TF as for the yeast cell cycle data, except we included targets whose regulation by a TF was supported by a single type of experimental evidence. The set included the regulators Zap1, Ndt80, Mcm1, Gcn4, Dal80, Rtg1, Pdr1, Met4, Ume6, Ste12, Mot3, Gln3, Cbf1, Mig1, Rlm1, Msn4, and Msn2 with 5 – 19 target genes each. ROC analysis looked for recovery of the correlated groups. While this is somewhat circular, it indicated whether the prior information was being used appropriately by the algorithm. Unfortunately, the genes with known coregulation from cell cycle studies [13] were not in this data set, as they do not vary across the deletion mutants.

Results

We present our results in three sections: 1) simulations of cell cycle data, 2) analysis of the Cho data with ROC analysis, and 3) analysis of the Rosetta data with ROC analysis.

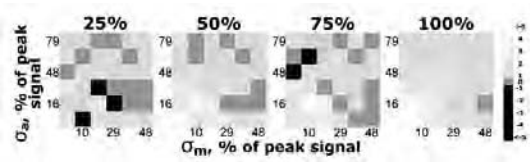


Figure 4 – χ^2 fits for simulation data

Simulations

In Figure 4, we show the $\log_2$ ratios of the χ^2 fits to the $\mathbf{A}_S$ matrix between BD-TF and BD. The heatmaps show the differences in the fits across different levels of noise and inclusion of prior information. In the figures, multiplicative noise increases to the right, from 0 to 48% of the peak signal; additive noise to the top, from 0 to 79% of peak signal; and the amount of prior information included (number of genes included from coregulation lists) increases as indicated by the percentages shown above each heatmap. Only the 36 simulations with the lowest noise levels are shown, as these cover levels of noise exceeding those in typical array experiments. Gray squares represent a neutral result (no improvement), while lighter squares represent improved fits with BD-TF and darker squares poorer fits. The advantage of using coregulation information increases as levels of prior information increase, as would be expected if the coregulation information is improving statistical inference by gaining power across genes.

Yeast cell cycle

In Figure 5 on the left, we present results of the application of BD-TF to the cell cycle data using ROC analysis based on known coregulation [13]. We compared the results using original BD (circles), BD-TF (squares), and shrinkage-based hierarchical clustering (triangles; performed previously [13]). BD-TF obtained an area under the curve of 0.82, compared with 0.83 for BD and 0.56 for hierarchical clustering. The lack of improvement from use of coregulation information reflected the lack of such data here, as we had TF data on only 67 of 788 genes, which was not adequate to improve inference over BD. However, we include the results to show the value in BD and BD-TF

that arises solely from the proper assignment of genes to multiple groups.

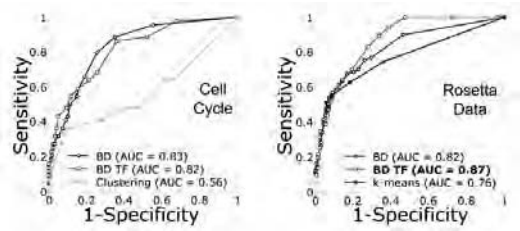


Figure 5 – ROC Curves for Cell Cycle and Rosetta Data

Rosetta compendium

The analysis of data from the Rosetta compendium data set using BD-TF (squares) was compared to K-means clustering (triangles) and the original BD analysis (circles) of the same data (Figure 5 right). Here, both the coregulation information and the "gold" standard gene lists were the same, so that the results demonstrated that the algorithm correctly used information about TF regulation and that such coregulation was reflected in the data. All techniques performed equally well at high specificity, however as sensitivity increased, BD-TF was superior due to a reduction of false positives with the inclusion of prior information on TF regulation.

Conclusion

This work demonstrates 1) the value of inclusion of prior knowledge on transcriptional regulation in the analysis of microarray data, and 2) the present limits of that knowledge. While the simulations showed a clear advantage in using this knowledge, especially at typical noise levels, the analysis of cell cycle data indicated that more prior information would be helpful. Nevertheless, the superiority of the BD-TF approach over clustering for microarray analysis is clear. Our knowledge of transcriptional regulation is rapidly increasing, and we expect improved statistical power with BD-TF over the next few years. This power will be critical to improved inference of biological process activity, especially with heterogeneous and limited samples typical in clinical settings. These samples introduce noise to an analysis focused on understanding biological response, such as in therapeutic interventions, and techniques to gain statistical power through use of existing biological knowledge will be critical to make progress.

Acknowledgements

This work was supported by NIH: LM008309 to mfo and CA06973 to mfo through the Sidney Kimmel Cancer Center.

References

[1] Cho RJ, Campbell MJ, Winzeler EA, Steinmetz L, Conway A, Wodicka L, et al. A genome-wide transcriptional analysis of the mitotic cell cycle. Mol Cell. 1998;2(1):65-73.

[2] Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature. 2000;403(6769):503-11.

[3] Vadigepalli R, Chakravarthula P, Zak DE, Schwaber JS, Gonye GE. PAINT: a promoter analysis and interaction network generation tool for gene regulatory network identification. Omics. 2003 Fall;7(3):235-52.

[4] Lukashin AV, Fuchs R. Analysis of temporal gene expression profiles: clustering by simulated annealing and determining the optimal number of clusters. Bioinformatics. 2001;17(5):405-14.

[5] Wingender E. TRANSFAC, TRANSPATH and CYTOMER as starting points for an ontology of regulatory networks. In silico biology. 2004;4(1):55-61.

[6] Alter O, Brown PO, Botstein D. Singular value decomposition for genome-wide expression data processing and modeling. Proc Natl Acad Sci U S A. 2000;97(18):10101-6.

[7] Moloshok TD, Klevecz RR, Grant JD, Manion FJ, Speier WFt, Ochs MF. Application of Bayesian Decomposition for analysing microarray data. Bioinformatics. 2002 Apr;18(4):566-75.

[8] Sibisi S, Skilling J. Prior distributions on measure space. Journal of the Royal Statistical Society, B. 1997;59(1): 217 - 35.

[9] Ochs MF, Moloshok TD, Bidaut G, Toby G. Bayesian Decomposition: Analyzing microarray data within a biological context. Annals of the New York Academy of Sciences. 2004;1020:212 - 26.

[10] Bidaut G, Suhre K, Claverie JM, Ochs MF. Determination of strongly overlapping signaling activity from microarray data. BMC Bioinformatics. 2006 Feb 28;7(1):99.

[11] Hughes TR, Marton MJ, Jones AR, Roberts CJ, Stoughton R, Armour CD, et al. Functional discovery via a compendium of expression profiles. Cell. 2000;102(1): 109-26.

[12] Rocke DM, Durbin B. A model for measurement error for gene expression arrays. J Comput Biol. 2001;8:557 - 69.

[13] Cherepinsky V, Feng J, Rejali M, Mishra B. Shrinkage-based similarity metric for cluster analysis of microarray data. Proc Natl Acad Sci U S A. 2003 Aug 19;100(17): 9668-73.

Address for correspondence

Michael F. Ochs,
Oncology Biostatistics and Bioinformatics,
The Sidney Kimmel Comprehensive Cancer Center
at Johns Hopkins, 550 N.
Broadway, Suite 1103,
Baltimore, MD 21205, USA
Email: mfo@jhu.edu

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Ensemble Stump Classifiers and Gene Expression Signatures in Lung Cancer

Lewis Frey[a1], Mary Edgerton[b], Douglas Fisher[c] and Shawn Levy[d]

[a] *Biomedical Informatics Department, University of Utah, USA*
[b] *Department of Anatomic Pathology, M.D. Anderson Cancer Center, USA*
[c] *Department of Electrical Engineering and Computer Science, Vanderbilt University, USA*
[d] *Department Biomedical Informatics, Vanderbilt University, USA*

Abstract

Microarray data sets for cancer tumor tissue generally have very few samples, each sample having thousands of probes (i.e., continuous variables). The sparsity of samples makes it difficult for machine learning techniques to discover probes relevant to the classification of tumor tissue. By combining data from different platforms (i.e., data sources), data sparsity is reduced, but this typically requires normalizing data from the different platforms, which can be non-trivial. This paper proposes a variant on the idea of ensemble learners to circumvent the need for normalization. To facilitate comprehension we build ensembles of very simple classifiers known as decision stumps – decision trees of one test each. The Ensemble Stump Classifier (ESC) identifies an mRNA signature having three probes and high accuracy for distinguishing between adenocarcinoma and squamous cell carcinoma of the lung across four data sets. In terms of accuracy, ESC outperforms a decision tree classifier on all four data sets, outperforms ensemble decision trees on three data sets, and simple stump classifiers on two data sets.

Keywords:

microarray, decision trees, ensembles, stumps

Introduction

Methods for finding robust mRNA signatures of cancer that remain consistent across experiments and microarray platforms (i.e., oligonucleotide and cDNA) have remained elusive in the bioinformatics literature. From a machine learning perspective this is expected, since many microarray data sets have a scarcity of sample (e.g., a few hundred); moreover, each sample has thousands of probes (i.e., continuous variables) resulting in a very pronounced curse of dimensionality. With thousands of variables from which to choose, the constructed classifier can overfit the specific data and cannot generalize to other data sets. It poses a challenge when applying machine learning techniques to discover a set of relevant probes that constitute a robust mRNA signature for the cancer. There are a number of papers describing the pitfalls of overfitting expression data and the failure of some classification models to do better than chance [1,2,3].

By combining data from different platforms, problems of data sparsity and overfitting can be mitigated. The microarray data sets available in repositories are growing at a rapid rate. The creation of data sharing initiatives such as oncomine.org and the Cancer Biomedical Informatics Grid (caBIG™) enable the combination of multiple data sets to find better classifiers.

However, combining data across platforms is challenging. First of all, there are multiple microarray platforms and these can differ in the types of probes arrayed (i.e. variables measured) for the specimens. A method for mapping the probes across platforms is first required to relate the results. The different means of getting expression levels from the platforms need to be consistently compared, which often requires normalizing the expression levels between platforms.

A methodology that supports the generation of classifiers that find easily interpretable, robust mRNA signatures of cancer that *generalize across experiments and platforms* is needed. By robust mRNA signatures what is meant is a set of probes that are consistently associates with a cancer type. This paper describes a novel method for combining data sets to discover classifiers that use robust mRNA signatures which generalize across experiments and platforms. This is done by using a classifier to focus on a limited number of predictive probes in the data that persist across data sets. *The result of our approach is a distinct classifier for each data set under investigation, but importantly the construction of each such classifier is informed by all the available data.*

Empirical results by Holte [4] showed that simple, single test classifier trees, referred to as *Stumps*, can be surprisingly close in accuracy to more complex decision tree classifiers[5] in many of the domains tested. Results on data sets used in machine learning have shown that combining multiple classifiers boosts classification accuracy by creating variance among the constituent classifiers. Our approach, Ensemble Stump Classifier (ESC), is a kind of subspace sampling. It combines different probes that have slightly dissimilar classification of the sample using a majority vote to boost accuracy due to the variance in the sample they accurately classify. Using Stump classifiers as the "base" classifier of the ensemble supports the creation of simple ensemble classifiers.

1 L.F. was supported in part by a fellowship from the National Library of Medicine (5T15LM007450). M.F. and D.F. were supported in part by a National Institute of Health grant (1R01LM008000) to M.E.

Materials and methods

Four classifier-methods -- decision trees (C5.0), ensemble decision trees, Stump rule and ESC -- are applied to the data sets. Ten-fold cross validation is used to derive the accuracy measure for the data sets. Each data set is divided into 10 mutually-exclusive folds, consisting of 10% of the data. One of the 10-folds is held out as a testing set and the remaining 90% is used as the training set. Then a different fold is selected as the testing set and the remaining 90% serves as the training set. The average testing accuracy is calculated across all 10-folds.

C5.0, ensemble C5.0 and stumps

C5.0 is a commercial implementation of Quinlan's decision tree approach [5]. The ensemble decision tree classifier is computed using C5.0's implementation of the boosting approach of Freund and Schapire [6]. Stumps are single variable-test decision trees. Each Stump is made up of one probe at the root and has the best threshold for the training set that separates the two classes. Stumps are obtained by running C5.0 rules on a single data set with pruning set to a 1% confidence level with only single variable rules allowed.

Ensemble stump classifier (ESC)

The ESC is a new classifier approach that bypasses the need to normalize across multiple data sets. The base classifier for the ensemble is the Stump classifier. *The outcome of our method is a separate classifier for each data set, but notably each classifier is instructed by all the available data.*

Basic representations

ESC's form of ensemble learning is different than boosting. The idea that learning is occurring over multiple data sets is key. To illustrate, consider three artificially constructed objects, one from each of three data sets (see Table 1a).

Probes P_{1j} (i.e., a *matching probe set*) represent "synonymous" probes across the three data sets. The three probes -- P_{11}, P_{12}, and P_{13} -- can be thought of as the same probe with different scales and labels. The probe P_{11} has an expression value for the probe in match Probe Set 1 and Data Set 1.

The ESC algorithm repeatedly finds the "probe" (though with different labels and scales) that is "collectively" best over all data sets. This step leads to one decision stump per data set, each of which uses the same decision variable/probe. For example, if probe P_1 were found to be "collectively best" over three data sets, then the three decision stumps for each data set would have the form expressed in Table 1b.

The threshold for match Probe Set 1 in Data Set 1 is equal to 0.7 in Table 1b. Given an expression value for probe P_{11} equal to or below 0.7 the Stump rule predicts Class 1, otherwise Class 2. For Probe Set 1 and Data Set 2 the threshold is different than Data Set 1 (compare threshold 0.7 for P_{11} and threshold 10 for P_{12}). Even though the probes are the same, the thresholds can be different between data sets. This occurs because the data sets are not normalized to each other. The threshold is only consistent within a data set.

Building an ESC

The following are the steps to generate a full ensemble stump classifier (as in Table 1b):

1. Determine which probes match each other and construct rows based on the matching (i.e., "synonymous") probes across the data sets. This was briefly described in the basic representations section, and is described in more detail in the match probes section.
2. Partition each of the three data sets into folds so that there are training and testing sets. In our experimental design we use 10-fold cross validation.
3. Learn the stump rules, one for each training set and matching probe set, from training folds. This was illustrated in the basic representations section (e.g., each P_{ij} in Table 1b), and follows the same process given for stump rules.
4. Reorder the stumps across the data sets based on the probes that are "collectively" best overall.
5. Determine the best number of probes to be in the ESC based on training accuracy.

Step 4 is implemented by ordering the stump rules generated in Step 3 according to the weighted average training accuracy across the data sets. The average accuracy is weighted by the size of the data sets. Step 5 is implemented by incrementally adding stump classifiers in order of quality (accuracy) until one fails to improve the accuracy of the ESC.

Classification with an ESC

Table 1a illustrates the data structures for three data sets with probes having expression values and each instance having a true class assigned. The true classes are assigned to the instance through an expert classifying the instance. For cancer tissues the expert is often a pathologist. Table 1b illustrates the data structure for an ESC consisting of three probes over three data sets. The thresholds and predictive classes are shown for the probes for each data set.

Table 1c is the substitution of the expression values for the probes from Table 1a into the data structures for ESC in Table 1b. The predicted classes are present for each probe. The cells that are grey in Table 1c are incorrect class assignments given the expression values and the thresholds. The bolded classes are correct assignments. For example in Table 1c, P_{11} is incorrectly assigned while P_{12} and P_{13} are correctly assigned. Two out of three correct gives an accuracy of 66.7%.

The ESC majority vote classifier is obtained by going down a column and choosing the majority class predicted by the rules. An odd number of probes are used in the ensemble for simplicity and to ensure there is always a winner. For example in Data Set 1 the majority class is 1 because P_{21} and P_{31} are both correctly assigned to Class 1. For Data Set 2 the Majority Class is 2 because P_{12} and P_{22} are both correctly assigned to Class 2. Going across the majority vote row gives three out of three correct for an accuracy of 100%. Table 2 is the accuracy of the probe sets across the data sets and the ESC accuracy. Because we have not normalized the classifiers in ESC, the column or data set against which the test instance is compared must be known; we will address this limitation in the discussion.

Table 1 - Obtaining accuracy given an instance and an example ESC.

(a) Test instance with expression values and class.			
	Data 1	**Data 2**	**Data 3**
Test Instance with expression values for probes	(P_{11}=0.8, P_{21}= 0.85, P_{31}=1.5)	(P_{12}=12, P_{22}=14, P_{32}=15.5)	(P_{13}=0.02, P_{23}=0.35, P_{33}= 0.55)
True Class	Class 1	Class 2	Class 1
(b) Example ESCs for the three probe sets and data sets:			
Match Probe	**Data Set 1**	**Data Set 2**	**Data Set 3**
P_{1j}	P_{11} <= 0.7 ->Class 1 P_{11} > 0.7 -> Class 2	P_{12} <= 10 ->Class 1 P_{12} > 10 -> Class 2	P_{13} <= 0.1 ->Class 1 P_{13}> 0.1 -> Class 2
P_{2j}	P_{21} <= 0.9 ->Class 1 P_{21} > 0.9 -> Class 2	P_{22} <= 12 ->Class 1 P_{22} > 12 -> Class 2	P_{23} <= 0.3 ->Class 1 P_{23} > 0.3 -> Class 2
P_{3j}	P_{31} <= 1.2 ->Class 2 P_{31} > 1.2 -> Class 1	P_{32} <= 15 ->Class 2 P_{32} > 15 -> Class 1	P_{33} <= 0.5 ->Class 2 P_{33} > 0.5 -> Class 1
ESC	Majority class predicted by P_{11}, P_{21} & P_{31} Stump rules	Majority class predicted by P_{12}, P_{22} & P_{32} Stump rules	Majority class predicted by P_{13}, P_{23} & P_{33} Stump rules
(c) Class predictions given expression values from 1(a) and classifiers from 1(b). Bold is correct and grey is incorrect as compared to true class in 1(a).			
Match Probe	**Data Set 1**	**Data Set 2**	**Data Set 3**
P_{1j}	0.8 > 0.7 -> Class 2	12 > 10 -> **Class 2**	0.02 <= 0.1 ->**Class 1**
P_{2j}	0.85 <= 0.9 ->**Class 1**	14 > 12 -> **Class 2**	0.35 > 0.3 -> Class 2
P_{3j}	1.5 > 1.2 -> **Class 1**	15.5 > 15 -> Class 1	0.55 > 0.5 -> **Class 1**
ESC	Majority class predicted by P_{11}, P_{21} & P_{31} Stump rules -> **Class 1**	Majority class predicted by P_{12}, P_{22} & P_{32} Stump rules -> **Class 2**	Majority class predicted by P_{13}, P_{23} & P_{33} Stump rules -> **Class 1**

Table 2 - Accuracy across the three data sets of the match probe sets and the ESC.

Match Probe	P_{1j}	P_{2j}	P_{3j}	ESC
Accuracy	66.7%	66.7%	66.7%	100%

Gene expression data

There are four data sets. Collectively, these data sets are made up of two affymetrix arrays and two cDNA microarrays data sets. There are two types of tumor tissue: adenocarcinoma and squamous cell carcinoma of the lung. The Su et al. [7] data set consists of 28 samples with 14 adenocarcinomas and 14 squamous cell carcinomas of the lung with 12,533 affymetrix probes (i.e., continuous variables). The Bhattacharjee et al. [8] data set consists of 160 samples with 139 adenocarcinomas and 21 squamous cell carcinomas of the lung with 12,600 affymetrix probes. The Yamagata et al. [9] data set consists of 20 samples with 9 adenocarcinomas and 11 squamous cell carcinomas of the lung with 4417 cDNA probes. The Garber et al. [10] data set consists of 52 samples with 39 adenocarcinomas and 13 squamous cell carcinomas of the lung with 24,192 cDNA probes.

Matched probes

Before obtaining an ESC we must identify matching probes across data sets. While the other classifier methods that we have described will be applied to each of the four data sets independently, an ESC will be learned from the "combined" data set that we are about to describe.

A probe corresponds to part of a gene. Thus, different probes can be associated with the same gene. The same gene may be referenced by different probes in different data sets obtained on different platforms. In order to use an ensemble method that generalizes across different data sets, the Affymetrix and cDNAs probes must be mapped to each other. For this paper all probes are matched via Affymetrix U95A probe names. For the ensemble method the probe sets are matched as follows. Bhattacharjee et al. and Su et al. are joined via their almost identical chips, which are

U95Av2 and U95A Affymetrix chips respectively. Thus they have total overlap of 12533 unique probes. Garber et al. and Yamagata et al. also have similar cDNA probes with a unique overlapping set of 2106 accession id probes.

Consequently, we have two pairs of similar platforms. It is trivial to map probes across platforms within each pair. To map across the pairs (i.e., across all four platforms) an online resource called ProbeMatchDB [11] is used to map these 2106 accession ids onto the U95A Affymetrix ids in a many-to-many mapping. The many-to-many mapping occurs because some of the 2106 cDNA probes have multiple U95A Af-fymetrix ids associated with them. All possible combinations of probes from the four data sets with matching U95A Af-fymetrix ids are used to construct the probe sets. This match-ing process results in the creation of 4491 probe sets. The probe sets are made up of four probes with one from each data set. A probe set can be thought of as a set of references for the same gene. In the discussion that follows we use "probe set" and "gene" synonymously.

Results

The four classification methods described above (i.e., ESC, Stump Rule, C5.0 and Ensemble C5.0) are compared using 10-fold cross validation testing accuracy, standard error and the average number of variables (Avg Var) for each classifier (See Table 3). Each C5.0 tree, Ensemble C5.0 forest and Stump Rule has been built in a way that is informed by a single data set. Thus, we will speak of classifications made by these approaches as using probes.

The ESC method uses the same 4,491 match probe sets or genes for each data set. As we have noted, the construction of ESCs are informed by multiple data sets, but to classify a datum with an ESC we must know the data set (e.g., what column in Table 1) that test instance is drawn.

The ESC does better than C5.0 on all four data sets, better than ensemble C5.0 on three out of four sets and better on two out of four for Stump Rule. Note that the ESC uses probes that are robust across the four data sets rather than using the best probes for each given data set. The Stump Rule, C5.0 and ensemble C5.0 are all using the best probes within a given data set, which gives them an advantage for computing accuracy, although not on generalization to other data sets.

A measure of the complexity of the classifiers is given in the mean number of variables across folds for a classifier. The ESC converges to a classifier of three variables (genes: BPAG1, KRT5 and ABCC5) for nine out of the 10-folds. As explained earlier, this number of genes does not result from a user-defined threshold, but to add more stumps (i.e., genes) to the ensemble would reduce training set accuracy. In the one remaining fold ESC converges to BPAG1, KRT5 and SIAT7B which results in a total of four variables used by the classifier. The thresholds in the rules are stable within each data set and hence there is consensus among the stumps in the cross-validation step Ensemble C5.0, C5.0 and Stump use varying number of probes for each data set, but they do not find probes that generalize across all four data sets.

Discussion

ESCs are not limited to genomic data per se, and we are interested in their characteristics from a machine learning standpoint. Ensembles boost accuracy by insuring variability in classification behavior among the base classifiers. In bagging [12] this variance stems from bootstrap sampling and the instability of the classifiers that are constructed with these differing sample. In random subspace selection the requisite variance comes from selection of differing variables [13] on which to form the classifier. To some extent such designs might be motivated by a desire to use off-the-shelf, greedy decision tree induction, which is a standard base classifier of ensemble approaches. Another way that variance could be achieved would be to modify classifier systems to directly return a set of sufficiently good and sufficiently different classifiers. In fact, this is our approach, though our base classifiers are stumps – we incrementally add "best" stumps (as assessed across multiple data sets) until performance drops.

The limitation of knowing what column in ESC to apply to a particular data set is moot within a lab or facility since they will always compare on the data they collect. The benefit comes from being informed by data collected at other labs and facilities. In practice it can be used to identify signature genes that when combined enable the construction of high accuracy classifiers, even in hold-out data sets.

Table 3 - Adenocarcinoma vs. Squamous Cell Carcinoma for 10-fold cross validation test accuracy for ESC, Stump classifiers, C5.0 decision tree, Ensemble C5.0 and SVMs with standard error (SE) and average number of variables across folds (Avg Var) in the classifier.

Data Set	ESC			Stump Rule			C5.0			Ensemble C5.0		
	Avg Acc %	SE	Avg Var	Avg Acc %	SE	Avg Var	Avg Acc %	SE	Avg Var	Avg Acc %	SE	Avg Var
Bhattacharjee	95.6	1.3	3	96.9	1.0	1	93.7	1.6	3	94.4	1.7	1.3
Su	93.3	4.4	3	83.3	7.5	1	91.7	5.7	1	90.0	7.1	1
Garber	91.8	3.4	3	92.7	3.1	1	87.0	4.8	1.4	88.3	4.4	1.2
Yamagata	95.0	5.0	3	40.0	10.0	1	80.0	11.1	1	90.0	6.7	1

Whenever ensembles succeed in boosting accuracy it can be argued that the base classifiers, by definition, must be overfitting or underfitting the data. The ensemble then results in a classifier that moves towards a "best" fit to the data. In the case of an ensemble of decision trees that boost accuracy, the move is probably from overfit to better fit. Overfitting may account for cases where C5.0 underperformed Stump Rule on

some data sets. In the case of ESCs the move is probably from underfit to better fit. ESCs are not only very simple, are formed from stumps that generalize well across data sets, and as a result may underfit any given data set.

One reason stumps that generalize well across data sets are desirable with microarray data is that artifacts can be introduced into the data by the lab that collects the data. Thus, the very best stump for a particular data set may exploit an environmental peculiarity of the lab that is collecting the data. When dealing in thousands of probes, any small laboratory bias may systematically influence the value of one or more probes [14]. Problems with data or facility bias are found in other contexts. For example, Evans and Fisher [15] found that a feature (i.e., printing press) that was highly predictive of a printing defect in a particular printing plant provided no insights to why similar problems occurred at other plants.

ESC builds classifiers that generalize across data sets (labs, facilities), and informs data collectors about probes that generalize beyond their data collection processes. When ESCs learned across data sets are contrasted with the best classifiers within a data set, our methodology can also point out lab biases that should be remedied.

ESCs were motivated initially by the desire to combine evidence from multiple sources of gene expression data, thereby mitigating the curse of dimensionality. We are interested then in what ESCs can tell in a biological domain. Notably, the ensemble method identifies the genes KRT5, BPAG1 and ABCC5 as informative across the four data sets that we examined. Importantly, the ESC method gives convergent support for the relevance of these probe sets relative to the findings in the original studies. Using hierarchical clustering Bhattacharjee et al. [8] found KRT5 and BPAG1 to be highly expressed in squamous cell carcinoma. Using hierarchical clustering Yamagata et al. [9] confirmed in their data that KRT5 and BPAG1 are highly expressed in squamous cell carcinoma. Using hierarchical clustering Garber et al. [10] also identified KRT5 and BPAG1 as highly expressed in squamous cell carcinoma. Using a SVM ranking method Su e al. [12] did not identify high expression of KRT5 or BPAG1 as predictive of squamous cell carcinoma. However, they did identify ABCC5 as predictive of squamous cell carcinoma. The convergent findings of these alternative methods provide additional support of the utility of the ESC method.

Conclusion

The ESC method does well with only three variables. These results suggest the existence of compact sets of genes with single thresholds, which can be measured using multiple modalities that consistently and accurately predict diagnosis. The building of data repositories and data exchange standards such as oncomine.com and the caBIG™ can assist in the discovering other robust mRNA signatures of cancer using ESC. The ESC finds in a greedy fashion the best available matched probes that can be used in an mRNA signature that generalizes across data sets.

References

[1] Simon, R. Diagnostic and prognostic prediction using gene expression profiles in high-dimensional microarray data. British Journal of Cancer 2003: 89; pp. 1599-1604.

[2] Simon, R., Radmacher, M.D., Dobbin, K. and McShane, L.M. Pittfalls in the Use of DNA Microarray Data for Diagnostic and Prognostic Classification. J. Natural Cancer Institute 2003: 95(1); pp. 14-18.

[3] Michiels, S., Koscielny, S. and Hill, C. Prediction of cancer outcome with microarrays: a multiple random validation strategy. Lancet 2005: 365; pp. 488-92.

[4] Holte, R. Very Simple Classification Rules Perform Well on Most Commonly Used Datasets. Machine Learning 1993: 11; pp. 63-91.

[5] Quinlan, J.R. Induction of decision trees. Machine Learning 1987: 1; pp. 81-106.

[6] Freund, Y. and Schapire, R.E. Experiments with a new boosting algorithm. In Machine Learning: Proceedings of the Thirteenth International Conference 1996; pp. 148 - 156.

[7] Su, A. et al. Molecular Classification of Human Carcinomas by Use of Gene Expression Signatures.* *Cancer Research* 2001: 61; pp. 7388-7393.

[8] Bhattacharjee, A, et al. Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. PNAS: 2001: 98 (24); pp. 13790-13795.

[9] Yamagata N. et al. A training-testing approach to the molecular classification of resected non-small cell lung cancer. Clin Cancer Res 2003: 9; pp. 4695-704.

[10] Garber, M.E., et al. Diversity of gene expression in adenocarcinoma of the lung. PNAS 2001: 98 (24); pp. 13790-13795.

[11] Wang,P. et al. (2002) ProbeMatchDB—a web database for finding equivalent probes across microarray platforms and species. Bioinformatics, 18, pp. 488–489.

[12] Breiman, L. Bagging Predictors. Machine Learning 1996: 24(2); pp 123-140.

[13] Breiman, L. Random Forests. Machine Learning 2001: 45(1); pp. 5-32.

[14] Shi, Leming et al. The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. Nature Biotechnology 2006: 24(9); pp. 1151 – 1161.

[15] Evans, B. and Fisher, D. Overcoming Process Delays with Decision Tree Induction 1994: 9(1); pp. 60-66.

Address for correspondence

Lewis Frey, Ph.D.
26 South 2000 East; Salt Lake City, UT 84112; USA

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

From "Glycosyltransferase" to "Congenital Muscular Dystrophy": Integrating Knowledge from NCBI Entrez Gene and the Gene Ontology

Satya S. Sahoo[1], Kelly Zeng[2], Olivier Bodenreider[2], Amit Sheth[1]

[1]Kno.e.sis Center, Department of Computer Science and Engineering, Wright State University, Dayton, OH, USA
[2]U.S. National Library of Medicine, NIH, Bethesda, Maryland, USA

Abstract

Entrez Gene (EG), Online Mendelian Inheritance in Man (OMIM) and the Gene Ontology (GO) are three complementary knowledge resources that can be used to correlate genomic data with disease information. However, bridging between genotype and phenotype through these resources currently requires manual effort or the development of customized software. In this paper, we argue that integrating EG and GO provides a robust and flexible solution to this problem. We demonstrate how the Resource Description Framework (RDF) developed for the Semantic Web can be used to represent and integrate these resources and enable seamless access to them as a unified resource. We illustrate the effectiveness of our approach by answering a real-world biomedical query linking a specific molecular function, glycosyltransferase, to the disorder congenital muscular dystrophy.

Keywords:

knowledge integration, Semantic Web, RDF, Entrez Gene, Gene Ontology

Introduction

A common scenario in biomedical research involves the correlation of genomic data with disease information, in other words, associating genotype and phenotype information. In the particular scenario illustrated in this paper, a researcher is interested in glycosylation and its implications for one disorder: congenital muscular dystrophy. The biological process of glycosylation results in the post-translational addition of glycosyl groups (saccharides) to proteins (and lipids). Various enzymes, namely glycosyltransferases, catalyze glycosylation reactions.

From the functional annotation of gene products with terms from the Gene Ontology (GO), a researcher can identify the genes having the molecular function of catalyzing the transfer of specific glycosyl groups (e.g., *hexosyltransferase*, for hexosyl groups). Known associations between these genes and diseases can then be mined from resources such as NCBI's Entrez Gene (EG), where phenotypic information is recorded as pointers to the Online Mendelian Inheritance in Man (OMIM) knowledge base [3]. (See the Materials section for a presentation of GO and EG.)

In order to validate the hypothesis of possible association between the molecular function *glycosyltransferase* and the disease *congenital muscular dystrophy*, a researcher could simply search EG for the term *glycosyltransferase*, and all records containing the string "glycosyltransferase" in GO annotations would be returned. This approach, however, is suboptimal for at least two reasons. First, the term *glycosyltransferase* might appear as a substring in other GO terms (e.g., in *UDP-glycosyltransferase*), possibly leading to false positives. Conversely, not all GO terms related to *glycosyltransferase* actually contain the string "glycosyltransferase" (e.g., *acetylglucosaminyltransferase*, a kind of *glycosyltransferase*), possibly leading to false negatives.

To avoid false positives and false negatives, a careful researcher would likely start exploring the Gene Ontology database to create a list of *glycosyltransferase*-related terms by selecting the term *glycosyltransferase* itself (GO:0016757) and all its descendants, including specialized types of *glycosyltransferase*, such as *acetylglucosaminyltransferase*. This researcher would then look for the genes annotated with any of the *glycosyltransferase*-related terms. Resources such as the web browser AmiGO [1] support such searches and can retrieve the genes associated with any descendant of a given GO term. Finally, each of the genes found associated with any of the *glycosyltransferase*-related terms must be searched individually in EG, looking for mentions of the disease *congenital muscular dystrophy* (as an OMIM phenotype) in the corresponding records.

The procedure described above is evidently inefficient, time consuming and error prone as several web interfaces need to be utilized (AmiGO and Entrez), and as the results of the search in one resource need to be copied and pasted as search terms in the other. The main reason for such inefficiency is that high quality resources such as GO and EG have been designed primarily for consultation by humans, not for automated processing by agents or integration in applications. Moreover, these resources have been developed by different groups, independently of each other and are therefore not interoperable. No system currently supports complex queries such as: *Find all the genes annotated with glycosyltransferase-related terms in GO and associated with the disease congenital muscular*

dystrophy in OMIM. Typically, querying across the different knowledge sources is accomplished manually through meticulous work or requires the development of complex and customized software applications.

In this paper, we propose an integrative approach to querying across knowledge sources. More specifically, we have applied Resource Description Framework (RDF) [4] standard developed by the World Wide Web Consortium (W3C) to integrate knowledge from GO and EG, and used this integrated resource to answer complex queries. We use the scenario presented earlier to illustrate the advantages of this approach. This work is a pilot contribution to the *Biomedical Knowledge Repository* under development at the U.S National Library of Medicine (NLM) as part of the *Advanced Library Services* project [8]. This repository integrates knowledge not only from structured resources (database and knowledge bases), but also from the biomedical literature (e.g., MEDLINE), in order to support applications, including knowledge discovery.

Background

Information integration is one of the most challenging areas of research in Computer Science [11]. The use of heterogeneous schemas for data storage, that are designed primarily to ensure optimization of storage space, makes it extremely difficult for users to query data sources in an integrated manner. (The interested reader is referred to [12] for a survey of approaches to information integration.) The Semantic Web provides a common framework that enables the integration, sharing and reuse of data from multiple sources. The use of a representation formalism based on a formal language enables software applications to 'understand' and reason over information. Recent research in Semantic Web technologies has delivered promising results to enable information integration across heterogeneous knowledge sources.

The Resource Description Framework (RDF) is a W3C-recommended framework for representing data in a common format that captures the logical structure of the data. This is in contrast to pure storage aspects addressed by traditional relational database schema. The RDF representational model uses a single schema in contrast to multiple heterogeneous schemas or Data Type Definitions (DTD) used to represent data in XML by different sources. Hence in conjunction with a single Uniform Resource Identifier (URI), all data represented in RDF form a single knowledge repository that may be queried as one knowledge resource. An RDF repository consists of a set of assertions or triples. Each triple is constituted of three entities namely, the *subject* – the triple pertains to this entity, the *object* – the entity that states something about the object and the *predicate* – the relationship between the *subject* and the *object*. For example, as shown in Figure 1, assertions such as *acetylglucosaminyltransferase* (GO:0008375) is a kind of *hexosyltransferase* (GO:0016758) and the gene *LARGE* (EG:9215) has molecular function *acetylglucosaminyltransferase* (GO:0008375) can be represented as RDF triples.

The RDF triples often share nodes, thus forming a graph. For example, the two triples shown in Figure 2 share the node *acetylglucosaminyltransferase* (GO:0008375). The resulting graph is shown in Figure 2. The graph structure created by RDF is key to information integration in the Semantic Web.

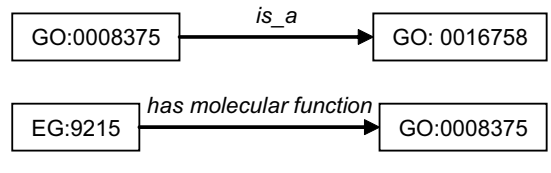
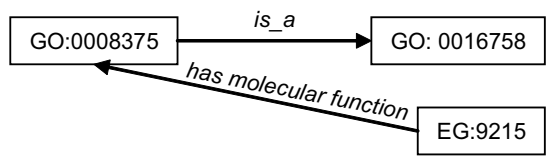


Figure 1 - Example of RDF triples



Fiigure 2 - Example of RDF graph

Materials

The **Gene Ontology** (GO) seeks to provide a consistent description of gene products [13]. GO consists of three controlled vocabularies for biological processes (9,234 terms), molecular functions (7,456 terms) and cellular components (1,804 terms). The GO monthly releases are made available on the GO website in various formats, including RDF. The version of GO used in this study is dated of September 2006.

The **Entrez Gene** (EG) database records gene-related information from sequenced genomes and of model organisms that are focus of active research [9], totaling about two million genes. EG contains gene information about genomic maps, sequences, homology, and protein expression among others [9]. In contrast to GO, EG is not available in RDF, but in XML (converted from ASN1 by the program *gene2xml* provided by NCBI), and can be downloaded from the NCBI website. The version of EG used in this study is dated of July 2006.

Methods

Our integration method can be summarized as follows and is illustrated in Figure 3. First, we extract manageable subsets from the two resources to be integrated. We then have to convert the EG subset from XML to RDF. Finally, we load both RDF resources in a common store, apply inference rules, and issue queries against it.

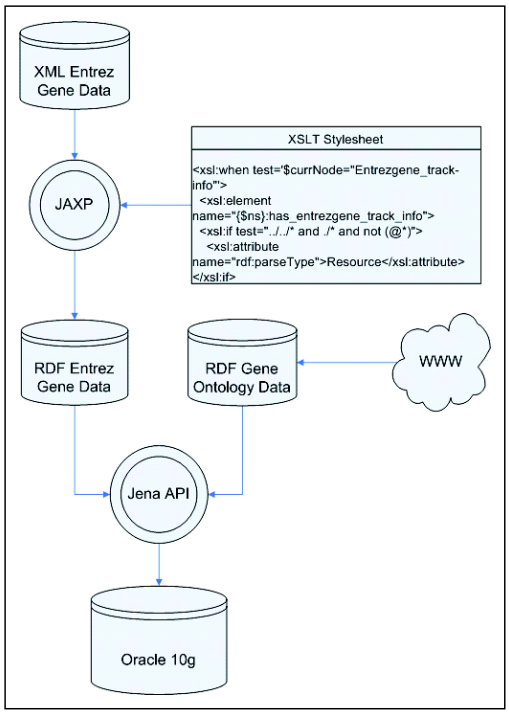


Figure 3 - Overview of the integration method

Creating subsets

The entire Entrez Gene data file (in XML format) is very large (50 GB) and unnecessarily difficult to manipulate. In order to obtain a manageable subset from EG, we restricted the gene records to two species: *Homo sapiens* (human) and *Mus musculus* (mouse). The resulting EG subset contains a total of 99,861 complete gene records (excluding obsolete records).

Converting XML format Entrez Gene data to RDF

A key element of our integration approach is the conversion of Entrez Gene from XML to RDF. There are many issues involved in the conversion of XML data into RDF format, including modeling the original semantics of the data, filtering redundant XML element tags, linking data entities using meaningful named relationships and identifying entities consistently within and across resources. Unlike traditional XML to XML conversion, XML to RDF conversion should exploit the advantages of the RDF model in representing the logical structure of the information.

We chose not to convert the element tags of the native EG XML representation mechanically into the *predicates* of the RDF triples. Instead, we manually converted the XML element tags into meaningful and standardized relationship names that convey explicitly the semantics of the connection between the *subject* and the *object*. For example, the element *<Org-ref_taxname>* was mapped to the more meaningful relationship named *has_source_organism_taxonomic_name*.

We selected the eXtensible Stylesheet Language Transformation (XSLT) [6] for converting the EG XML information into RDF, because this approach allows for a clean separation between the application (using Java API for XML Processing (JAXP)) and the conversion logic (using XSLT stylesheet). Once the stylesheet is created, it can serve as an auxiliary file for existing programs realizing the XML to RDF conversion. In other words, the major interest of this approach is that no specific code is required for the conversion, because the transformation logic resides entirely in the stylesheet.

Loading the two resources into a single data store

Some of the requirements for our RDF store include native support for the RDF graph data model, support for persistence and indexing of the RDF triples, support for extensive collections of triples, and availability of a query language for the RDF graph. After surveying available RDF storage solutions, we decided to use Oracle Spatial 10g [7] as the RDF storage system.

The RDF file resulting from the XSLT conversion of the original XML file for EG and the downloaded RDF version of GO are both loaded into a single RDF store. More precisely, the RDF resources are first converted to the NTriple format using the Jena API [10] and loaded into the RDF database using a utility program provided by Oracle.

Applying inference rules

Unlike the Web Ontology language OWL, RDF provides no direct support for inference. However, inference rules can be implemented in the RDF store to make explicit the semantics of some predicates. For example, the relationships *is_a* and *part_of* used in GO are partial order relations, thus being reflexive, antisymmetric and transitive. The inference rules we created for implementing the transitivity and combination of these two relationships are shown in Table 1. The inference rules are stored in a rule base created in Oracle 10g.

Table 1 - Inference rules for is_a and part_of in GO

Relation	is_a	part_of
is_a	IF <**x** *is_a* **y**> & <**y** *is_a* **z**> THEN <**x** *is_a* **z**>	IF <**x** *is_a* **y**> & <**y** *part_of* **z**> THEN <**x** *part_of* **z**>
part_of	IF <**x** *part_of* **y**> & <**y** *is_a* **z**> THEN <**x** *part_of* **z**>	IF <**x** *part_of* **y**> & <**y** *part_of* **z**> THEN <**x** *part_of* **z**>

Querying the RDF Graph with SPARQL

SPARQL [5] is a query language for RDF graphs, equivalent to SQL, the Structured Query Language, for relational databases. Unlike SQL, SPARQL does not require users to be familiar with the data model (e.g., tables, foreign keys), but simply to indicate how entities of interest relate to each other. For example, the structure of the query: *Find all the genes annotated with the GO molecular function glycosyl-*

*transferase (*GO:0016757) *or any of its descendants and associated with any form of congenital muscular dystrophy* is represented in Figure 4.

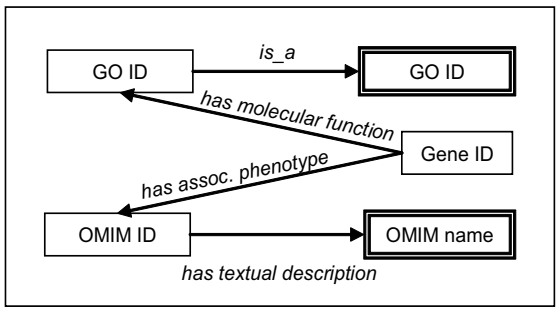


Figure 4 - RDF graph corresponding to the query above

```
SELECT distinct t,g,d
FROM TABLE(SDO_RDF_MATCH(
'(?t is_a GO:0016757)
 (?g has molecular function ?t)
 (?g has_associated_phenotype ?b2)
 (?b2 has_textual_description ?d)',
SDO_RDF_Models('entrez_gene'),
SDO_RDF_Rulebases('entrez_gene_rb'),
SDO_RDF_Aliases(SDO_RDF_Alias('','')), null) )
where (
    REGEXP_LIKE(LOWER(d), '((.)*(congenital)(.)*)')
AND REGEXP_LIKE(LOWER(d), '((.)*(muscular)(.)*)')
AND REGEXP_LIKE(LOWER(d), '((.)*(dystrophy)(.)*)'));
```

Figure 5 - Example of SPARQL query (simplified)

The query can be understood as finding a path in the RDF graph using a predetermined set of semantic relationships and would be formulated as follows. Because of the inference rules implementing the transitivity and reflexivity of the *is_a* relationship, the condition on the GO annotation "*glycosyltransferase (*GO:0016757) *or any of its descendants*" is easily expressed by '`?t is_a GO:0016757`'. The link between genes and GO terms is expressed by '`?g has_molecular_function ?t`'. Similarly, the link between genes and OMIM diseases is expressed by '`?g has_associated_phenotype ?b2`' (OMIM ID) and '`?b2 has_textual_description ?d`' (disease name). Finally, direct constraints are put on the GO term on the one hand ('`?t is_a GO:0016757`', to select *glycosyltransferase* (GO:0016757)) and on disease names on the other (where a regular expression is used to select disease names containing the strings "congenital", "muscular" and "dystrophy"). The actual (but simplified) SPARQL query is shown in Figure 5.

Results

One integrated RDF repository for Entrez Gene and GO

The subset of Entrez Gene restricted to *Homo sapiens* (human) and *Mus musculus* (mouse) as biological sources comprises 99,861 gene records. Once converted to RDF, it consists of 772,530 triples. The RDF version of GO contains 293,798 triples. Overall, there are over one million triples in the store created for this experiment, which is relatively small in comparison to the 411 million triples resulting from the conversion of the entire EG to RDF [2].

Biological query result: extended example

The SPARQL query presented above returned one result, corresponding to one path in the graph between the GO term *glycosyltransferase (*GO:0016757) and OMIM disease names containing (variants of) the string "*congenital muscular dystrophy*".

This path involved the human gene *LARGE like-glycosyltransferase* (EG:9215), annotated with the GO term *acetylglucosaminyltransferase* (GO:0008375), a descendant of *glycosyltransferase (*GO:0016757). Also involved in this path is the OMIM disease identified by MIM:608840. The name (textual description) of this disease is *Muscular dystrophy, congenital, type 1D* and contains the required substrings "congenital", "muscular" and "dystrophy". The instantiated RDF graph with path between *glycosyltransferase* (GO:0016757) and *Muscular dystrophy, congenital, type 1D* is shown in Figure 6.

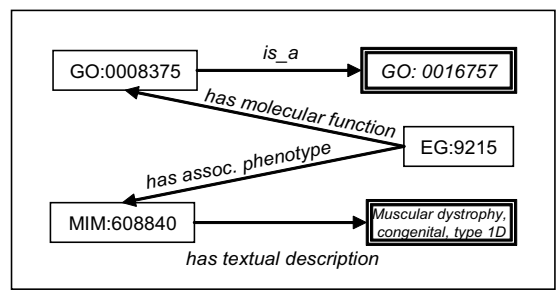


Figure 6 - Instantiated RDF graph

This simple SPARQL query provides an easy way of testing the biological hypothesis under investigation, i.e., the existence of a possible link between glycosylation and *congenital muscular dystrophy*. On manual inspection of the Entrez Gene record, we also note that the given gene may be involved in the development and progression of meningioma through modification of ganglioside composition and other glycosylated molecules in tumor cells.

Discussion

Significance

In this study, we demonstrated the feasibility of integrating two biomedical knowledge resources through RDF. We also provided anecdotal evidence for the benefits of such integration by showing how *glycosyltransferase* can be linked to *congenital muscular dystrophy*. The integrated resource is greater than the sum of its parts as it supports complex queries that could typically not be handled otherwise without tedious manual intervention or customized software applications.

Integrated resources based on a graph model are particularly important in an exploratory context where researchers need to "connect the dots" in order to validate an hypothesis. This approach also facilitates intuitive

hypothesis formulation and refinement. For example, after verifying that glycosyltransferase is linked to congenital muscular dystrophy, our researchers may narrow the focus of their wet lab experiments to only hexosyltransferase out of the potential seven glycosyltransferases. Analogously, they can focus their research on Muscular dystrophy, congenital, type 1D, out of several other diseases.

Arguably, the graph data model of RDF resources is more intuitive than the database schemas. In fact, the RDF data model enables us to model the inherent logical relations between entities that mirror the human cognitive model of the real world. Additionally, the RDF data model offers more flexibility than database schemas for accommodating changes to the underlying model.

Generalization

The integration approach demonstrated in this study can be generalized to more complex queries and to additional information sources. For example, many additional constraints can be easily added to the query presented earlier by exploiting other properties represented in GO or EG. Examples of such constraints include restricting the annotations to specific evidence codes (e.g., *TAS*) and narrowing the query to a specific model organism.

Only two resources are currently integrated in our RDF store. However, this approach can be generalized to other resources including pathway databases, microarray resources, disease ontologies and virtually all the structured knowledge bases currently under the umbrella of the Entrez system, including UniGene and HomoloGene. Knowledge extracted from unstructured sources such as the biomedical literature can also be integrated. Creating such an extensive repository of biomedical knowledge is one of the goals of the *Advanced Library Services* project under development at NLM.

Unresolved issues and challenges

In addition to scalability issues, which can be addressed by mature software and the next generation of hardware, challenges include the identification and organization of entities and relationships. Heterogeneous resources can interoperate in a RDF graph only if the entities shared by these resources are identified consistently. The namespace provided by the UMLS is expected to play an important role for the permanent identification of biomedical entities. In contrast to entities for which organizational schemes currently exist (terminologies and ontologies), the named relationships used to connect data entities during the conversion of EG from XML to RDF are currently not formalized in an ontology of relationships. As a consequence, only limited reasoning can be supported by the RDF graph. As sizeable ontologies of relationships become available, they too will be used for normalizing knowledge in our repository. RDF schemas and OWL will also be investigated.

Acknowledgments

This research was supported in part by the Intramural Research Program of the National Institutes of Health (NIH), National Library of Medicine (NLM) and by the Integrated Technology Resource for Biomedical Glycomics (5 P41 RR18502), funded by the National Institutes of Health National Center for Research Resources.

References

[1} AmiGO: Gene Ontology browser [cited 11/29/06; Available from: http://www.godatabase.org/]

[2] BioRDF subgroup: Health Care and Life Sciences interest group [cited 11/29/06; Available from: http://esw.w3.org/topic/ HCLSIG_BioRDF_Subgroup]

[3] Online Mendelian Inheritance in Man, OMIM (TM). McKusick-Nathans Institute for Genetic Medicine, Johns Hopkins University (Baltimore, MD) and National Center for Biotechnology Information, National Library of Medicine [cited 11/29/06; Available from: http://www.ncbi.nlm.nih.gov/omim/]

[4] Resource Description Framework (RDF), [cited 11/29/06; Available from: http://www.w3.org/TR/2004/REC-rdf-primer-20040210/]

[5] SPARQL Query Language for RDF. W3C Working Draft 2006 [cited 11/29/06; Available from: http://www.w3.org/TR/rdf-sparql-query.]

[6] XML Schema Language Transformation (XSLT) [cited 11/29/06; Available from: http://www.w3.org/TR/xslt]

[7] Alexander, N., Ravada S., "RDF Object Type and Reification in Oracle"—Technical White Paper [cited 11/29/06; Available from: http://download-east.oracle.com/otndocs/tech/semantic_web/pdf/rdf_reification.pdf]

[8] Bodenreider O, Rindflesch TC. Advanced library services: Developing a biomedical knowledge repository to support advanced information management applications. Technical report. Bethesda, Maryland: Lister Hill National Center for Biomedical Communications, National Library of Medicine; September 14, 2006.

[9] Maglott D, Ostell J, Pruitt KD, Tatusova T. "Entrez Gene: gene-centered information at NCBI", Nucleic Acids Res. 2005 January 1; 33(Database Issue): D54–D58.

[10] McBride, B.. Jena: A Semantic Web Toolkit. IEEE Internet Computing 2002;6, 6 (Nov. 2002), 55-59.

[11] Sheth AP. "Changing Focus on Interoperability in Information Systems: From System, Syntax, Structure to Semantics", in Interoperating Geographic Information Systems. M. F. Goodchild, M. J. Egenhofer, R. Fegeas, and C. A. Kottman (eds.), Kluwer, Academic Publishers, 1999, pp. 5-30.

[12] Shvaiko P, Euzenat J. 2005. A survey of schema-based matching approaches. Journal on Data Semantics 4: 146-71.

[13] The Gene Ontology Consortium. Gene Ontology: tool for the unification of biology. Nature Genetics 2000;25: 25-29.

Address for correspondence

Olivier Bodenreider, National Library of Medicine
8600 Rockville Pike, MS 3841, Bethesda, MD 20894, USA.
Email: olivier@nlm.nih.gov. Phone: (301) 435-3246.

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Building a Research Model for Human Genetic Variation Knowledge Management

Yulong Gu[a], James Warren[a]

[a] *Department of Computer Science, University of Auckland, Auckland, New Zealand*

Abstract

Organizational knowledge management (KM) research studies the nature of knowledge, the scope of KM, the factors and mechanisms that affect KM outcomes, as well as theoretical KM frameworks. This paper discusses the implications of past studies for the KM efforts in the human genetic variation (HGV) research domain and presents a HGV-KM research model. This model identifies the context of HGV KM studies, the predispositions and factors that may impact KM outcomes, and important KM processes. It also represents the relationships among these issues. Applying the model, further studies will point the way for improved capture and dissemination of HGV knowledge from routine HGV research activities to contribute to the global genetics knowledgebase.

Keywords:

Knowledge Management (KM); knowledge processing; human genetic variation (HGV) research; KM approach; KM framework; KM process; HGV-KM research model

Introduction

Knowledge Management (KM) theories

Knowledge is a fluid mix of framed experience, values, contextual information and expert insight; it includes both explicit knowledge that is transmittable in formal systematic language and tacit knowledge that has a personal quality and is hard to formalize and communicate [1-3]. Different views of knowledge lead to different perceptions of knowledge management (KM) [4, 5]. For instance, (i) if knowledge is an object or information access, KM should focus on building and managing knowledge stocks; (ii) if knowledge is a process, the focus of KM is knowledge flow and knowledge processing – knowledge creation, acquisition, codification, retention, storage/ retrieval, integration, coordination, transfer, sharing, distribution, application, valuation and use; (iii) if knowledge is viewed as an organizational capability, then KM centers on building core competencies, understanding the strategic advantage of know-how, and creating intellectual capital [6-8].

Human Genetic Variation (HGV) context

Human genetic variation (HGV) research aims to characterize the nature, distribution and evolution of genetic variations in humans and to study the relationship *between* genetic variation *and* environment in the origins *and* characteristics of human populations *and* causes, diagnoses, treatments and prevention of the disease; HGV scientists come from various domains, e.g. genomics, proteomics and clinical sciences [9]. The clinical and laboratory genetic services test and interpret the HGV of patients and/ or families [10]. There were over 500 HGV testing laboratories in the United States at 1997 [11]. The HGV testing is becoming a routine procedure in clinics and research, with at least 751 active laboratories and 936 clinical chemistry/ hematology centers at 2004 in the European Union alone [12]. An international survey at 2005 shows that 45% of early breast cancer patients discuss genetic testing with their physician and/or are referred to see a genetic counselor and 16.7% are then tested [13]. However, despite the significant growth of HGV knowledge, KM success is seldom reported:

i) for managing high-quality *knowledge stocks*. Enormous efforts are put into storing HGV data in Locus Specific Databases (LSDB) and general databases, such as HGVbase [14], UMD LSDB [15], OMIM Database [16] and the proposed Central Database plus WayStation Submission tool [17-20]. However, the validity of data in these databases is of some concern, possibly due to inadequate data curation [21, 22].
ii) for supporting *knowledge flow* and enhancing *knowledge processes*. For example, although HGV testing laboratories collect and produce a lot of data with HGV details, knowledge flow from these laboratories to HGV research community is not occurring often [10, 23].
iii) for building core *competencies* and creating *intellectual capital*. With no attempt to improve the present "pattern" [24] of knowledge processing in HGV research facilities, the research ability and intellectual capital are not being managed or changed by KM efforts.

With above KM problems, we apply KM theories to identify all the relevant factors; then, we develop a HGV-KM research model to present the anatomy among the significant issues.

Implications of past literature

We synthesize past literature from leading Organizational Management journals (e.g., Organization Science and Harvard Business Review), Information Systems (IS)

periodicals, (e.g., MISQ and ISR), HGV journals (e.g., Human Mutation and Nucleic Acids Research), KM or HGV proceedings and books, then present the significant implications in this section.

KM approaches in practice

Most recorded KM practice took the *product-centric* or *process-centric* approach, reflecting type (i) or (ii) KM focus [6]; however, *capability-centric* KM exercise (type three) is rarely reported. The *product-centric* KM manage knowledge as an objective organizational asset [4, 25]. It relies on the transformation of implicit or explicit knowledge from employees' heads to written information in documents and the subsequent management of these documents [26]. Accordingly, by using searchable document repository and content management systems [7], HGV knowledge can be captured, stored, retrieved and distributed in well-organized research documentations.

On the other hand, *Process-centric* KM views knowledge as residing with a person and/or a business process. It provides pointers to experts [25] and implements business process management [27], by adopting database of experts, decision aids and expert system, workflow management system, groupware, the systems supporting 'Community of Practice' and 'hardwiring' of social networks, etc. [6, 28]. HGV studies require profound knowledge on the subjects and methodologies. Therefore, research done around the globe is frequently referred; and international collaborations are often performed. Taking these natures of HGV research into account, the key to managing HGV knowledge is to share it among the researchers [26], applying process-centric KM. Meanwhile, such KM approach may trigger the benchmarking, reengineering and optimizing of the HGV research processes. By tracking and sharing HGV research activities, instance decision making in single variant interpretation may be transited into best practice in studying the gene. This transition may eventually enhance an HGV research methodology and become a valuable intellectual capital, since medical data analysis may discover new models, into which available knowledge could be incorporated [29]. Thus, approaches in managing HGV knowledge may take all product-, process-, and capability-centric endeavors.

KM frameworks in literature

Following KM approach review, this subsection extracts the implications of KM models as they offer best KM practices.

The result of knowledge creation, retention and transfer is affected by the properties of organization units, of unit relationships and of knowledge itself; and this effect is moderated by three key causal mechanisms – the three important KM processes: ability, motivation and opportunity [8].

Based on organizational capability perspective theory [30] and contingency perspective theory [31], two more preconditions for effective KM are discovered – the 'knowledge infrastructure capability' and 'knowledge process capability' – with the latter being influenced by contingent knowledge tasks [32].

Knowledge management systems (KMS) are a class of IS to manage organizational knowledge, and to enhance knowledge processes [6, 33]. Developed upon IS Success Model [34, 35], Figure 1 shows how the individual's and organization's performance at workplace are improved from using KMS [36-38].

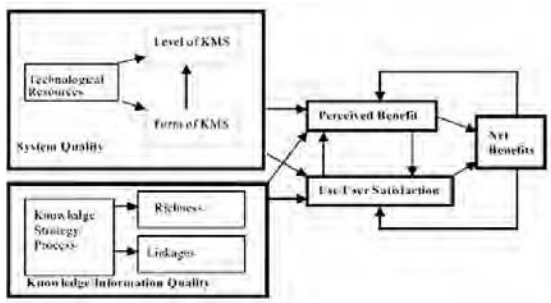


Figure 1 – KMS Success Model [36]

HGV-KM research model

To manage knowledge in HGV context, the practical questions are "which activities are the promising targets for KM support?" and "what are the nature and benefits of effective KM support?" An HGV KM study is to point the way for improved knowledge capture and dissemination from routine research activities to contribute to the global genetics knowledgebase. Drawing on past literature, we develop a research model (Figure 2) with nine significant KM issues: I. KM context, II. KM process, III. knowledge process capability, IV. contingent task characteristics, V. technology and system quality, VI. knowledge and information quality, VII. perceived benefits and user satisfaction, VIII. knowledge infrastructure capability, and IX. KM outcome. The construct relationships are direct impacts (as arrows from cause to result), moderating effects on a relationship's strength (by broken arrows) and weak connections (e.g. the quality of KM technologies as the extent to which knowledge processes and KM processes are computerized and integrated [36], represented by curves in the figure). In detail:

I. Context of a KM initiative or project [8]

According to [8], the predispositions of a HGV KM project include (i) the properties of units, e.g. a researcher's academic status and an organization's social status; (ii) the properties of relationships between units, e.g. the contact frequency between dyadic units and the connection pattern among multiple units; and (iii) properties of knowledge, including explicit HGV data and tacit know-how in HGV research, external and internal knowledge, and uniquely possessed and public knowledge.

II. KM process [8]

A KM process presents what a KM project offers and how well it functions to enhance an organization's (i) KM abil-

ity to codify implicit knowledge and produce information that makes sense to people other than the author [28], (ii) knowledge sharing motivation – the rewards and incentives, (iii) opportunity, and (iv) from process-centric KM perspectives, the management of HGV research activities and processes.

III. Knowledge process capability [32]

This organizational competency is the result of contextual factors and efforts (Constructs I and II) [8]; it then decides the quality of processed knowledge (Construct VI), according to product-centric KM theories and KMS Success Model [36]. It also has direct impact on overall KM effectiveness and this impact is moderated by knowledge-related tasks [32].

IV. Contingent knowledge tasks' characteristics [32]

The characteristics of knowledge tasks (such as task content and task domain) decide if the right knowledge is captured and used [38]. An example of task domain is about knowledge creation tasks that may belong to any of the four modes of socialization, externalization, combination and internalization [39]. The moderating role of this construct suggests that the KM efforts that precisely suit a task provide more effective results than those that don't [31, 32].

V. Technology and system quality [36]

In addition to the quality of general IS, such as 'ease of learning', 'integration of systems', and quick 'system response time' [34], KMS quality has three more dimensions: (i) the technological resources – the ability to develop, operate, and maintain a KMS, (ii) KMS form – the extent to which organizational memory and KM processes are computerized and integrated, and (iii) KMS level – the ability to bring past information to bear upon current activities [36, 37]. Given the teamwork nature of HGV research, quality KM technology should also support an HGV research facility's social capital (Construct VIII) by facilitating collaborations, distributed learning, knowledge mapping and opportunity generation [30, 40, 41].

VI. Knowledge and information quality [36]

High-quality knowledge and information are complete, accurate, current (of linkages), informative, rich in expression and in detail [34, 37]. For example, a valid HGV result has to offer reliable and sufficient evidence for variant interpretation based on accurate and unambiguous variant description.

VII. Perceived benefits and user satisfaction [36]

Perceptions on the benefits of KM technologies include perceived usefulness and ease-of-use, both of which are significant predictors of technology acceptance – the actual levels of system usage and user satisfaction [42]. This perspective of users is a result of knowledge quality; and it delivers ultimate KM outcomes in the organization [36, 38].

VIII. Knowledge infrastructure capability [32]

This capability represents the organization's social capital – the network of relationship; and it delivers KM results through knowledge sharing via the network [32]. It is operationalized by (i) technologies (Construct V), (ii) the organizational structure that provides the relationships (i.e. is an organization's property within Construct I) and (iii) the culture that provides a shared context (as a relationship's property in Construct I).

IX. Ultimate KM outcomes [36]

KM outcomes include the KM project-improved organizational effectiveness – such as the ability to innovate and coordinate [30] – and individual/organizational performance [38]. The individual KM performance, as measured by 'correctness of decision' and 'confidence in decision', will in turn have an impact on the organization's performance, e.g. on product quality [34, 36]. As a KM project's outcomes, the KM efforts may cause positive or negative consequences that will trigger more or less use of the knowledge and the KMS [38].

Discussion

From past KM approaches and frameworks that recorded the nature of knowledge, KM scope, KM factors and mechanisms, we identify nine categories of contextual,

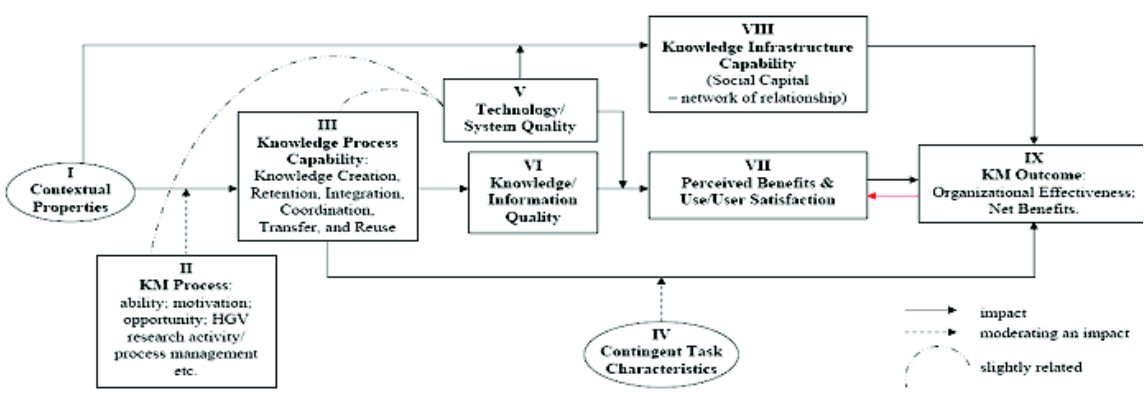


Figure 2 – HGV-KM Research model

cultural, structural, managerial, cognitive and technological issues that might be critical to the genetics research domain. Then we develop an HGV-KM research model presenting the dynamics among these nine constructs. There are a number of KM models proposed by both academics and practitioners that are not included in this paper, e.g. [43], [27], [24]. The reason for not covering them is that they focus more on single KM approach (product/process/capability-centric), but the HGV setting may have to apply all three strategies. Another limitation of our model is that it still needs validation, for instance, by empirical testing. However, it has revealed a promising research area that is to seek resolutions for the KM issues in genetics.

Our current research direction is to validate the model proposed herein through iterative action research. We plan to evaluate the impact of IT on various KM processes and capabilities, as well as on overall KM outcomes. This will also validate the model constructs and the model dynamics in terms of construct relationships; for instance, we are expecting to discover significant correlations between the features of a KM project (including implementation of a KMS) and resulting changes in KM performance. In addition to triggering more or less use of the knowledge and the KMS, feedbacks from KM effectiveness (Construct IX) might contribute to subsequent increases of user satisfaction, and possible establishments of KM-related organizational capabilities, and even changes in the contextual properties, such as the contact frequency and connection pattern in the social network. In conclusion, longitudinal studies on HGV KM practice may further refine our research model and add more insights on its anatomy.

Conclusion

Organizational knowledge management (KM) aims at effectively building and managing knowledge stocks, supporting knowledge flow and knowledge processes, building core competencies and creating intellectual capital. In the human genetic variation (HGV) research domain, KM efforts might improve the capture and dissemination of knowledge from routine HGV research activities to contribute to the global genetics knowledgebase. By synthesizing past literature, we have developed a KM research model with nine significant constructs; and we hope this paper will cast a light on future knowledge management research in the genetics domain.

Acknowledgments

We would like to thank many people for their support to this PhD project in the University of Auckland (UoA), especially the inspirational co-supervisor – Dr. Alexei Drummond, and our dearest colleagues Ms Karen Day, Dr. Martin Orr and Dr. Rekha Gaikwad. We are also indebted to Dr. Graeme Suthers, Prof. Andrew Shelling, Ms Deborah Prendergast, Ms Clare Stones and all the research students in the UoA Medical Genetics Group for their precious information and participation.

References

[1] Davenport TH, and Prusak L. Working Knowledge: How Organizations Manage What They Know, Cambridge, MA: Harvard Business School Press, 1997.

[2] Polanyi M. The Tacit Dimension, London: Routledge & Kegan Paul, 1966.

[3] Nonaka I. A dynamic theory of organizational knowledge creation. Organization Science 1994; 5 (1): 14-37.

[4] Schultze U, and Leidner DE. Studying knowledge management in information systems research: discourses and theoretical assumptions. MIS Quarterly 2002; 26 (3): 213-242.

[5] Carlsson SA, El Sawy OA, Eriksson I, and Raven A. Gaining competitive advantage through shared knowledge creation: in search of a new design theory for strategic information systems. In Proceedings of the Fourth European Conference on Information Systems, Lisbon, 1996; pp. 1067-1075.

[6] Alavi M, and Leidner DE. Review: knowledge management and knowledge management systems: conceptual foundations and research issues. MIS Quarterly 2001; 25 (1): 107-136.

[7] Sambamurthy V, Bharadwaj A, and Grover V. Shaping agility through digital options: reconceptualizing the role of information technology in firms. MIS Quarterly 2003; 27 (2): 237-263.

[8] Argote, L, McEvily B, Reagans R. Managing knowledge in organizations: An integrative framework and review of emerging themes. Management Sci 2003; 49(4 Special Issue) 571–582.

[9] The Human Genome Organisation (HUGO). Mission Statement. 2006; http://www.hugo-international.org/mission.htm, accessed at 2006-11-20.

[10] ITPerspectives Inc. NHS genetic service information systems output based specification. 2005; http://www.ngrl.org.uk/Manchester/Pages/Downloads/ITP-GEN3-OBSv05.pdf, accessed at 2005-11-03.

[11] Holtzman NA, Watson MS, eds. Promoting safe and effective genetic testing in the United States. Bethesda (MD): National Institutes of Health; 1997; http://www.genome.gov/10001733, accessed 2006-12-3.

[12] Ibarreta D, Elles R, Cassiman JJ, Rodriguez-Cerezo E, Dequeker E. Towards quality assurance and harmonization of genetic testing services in the European Union. Nat Biotechnol 2004; 22:1230-1235.

[13] Brown KL, Hutchison R, Zinberg RE, and McGovern MM. Referral and experience with genetic testing among women with early onset breast cancer. Genetic Testing Dec 2005; 9 (4): 301-305.

[14] The Karolinska Institute (Sweden) and the European Bioinformatics Institute (UK). Human Genome Variation Database. 2004; http://hgvbase.cgb.ki.se/, accessed at 2005-11-18.

[15] Laboratory of Human Genetics Montpellier (France). The Universal Mutation Database (UMD) Locus Specific Databases. 2005; http://www.umd.be/LSDB.html, accessed at 2005-11-18.

[16] National Center for Biotechnology Information (NCBI). OMIM – Online Mendelian Inheritance in Man. 2005; http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM, accessed at 2005-11-18.

[17] Horaitis O, and Cotton RGH. The challenge of collecting mutations across the genome: The Human Genome Variation Society approach, Hum Mutat 2004; 23: 447-452.

[18] Scriver CR, Nowacki PM, and Lehväslaiho H. Guidelines and recommendations for content, structure and deployment of mutation databases. Hum Mutat 1999; 13: 344-350.
[19] Scriver CR, Nowacki PM, and Lehväslaiho H. Guidelines and recommendations for content, structure, and deployment of mutation databases II: Journey in progress. Hum Mutat 2000; 15: 13-15.
[20] Genomic Disorders Research Centre. Welcome to the Waystation. 2006; http://www.centralmutations.org, accessed at 2006-11-20.
[21] Gu Y, Stanek J, and Warren J. Knowledge management (KM) technologies in medical genetic testing (MGT) laboratories – a literature review. In Proceedings of Health Informatics Conference (HIC 2006), Sydney, Australia, August 20-22, 2006.
[22] Fredman D, Munns G, Rios D, Sjöholm F, Siegfried M, Lenhard B, Lehväslaiho H, and Brookes AJ. HGVbase: a curated resource describing human DNA variation and phenotype relationships. Nucleic Acids Research 2004; 32(Database-Issue): 516-519.
[23] Gu Y, Warren J, Stanek J, and Suthers G. A system architecture design for knowledge management (KM) in medical genetic testing (MGT) laboratories. In Proceedings of the 10th International Conference on Computer Supported Cooperative Work in Design (CSCWD 2006), Nanjing, China, May 3-6, 2006.
[24] Firestone JM, and McElroy MW. Doing knowledge management. The Learning Organization 2005; 12 (2): 189-212.
[25] Dennis AR, and Vessey I. Three knowledge management strategies: knowledge hierarchies, knowledge markets and knowledge communities. MIS Quarterly Executive 2005; 4 (4).
[26] Bossen C, and Dalsgaard P. Conceptualization and appropriation: the evolving use of collaborative knowledge management. In Proceedings of the 4th decennial conference on Critical computing: between sense and sensibility CC '05, August 2005, pp. 99-108.
[27] Massey AP, Montoya-Weiss MM, and O'Driscoll TM. Knowledge management in pursuit of performance: insights from Nortel Networks. MIS Quarterly 2002; 26 (3): 269-289.
[28] Brown JS, and Duguid P. Organizing knowledge. California Management Review 1998; 40 (3): 90-111.
[29] Zupan B, Holmes JH, and Bellazzi R. Knowledge-based data analysis and interpretation. Artificial Intelligence in Medicine 2006; 37 (3): 163-165.
[30] Gold AH, Malhotra A, and Segars AH. Knowledge management: an organizational capabilities perspective. Journal of Management Information Systems 2001; 18 (1): 185-214.
[31] Becerra-Fernandez I, and Sabherwal R. Organizational knowledge management: a contingency perspective. Journal of Management Information Systems 2001; 18 (1): 23-55.
[32] Lindsey, K., "Measuring Knowledge Management Effectiveness: A Task-Contingent Organizational Capabilities Perspective," Eighth Americas Conference on Information Systems, 2002, pp. 2085-2090.
[33] Ruggles RL. Knowledge tools: using technology to manage knowledge better. Cambridge, MA: Cap Gemini Ernst & Young, Center for Business Innovation, 1997; http://www.cbi.cgey.com/pub/docs/KnowledgeTools.pdf, accessed at 2001-09-01.
[34] DeLone WH, and McLean ER. Information systems success: the quest for the dependent variable. Information Systems Research 1992; 3 (1): 60-95.
[35] DeLone WH, and McLean ER. Information Systems Success Revisited. In Proceedings of HICSS, 2002, pp. 238-248.
[36] Jennex ME, and Olfman L. A knowledge management success model: an extension of DeLone and McLean's IS success model. In Proceedings of Ninth Americas Conference on Information Systems, August 2003.
[37] Jennex ME, and Olfman L. Organizational memory/knowledge effects on productivity, a longitudinal study. In Proceedings of 35th Hawaii International Conference on System Sciences, HICSS35, IEEE Computer Society, January 2002.
[38] Jennex ME, and Olfman L. Assessing knowledge management success/effectiveness models. In Proceedings of the 37th Hawaii International Conference on System Sciences, 2004.
[39] Nonaka I. The knowledge-creating company. Harvard Business Review 1991; 96-104.
[40] Grant R. Toward a knowledge based theory of the firm. Strategic Management Journal, Winter 1996; 17: 109-122.
[41] Leonard D. Wellsprings of Knowledge: Building and Sustaining the Source of Innovation. Boston: Harvard Business School Press, 1995.
[42] Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of infoamtion technology. MIS Quarterly 1989; 13 (3): 319-340.
[43] Bots PWG, and de Bruiin H, Effective Knowledge Management in Professional Organizations: Going by the rules. In Proceedings of the 35th Hawaii International Conference on System Sciences, 2002; IEEE Computer Society Press

Address for correspondence

Yulong Gu,
Department of Computer Science (Tamaki),
University of Auckland,
Private Bag 92019, Auckland,
New Zealand.
ygu@cs.auckland.ac.nz

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

ECTracker – An Efficient Algorithm for Haplotype Analysis and Classification

[1]Li Lin, [2]Limsoon Wong, [3]Tze-Yun Leong, [4]Pohsan Lai

[1;2;3] School of Computing, National University of Singapore
[4]Dept of Pediatrics, National University Hospital, National University of Singapore

Abstract

This work aims at discovering the genetic variations of hemophilia A patients through examining the combination of molecular haplotypes present in hemophilia A and normal local populations using data mining methods. Data mining methods that are capable of extracting understandable and expressive patterns and also capable of making predictions based on inferences made on the patterns were explored in this work. An algorithm known as ECTracker is proposed and its performance compared with some common data mining methods such as artificial neural network, support vector machine, naive Bayesian, and decision tree (C4.5). Experimental studies and analyses show that ECTracker has comparatively good predictive accuracies in classification when compared to methods that can only perform classification. At the same time, ECTracker is also capable of producing easily comprehensible and expressive patterns for analytical purposes by experts.

Keywords:

datamining, classification, hemophilia A, genetic variations, haplotypes

Introduction

In this paper, we propose a new algorithm, called ECTracker[1], for pattern extraction and classification of a specific type of biological dataset known as haplotypes. A total of 47 patients affected by hemophilia A and 47 matched normal controls from Singapore were genotyped with a set of markers located on chromosome Xq28 which tags the hemophilia A disease gene. Hemophilia A is an X-linked recessive bleeding disorder that results from deficiency and/or abnormality of coagulation factor VIII (FVIII) [1]. The FVIII gene spans 186 kb of DNA and resides on 0.1% of the X chromosome (band Xq28).

We are interested in methods that are capable of performing the two tasks efficiently – first to extract expressive patterns for descriptive analysis, and second to perform classification. Intuitively, expressive haplotype patterns (or genetic variations) need to be extracted to provide medical practitioners with insights about the genetic manifestations of patients affected by hemophilia A. The extracted patterns are used for predictive inference (or classification) to help in carrier detection, which is useful for medical prognosis and decision making.

In this paper, we present the design and implementation of the ECTracker method. We also examine its performance as compared to common data mining methods in supporting the targeted tasks. Specifically, we compared the expressiveness of the haplotype patterns discovered using ECTracker with the haplotype patterns discovered using the Decision Tree method (C4.5). Furthermore, we also compared the classification predictive accuracy of ECTracker with existing classification methods including Artificial Neural Network, Naïve Bayesian Network, Support Vector Machine and Decision Tree (C4.5) [2][3].

The ECTracker method

There are two main steps in ECTracker. First, it identifies the genetic variations (or haplotype patterns) of hemophilia A patients to help analyze FVIII gene polymorphism for linkage analysis. Second, the haplotype patterns found in the first step are used to perform classification to facilitate carrier screening by medical practitioners. Details of the hemophilia A dataset will be introduced in the next section.

Step 1 – Finding interesting patterns

The first step of the ECTracker algorithm uses a level-wise neighborhood search method to enumerate all possible marker patterns of length one, two, and three etc, and then computes the statistical *odds ratio* of each of the patterns. Only those patterns that are significant are selected. The significance of a potential/candidate pattern is determined by computing its p-value. P-value calculates the probability due to chance alone of getting a difference larger than or equal to that actually observed in the data [4] [5]. A small p-value means it is difficult to attribute the observed difference to chance alone, and this can be taken as evidence against the null hypothesis of non-significance.

Odds ratio is a test statistic that has been widely used in the biomedical arena to measure the magnitude of association between two categorical variables based on some data collected [6] [7]. Given a pattern x, odds ratio computes the ratio of non-association between x and the label L, to the association between x and L based on a set of data. For example, given a pattern, say (1,3), and there are σ number of such pattern found in a dataset D associated with the class label *Abnormal* and π number of such pattern found in D associated with class label *Normal*. We are interested

1 Initial findings of this work were presented as a poster in *Asia-Pacific Conference on Human Genetics,* 2004.

in finding out whether the marker pattern (1,3) is strongly associated with the label *abnormal*. Table 1 shows the contingency table for our example where *P* is the number of samples in the dataset associated with the class label *Abnormal* and *N* is the number of samples in the dataset with class label *Normal*. The odds ratio is computed based on equation 1 defined below.

Table 1 – 2x2 contingency table

	Abnormal	**Normal**
not(*1,3*)	$P - \sigma$	$N - \pi$
(*1,3*)	σ	π

$$\text{Odds Ratio, } \theta = \frac{(P-\sigma)\pi}{(N-\pi)\sigma} \qquad (1)$$

Step 2 – Predictive inference / classification

The following describes the algorithm for predictive inference using the patterns derived from the previous step. Before presenting the algorithm, let us define the order of precedence of the derived patterns. This is used in selecting patterns for our classifier.

Definition: Given two patterns, r_i and r_j, $r_i >> r_j$ (also called r_i precedes r_j or r_i has a higher precedence than r_j) if

1. The p-value of r_i is less than the p-value of r_j, the smaller the p-value of a pattern the greater the statistical significance of that pattern.
2. Both patterns have the same p-values and $r_i \subset r_j$, the pattern length of r_i is shorter than the length of r_j. The pattern with shorter pattern length that can correctly classify an unseen case is preferred.
3. Both patterns have the same p-values and $r_i \not\subset r_j$, but r_i is generated earlier than r_j.

Let *R* be the set of patterns derived in step 1, and *D* be the training data used to derive *R*. The basic idea of the algorithm is to choose a set of high precedence patterns in *R* as our classifier. The classifier is of the following format: $<r_1, r_2, \ldots, r_n, default\_class>$, where $r_i \in R$, $r_a >> r_b$ if $b > a$. The *default_class* is the chosen class for an unseen case when no pattern in the classifier could classify the unseen case. The *default_class* can be selected by the user. However, if the user decides to let the classifier select the *default_class*, then the majority class in the data *D* will be chosen as the *default_class*.

The algorithm for building the classifier consists of five steps:

$$\text{Total score for class } C_x, \ \Omega c_x = \sum \omega_{C_x} \qquad (3)$$

Step 1: Sort the set of generated patterns *R* according to the relation ">>". This is to ensure that we will choose the highest precedence patterns for our classifier.

Step 2: For each pattern *r* in sorted *R*, if there exist another pattern *r*' such that the p-values of both *r* and *r*' are the same, and $r' \subset r$, then remove *r* from sorted *R*. This ensures that we choose the pattern with the shortest pattern length for each p-value.

Step 3: Select the first *n* patterns from sorted *R* following the sorted sequence to form the set $\Re$ for classification.

Step 4: Perform classification on the training data *D* using the *n* pattern classifier $\Re$ and compute the true positive rate of the prediction.

Step 5: If the true positive rate is less than the user defined minimum true positive rate, then repeat Step 3 and Step 4 with a different *n* value.

In classifying an unseen case in Step 4, the first pattern that satisfies the case will classify it. If no pattern applies to the case, a scoring method will be used for each of the classes, where the class with the highest score classifies the case. However, if the scoring method produces the same score for each of the available classes, then the unseen case will take on the default class. The user is able to set the default class to "unknown" to allow the classifier to make no prediction when no pattern applies to the case. This is useful when there are samples that are identical in attribute values but belonging to different classes. Figure 1 shows the pseudocode for scoring the classes.

1. for each class C_x do
2. $Score(C_x) = 0$
3. for each pattern $r_i \in \Re$ do
4. if $r_i.class == C_x$
5. compute ω_{C_x}
6. $Score(C_x) = Score(C_x) + \omega_{C_x}$
7. end
8. end
9. end

Figure 1 - Pseudocode for computing score of each class

For each pattern r_i in $\Re$ that classifies a class C_x, computes the individual pattern score using equation 2 as follows:

$$\text{Individual pattern score of } r_i \text{ for class } C_x, \ \omega_{C_x} = \frac{(nummatch)^2}{patternlength - casepatternlength} \qquad (2)$$

Where *patternlength* refers to the pattern length of r_i and *casepatternlength* refers to the pattern length of the case to be classified, and *nummatch* refers to the number of attribute matches between r_i and the case pattern. The total score for a class C_x is computed as shown in equation 3 as follows:

The unseen case will take on the class with the maximum Ω value. The objective of counting partial matches is for better noise handling.

We now describe how the scoring scheme handles noise with an example. Given that a pattern ABC is significant, its subset AB may or may not be significant since the odds ratio value is neither upward nor downward closed. However, if at a different odds ratio value, it is found that both ABC and AB are the shortest significant patterns for a class C1, this would mean that the attributes A and B are

important for determining the attributes for the class C1. Now, if we have some other patterns say ADE and DEF that are significant for another class C2. If we now have a case pattern to classify, say A, then A will have a higher score for class C1. This is the desired effect since the pattern ADE for class C2 may become significant due to noise, but it is less likely for AB and ABC to become significant due to noise.

The hemophilia dataset

A set of five common PCR-based polymorphisms located on chromosome Xq28 which tags the hemophilia A disease gene were collected and analyzed from 47 patients and 47 matched normal controls. The five polymorphisms collected are two microsatellite repeats in introns 13 and 22, and three Restriction Fragment Length Polymorphisms (RFLPs), namely BclI-intron 18, HindIII-intron 19, and XbaI-intron 22.The exact location of the markers are shown in Figure 2.

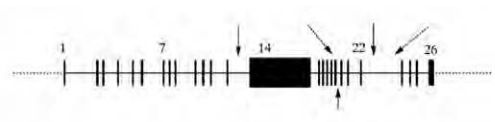


Figure 2 – Factor VIII gene

In the next sub-section, we describe the allelic frequencies of Factor VIII gene observed in our local population and the allelic frequencies reported by the authoritative resource website [8] for hemophilia A disease. The reporting of the allelic frequencies of our local population is useful for other medical practitioners not located in Singapore to decide whether they could make use of our discovery of the genetic variations for prognosis and counseling of their patients.

Allelic frequencies

The allelic frequencies observed in this study and those reported by *Hemophilia A Mutation, Structure, Test and Resources Site* [8] are tabulated in Tables 2, 3, and 4. Our results for *Bcl*I, *Hind*III, *Intron-13(CA)n* and *Intron22(GT)n(AG)n* are significantly similar to those reported in [35] with $\chi 2 < 3.841$ for *Bcl*I and *Hind*III, and $\chi 2 < 12.59$ for *Intron-13(CA)n* and *Intron22(GT)n(AG)n*. They are all within 95% confidence interval. However, the frequency for *Xba*I is significantly different from those reported by [35] with $\chi 2 > 3.841$.

Table 2 – Allelic frequencies of RFLPs

RFLPs	**Allele Frequencies (This Study)**		**Allele Frequencies (Reported by [35])**	
	(-)	(+)	(-)	(+)
	1	2	1	2
*Bcl*I	0.22	0.78	0.29	0.71
*Hind*III	0.78	0.22	0.75	0.25
*Xba*I	0.56	0.44	0.41	0.59

Table 3 – Allelic frequencies of Intron 13 (CA)n Repeats

	Allele Frequencies						
Intron 13 (CA)n Repeats	24	23	22	21	20	19	15
	1	**2**	**3**	**4**	**5**	**6**	**10**
This Study	0.01	0.10	0.06	0.26	0.52	0.04	0.01
Reported by [35]	0.013	0.05	0.11	0.29	0.45	0.07	0

Table 4 – Allelic frequencies of Intron 22 (GT)n/(AG)n repeats

	Allele Frequencies						
Intron 22 (GT)n/ (AG)n Repeats	31	30	29	28	27	26	25
	1	**2**	**3**	**4**	**5**	**6**	**7**
This Study	0.01	0.01	0.04	0.03	0.09	0.63	0.19
Reported by [35]	0	0	0	0.013	0	0.667	0.307

It was observed that samples with *Bcl*I-intron 18 allele 1 were always associated with *Hind*III-intron 19 allele 2 with $\chi 2$ p-value < 0.001. This is an expected observation as there are reported linkage disequilibrium between *Bcl*I and *Hind*III alleles from literature such as Ahrens et. al. [9] and EL-Maarri et. al. [10]. The *Hind*III marker was thus excluded since *Bcl*I and *Hind*III are in linkage disequilibrium, we could easily predict the value of the other attribute base on the value of one attribute, and hence 4 markers are sufficient in for the analysis.

It was further found that 70% of the samples had exactly the same allele values in all the markers in both patient and normal controls, which means that the 5 markers/attributes in the dataset is insufficient for separating 70% of the samples. After removing those samples whose disease and normal haplotypes cannot be distinguished, there are 28 samples remaining – 18 samples belonging to the disease phenotype and 10 samples belonging to the normal/control phenotype. Tables 5 and 6 show the frequencies of the disease and normal/control haplotypes respectively.

For descriptive analysis, we mainly report on the expressive and interesting patterns extracted from the remaining 30% of the dataset. Whereas for classification or predictive analysis, we divide the experiment into two parts: The first part compares the accuracies of the five classifiers based on the full hemophilia dataset. The second part we concentrate our study on the 30% of the dataset where those samples whose disease and normal haplotypes cannot be distinguished were removed.

Table 5 – Haplotype frequencies of probands with disease phenotype

Marker	Disease Haplotypes											Total
Intron-13 (CA)n	3	4	4	4	4	4	5	5		5	10	
*Bcl*I	1	2	2	2	2	1	2	2	2	1	2	
*Xba*I	1	1	1	2	2	1	2	1	2	1	1	
Intron-22 (GT)n/(AG)n	3	1	3	3	5	7	2	4	5	6	6	
No. of Probands	1	1	2	1	1	6	1	2	1	1	1	18

Table 6 – Haplotype frequencies of probands with normal/control phenotype

Marker	Normal/Control Haplotypes									Total
Intron-13 (CA)n	1	2	2	3	3	4	4	5	6	
*Bcl*I	1	1	2	1	1	2	2	2	2	
*Xba*I	1	1	2	1	1	1	1	1	2	
Intron-22 (GT)n/(AG)n	7	5	6	5	7	4	5	7	6	
No. of Probands	1	1	1	1	2	1	1	1	1	10

Results

Interesting pattern extraction

Expressive patterns derived by C4.5

C4.5 deduced that haplotype patterns (or genetic variations) of 4–*–*–*, 5–*–*–*, or 10–*–*–* (Intron13(CA)n–BclI–XbaI–Intron22(GT)n(AG)n) are highly associated with the disease phenotype. This derivation is not very useful as we could see from Table 6 that there are 3 probands with normal/control phenotype having intron-13 (CA)n allele values 4 and 5. Moreover, allele value 10 in intron-13 (CA)n only occurs once in the proband with disease phenotype (from Table 5), hence it is not able to give a generalize conclusion base only on allele value 10 of intron-13 (CA)n.

The possible reason for such a deduction by C4.5 may be due to the problem that the dataset is very small, and as a result the selection for partitioning attribute becomes biased for those attributes with more attribute values. Hence attributes with more attribute values will be assigned higher information gain as compared to attributes with fewer attribute values.

Expressive patterns derived by ECTracker

The longest most significant pattern associated with the disease phenotype derived by ECTracker is 4–1–1–7 (Intron13(CA)n–*Bcl*I–*Xba*I–Intron22(GT)n(AG)n). This is an interesting observation as the haplotype occurs in 33.3% of the disease phenotype and 0% of the normal/control phenotype with $\chi 2 > 3.841$ and odds ratio $\theta = 0$, which means that such observation occurs significantly greater than by chance. From Table 5, the haplotype occurs in 6 probands with disease phenotype as compare to other haplotypes which occur in no more than 2 probands. The shortest most significant patterns derived by ECTracker are 4–*–*–7 or 4–1–*–* with $\chi 2 > 3.841$ and odds ratio $\theta = 0$. This means that two markers alone were sufficient to define the disease haplotype, however, the longest most significant pattern provides a useful insight for the medical practitioners or scientists who seek to better understand the genetic variations of the disease.

This experiment shows that as compared to the decision tree approach of C4.5, ECTracker is capable of deriving useful patterns even when the dataset is very small.

Classification of the hemophilia A dataset

There are a total of 94 records in the hemophilia dataset, 47 records belonging to the class patient and 47 records belonging to the class normal. The classification methods that we examined include C4.5, Naïve Bayesian Classifier, Neural Network, Support Vector Machine and ECTracker. Except for ECTracker, all the other four classification algorithms are available from the package WEKA. WEKA is an open source data mining and machine learning software [2]. Table 7 shows the performance of various classifiers when applied to the full hemophilia dataset, and Table 8 shows the performance of the classifiers when applied to the pruned hemophilia dataset.

Table 7 – Analysis of classifiers based on full hemophilia dataset

	Accuracy	Precision of Class *Patient*	Recall for Class *Patient*	Precision for Class *Normal*	Recall for Class *Normal*
C4.5	71.43%	0.708	0.944	0.75	0.3
Naïve Bayesian Network	64.29%	0.7	0.778	0.5	0.4
Artificial Neural Network	78.57%	0.833	0.833	0.7	0.7
Artificial Neural Network	71.43%	0.75	0.833	0.625	0.5
Support Vector Machine	82.14%	0.842	0.889	0.778	0.7

Table 8 – Analysis of classifiers based on pruned hemophilia dataset

	Accuracy	Precision for Class *Patient*	Recall for Class *Patient*	Precision for Class *Normal*	Recall for Class *Normal*
C4.5	62.77%	0.615	0.681	0.643	0.574
Naïve Bayesian Network	62.77%	0.615	0.681	0.643	0.574
Artificial Neural Network	64.89%	0.675	0.574	0.63	0.723
Support Vector Machine	63.83%	0.623	0.702	0.659	0.574
ECTracker	67.02%	0.674	0.660	0.667	0.681

The classifiers were evaluated using 5-fold cross validation for the full dataset and leave-one-out for the pruned dataset. Since for each fold, different data samples are selected it is necessary run both steps in Section 2 fresh to avoid bias. Therefore, the patterns used for classification will vary for each fold.

ECTracker outperformed other classifiers with higher predictive accuracies on both full and pruned hemophilia A datasets.

On the full hemophilia dataset, the predictive accuracy of ECTracker is 67.02%, followed by Artificial Neural Network and Support Vector Machine with predictive accuracies of 64.89% and 63.83% respectively. Both Naïve Bayesian Network and C4.5 have the same predictive accuracy of 62.77%.

On the pruned dataset, ECTracker was able to accurately predict the phenotype of a sample given its polymorphic markers 82.14% of the time, followed by Artificial Neural Network at 78.57% of the time. C4.5 and Support Vector Machine were able to make accurate predictions 71.43% of the time.

There are 94 records in the unprune hemophilia A dataset, which is reasonably large, and there are 23 records in the pruned hemophilia A dataset which is rather small. Our experiments show that ECTracker is capable of providing good classification accuracy on both small and large datasets when compared to other classification methods.

Conclusion

In this work, we explored methods that are capable of extracting understandable and useful patterns, and also capable of performing inference on the patterns to make prediction. We applied these methods to find the genetic variations of a real dataset consisting of patients affected by hemophilia A to facilitate haplotype analysis by medical practitioners. We examined the issues of descriptive and predictive analyses using our proposed method called ECTracker.

In descriptive analysis, ECTracker is capable of extracting comprehensible and useful patterns from the hemophilia A dataset. Comparing with the patterns derived by C4.5, the patterns derived by C4.5 are less useful, as described earlier.

In predictive analysis or classification, ECTracker is capable of producing good predictive accuracies in classification that are comparable to those methods that only perform classification such as Artificial Neural Network and Support Vector Machine.

The experiments have indicated that ECTracker is potentially an effective method for both pattern extraction and classification for biomedicine in particular, and datamining in general.

The approach proposed here provides analysis and classification based on mainly the disease status of an individual. Continuously distributed quantitative traits such as blood pressure and choltesterol level may also be of significance to the clinicians. ECTracker can be extended to perform analysis and classification based on continuously distributed quantitative traits by defining a new scoring method for the interesting patterns. Further investigation will need to be done to assess the feasibility of such extension.

References

[1] Antonarakaris S, Kazazian H, Tuddenham E. Molecular etiology of factor VIII deficiency in hemophilia A. Human Mutation 1995; 5:1-22.

[2] Witten IH, Frank E. Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations. Morgan Kaufmann; 2000.

[3] Michie D, Spiegelhalter D, Taylor C. Machine Learning, Neural and Statistical Classification. Ellis Horwood: New York; 1994.

[4] Ott L. An Introduction to Statistical Methods and Data Analysis. PWS-Kent Publishing Company, Massachussetts; 1988.

[5] Theodorou, S.J., Theodorou, D.J. and Kakitsubata, Y. Statistical Analysis in Clinical Studies: An Introduction to Fundamentals for Physicians. Internet Medical Journal, 2004.

[6] Le Moual N, Siroux V, Pin I, Kaufmann F, Kennedy SM. Asthma Severity and Exposure to Occupational Asthmagens. American Journal of Respiratory Critical Care Medicine 2005.

[7] Deheinzelin D, Lourenco MT, Costa CL, Younes RN. The Level of Nicotine Dependence is an Independent Risk Factor for Cancer: A Case Control Study. Clinics 2005; 60(3):221-226.

[8] HAMSTeRS *Haemophilia A Mutation, Structure, Test and Resource Site* http://europium.csc.mrc.ac.uk/WebPages/Main/main.htm.

[9] Ahrens P, Kruse TA, Schwartz M, Rasmussen PB, Din N.A. New HindIII Restriction Fragment Length Polymorphism in the Hemophilia A Locus. Human Genetics 1987; 76(2):127-128.

[10] EL-Maarri O, Kavakli K, Caglayan H. Intron 22 Inversions in the Turkish Haemophilia A Patients: Prevalence and Haplotype Analysis. Haemophilia 1999; 5(3):169-173.

Address for correspondence

Email: hi.linli@gmail.com

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

A Dynamic Query System for Supporting Phenotype Mining in Genetic Studies

Angelo Nuzzo, Daniele Segagni, Giuseppe Milani, Carla Rognoni, Riccardo Bellazzi

Department of Computer Science and Systems, University of Pavía, Pavía, Italy

Abstract

This paper describes an information technology infrastructure aimed at supporting translational bioinformatics studies that require joint management of phenotypic and genotypic data. In particular, we integrated an electronic medical record with an open-source environment for data mining to create a flexible and easy to use query system aimed at supporting the discovery of the most frequent complex traits. We propose a logical formalization to define the phenotypes of interest; this is translated into a graphical interface that allows the user to combine different conditions relative to the electronic medical record data (e.g., the presence of a particular pathology). The phenotypes are then stored in a multidimensional database. Then, the data mining system engine reads the filtered data from the database and executes dynamic queries for analyzing phenotypic data, presenting the results in a multidimensional format through a simple web interface. The system has been applied in a study on genetically isolated individuals, the Val Borbera project.

Keywords:

phenotype mining, complex traits, intelligent query, clinical data warehouse.

Introduction

A specific characteristic of the post-genomic era will be the correlation of genotypic and phenotypic information [1][2]; the emerging discipline of Biomedical Informatics may provide knowledge and tools for dealing with such an ambitious goal [3]. In this context, the studies aimed at the so-called genetic dissection of complex traits represent a first crucial benchmark for Biomedical Informatics and for translational bioinformatics.

The definition of an Information Technology infrastructure to support this kind of studies, and in particular the studies aimed at the analysis of large sets of phenotypes to discover the most prevalent diseases and then to integrate genotypic information, is a challenge which can be considered a paradigmatic goal of Biomedical Informatics. As a matter of fact, research on phenotypes requires the definition of an architecture for data collection, the implementation of an electronic medical record, the development of a system for the definition of the phenotypes of interest, and the design and implementation of a data warehouse system for analyzing phenotypic data. Moreover, selecting which are the phenotypes of interest determines the subsequent genotyping choices, especially when a genome-wide scan is not feasible or suitable.

Once clinical data are collected, it is crucial to perform a series of queries and data aggregation steps to characterize the population and extract the most prevalent phenotypes. However, clinicians, biologists and epidemiologists are usually unable to explore the collected information, because the use of general query languages requires substantial technical skill, as well as knowledge of the underlying database structures. On the other hand, the need of performing "dynamic" queries hampers the implementation of a "standard" user interface for pre-defined queries. To address this problem, we are defining a dynamic query system based on data warehouse and mining concepts. The clinical data are copied into a data mart oriented to data analysis. Thanks to the integration in the overall system of open-source environment for data mining it is possible to design a simple interface for performing aggregation, counting and simple statistics on the majority of the variables contained in the clinical database.

The implementation of the system has a number of steps which follows a workflow targeted at identifying the most important phenotypes which characterize a particular population. Such workflow includes the following steps:

1. The development of a relational database collecting clinical data on the target population
2. The translation of the database structure into a multidimensional data-base (data mart) oriented for query and reporting
3. The formal definition of the phenotypes to be searched and studied and their mapping in the database
4. Finally, the design and implementation of the data mining tool to easily extract the phenotypes and analyze their relationships

The aim of the paper is to describe the IT infrastructure of the system and the biomedical informatics challenges we are dealing with. Preliminary results on the use of the system to support a study on genetically isolated individuals, the Val Borbera project [4], will also be reported.

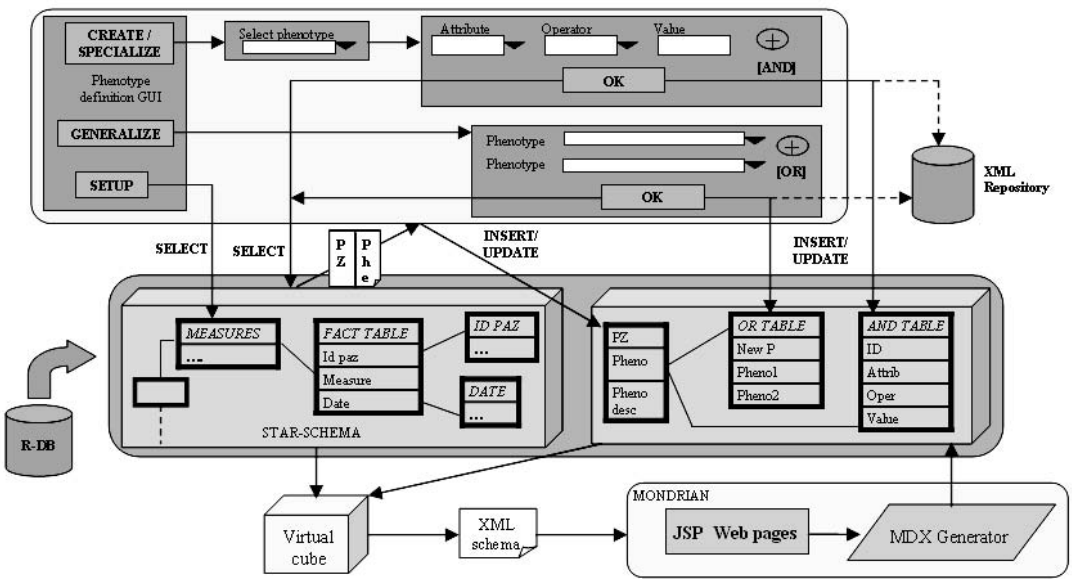


Figure 1 - Overview of system's components an the interactions among the 3 layers

Materials and methods

Objective

The final purpose of the system that we have developed is to provide a versatile and easy to use data inspection tool to identify which phenotypes may be more successfully investigated in the population under study, so that suitable genotyping choices may be made subsequently. The system accomplishes this purpose by providing tools for two main tasks:

1. for formally define the phenotypes to be investigated by a graphical user interface
2. for exploring clinical data to extract and analyze individuals with the same phenotypes (as they have been previously defined)

Both aspects can be performed in a database where the information is stored in a non-normalized data structure, which is often referred to as data mart. As a matter of fact, while a normalized structure is always required for correct management of the database (as it concerns data security, integrity, reliability), the database is greatly facilitated by the use of query-oriented data schema. That's why the basic common layer of our implemented system is a non-normalized data structure, the so-called "star schema", to which the two developed applications for query definition and data assessment are interfaced. Details of the components and their interactions are described in the following paragraphs, while a global overview is given in Figure 1.

From the relational DB to the "star-schema"

The typical data structure that has become a standard for all data warehouse applications is a multi-dimensional model called the "star join schema". Unlike the Entity-Relation model, the dimensional model is asymmetric. There is one large dominant table in the center of the schema, called the fact table. It is the only table in the schema which is connected to the other tables with multiple joins. Such other tables, called the dimensional tables, only require a single join to be referenced by the fact table [5].

Typically a clinical database can be modeled by a star schema in which each record in the fact table represents a combination of a clinical measure and its values on a specific date for a specific patient. So the dimensions are individuals, measurement time and measurement values: all of them can be further specified using a snowflake model[1].

The adaptation of the star-schema and of the snowflake models to the clinical context requires however several efforts. In fact, when taking into account the phenotype information in the analysis, we cannot model it as a dimension of the fact table, because it would be a *non-additive* dimension with respect to the others. *Additivity* is the ability to use an aggregate operator (summation, counting, average) along the dimensions of the same fact [6]: in our case, the phenotype dimension would be additive only along the patient dimension, but not along the others (measurements and time), as it is defined by a set of measurements. To overcome this problem, we have defined a new fact table to model the relationship between phenotypes and individuals.

The star schema and the phenotype tables are the physical models of the new multidimensional database. They represent two multidimensional "cubes" that together form the logical model of the database. The cubes may be merged in a single "virtual" cube, so that it is possible to use an Online Analytical Processing (OLAP) engine to perform

1 A *snowflake model* is a model in which a given dimension has relationships to other levels of the same dimension. It is used to re-normalize complex dimensions to eliminate redundancy.

data analysis (described in detail in the following paragraphs).

The phenotype definition tool

Clinicians and biologists usually define a phenotype by a set of variables and the values they may take. In order to select (and then to analyze) the individuals satisfying that set of rules, it is necessary to write a suitable SQL statement to run a query to retrieve them. However, as the users may have no expertise in the use of a query scripting language, we provide a tool that automatically generate the proper SQL script to select individuals with the defined phenotype.

To perform this task, a formalization of the phenotype definition is needed. The basic assumption is to consider a phenotype as a set of conditions in the form of attribute/value pairs. Then using logical operators (AND, OR) it is possible to combine different conditions to define more and more complex phenotypes. In particular, the AND operator allows the specialization of a defined phenotype, while the OR operator is used to merge different phenotypes into a single, more comprehensive one. This procedure corresponds to a logical tree construction, in which the nodes are the conditions, the AND operator is used to go from the top to the bottom and the OR operator is used to add an upper node from the bottom to the top (figure 2).

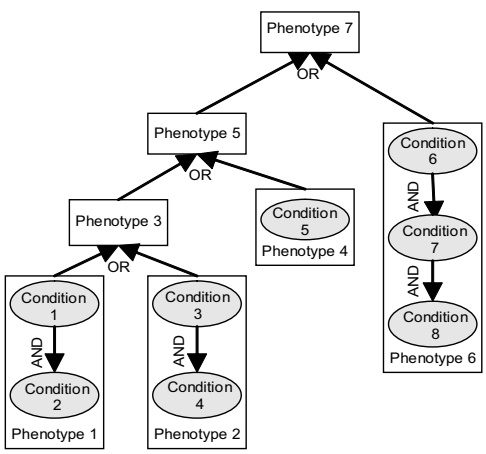


Figure 2 - Formalization of the phenotype definition

The limitation of this model is that it is not possible to define any kind of AND/OR combination of the conditions under investigation, as in the corresponding SQL string, it will not be possible to interpret the correct order of the subgroups (i.e., the correct position of brackets in the query string). So it is necessary to define the AND-conditions first, and then combining them using the OR operator.

The end users can create these definitions using a graphical wizard developed in the JAVA programming language. It interacts with the two sections of the non-normalized database, i.e., the star schema of the clinical data and the phenotype definition tables. The wizard is automatically setup by reading an XML file in which the two data structures are encoded. The conditions (attributes and their values) may be defined by combo boxes, which provide lists of attributes according to the measures table of the star schema, suggesting the admissible ranges of values for each attribute. Once the rules are defined, the corresponding SQL string is created by merging conditions by AND operators, in order to: i) store the rules in the phenotype section tables, and ii) select the subgroup of individuals satisfying that conditions and storing the relation individuals-phenotype. In the same way, it is possible to choose another graphical panel to selected some defined phenotypes to be merged together by the OR operators to create a new phenotype, which is stored in the phenotype section.

The data analysis tool

Dealing with clinical data to analyze phenotypic information implies taking into account heterogeneous data and viewing them at the same time. This means that it should be possible to perform a multidimensional inspection of the dataset. Whereas a relational database stores all data in the form of rows and columns, a multidimensional dataset consists of axes and cells organized in multidimensional "cubes", the dimensions of which are the directions of drill-down investigations.

The technique of multidimensional analysis is implemented in software tools called online analytical processing (OLAP) engines., OLAP, in fact, means analyzing large quantities of data in real-time. Unlike Online Transaction Processing (OLTP), where typical operations read and modify individual and small numbers of records, OLAP deals with data in bulk, and operations are generally read-only. The term "online" implies that even though huge quantities of data are involved — typically many millions of records, occupying several gigabytes — the system must respond to queries fast enough to allow an interactive exploration of the data.

As described above, the logical model of the star schema and the phenotype tables consist of two virtual cubes that may be merged in a single one, which is the input for the OLAP engine. In our system we use an OLAP engine written in the Java programming language: Mondrian [7]. It executes queries written in the MDX language (that has actually become a standard for data warehouse applications) [8], reads data from a relational database, and presents the results in a multidimensional format through a Java API, so that the presentation layer may be chosen by the final user. JSP pages are provided by default, so that the user can simply use a web browser for data visualization. The MDX queries have to be defined by the user, so we have developed a specific module (the "MDX generator" box in figure 1) that automatically create the MDX scripts directly from the attributes of interest chosen by the check box lists of the main page.

Sharing and generalizing issues

In order to make the graphical wizard usable on different databases, it has been made configurable via an XML file containing the star schema description. So the only prereq-

uisite for using it is to provide a star schema that is compliant to the model described above. Then the phenotype tables are automatically generated and populated by the GUI. Moreover, the phenotype definitions are also stored in XML files, so that existing phenotypes loading is performed by reading the XML files instead of the tables, and the XML repository may be shared with other scientists interested in analogue analysis. On the other side, whichever OLAP engine will be chosen, the only manual task needed is to code the virtual cube in the specific format required as input by the engine. Using Mondrian, it means to create the XML file containing the definition of the cube.

Results

The tools described in the previous section have been tested for the exploration of the clinical database of the Val Borbera genetically isolated population project [4]. This study is conducted in collaboration with the DIBIT of San Raffaele Scientific Institute of Milan, for which we have provided the architectural IT infrastructure for data collecting and storing. The clinical data have been collected in a relational database in a high normal form, actually containing about one hundred clinical measures relative to more than 4000 individuals. So the first step was to create the correspondent star-schema for the multidimensional analysis. Here we present an example regarding the analysis of dysfunctions related to the thyroid.

Phenotype definition

Before the use of a multidimensional analysis approach, the biologists had to ask a technician to extract individuals with the traits of interest by writing the SQL statement to be executed on the relational database. The statements are often some hundreds of lines long, due to the large number of join to be performed. Using the developed infrastructure, in contrast, the user has only to define the conditions by the graphical window shown in figure 3, that will be merged by the AND operators. The combo box are automatically filled in by reading the XML file in which the star schema is encoded.

Figure 3 - Creation/specialization panel

The defined phenotypes are summarized in the main panel of the GUI. The phenotype list shows the XML files that have already been created; details of the rules applied to define it are given below by selecting one of them (figure 5).

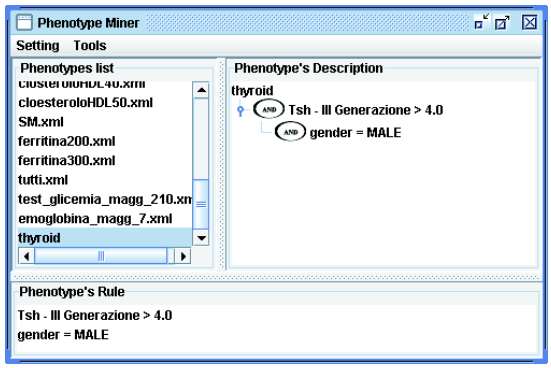


Figure 4 - Main panel: existing phenotypes are shown and details for the selected one

So if the user wants to specialize an existing phenotype, he can select it from the "Phenotype's Rules" window (figure 3) and add conditions using the AND operators.

Otherwise, if some defined phenotypes have to be merged together by OR operators to create a new phenotype definition, they can be selected using the "generalization panel", as shown in figure 4. In this case the combo box are filled by reading the XML phenotype files stored in the suitable repository.

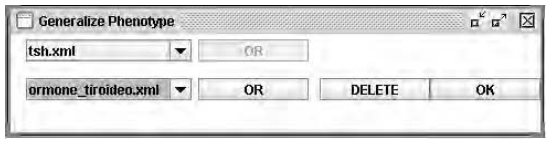


Figure 5 - The generalization panel

In both cases, the selection of individuals with the specified phenotypes is performed, and the relations are stored in the proper table. So the phenotype becomes a dimension which can be used to explore individual's data.

Data analysis and validation

When the OLAP engine starts up, it reads the XML file containing the cube definition. The first page shows a set of check boxes containing the fields of the underlying tables, so that the user may choose the variable to be investigated (the phenotypes are among them). Once the features have been chosen, the engine loads information related to the individuals having that phenotype. Then a visual inspection of the measurements values can be done expanding or collapsing cells of the resulting table, so that the analysis can be executed at different levels of detail (figure 6). Automatic graphical reports can also be generated.

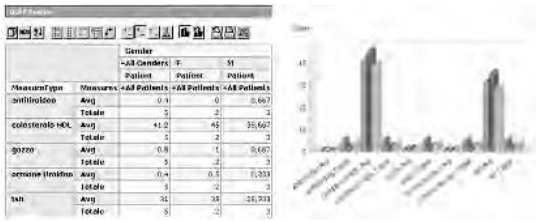


Figure 6 - The resulting dynamic table provided by Mondrian: the user may expand or collapse the cells to view the data at different levels of detail

A first validation of the system is now undergoing; it consists of verifying how it improves the phenotype searching process. Biologists formulate several verbal requests, we then compared the results obtained by themselves using the system and the ones obtained by a technician performing a SQL statement (which is typically quite a lot of instruction lines long). The first initial report table is always the same for both tasks and for any condition exploited (many different phenotypes corning thyroid diseases, hypertension and diabetes have been defined and searched): this means that the query editor correctly runs. The advantage provided by the OLAP engine is then the capability of allowing a dynamic interaction with the data to perform a more detailed exploration.

Conclusion

Current genetic studies are characterized by the collection of huge quantity of both clinical and genotypic data. The final goal of such effort is the "genetic dissection" of complex phenotypes; the first challenge of data analysis is therefore to identify which phenotypes must be investigated. However, this task may be difficult to be performed using "standard" tools for database navigation, such as SQL query, as they require technical skills for the end user to extract interesting information. In order to solve this problem, in this paper we present a dynamic query system based on data warehouse and mining concepts, which allows phenotypes definition by a graphical user interface and data exploration using OLAP tools. Both applications are based on a common underlying data layer, the structure of which is a data mart oriented to data analysis. The phenotype definition is based on a logical formalization and it is properly stored to be processed by the OLAP engine. Once the phenotypes have been defined, the OLAP engine allows the user to perform a visual inspection of the data through a set of results dynamically created. We have chosen an open source OLAP engine, Mondrian, and we integrated it with new components (the MDX generator) in order to automate other specific technical operations.

All the modules of the system are configurable via XML files, so that they can be reused to analyze other clinical databases. The only requirement is to translate the data structure into the data mart described in this paper, and to codify it in an XML file. Moreover, the phenotype definitions are also stored in an XML repository, so they may be reused and shared with other users to compare results.

The system as been tested on a real dataset, the clinical database of the Val Borbera project, showing that it is easy to use and time-saving. Future development of the system will improve the graphical user interface. The phenotype definitions will be shown by graphs, corresponding to the logical model used to create it, so that the user may expand it directly by adding or removing nodes.

Acknowledgments

We gratefully acknowledge Cristiana Larizza for software development supervision and DIBIT group (D. Toniolo, C. Sala, C. Camaschella) for providing data of the Val Borbera project.

References

[1] Lander ES, Schork NJ. Genetic dissection of complex traits. Science. 1994 Sep 30; 265(5181):2037-48.

[2] Botstein D, Risch N. Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. Nature Genetics, 2003 Mar; 33:228-237.

[3] Martin-Sanchez F, Iakovidis I, Norager S, Maojo V, de Groen P, Van der Lei J, Jones T, Abraham-Fuchs K, Apweiler R, Babic A, Baud R, Breton V, Cinquin P, Doupi P, Dugas M, Eils R, Engelbrecht R, Ghazal P, Jehenson P, Kulikowski C, Lampe K, De Moor G, Orphanoudakis S, Rossing N, Sarachan B, Sousa A, Spekowius G, Thireos G, Zahlmann G, Zvarova J, Hermosilla I, Vicente FJ. Synergy between medical informatics and bioinformatics: facilitating genomic medicine for future health care. JBI, 2004 Feb; 37(1):30-42.

[4] Sala C, Bione S, Crocco L, Gatti M, Poggiali E,. Bellazzi R, Buetti I, Rognoni C, Camaschella C and Toniolo D. The Val Borbera Project: epidemiological and genealogical analysis of an isolated population in Northern Italy. Submitted to European Society of Human Genetics, 2006.

[5] Kinball R, Ross M. The Data Warehouse Toolkit, Second Edition. Wiley and Sons, Inc., 2002.

[6] Wyderka KC. Data Warehouse Technique for Outcomes Management. Health Management Technology, Nov. 1999; 20 (10):16-7.

[7] Hyde J, Mondrian OLAP project, Pentaho Analysis Service. http://mondrian.pentaho.org/.

[8] Spofford G, Harinath S, Webb C, Huang DH, Civardi.F. MDX Solutions, Second Edition. Wiley Publishing, Inc., 2006

Address for correspondence

angelo.nuzzo@unipv.it

This page intentionally left blank

Chapter 10.

Biomedical Image and Signal Processing

This page intentionally left blank

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Identifying QT prolongation from ECG impressions using Natural Language Processing and Negation Detection

Joshua C. Denny, MD[a,b], Josh F. Peterson, MD[a,b]

Departments of [a]Biomedical Informatics and [b]Medicine, Vanderbilt University Medical Center, Nashville, TN, USA

Abstract

Electrocardiogram (ECG) impressions provide significant information for decision support and clinical research. We investigated the presence of QT prolongation, an important risk factor for sudden cardiac death, compared to the automated calculation of corrected QT (QTc) by ECG machines. We integrated a negation tagging algorithm into the KnowledgeMap concept identifier (KMCI), then applied it to impressions from 44,080 ECGs to identify Unified Medical Language System concepts. We compared the instances of QT prolongation identified by KMCI to the calculated QTc. The algorithm for negation detection had a recall of 0.973 and precision of 0.982 over 10,490 concepts. A concept query for QT prolongation matched 2,364 ECGs with precision of 1.00. The positive predictive value of the common QTc cutoffs was 6-21%. ECGs not identified by KMCI as prolonged but with QTc>450ms revealed potential causes of miscalculated QTc intervals in 96% of the cases; no definite concept query false negatives were detected. We conclude that a natural language processing system can effectively identify QT prolongation and other cardiac diagnoses from ECG impressions for potential decision support and clinical research.

Keywords:

electrocardiogram, Unified Medical Language System, natural language processing, concept identification, decision support.

Introduction

Electrocardiograms (ECG) provide significant medical information that has been largely untapped in medication decision support interventions. ECGs are commonly used to help diagnose cardiac diseases such as myocardial infarction and ischemia, arrhythmias, and cardiomyopathies as well as some extracardiac diagnoses. Recent studies have highlighted the risk of sudden cardiac death due to medications known to increase QT intervals [1, 2]. Medications can also affect the risk of myocardial infarction [3, 4], induce second and third degree atrioventricular block [5], and cause other, potentially fatal, arrhythmias [1]. A decision support system that warns providers when prescribing medications to patients with existing risk factors, such as QT prolongation or atrioventricular block, may be valuable in guiding prescription choice.

QT prolongation is a key risk factor for development of Torsades de Pointes, a potentially fatal cardiac dysrhythmia. Since the QT interval, the measurement of the time between ventricular contraction ("QRS complex") and its repolarization ("T wave"), varies with heart rate, QT prolongation is typically assessed via a QT corrected for rate (QTc) [6]. A value greater than 440-460 ms is typically considered prolonged. Many drugs are known to prolong the QT interval; it is the most common reason for a drug to be removed from the market [2, 7]. However, many other factors can influence measurement of the QT (and thus QTc), including arrhythmias, intraventricular conduction disturbances, ECG measurement technique, and morphological changes in the ECG.

ECG findings consist of two types: morphologic descriptions (e.g., QT prolongation or widened QRS) and interpretations of those findings (e.g., myocardial infarction, atrial fibrillation, or ventricular hypertrophy). While many have developed automated feature extraction programs based on ECG waveforms, automated algorithms are imperfect, with accuracies of 42-96% [8, 9]. These algorithms are generally superior for morphological descriptions than their interpretations [8, 9]. However, many factors, such as an arrhythmia or ischemia, can alter the accuracy of morphological descriptions as well. For these reasons, cardiologists' interpretations of ECGs remain the consensus gold standard [8, 10, 11].

Natural language processing and concept-based indexing to standardized vocabularies such as the Unified Medical Language System (UMLS) have been applied to radiology reports [12, 13], clinical notes [14, 15], and medical education documents [16], among others. Previously, we reported the use of the KnowledgeMap concept indexer (KMCI) to identify UMLS concepts from cardiologist-generated ECG impressions [17]. KMCI is a general purpose concept identifier, using rigorous, scored-based algorithms to identify concepts from free text [16]. It accurately identifies unknown abbreviations, acronyms, and underspecified concepts (e.g., document phrase "1st degree block" for the closest UMLS match "1st degree *atrioventricular* block"). KMCI scores ambiguous concept matches using the context of other concepts matching around it to favor candidates that are likely to co-occur. Previously, we optimized and evaluated its performance on ECG impressions, finding a recall of 0.90 and precision of 0.94 [17]. It was especially accurate for myocardial perfusion changes, ECG rhythms, and extracardiac manifestations (recall and precision in excess of 0.98). This system, however, did not have the ability to detect negated or possible findings.

Table 1 – Comparison of KnowledgeMap Concept Identifier (KMCI) identification of negation to gold standard physician review

	Gold standard		
KCMI	Negated or possible findings	Positive or Probable findings	**Total concepts**
Identified as negated or possible	722 (6.9%)	13 (0.1%)	735 (7.0%)
Identified as positive	20 (0.2%)	9725 (92.8%)	9745 (93.0%)
Total Concepts	742 (7.1%)	9738 (92.9%)	10480

In this paper, we report on the integration of negation detection algorithms into KMCI and its initial application on a four-year collection of ECGs to identify QT prolongation. The ultimate goal is a codified database of ECG impressions for the development of a medication decision support system.

Methods

Creation of ECG database

Vanderbilt University Medical Center has developed an anonymized database of all orders, laboratory results, and ECGs for all inpatients admitted for 2-30 days from 1999-2003 as part of an ongoing research study investigating drug effects. The ECGs were imported in an XML format from an ECG management system. Every ECG includes machine calculated intervals and heart rate as well as a cardiologist-generated free-text impression. Cardiologists create an impression for all ECGs by selecting among personalizable stock phrases ("normal sinus rhythm") and editing as necessary ("normal sinus rhythm *with rare PVCs*"), or typing comments *de novo* ("LA abnormality, PVCs, and inferolat ST-T changes"). Finally, cardiologists code each ECG with a standard severity: normal, otherwise normal, borderline normal, or abnormal. We extracted all ECGs from our research repository and loaded them into a relational database. There were 44,808 ECGs in the database with more than 155,000 sentences.

From this dataset, we randomly selected a test set of about 5,000 sentences for development of our negation algorithm. Another 5,000 randomly selected sentences were reserved as a validation set. All subsequent data analysis was performed on the entire dataset of ECGs.

Negation tagging algorithm and evaluation

Many authors have developed negation tagging algorithms [18-20]. We applied a modified version of the NegEx algorithm [18] using regular expressions. We marked phrases as "negated", "possible", or "asserted." For the purposes of decision support, we included findings indicated as "probable" or "likely" as asserted findings. We used a total of 205 phrases indicating negation or possibility, including symbols such as "?" and "r/o" as indictors of "possible." We used a window of 8 words before or after a negating phrase. In this dataset, we found that most periods marked abbreviations rather than sentences. Semicolons, unmatched closing parentheses, and other negating phrases terminated the current negation window.

One author (JP, an internist), unfamiliar with the NegEx algorithm or its implementation, scored 5000 sentences from randomly selected ECGs via a color-coded HTML interface that highlighted the negating phrase and words modified by them. Only medical concepts or medical modifiers were considered for scoring. Concepts were marked as a correctly identified negated concept (i.e., a true positives, TP), false positives (FP), true negatives (TN), and false negatives (FN). We calculated recall of negated concepts as TP/(TP+FN); precision as TP/(TP+FP); and negative predictive value, the probability that the concept is not negated (i.e., an asserted finding), as TN/(TN+FN). Following the evaluation, three new negation phrases were added and validated over several thousand sentences with an improvement in matching before application to the entire dataset.

Development of concept-based ECG database

We applied KMCI to identify Unified Medical Language System (UMLS) concepts from the free text ECG impres-

Table 2 – Concepts employed to identify QT prolongation within ECG impressions.

Concept query (Asserted and Possible Concepts only)				**ECG matches**	**True Positives**
1	C0151878 Prolonged QT interval			2294	2193
2	C1560305 Prolonged QTc interval			55	55
3	C0023976 Long QT syndrome			2	2
4*	C0429028 QT interval C0489625 QTc interval C0860814 QTc	with any of	C0205166 Long C0439590 Prolonged C0392744 Lengthened	100	100
Total unique ECGs				**2,364**†	**2,363**†

*For query 4, one concept from each list must occur within the same sentence to be considered a match.
†Some ECGs matched more than one concept query (e.g., "QT prolongation. Compared with [date], the QT has lengthened.")

Table 3 – Comparison of QT prolongation identified in ECG impression to automated QTc by ECG machine.

	Concept query	QTc > 400	QTc > 450	QTc > 500	QTc > 550
ECGs matching criteria	2,364	34,059	11,804	2,518	620
KMCI positives	2,364	2,357	2,304	539	117
Sensitivity		1.00	0.98	0.23	0.05
Specificity		0.19	0.77	0.95	0.99
Positive predictive value		0.06	11,804	2,518	620
Negative predictive value		1.00	0.20	0.21	0.19

sions, using the optimizations identified in prior study [17]. We added synonyms and derivational transformations to KMCI's lexicon and modified the sentence-identification algorithm to ignore most spaces and periods when determining sentence breaks. We used the 2006AC version of the UMLS [21]; the only restriction on concept matching is favoring underspecified concepts with words such as "heart" and "electrocardiogram" (see [17] for a full list). Candidate UMLS concepts with these words are penalized less than candidates with other words when the words do not match a document word. We applied the negation algorithm following the concept identification step to mark each concept as positive, possible, or negated. The concept-identified ECGs are linked to the original ECG impressions and calculated intervals, forming the identified ECG dataset.

Analysis of QT prolongation

Through perusal of the ECG dataset, the authors identified the UMLS concepts representing "QT prolongation," including any text indicating a probable or possible QT or QTc prolongation. To verify that we had found all concepts representing QT prolongation, we also did text searches for all string matches with "QT" or "QTc." Each was manually verified by analyzing all unique strings containing the concept. We extracted these concepts, ignoring negated concepts, along with the automated QT and QTc intervals identified by the ECG management system. We compared the predictive value of the computer calculated QT and QTc intervals (a continuous number) via the area under the receiver operator characteristic curves (AUC) using the cardiologist interpretation as the gold standard and also by the positive predictive value of the commonly-used cutoff values of 400ms, 450ms, 500ms, and 550ms.

To assess the negative predictive value of our query, we reviewed a random sampling of 100 ECGs with calculated QT or QTc intervals > 450 for potential reasons why the QT calculations may be incorrect and to ascertain if any positive ECGs were missed by our concept query. Since the ECG images are not stored in our anonymized database, we were only able to evaluate the ECG intervals and raw text of the cardiologist impression. Student t-tests were used to compare parametric data. AUC and statistical analyses were performed with Stata, version 9.2 (StataCorp LP, College Station, TX).

Description of identified ECG database

To investigate the possibility for the database for decision support, we developed preliminary queries for a number of diagnoses that may be of interest for decision support (see Table 4). Each of these topics requires an interpretation of the free text. Topics were selected by finding the UMLS concepts representing the topic of interest in the database of matched concepts. For myocardial infarction, this involved the tree of concepts related to "myocardial infarction" and "infarct."

Results

Table 1 shows the results of the negation analysis. The 5,000 sentences in the negation test set contained a total of 10,480 UMLS concepts. Overall recall was 0.973 and overall precision 0.982. The negative predictive value of finding negation (probability that a statement was positive given it was identified as positive) was 0.998. All false negatives were due to three phrases not present in the regular expression list: "replaced <negated concept>", "<negated concept> replaced by", and "<negated concept> is/are gone." Several of the false positives were instances in which negating phrases were amid multiple concept words (e.g., "ST *no longer* depressed," in which the negated concept "ST depression" is separated by the negation phrase "no longer"). KMCI typically identified the correct UMLS concept for these phrases. Misspellings also caused some errors.

KMCI identified 375,838 concepts from the 44,080 ECGs in the database mapping to 23,080 unique admissions. Cardiologists identified 70% percent of the ECGs as "abnormal," 12% as "borderline," and 18% as "normal" or "otherwise normal." Of identified concepts, 339,554 (90.3%) were asserted, 29107 (7.7%) possible, and 7177 (1.9%) negated.

Table 2 shows the concepts used for the QT prolongation query and frequency of each in the database. There were 254 unique strings with a range of 2-18 words (median 11 words, weighted median 5 words) matching the QT prolongation query; 15 of these (e.g., "QT interval long for rate") accounted for nearly 90% of all matching impressions. The overall precision was 2363/2364 (1.00). Table 3 shows the results of different methods of predicting prolonged QT intervals; 2,364 ECGs (5.3% of all ECGs) were identified as representing QT prolongation by our concept query. The average QTc interval for those with prolonged QT intervals was 487 ms (range 363-716 ms); ECGs without mention of QT prolongation averaged 429 ms ($p<0.001$, range 46-785 ms). Overall, the calculated QT interval had an AUC of 0.73 for predicting QT prolonga-

Table 4 – Number of ECGs expressing potential targets for decision support

Concept	Total (%)	Positive	Possible	Negated
Myocardial infarction (MI)	6,355 (14)	3,381	2,919	55
Acute MI	208 (0.5)	167	36	5
Myocardial ischemia	2,312 (5.2)	1,855	444	13
ST segment elevation	1,015 (2.3)	895	92	28
Wolff-Parkinson-White bypass tract	107 (0.2)	93	7	7
Atrioventricular block, 1st degree	2,461 (5.5)	1,876	571	14
Atrioventricular block, 2nd degree	61 (0.1)	58	1	2
Atrioventricular block, 3rd degree	24 (0.05)	23	1	0
Pericarditis	105 (0.2)	43	62	0
Atrial fibrillation	1,748 (4.0)	1,719	16	13
Atrial flutter	397 (0.9)	373	20	4
Total number of ECGs	**44,080**			

tion; the QTc interval AUC was 0.91. Of the 100 manually-reviewed ECGs with QTc intervals longer than 450 ms but that were negative by concept query, 32% had a bundle branch block; 24% had various ST segment or T wave abnormalities; 24% had an arrhythmia, aberrant complexes, or a pacemaker; and 23% had myocardial ischemia or infarct. No ECGs contained comments suggesting QT prolongation. Only 4 ECGs had no significant electrocardiographic abnormalities that could not alter calculation of the QT interval.

Table 4 shows concept findings over the entire database.

Discussion

We studied the application of a concept-based, natural language processing system to identify QT prolongation within cardiologist-generated ECG reports. Commonly used cutoffs for diagnosing QT prolongation from the QTc intervals calculated by ECG machines were poor, with positive predictive values of only 6-21% when compared with the NLP-based approach. Physician review of ECGs with long QTc by calculation but not by the NLP query found 94% of these ECGs had findings that would confound automatic QTc calculation. No ECGs not identified by KMCI were identified as prolonged by the cardiologist. KMCI is a more accurate means of identifying QT prolongation than automated interval analysis. Due to the high provider override rates in most medication decision support systems, due in part to poor specificity [22], a medication decision support system for QT prolongation requires use of the cardiologist-generated impression.

A modified version of the NegEx negation algorithm performed well in detected negation within this dataset with a recall of 97.3% and a precision of 98.2%. Negation within ECGs was rare; overall only 1.9% of all concepts were marked as negated and 7.7% as possible. Thus, the probability of a concept the system identified as positive truly being positive was 0.998. The high recall and precision of negation detection in this dataset is likely due to a constrained vocabulary and the relative simplicity of the ECG impression sentences compared to prior studies in other clinical document types, which had recalls of 78-97% and precisions of 85-91%[18-20].

We have grouped terms such as “questionable”, “cannot rule out”, and “borderline” together. A more granular approach to negating terms may be more predictive of specific clinical outcomes. Also, negating phrases such as “no longer” indicate that a patient has a history of a finding as well as the absence of it now; the current algorithm only identifies the latter. Such information may help determine treatment efficacy. Finally, we used a rather simple “negation window” technique consisting of a certain number of words before or after a negation phrase. Some commas separate clauses while others represent a list of concepts. A more advanced algorithm could use a parsed sentence and the presence of prepositions or coordinating conjunctions to correctly size of the negation window.

The handling of “possible” findings, included in our QT prolongation query, would vary depending on application. For the purposes of decision support, inclusion of potential findings may help prevent adverse events. For example, one would want to avoid starting a drug known to prolong the QT interval in a patient that had a “borderline long QT.” In addition, many uncertain ECG findings require further workup. A patient with potential ischemia requires further evaluation, and one would likely discontinue cyclooxegenase-2 inhibitors in this patient. For research investigations, however, one may desire to exclude uncertain findings, as many ECGs indicating “cannot rule out” may represent benign changes.

The next step in this investigation involves building a decision support tool. Many resources list possible QT prolonging medications, such as the medication registry maintained at the University of Arizona [7]. A decision support system for QT prolongation could intercept orders for high-risk medications. By investigating the occurrences of medication orders in our retrospective database of inpatient admissions in individuals with QT prolongation, we may be able to identify potential medications interactions and new causes of QT prolongation. Some of these interactions may be indirect, such as the addition of a

potent cytochrome P450 inhibitor that raises serum concentrations of a known offending agent. Ideally, a medication intervention could not only intercept medications that prolong the QT interval but also those that significantly interact with those already prescribed.

Since this is a full-text concept index over a large dataset of ECGs, we also have the ability to support many other types of decision support and research venues. While a string-matching algorithm could potentially replicate the performance of the QT prolongation concept query, it would lack flexibility and scalability. By fully concept indexing, we can quickly assess multiple queries, enabling a broader range of research and decision support tools. In addition, the parent-child relationships between concepts in the UMLS ameliorate querying across broad concepts, such as myocardial infarction.

In this study, we used a general purpose concept-identification program with the entire UMLS. We optimized the algorithm to enhance synonymy and favor underspecified matches that match cardiology-related concepts. By processing these ECGs in bulk, KMCI is able to "learn" a corpus of frequently-occurring concepts and thus favors those concepts and their related concepts when encountering ambiguous matches. Other general purpose concept identification algorithms may not perform as well without similar optimizations.

The interpretation of our findings is limited. The performance of the negation algorithm and concept identifier may not translate to other repositories of ECG impressions; however, no specific optimizations were made that correspond to our institution. Second, while we evaluated the QT prolongation algorithm against a random set of ECGs for false negatives, we did not have access to the original ECG waveforms to reevaluate for potential false negatives due to errors in the cardiologists' reading; however, we expect that this would be unlikely to dramatically alter our results. Furthermore, the prevalence of QT prolongation in the KMCI-identified dataset (5.3%) carries more face validity than the QTc interval prevalence of 26.6%. Third, since cardiologists read all ECGs in our institution in a timely manner, decision support based on the textual interpretation of ECGs is feasible. It may not be practical at other institutions in which formal ECG impressions are not rapidly available. Fourth, while we have accurately identified concept matches and their negation status, this is not the same as asserting normality. Our algorithm tells the presence or absence of "atrial fibrillation," for instance, but cannot tell that there were no arrhythmias. These questions may be addressed by classifying concepts by type (e.g., "rhythm" or "perfusion abnormalities") and defining normal status. Finally, our exploratory list of concepts in Table 4 has not been formally assessed for false negatives and provides only a rough incidence of these findings in set of ECGs.

Conclusion

The combination of a negation tagging algorithm with an effective concept identifier allows rapid assessment of clinical syndromes such as QT prolongation, myocardial ischemia, or atrioventricular conduction disturbances. We believe this technique may enable large-scale research on drug adverse events and development of new decision support tools to improve cardiovascular medication safety.

Acknowledgements

We would like to thank Mr. Mark Arrieta for his assistance in assembling the database of ECGs. This work is supported by two National Library of Medicine grants, T15 LM007450 and R01 LM007995.

References

[1] Ray WA, Murray KT, Meredith S, Narasimhulu SS, Hall K, Stein CM. Oral erythromycin and the risk of sudden death from cardiac causes. N Engl J Med 2004 Sep 9; 351(11): 1089-96.

[2] Roden DM. Drug-induced prolongation of the QT interval. N Engl J Med 2004 Mar 4;350(10):1013-22.

[3] Bresalier RS, Sandler RS, Quan H, Bolognese JA, Oxenius B, Horgan K, Lines C, Riddell R, Morton D, Lanas A, Konstam MA, Baron JA. Cardiovascular events associated with rofecoxib in a colorectal adenoma chemoprevention trial. N Engl J Med 2005 Mar 17;352(11):1092-102.

[4] Solomon SD, McMurray JJ, Pfeffer MA, Wittes J, Fowler R, Finn P, Anderson WF, Zauber A, Hawk E, Bertagnolli M. Cardiovascular risk associated with celecoxib in a clinical trial for colorectal adenoma prevention. N Engl J Med 2005 Mar 17;352(11):1071-80.

[5] Zeltser D, Justo D, Halkin A, Rosso R, Ish-Shalom M, Hochenberg M, Viskin S. Drug-induced atrioventricular block: prognosis after discontinuation of the culprit drug. J Am Coll Cardiol 2004 Jul 7;44(1):105-8.

[6] Bazett H. An analysis of the time relationship of electrocardiograms. Heart. 1920;7:353-70.

[7] Drugs that prolong the QT interval and/or induce Torsades dr Pointes ventricular arrhythmia. [cited 2006 12/2]; Available from: http://www.arizonacert.org/medical-pros/drug-lists/drug-lists.htm

[8] Tsai TL, Fridsma DB, Gatti G. Computer decision support as a source of interpretation error: the case of electrocardiograms. J Am Med Inform Assoc. 2003 Sep-Oct; 10(5):478-83.

[9] Willems JL, Abreu-Lima C, Arnaud P, van Bemmel JH, Brohet C, Degani R, Denis B, Gehring J, Graham I, van Herpen G, et al. The diagnostic performance of computer programs for the interpretation of electrocardiograms. N Engl J Med 1991 Dec 19;325(25):1767-73.

[10] Olsson SE, Ohlsson M, Ohlin H, Dzaferagic S, Nilsson ML, Sandkull P, Edenbrandt L. Decision support for the initial triage of patients with acute coronary syndromes. Clin Physiol Funct Imaging 2006 May;26(3):151-6.

[11] Paoletti M, Marchesi C. Discovering dangerous patterns in long-term ambulatory ECG recordings using a fast QRS detection algorithm and explorative data analysis. Comput Methods Programs Biomed 2006 Apr;82(1):20-30.

[12] Fiszman M, Chapman WW, Aronsky D, Evans RS, Haug PJ. Automatic detection of acute bacterial pneumonia from chest X-ray reports. J Am Med Inform Assoc 2000 Nov-Dec;7(6):593-604.

[13] Huang Y, Lowe HJ, Hersh WR. A pilot study of contextual UMLS indexing to improve the precision of concept-based representation in XML-structured clinical radiology reports. J Am Med Inform Assoc 2003 Nov-Dec;10(6):580-7.

[14] Friedman C, Shagina L, Lussier Y, Hripcsak G. Automated encoding of clinical documents based on natural language processing. J Am Med Inform Assoc 2004 Sep-Oct;11(5):392-402.

[15] Nadkarni P, Chen R, Brandt C. UMLS concept indexing for production databases: a feasibility study. J Am Med Inform Assoc 2001 Jan-Feb;8(1):80-91.

[16] Denny JC, Smithers JD, Miller RA, Spickard A, 3rd. "Understanding" medical school curriculum content using KnowledgeMap. J Am Med Inform Assoc 2003 Jul-Aug;10(4):351-62.
[17] Denny JC, Spickard A, Miller RA, Schildcrout J, Darbar D, Rosenbloom ST, Peterson JF. Identifying UMLS concepts from ECG Impressions using KnowledgeMap. AMIA Annu Symp Proc 2005:196-200.
[18] Chapman WW, Bridewell W, Hanbury P, Cooper GF, Buchanan BG. A simple algorithm for identifying negated findings and diseases in discharge summaries. J Biomed Inform 2001 Oct;34(5):301-10.
[19] Elkin PL, Brown SH, Bauer BA, Husser CS, Carruth W, Bergstrom LR, Wahner-Roedler DL. A controlled trial of automated classification of negation from clinical notes. BMC Med Inform Decis Mak 2005;5(1):13.
[20] Mutalik PG, Deshpande A, Nadkarni PM. Use of general-purpose negation detection to augment concept indexing of medical documents: a quantitative study using the UMLS. J Am Med Inform Assoc 2001 Nov-Dec;8(6):598-609.
[21] Unified Medical Language System. [cited 2006 12/2]; Available from: http://www.nlm.nih.gov/research/umls/
[22] van der Sijs H, Aarts J, Vulto A, Berg M. Overriding of drug safety alerts in computerized physician order entry. J Am Med Inform Assoc.2006 Mar-Apr;13(2):138-47.

Address for correspondence:

Josh Denny, MD
4th floor, Eskind Biomedical Library
2209 Garland Ave
Nashville, TN 37232, USA
josh.denny@vanderbilt.edu

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

A Comparison of Supervised Classification Methods for Auditory Brainstem Response Determination

Paul McCullagh, Haiying Wang, Huiru Zheng, Gaye Lightbody and Gerry McAllister

Department of Computing and Mathematics, University of Ulster, United Kingdom

Abstract

The ABR is commonly used in the Audiology clinic to determine and quantify hearing loss. Its interpretation is subjective, dependent upon the expertise and experience of the clinical scientist. In this study we investigated the role of machine learning for pattern classification in this domain. We extracted features from the ABRs of 85 test subjects (550 waveforms) and compared four complimentary supervised classification methods: Naïve Bayes, Support Vector Machine, Multi-Layer Perceptron and KStar. The ABR dataset comprised both high level and near threshold recordings, labeled as 'response' or 'no response' by the human expert. Features were extracted from single averaged recordings to make the classification process straightforward. A best classification accuracy of 83.4% was obtained using Naïve Bayes and five relevant features extracted from time and wavelet domains. Naïve Bayes also achieved the highest specificity (86.3%). The highest sensitivity (93.1%) was obtained with Support Vector Machine-based classification models. In terms of the overall classification accuracy, four classifiers have shown the consistent, relatively high performance, indicating the relevance of selected features and the feasibility of using machine learning and statistical classification models in the analysis of ABR.

Keywords:

Auditory Brainstem Response, wavelet decomposition, feature extraction, classification, decision support

Introduction

The Auditory Brainstem Response (ABR) is evoked when a stimulus click is applied to a subject's ear to determine hearing acuity and integrity of the auditory pathways. If the stimulus is perceived, a response changes their electroencephalogram (EEG) within 10ms from stimulus onset (Figure 1). The amplitude of the ABR signal is approximately 1 µVolt -5µVolts and is hidden behind the background EEG and noise (approximately 50µVolts). The components of the ABR are swamped by the endogenous electrical activity of the brain and the determination of a response can be difficult particularly at low levels of acoustic stimulation, as hearing threshold is reached. The ABR waveform is extracted by coherent averaging which exploits the deterministic nature of the signal to enhance the waveform while suppressing the uncorrelated EEG, extraneous noise and artifact. It is necessary to average approximately 1000-2000 trials before the noise is sufficiently suppressed, with signal to noise ratio enhanced proportional to the square root of the number of trials.

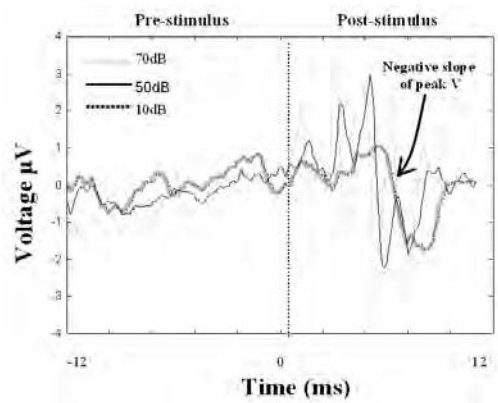


Figure 1 - ABR waveform at a 70, 50 and 10 dB Stimuli

Figure 1 shows responses for a healthy adult ranging from 70dB (normal hearing level, nHL) stimulus to threshold (10dB nHL). The characteristic shape is referred to as the Jewett waveform, comprising up to seven peaks, labeled I-VII. As the level of stimulus is reduced, the different peaks of the waveform become less defined, and their latency is increased. The shape of the waveform from the long slope at the top of peak V is the strongest part of the waveform to remain as the stimulus diminishes. When the stimulus level is set near the subject's hearing threshold the identification of wave V and its following negative slope assist with classifying the presence of the response. The shape of the ABR response will differ between subjects depending on a number of factors such as, electrode placement, filter settings, stimulus intensity, ear used, age, sex and subject's skull thickness. For this reason the range of factors need to be taken into consideration before a clinical expert can make an interpretation. In particular the expert checks the latencies of waves I, III and V and examines the morphology of the waveform. Often sub-averages based on successive recording sessions are compared for consistency and a cross-correlation may also be taken between repeat recordings to assist in classification.

The interpretation of the ABR is subjective, thus clinical experts may not always draw the same conclusion [1], particularly at auditory threshold. Artificial intelligence methods have been used to provide objective assistance in response interpretation [2]. Useful information may be extracted from the EEG recordings using features extracted from the time and frequency domains [3]. Davey [4] showed that the ratio of the power of the post over pre stimulus time domain waveforms could be used as an effective method to classify 'strong' responses with an accuracy of 98.6%. Remaining ABR waveforms, i.e. those which did not have a strong response, were passed to a second stage of analysis, whereby repeat recordings were used to derive features based on cross-correlation parameters using both the time and frequency domains. Lower accuracies ranging from 65% to 82% were achieved, dependent upon features used. The lower accuracies reflect the more difficult classification process.

Lightbody [5] used wavelet decomposition [6, 7] on the same dataset following a two-stage classification process. Power ratios of the post over pre stimulus wavelet coefficients were used to classify 'strong' responses. Those remaining were then classified using correlation features of repeated tests derived from the wavelet domain. Overall accuracy of 76.4% was obtained using a C5.0 decision tree classifier. Strong responses were classified without error. The criterion to determine a 'strong' response required a combination of time and wavelet post stimulus to pre stimulus power ratios (time domain power>2 and wavelet domain power>1.6, determined heuristically). By combining features using the Demster-Shafer method, as used in evidential reasoning [8], Davey achieved a classification accuracy of 95.6% for 'strong' responses and 85% for lower level responses [9].

In this study we compared four additional classification techniques which are in widespread use by machine learning researchers. The aim was to determine whether one of these techniques provides superior classification accuracy for the quasi-stationary ABR evoked response. Three additional statistical descriptors, namely precision, sensitivity and specificity were used to further explore the data set. The data set comprised both high and low level recordings. A secondary aim was to determine whether a single ABR recording, irrespective of level, provided sufficient information for decision support. Hence the classification process does not involve correlation parameters from repeated recordings.

Methods

The study was performed on a database of 85 test subjects, provided by the Audiology Department of the Royal Group of Hospitals in Northern Ireland. Each subject had a range of test stimulus levels applied providing a mix of good, weak and non-response waveforms, all of which were classified by a clinical expert. There were 550 waveforms in total, 396 recordings with a YES classification and 154 with a NO classification.

The data were pre-processed by band-pass filtering (100Hz-3kHz), sampled at 20kHz and then de-noised using a wavelet filter. Between 1000 and 2000 ensembles were averaged to provide one ABR waveform. Each waveform consisted of 480 data samples, half before the stimulus and half after, which related to 12ms before the stimulus and 12ms after the stimulus. The ABR waveform appears within 10ms of stimulus onset, which after band-pass filtering and sampling relates to 200 data points. A number of features were extracted from both time and wavelet domains. The Daubechies wavelet has been used to de-noise biosignals [10].

We computed a range of scaling lengths performed to 6 levels (A6: 0–156Hz, D6: 156–312Hz, D5: 312–625Hz, D4:625-1250Hz, D3:1250-2500Hz, D2:2500-5000Hz, D1:5kHz-10kHz). Levels D6, D5 and D4 showed to be of significant interest as they related to the key frequencies (200, 500 and 900Hz [11]) contained within the ABR. This is depicted in figure 2 which shows the Fast Fourier Transform (FFT) amplitude values for post-stimulus ABR waveforms at 70dB and 30dB stimulation intensities. It highlights the frequencies covered by the D6, D5 and D4 wavelet coefficients.

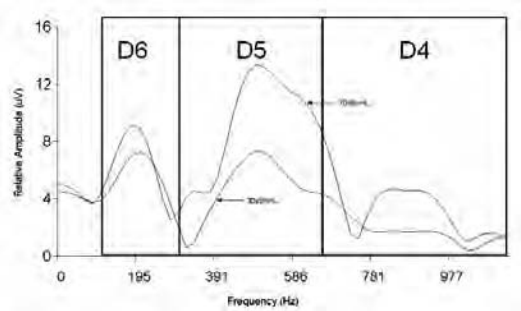


Figure 2 - FFT amplitude values for post-stimulus ABR waveforms at 70dB and 30dB stimulation intensities

Although in figure 2 the D4 coefficients seemed to show the least promising power peaks they have double the number of coefficients than the D5 range helping to reveal more feature consistency. Figure 3 illustrates the D4 component for pre and post stimulus activity for decomposition on 256 data points.

The decomposition was carried out on the full pre stimulus section and only waves I to V of the post section. Davey [4] showed that by using data from 1.5 to 9.5 ms post stimulus then the classification accuracy was improved. By focusing in on the ABR where waves I to V where most likely located removed the samples early in the waveform that may be more likely to contain artifacts and influence the classification. This same process was used for the wavelet decomposition of the post stimulus section. For both the pre and post stimulus waveforms data extension was required to extend the dataset to 256 which is the nearest dyadic number [5] so to support wavelet decomposition. Different methods to extend the data had been looked at to determine the method least likely to incur boundary issues [12].

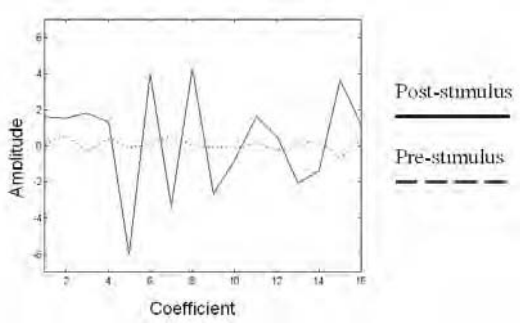


Figure 3: Extracted D4 wavelet coefficients (70dB stimulus), for pre and post-stimulus regions

An alternative partitioning of the data has been considered by performing wavelet decomposition on a smaller section of the data containing 128 data points, therefore removing the need for data extension [13]. The window of data relates to a 6.4ms section of the waveform, so the choice of window within the full post stimulus section is vitally important to ensure that the subject's response has been captured. This is crucial because as the strength of a response diminishes the peaks fade and latencies alter until peak V remains. Also, the exact position of the response differs from subject to subject. A sliding window of 128 data points was used to derive features between 1.5ms and 9.5ms of the post stimulus waveform, which was shown to be the time frame with the strongest waveforms in an ABR response [4].

Using a moving window of 128 samples from stimulus onset, the window containing the peak time domain power ratio was determined. The ABR response is expected to be at its maximum within 1.5ms and 9.5ms post stimulus. It was considered that a peak power before the 1.5ms mark could signify that the power ratio might be a result of something other than the ABR. A weighting based on the position of the peak power was used to numerically represent this doubt in the power ratio as a feature for classification. This weighting was applied to features DataAbs and DataPwr resulting in features wDataAbs and wDataPwr respectively. Whereby, DataAbs is a feature derived from the ratio of the post and pre stimulus D4 coefficients [5] (Figure 3) and DataPwr is the ratio of the post and pre stimulus time domain waveform [4]. In both cases only waves I to V in the post stimulus waveform were considered and the wavelet decomposition was performed on 256 data points resulting in 16 D4 coefficients (Figure 3).

Each 128 sample window starting from sample 35 to starting from sample 75 were independently analyzed. Power ratios were calculated labeled RNxx, where R represents the 'response' post stimulus region in the waveform and N represents the 'no response' pre stimulus region of the waveform, and 'xx' represents the starting sample for the window of data. Similar ratios for the D4, D5 and D6 wavelet coefficients were calculated for each of the 128 sample windows. Using all the D4-D6 coefficients was first considered since a 6-level decomposition of a 128 sample waveform results in a small number of coefficients (8-D4, 4-D3 and 2-D6 coefficients). Then the individual bands were considered. The last features were derived by averaging the calculated features for each of the 128 sample windows, so to capture any overall trends.

The features were used as inputs to four classification models Naïve Bayes (NB), Support Vector Machine (SVM), Multi-Layer Perceptron (MLP) and KStar. NB is a probalilistic classifier based on the Bayes rule of conditional probability. It assumes independence between features. It uses the normal distribution to model numeric attributes by calculating the mean standard deviation for each class [14]. SVM is a kernel based classifier. The basic training for SVMs involves finding a function which optimizes a bound on the generalization capability, i.e., performance on unseen data. By using the kernel trick technique, SVM can apply linear classification techniques to non-linear classification problems [15]. A MLP is a non-linear classification approach that may be trained using the back propagation algorithm. A MLP consists of multiple layers of computational units (an input layer, one or more hidden layer and one output layer). For the MLP model, the results were obtained using a model consisting of one hidden layer with six nodes when evaluating the top ten features, four nodes when evaluating the top six features and two nodes when considering the top three features (the choice of feature subsets is discussed later in the paper). Each MLP was trained for 500 epochs and the learning rate was set to 0.3. The final classification is KStar, a relatively simple instance based method. The class of a test instance is based upon the class of those training instances similar to it, using an entropy-based distance function to compute the similarity between two cases. The underlying assumption is that similar instances will have similar classes. All four classification models were implemented within the framework provided by the open-source platform Weka package [16]. This provides a collection of machine learning algorithms for data mining tasks, which may be implemented directly or called from Java code.

In the evaluation of each classifier we used ten-fold cross validation, i.e. the entire dataset is partitioned into 10 subsets of approximately equal size. Each subset in turn is used as the test set while the other subsets are combined to form a training set. The quality of each classifier was assessed by the extent to which the correct class labels have been assigned. Each classifier is assessed by three statistical measures: Precision (Pr), Sensitivity (Se), Specificity (Sp). All the features were ranked using correlation coefficient-based ranking.

Results

The top five features are illustrated in Table 1. Table 2 details accuracy, precision, sensitivity and specificity of the classifiers in separating the two classes, based on the top 5 discriminative features. (Note as this is a two-class problem, Se and Sp are reversed for Response='NO' and Response='YES').

Table 1 - Top five predictive features

Feature	Name	Description
1	wDataAbs	Wavelet power feature: D4 (weighted)
2	DataAbs	Wavelet power feature: D4
3	RN75	Power ratio: 128 samples from 75
4	RN70	Power ratio: 128 samples from 70
5	RN65	Power ratio: 128 samples from 65

Table 3 details accuracy, precision, sensitivity and specificity of the classifiers in separating the two classes, based on the top 2 discriminative features. The accuracy and precision in determining a NO response are slightly reduced, implying that most discrimination is harnessed from the relative power measure from the wavelet domain, i.e. the D4 coefficient. Time may have a small role to play in the overall morphology.

Table 2 - Prediction results for four classifiers using 10-fold cross validation with top 5 features (wDataAbs, DataAbs, RN75, RN70, and RN65)

Method	Ac (%)	Response='NO'			Response='YES'		
		Pr (%)	Se (%)	Sp (%)	Pr (%)	Se (%)	Sp (%)
NaiveBayes	83.4	93.5	82.2	86.3	66.8	86.3	82.2
SVM	81.9	83.2	93.1	54.9	76.8	54.9	93.1
MLP	81.9	87.9	86.2	71.4	68.3	71.4	86.2
Kstar	82.2	86.9	88.1	68.0	70.4	68.0	88.1

Table 3 - Prediction results for four classifiers using 10-fold cross validation with top 2 features (wDataAbs and DataAbs)

Method	Ac (%)	Response='NO'			Response='YES'		
		Pr (%)	Se (%)	Sp (%)	Pr (%)	Se (%)	Sp (%)
NaiveBayes	81.4	86.7	86.9	68.0	68.4	68.0	86.9
SVM	80.2	81.5	93.1	49.1	74.8	49.1	93.1
MLP	80.7	84.5	89.1	60.6	69.7	60.6	89.1
KStar	81.2	85.0	89.1	62.3	70.3	62.3	89.1

Discussion

The ABR test is routinely performed to detect hearing loss (Response='NO') in the Audiology clinic. Where the ratio of post-stimulus power to pre-stimulus power is high (>5), the classification decision can be made to high accuracy (98.6%). This is manifest as a clear Jewett response. Misclassification may be due to artifact attributed to stimulus, myogenic (muscle) activity, or eye blink artifact which corrupts the ABR. Such activity may be detected by an expert due to latency; stimulus artifact occurs within the first msec post-stimulus, and myogenic artifact normally occurs after wave VII. Eye blinks may be harder to eliminate as they may be more randomly distributed in the pre and post stimulus activity and hence overlap with the ABR components. The strategy is normally to remove eye blinks at source, as much as possible, by rejecting a contaminated trial.

ABR responses with a lower post-stimulus power to pre-stimulus power ratio (<5), provide a more difficult classification task, with a classification accuracy of 76.4% reported using the C5.0 decision tree classifier. This second stage classification process required the use of correlation features, from repeated trials, to assess consistency of features. In the context of decision support this would provides a complex implementation. In this research we have utilized the same data set, to assess four additional classifiers, used by data mining researchers. The classification is based only on single ABR waveforms, which provides a more difficult classification task, but is more straightforward to implement. In the context of decision support, a classifier with high accuracy and high sensitivity is required. The best overall accuracy of 83.4% was obtained for the statistically based Naïve Bayes method, using the top five features. This has a good balance between sensitivity (82.2%) and specificity (86.3%). The SVM classifier has a higher sensitivity (93.1%), but at the expense of a lower specificity (54.9%). The SVM also maintains its sensitivity as the number of features reduces to two.

In terms of mathematical complexity, KStar, an instance-based classifier, and Naive Bayes (NB), a Bayesian theorem-based model, are relatively simple classifiers, but KStar requires additional space to store training datasets. In the case where the dataset is large, using KStar could be computation-intensive. In terms of algorithm configuration, the implementation of NB is quite straightforward. It does not require any predefined parameters. MLP and SVM have more complicated parameters settings to tune. For example, MLP requires users to specify the learning rate and the number of learning epochs in advance; and SVM requires users to predefine an appropriate kernel function. NB has been proven to be one of most proficient classification algorithms in most cases. Nevertheless, given the size of datasets studied and current advanced computational resources, the differences between four classifiers in terms of computational time are not obvious in this study. Each classifier has its own advantages. For example, NB can achieve the best results in terms of specificity (86.3%). The best sensitivity was obtained by using SVM-based classification models (93.1%). Therefore, it is

hard to say which one is best in this case. Nevertheless, it is expected that a better result could be obtained by combining different classifiers. A subsidiary aim of this study was to assess the suitability of the classifiers in evoked potential analysis. However there was little difference in the overall accuracy and the choice of classifier may be more influenced by ease of implementation. A further analysis is required to assess the agreement between the classifiers, and this can be undertaken by analyzing the classification/ misclassification of single response. This study illustrates the need for further validation of the approach by applying unseen data set to the classifiers.

Conclusions

The overall classification accuracy is quite similar for the four methods, ranging from 81.9% to 83.4% with the top 5 features and 80.2% to 81.4% with the top 2 features. These accuracies exceed the C5.0 (76.4%) determined in a previous study, but may be inflated because 'strong' responses are also included in this study as we adopt classification on a single averaged trail. A valid between-study comparison would require a 2 stage classification, as used in the previous study. Currently learning parameters required for each classifier are based on a trial-and-error approach. Combining other machine learning techniques such as the genetic algorithm to automatically determine the optimal learning parameters would enhance our methodology. All the classifiers show the consistent high accuracy which indicates the relevance of selected features and the feasibility of using machine learning and statistical classification models in the analysis of ABR. Additionally the feature ranking technique used is not optimal and could be further enhanced.

Acknowledgment

The authors wish to thank Mr HG Houston, Royal Group of Hospitals for providing classification on the ABR data set.

References

[1] Pratt TL, Olsen WO. Four-channel ABR recordings: Consistency in interpretation. Am J of Audiology 1995, 4 (2): 47-54.

[2] Delgado RE and Ozdamar O. Automated Auditory Brainstem Response Interpretation. IEEE Engineering in Medicine and Biology 1994, 227-237.

[3] Aoyagi M, and Harada J. Application of fast Fourier transform to auditory evoked brainstem response. Med. Inform., 1988, 13(3), 211-220.

[4] Davey R, McCullagh PJ, McAllister HG, Houston HG. The use of artificial neural networks for the objective determination of hearing threshold, using the auditory brainstem response. In: Neural Networks in Healthcare: Potentials and Challenges, IDEA Group Inc, 2006.

[5] Lightbody G, McCullagh PJ, and McAllister HG. Feature Extraction in Auditory Brainstem Responses: Use of the Daubechies Wavelet. EMBEC Conf 2005.

[6] Bradley A and Wilson W. On Wavelet Analysis of Auditory Evoked Potentials. J Clin Neurophysiology 2004, 115: 1114-1128.

[7] Mallat SG. A Theory for Multiresolution Signal Decomposition: the Wavelet Representation. IEEE Trans. on Pattern Anal. Machine Intelligence 1989, 11: 674-693.

[8] Shafer G. and Logan R. Implementing Dempster's rule for hierarchical evidence. Artificial Intelligence 1987, 33:271-298.

[9] Davey R, McCullagh P, Lightbody G and McAllister G. Auditory Brainstem Classification: a Hybrid Model using Time and Frequency Features, Artificial Intelligence in Medicine 2006. [Epub ahead of print].

[10] Zhang R, McAllister G, Scotney B, McClean S, Houston G. Combining wavelet analysis and Bayesian networks for the classification of auditory brainstem response. IEEE Trans Inf Tech Biomed 2006, Jul; 10(3):458-67.

[11] Ozdamar O, Delgado RE. Automated electrophysiologic hearing testing using a threshold-seeking algorithm. J Am Acad Audiol 1994, 5, 77-88.

[12] Gutierrez A, Somolinos A. Influence of wavelet boundary conditions on the classification of biological signals. Proceedings of the IEEE 26th Annual Northeast Bioengineering Conf 2000, 25-26.

[13] Lightbody G, McCullagh PJ, and McAllister HG. Feature extraction in auditory brainstem responses using wavelet decomposition on a moving window of waveform Data, IEE Irish Signals and System Conf 2006, 321-326.

[14] Irina R. An empirical study of the naive Bayes classifier. IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence, 2001.

[15] Burges C. A Tutorial on Support Vector Machines for Pattern Recognition. Data Mining and Knowledge Discovery 1998: 2: 121-167.

[16] Witten IH and Frank E. Data Mining: Practical machine learning tools and techniques, 2nd Edition, Morgan Kaufmann, San Francisco, 2005.

Address for correspondence

Dr PJ McCullagh,
School of Computing and Mathematics,
University of Ulster,
Shore Road,
Newtownabbey,
BT370QB,
N Ireland,
United Kingdom

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Non-linear Analysis for the Sleepy Drivers Problem

Ioanna Chouvarda[a], Christos Papadelis[a], Chrysoula Kourtidou-Papadeli[b], Panagiotis D Bamidis[a], Dimitris Koufogiannis[a], Evaggelos Bekiaris[c] and Nikos Maglaveras[a]

[a] *Lab of Medical Informatics, The Medical School, Aristotle University of Thessaloniki, Greece*
[b] *Greek Aerospace Medical Association, Thessaloniki, Greece*
[c] *Institute of Transport, CERTH, Thessaloniki, Greece*

Abstract

The problem addressed in this work is sleepiness during driving, which often leads to accidents in the streets. Experiments with sleepy drivers took place and the EEG data were analysed in terms of non-linear methods. Sample entropy and phase synchronization variations were investigated within the signal sections corresponding to "driving events", i.e. driving mistakes or loss of control, as well as to periods of drowsiness and sleepiness, as compared to the periods of normal driving. Decreased sample entropy, indicating loss of complexity, and an increased phase synchronisation have been found in the preliminary study presented. The results are encouraging towards developing an alerting system for predicting and preventing driving accidents.

Keywords:

sleep disorders, nonlinear dynamics, automobile driving

Introduction

Sleep is necessary for human beings and its gradual loss can lead from extreme short-term sleepiness to chronic sleepiness. Yet, the conditions of modern life increase stress and often lead people into sacrificing sleep despite the negative effects on health and abilities. In addition, as much as 20% of the general population suffers to some degree from sleep disorders which may cause extreme tiredness, loss of concentration and a pronounced inability to function normally, raising the risk of causing traffic and work accidents.

Sleepiness reduces reaction time, vigilance, alertness and concentration. Lack of adequate sleeps is associated with excessive fatigue, hypovigilance, stress and impairment of attention, information processing speed and decision-making quality, which are among the key causes of serious industrial accidents, including most nuclear accidents (Chernobyl, Three-mile Island) and other large scale accidents (such as the Bhopal chemical disaster) [1]. In addition, about 40% of fatal accidents on US highways are fatigue-related, whereas sleep-related accidents account annually for as many as 240,000 motor-vehicle accidents. UK statistics show that 25% of motor accidents are associated with driver fatigue [2]. According to NASA's Aviation Safety Reporting System, approximately 21% of the aviation incidents are fatigue-related.

As sensor, communication and artificial intelligence technologies are evolving, an effort is being paid to apply and combine such technologies in ambient intelligence systems for a more secure environment that detects and copes with the effects of sleepiness. An important challenge in this effort is the development of methods to define sleep/wakefulness involuntary transitions by physiological EEG measurements. Such predictions made feasible pervasively and unobtrusively can help decrease drastically the number of fatigue or sleepiness-related accidents and make work and streets a safer place. The European IST project SENSATION is involved in this field. A series of experiments concerning involuntary sleepiness have taken place, along with development of new miniaturized relevant sensors and new tools for processing the derived data in a meaningful way.

Within this scope, a sleepy driver experiment has taken place aiming at deriving new information regarding the driver's physiological status when losing driving control. In a previous work [3], frequency analysis of these experimental data lead to findings such as an increase of slowing activity and an acute increase of the alpha waves 5-10 seconds before driving events, accompanied by A rapid decrease of both Shannon and K-L entropies have been observed just before the driving events, i.e. loss of driving control.

A more extensive quantitative analysis is presented in this work, focusing on non-linear analysis. The hypothesis tested here that EEG "systemic" features, like complexity and synchronization, as calculated during incidences of sleepiness, and mainly during a driving event, decline from the ones calculated during driving without events. The detection of "patterns of sleepiness" among such features may help predict dangerous driving events.

Materials and methods

Subjects and experimental protocol

Five subjects participated in the present study. The subjects were average drivers (mean driving experience: 8.3 years), with a mean age of 26.5 years. The experiments

were performed in CERTH, Thessaloniki, Greece, from 6 June till 27 July 2005.

The subjects were asked to stay awake for at least 24 hours, and then to arrive at CERTH premises at around 22.00. Upon arrival and after passing the standard medical examination test, the subjects' level of sleepiness was estimated by using the Karolinska Sleepiness Scale (KSS) test, and their sleepiness behavior was scaled by the M.D. by using the Epworth Sleepiness test (1 = 'extremely alert', 5 = 'neither alert nor sleepy', 9 = 'extremely sleepy – fighting sleep'), resulting in two subjects with score 7 (very sleepy), two with grade 6, and one with 5 (neither sleepy nor awake).

The measurements were performed in the CERTH experiment car. The subject had to seat at the driver seat and the attached electrodes were connected to an ambulatory EEG monitoring system. EEG channels Fp1, C3, P3, O1, Fp2, C4, P4 and O2 were recorded, along with two horizontal and two vertical EOG channels for bipolar signals, two EMG channels and two ECG channels. For the data acquisition, a sampling rate of 200Hz, for each channel of the recording. The monitoring system hardware filters were adjusted to the band pass filtering option with a frequency range of 0.5 to 70Hz for EEG, with a notch filter at the 50Hz power supply component. A battery box with power supply independence of approximately 3 hours supported the monitoring system.

An experienced driving instructor was seated at the co-driver's seat, as shown in Figure 1. At the back there was a technician monitoring the functioning of the recording equipment and a medical doctor monitoring the EEG.

Figure 1 – The experimental car with the subject and the driving instructor

Every subject drove the research vehicle for a maximum of 1 hour on a motorway. In eight cases, subjects' sleepiness level during driving was very high, and the driver instructor stopped the measurements after three or more consecutive sleepiness events (unintentionally cross the lane border). The traffic on the motorway was very low, and the task was monotonous enough to stimulate hypovigilance. The driver's behavior and the sleepiness events for each subject were manually recorded. In parallel, the experimental car was equipped with the Eyelid Sensor System (ELS), i.e. a camera and appropriate commercial software capable of sensing the eyelid movements and detecting periods of drowsiness and sleepiness. In Table 1 there is an example of such a series of driving behavior annotations.

Table 1 – Example of annotations of the driver's condition and behavior

sec	sample	code	Notes
1170	234000	–1	ELS sleepy
1890	378000	–2	Theta waves in EEG
1895	379000	–3	ELS and EEG drowsy and sleepy
2575	515000	1	driver's error, wrong breaking
3185	637000	2	driver's error, serious driving event, wrong deviation from straight line, instructor's intervention
3360	672000	2	wrong deviation from straight line, instructor's intervention
3545	709000	2	instructor's intervention
3705	741000	3	end of measurement

Preprocessing

The EEG data were firstly filtered off-line by using a 3rd order Butterworth filter (band pass range: 0.5-45 Hz), and then the Infomax Independent Component Analysis (I-ICA) technique was used in order to remove eye movements and eye blinks [4]. ICA decomposition was performed on the EEG+EOG signals by using EEGLAB software [5]. Components contaminated by artifacts were rejected, and the remaining components were mixed and projected back onto the scalp channels. The analysis was performed on the artifacts-free EEG data.

Non-linear EEG analysis

A preliminary analysis took place for five subjects with various initial KSS scores. EEG channels C3, C4, P3 and P4 have been considered and analysis has taken place in time windows corresponding to 10 secs, with 25% overlap. Within each such EEG data segment the following features have been calculated:

- Sample entropy for the assessment of complexity of each channel, and

- Phase synchronization between two channels based on Hilbert Transform. Specifically C3-C4, C3-P3, C4-P4 and P3-P4 channel pairs have been considered.

The main idea behind this analysis was that both of these non-linear features can have relation to the systems dynamics and reveal changes during driver's gradual sleepiness and during incidents of driver's loss of control.

Sample entropy

A measure of the regularity of a time series is informative about the underlying complexities in the processes giving rise to it. Pincus [6] developed a regularity/complexity measure known as the Approximate Entropy (ApEn), which assigns a nonnegative number to a time series with larger values corresponding to greater apparent randomness of the process underlying the data while smaller values point to regular features in the data. Given groups of N points in a series, the approximate entropy is related to the probability that two sequences which are similar for N points remain similar at the next point.

A modification of the Approximate Entropy has been introduced as Sample Entropy[1] by Richman and Moorman [7], who claim that approximate entropy as defined by Pincus is biased. Larger Sample Entropy values indicate greater independence, less predictability, hence greater complexity in the data. This, in turn, may imply that decreased complexity or greater regularity in the time series is associated with disease or not regular function.

Considering two sequences $y_m(i)$ and $y_m(j)$ of the form $y_m(i)=[x(i), x(i+1),\ldots,x(i+(m-1))]$, then $B^m(r)$ gives the probability that the two sequences match for m points and $A^m(r)$ gives the probability of match for m+1 points and Sample Entropy is defined according to Equation 1.

$$SampEn(m,n,r) = [-\ln(\frac{A^m}{B^m})] \quad (1)$$

Phase synchronisation

Besides complexity measures, spatial features related to the dependence among EEG channels can be considered. Another approach is to consider the time-series as the output of oscillators and quantify their interaction by measuring the phase difference between these signals. In the present work, the analytic signal approach based on Hilbert transform has been adopted [8]. Hilbert transform can be defined as shown in Equation (2), where p.v. denotes the Cauchy principal value integral (which avoids the singularities at = t, and =±∞).

$$\tilde{s}(t) = \frac{1}{\pi} p.v. \int_{-\infty}^{\infty} \frac{s(\tau)}{t-\tau} d\tau \quad (2)$$

By use of Hilbert transform, the analytic $s_A(t)$ signal is defined from the initial signal s(t), as depicted in Equation (3)

$$s_A(t) = s(t) + j\tilde{s}(t) \equiv A(t)e^{j\varphi(t)} \quad (3)$$

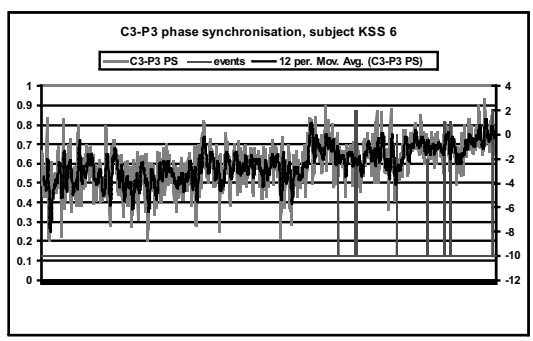
The instantaneous phase of the analytic signal is defined as shown in Equation (4)

$$\varphi(t) = \arctan\frac{\tilde{s}(t)}{s(t)} \quad (4)$$

Since instantaneous phase is available, the phase difference between channels can be calculated for each time sample. The measure of phase synchronization (PS) between two channels i and j within a time window of N samples is formulated as shown in Equation (5)

$$PS_{ij} = \frac{1}{N}\sum_{n=1}^{N} e^{j(\varphi_i(n)-\varphi_j(n))} \quad (5)$$

This is a normalized measure can vary between 0 and 1. Two channels which are synchronized for some time period will have X index close to 1 for that time window.



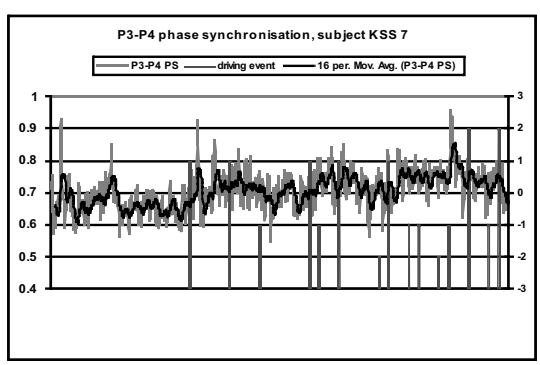


Figure 2 - Phase Synchronisation for two subjects. Synchronisation tends to increase. Events are related to periods of increased synchronization. The annotations follow the codes depicted in Table 1.

1 Matlab code for calculating Sample Entropy available at http://www.physionet.org/physiotools/sampen/

Results

The non-linear measures described above have been calculated for the five subjects with different initial sleepiness status. Since a series of such entropy and synchronization values was calculated, it was intended to seek for intra-subject differences in these features between the start and the end of the experiment, or between periods when a driving incident happened, or heavy sleepiness was reported and the rest of driving time.

Figure 2 shows the evolution of phase synchronization index and Figure 3 the evolution of Sample entropy for two subjecta, along with the annotated eventa, following the code depicted in Table 1. It can be observed that phase synchronization among the hemispheres is gradually increasing and sample entropy is decreasing, both pointing at a gradual increase of sleepiness related to loss of EEG complexity. Driving incidents seem to occur in relation to a local increase in phase synchronization and decrease in sample entropy.

In order to further evaluate possible variations of these features reflecting the subject's different condition, statistical analysis was performed. Specifically, different sets of feature data were considered, corresponding to: a) serious driving event, b) driving event, c) sleepy, d) drowsy areas, e) initial driving period without incidents and f) final driving period without incidents. The differences among some pairs of sets, i.e. serious driving event vs. initial driving period, were statistically assessed. The Wilcoxon ranksum test was applied, testing the hypothesis that populations generating two independent samples, x and y, are identical. Some results regarding these differences are depicted in Figure 4(a).

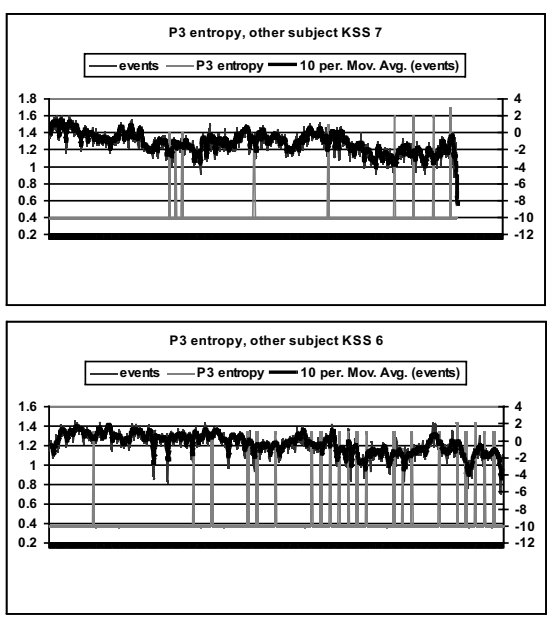


Figure 3- Sample entropy for two subjects. Events are related to periods of decreased entropy, and thus decreased complexity.

The statistical analysis showed that when there was statistical significance in the comparison, sample entropy for the aforementioned EEG channels was always lower during driving incidents than during normal driving periods. Moreover, it is interesting to notice a gradual change of the sample entropy group median values among the normal-sleepy-drowsy-event groups, as shown in Figure 4(b).

When the whole broadband signals were used in the analysis, a statistical difference in the PS feature between the normal driving period and the event period occurred for 4 out of 5 subjects. However, the expected under our hypothesis phase synchronisation elevation was evident for some cases, but not in all subjects. Specifically, an elevation clearly occurred in three out of five subjects, a decrease in two. For the sleepy/drowsy periods, the results were similar. It is possible that inter-subject variation is also related to the difference in sleepiness condition, as quantified by the KSS test. For example, the least number of statistically significant differences was found in the KSS 5 subject.

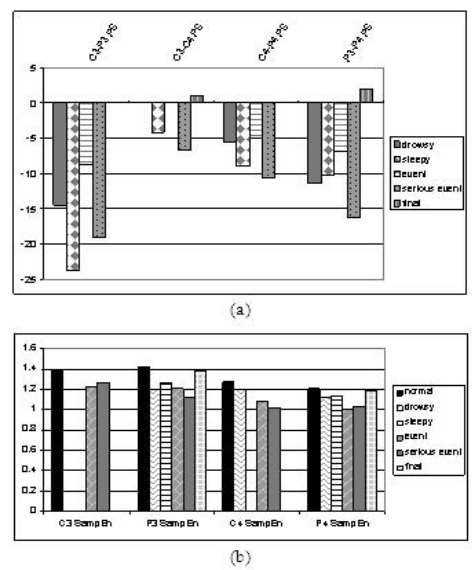


Figure 4 - a) For a subject KSS 7, the relative difference (1-value/initial_value)% between medians of the sample entropy calculated in the initial period and in the periods of various annotations. Calculations take place only when there is statistical significance. b) For another KSS 7 subject, median sample entropies for different channels and event periods than have statistical difference with the initial driving period.

When analyzing synchronization in different frequency bands (alpha and theta bands), there were some small differences, which were expected since more narrowband signals were analysed. In theta band, the statistically significant features showed a synchronisation increase for the driving event periods (4 out of 5 subjects), but that was not so persistent for the sleepy and drowsy periods. In alpha band, a phase synchronisation increase occurred for 3 out

of 5 subjects in the driving event and sleepy/drowsy periods.

It has to be noted that driver's loss of control and other annotated events do not occur momentarily, i.e. they refer to a period of time, for example not a specific second but a few seconds. Furthermore, they are manually reported by the co-driver, and their exact duration is unknown. Therefore, some inter-subject variation in the results may be explained by the variation in the definition and exact location of events. The accurate time annotation of driving events would allow the analysis or differences in pre and post event periods.

Discussion

An experiment with sleepy drivers has taken place and preliminary results of EGG data non-linear analysis have been presented. The decrease in Sample entropy and increase in Phase Synchronisation reported in this work can be regarded as measures of the system's complexity and spatial organization, complementary to the frequency band measures, traditionally applied in the sleep analysis. While these findings are somewhat macroscopic, a more extended study can seek more detailed patterns of sleepiness, to be used for prediction. A better elaboration of the driving event annotations would then be necessary. The correlation of polysomnographic signals or derived features with automatically generated driving data, as produced by the vehicle logging mechanism would be of use. This analysis can lead to an understanding of the mechanisms of alertness of drowsiness, and eventually to the prediction of dangerous driving events. The incorporation of such alertness predicting facilities in the cars can be extremely important for the driver safety. Such sensing platforms, as the one aimed by SENSATION project, must be real-time, autonomous, non-supervised, unobtrusive, and must be able to operate in non-constrained environment, able to monitor, detect and predict human physiological state in relation to alertness, fatigue and stress anytime, everywhere and for everybody.

Conclusions

A study of the intra-subject variation in sample entropy and phase synchronization in relation to driving accidents has been presented here. The non-linear measures, related to system's complexity and spatial organisation, can reveal overall system changes related to subject's vigilance, sleepiness, and hopefully predict conditions such as loss of control. The initial steady values can be used as reference, eliminating the needs for absolute thresholds. The elaboration of more detailed biosignal patterns related to the subject's driving behavior and the extended validation of our methodologies are among the future tasks of this research work.

Acknowledgments

This work has been partly supported by the IST - 2003- 507231 project SENSATION.

References

[1] Milter MM, Carskadon MA, Czeiler CZ, et al. Catastrophes. sleep and public policy: Consensus report. Sleep 1988: 11:100-09.

[2] Horne JA., Reyner LA., Sleep related vehicle accidents, BMJ 1995: 310(6979):565-7.

[3] Papadelis C, Kourtidou-Papadeli Ch, Bamidis P, Chouvarda I, Koufogiannis D, Bekiaris E, and Maglaveras N. Indicators of Sleepiness in an Ambulatory EEG Study of Night Driving. 28th IEEE EMBS Annual International Conference, Aug 30-Sept. 3, 2006, New York City, New York, USA.

[4] Jung TP, Makeig S, Humphries C, Lee TW, Mckeown MJ, Iraqui V and Sejnowski TJ. Removing electroencephalographic artifacts by blind source separation. Psychophysiology 2000: 37: 163-178.

[5] Delorme A and Makeig S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J Neurosci Methods 2004: 134: 9–21.

[6] Pincus SM. Approximate entropy as a measure of system complexity. Proc Natl Acad Sci USA 1991: 88: 2297-2301.

[7] Richman JS and Moorman JR Physiological time-series analysis using approximate entropy and sample entropy. Am J Physiol Heart Circ Physiol 2000: 278(6): H2039-H2049.

[8] Rosenblum, M, Pikovsky A, and Kurths, J. Phase synchronization of chaotic oscillators. Phys. Rev. Lett. 1996: 76: 1804–1807.

Address for correspondence

Nicos Maglaveras
E-mail address nicmag@rned.auth.gr

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Identification and Genotype Related Classification of Children with Long QT-Syndrome Using 24h Holter Recordings

Matthias Bauch[a], Annamaria Siegler[a], Markus Khalil[b], Jörg Zehelein[c], Herbert E Ulmer[b], Hartmut Dickhaus[a]

[a] *Department of Medical Informatics, Heidelberg University, Germany*
[b] *Department of Paediatric Cardiology, University Medical Centre, Heidelberg, Germany*
[c] *Department of Internal Medicine III Cardiology, Angiology and Pneumology, University Medical Centre, Heidelberg, Germany*

Abstract

The long QT syndrome (LQTS) belongs to the family of hereditary diseases and can cause life-threatening arrhythmias and leads to sudden cardiac death. Mutations on six genes are responsible for changes in the electrophysiological properties of myocardial cells that are involved in the repolarization phase. In the surface ecg this is expressed by a prolonged QT interval and genotype-specific shapes for the T-Wave. The aim of the study was to find parameters that quantify properties of the repolarization phase which can be used in addition to the established Schwartz score in the process of diagnosing LQTS. Furthermore, ecg features were evaluated for the separation of the LQT subtypes LQT1, LQT2 and LQT3. The combination of the features PtA50 and QTc yielded with 93% sensitivity and 100% specificity the best results in the field of patient identification. Despite the small dataset consisting of 14 patients that was available for the second aim, the achieved results for the morphology indices motivate further research in this field.

Keywords:

Electrocardiography, Holter, long QT-syndrome.

Introduction

The long QT syndrome (LQTS) [1] belongs to the family of hereditary diseases which can cause life-threatening arrhythmias. Most of the patients suffer from first cardiac events like syncope or cardiac arrest during childhood. Approximately 54% of them die before they reach the age of twenty if no appropriate therapy was applied.

In clinical routine the diagnosis of the LQTS is essentially based on two aspects: information from the medical history and the evaluation of the ecg. The probability for the existence of a LQTS is given by a score proposed by Schwartz et al. in 1985 and 1993 [2]. The score is expressed by a number of points that are given for a prolonged QT interval, furthermore the former presence of Torsade de pointe arrhythmias, T wave alternans, notched T waves in at least three leads in the ecg, a lowered heart rate, former syncopes, congenital deafness, the fact that other family members have a LQTS and the occurrence of sudden death in the family. In addition to the Schwartz score the molecular genetic testing becomes more and more important. But this examination method is not suitable for screening tests because of the considerable costs and time. Nevertheless this is the only way to get evidence for the existence of a LQTS.

Particularly when applied to young children the Schwartz score shows some drawbacks. Especially for young patients most points that are related to the medical history are not achievable. Furthermore the widely used frequency correction by Bazett's formula often leads to unusable values for QTc if the original heart rate is very high which is a common property for young children. So a bigger part of the possible aspects in the Schwartz score are only partially suitable for the diagnostics of affected children. Therefore additional parameters are desired to identify patients correctly.

The LQTS encompasses a disease pattern that is based on mutations on six genes (KCNQ1 → LQT1, HERG → LQT2, SCN5A → LQT3, ANK2-ß → LQT4, KCNE1 → LQT5 and KCNE2 → LQT6) which cause functional changes in selected ion channels of the myocardial cells. This leads to prolonged action potentials of the cells and is represented by a longer QT interval in the surface ecg. Furthermore the changes on the different genes are responsible for different shapes of the repolarization phase in a heart cycle as various studies have shown [3-5].

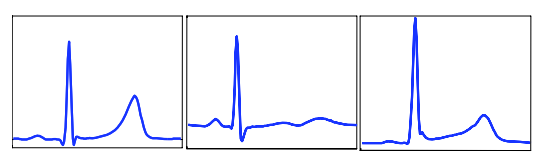


Figure 1 - Characteristic T wave shapes for LQT1 (left), LQT2 (middle) and LQT3 (right)

The drawbacks of the Schwartz score and the described T wave shape for the LQTS subtypes brings up the following

question: Do ecg parameters other than those mentioned in the Schwartz score improve the diagnostic process of LQTS? Furthermore it would be desirable that these features would also support the genotype based classification of identified patients. These two aspects might help in the diagnostic process of LQTS patients which are not identified by the Schwartz score. And additionally the time consuming process of a molecular genetic testing could be shortened if a probability for an affected gene is known.

Materials and methods

Materials and data preprocessing

24h Holter recordings from 14 genetically identified children with LQTS (age: 11,01±7,30 years; gender: f=7, m=7) and 22 healthy control persons (age: 11,31±3,75 years; gender: f=7, m=15) were used for the analysis. The group of patients is divided into three subgroups according to the genotype: LQT1 (9), LQT2 (3) and LQT3 (2).

The ecg recordings were performed using digital twelve channels recorders from Mortara Instrument Inc (Milwaukee, USA). In a first stage the recordings were processed using the Mortara Holter software H-Scribe for the identification and annotation of normal beats. All further processing was carried out using a self developed software framework for biosignal analysis [6]. An ecg viewer which is part of the framework was used to manually add missing beats that were not detected by the Holter software. Superimposed noise was removed using a highpass filter with a cut off frequency of 0,5 Hz and a lowpass filter with a cut off frequency of 50 Hz. Baseline wander was corrected using a cubic spline interpolation.

All further processing of the data except for the morphological indices was carried out using only lead V5. Due to the most distinct depiction of the repolarization phase lead V5 was chosen. Another reason is that lead V5 has the best signal to noise ratio in the considered population. To take daytime dependencies into account, three sections of the recordings were analyzed. These segments were taken from a time window in the morning starting at 8 am, from the evening starting at 6 pm and at night starting at 1 am. From these segments sequences of 10, 20, 50 and 100 consecutive normal beats are used for the analysis.

In all selected segments a self developed delineation routine to find the correct positions of Q, R, R_{Peak}, S, J, T_{Peak} and T_{End} was applied [7].

Methods

Because of the importance of the repolarization phase for the LQTS particular T wave parameters were quantified which can be assigned to the categories time interval measurements, area measurements, and morphology indices.

Time interval measurements

This category contains the features that are widely used in clinical routine. These are the intervals between prominent marker positions which were automatically inserted during the delineation process. Figure 2 demonstrates the considered intervals. All measures of this category were calculated as absolute and frequency corrected values using Bazett's formula.

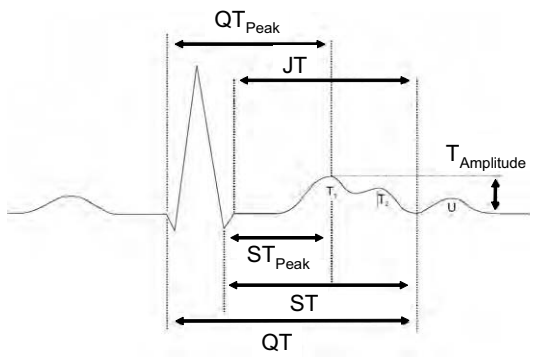


Figure 2 - Time interval measurements

Area measurements

The following area-derived parameters contain also information about the repolarization phase [8]. Table 1 summarizes the parameters in the area measurement category.

Table 1 - Area measurements

Feature	Description
A_{tot}	Total area of the repolarization phase
t_{A50}	Time interval to accumulate 50% of the area.
Pt_{A50}	Percentage of repolarization time needed to accumulate the first 50% of the area.
$t_{A25\text{-}50}$	Time interval to accumulate the mid 50% of the repolarization area from 25% to 75%.
Pt_{A25_75}	Percentage of repolarization time needed to accumulate the mid 50% of the repolarization area from 25% to 75%.
t_{A97}	Time interval to accumulate 97% of the total repolarization area
T_{amp}	Maximum amplitude of the T wave

Morphology indices

The calculated morphology indices are applied as proposed by Acar et al [9]. They characterize the spatial and temporal variations of the T wave morphology and wavefront direction differences between the depolarization and repolarization phase.

In contrast to the already described parameters the morphology indices are applied to a vector curve. It is constructed by applying a singular value decomposition on the data matrix whose rows correspond to the ecg leads I,

II, V1-V6. A summary of the morphology indices is given in table 2.

Table 2 - Morphology indices

Feature	Description
TMD	T wave morphology dispersion. Represents the variation of the morphology of the repolarization phase between different leads
TMD_{pre}	The same as TMD, but takes only the rising edge of the T wave into account
TMD_{post}	The same as TMD_{pre} with respect to the falling edge of the T wave
TCRT	Total cosine R to T. Quantifies the angle between the R- and T-wave loop
PL	Percentage of the inner area in a 2D projection that is covered by the T wave
PO	Percentage of the outer area in a 2D projection that is covered by the T wave
LD_1, LD_2	Lead dispersion 1 and 2: variation of the ecg vector during the repolarization phase

Statistical evaluation

The discriminative power of single parameters was evaluated using receiver operating characteristics (ROC) plots. Combinations of up to three parameters were used as inputs for a second order polynomial classifier. The selectivity of all results is given by values of sensitivity and specificity.

Results

As mentioned above all features were calculated for sequences of consecutive normal beats at specific times of the day. By means of this approach time dependent effects should be considered. All parameters were calculated for each single beat in the corresponding segments. To aggregate the results mean values for the parameters were calculated for the first 10, 20, 50 and 100 consecutive beats.

Patient identification

The best results for single features are given by the time interval measurements with rates of correct classified datasets of up to 97%. For two reasons these results were excluded in the further investigations. First of all the widely used Bazett formula for the frequency correction yields erroneous data sets if the heart rate is too high which is very common in young childrens. The second reason for the exclusion is the strong correlation between the parameters in this category and the fact that QTc is already part of the Schwartz score. For the remaining features those from the category of area measurements yield the best results. Table 3 lists the best features to discriminate the LQTS patients from the healthy subjects.

As table 3 shows, the best separation of the two groups is achieved in the evening segment using the parameter Pt_{A50} with values of 86% for the sensitivity and 100% for the specificity. This yields a rate of correct classifications of 93%.

Table 3 - Best single features for the identification of LQTS patients. The discriminative power is given by values for sensitivity and specificity in percentage

	10 Beats	20 Beats	50 Beats	100 Beats
Morning 09 am	t_{A50}: 86 / 91	t_{A50}: 79 / 91	t_{A50}: 79 / 91	t_{A50}: 79 / 95
Evening 06 pm	Pt_{A50}: 86 / 95	Pt_{A50}: 86 / 100	Pt_{A50}: 86 / 100	Pt_{A50}: 86 / 95
Night 01 am	Pt_{A50}: 71 / 86	t_{A25_75}: 64 / 91	t_{A25_75}: 64 / 95	Pt_{A50}: 64 / 91

The results in table 3 also indicate that the number of feature values used in the averaging process is not significant. Reliable values were achieved with a minimum number of 20 beats. The results for the morphological parameters are not suitable to separate the two groups of the study.

Combinations of two features were also evaluated. The criteria for parameters to be used in this task is given by values for the sensitivity and specificity greater than 70%. Table 4 lists the best combinations of two parameters.

Table 4 - Best feature combinations for the identification of LQTS patients. The discriminative power is given by values for sensitivity and specificity in percentage

	10 Beats	20 Beats	50 Beats	100 Beats
	2 Features			
Morning 09 am	Pt_{A50}, QTc: 86 / 100	Pt_{A50}, SPTc: 86 / 100	Pt_{A50}, QTc: 86 / 100	Pt_{A50}, QTPc: 86 / 100
Evening 06 pm	Pt_{A50}, SPTP: 86 / 100	Pt_{A50}, QTc: 86 / 100	Pt_{A50}, QTc: 93 / 100	Pt_{A50}, QTc: 93 / 100
Night 01 am	Pt_{A50}, SPTP: 93 / 91	Pt_{A50}, SPTP: 93 / 91	Pt_{A50}, SPTP: 93 / 91	t_{A25_75}, QTPc: 71 / 100

As for single features the best results for parameter combinations are achieved in the evening. A rate of 97% correctly classified datasets was achieved for the combination of Pt_{A50} and QTc.

Genotype related classification

Due to the small number of patients in the three target groups LQT1 (9 children), LQT2 (2 children) and LQT3 (3 children) a reliable classification was not possible. Nevertheless single features were evaluated with respect to their possible application for this task.

The most promising results were achieved by the area measurement t_{A50} and the morphology index TCRT. For t_{A50} the rate of overall correctly classified patients was 79% whereas TCRT yielded 77%. Despite the small sample size these result can be used as a first indicator for a genotype related classification of LQTS patients.

Discussion

The established gold standard in the diagnosis of the Long QT-syndrome is the Schwartz score as mentioned in the introduction. But especially in the diagnosis of affected young children and babies this approach often gives non-optimal results due to two limiting factors. The first one is that the frequency correction according to the Bazett formula gives erroneous values for recordings with high frequencies. Second the aspects from the previous medical history can not be taken into account as they are often not available for the young patients. Therefore the accomplished scores are often to low for the identification of affected patients. As mentioned before our work addresses these deficits by adding additional information derived from the analysis of ecg recording to the diagnostic process. These new features are not intended to replace the Schwartz score but to be used in addition especially when applied to children.

Besides of the QT interval the prolongation of the T wave can be clearly quantified by the area derived measurements as our results indicate. Approximately 68% of the whole time for the repolarization phase is needed to fill the first 50% of the covered area (Pt_{A50}) whereas healthy test persons needed only 62% of the time.

Interesting is also the fact that the used morphology indices are not suitable to describe the characteristic T wave patterns that were qualitatively described in other studies [3-5].

As our findings indicate, the area measurements and partly the morphology indices are suitable to identify LQTS patients. In combination with the traditional time interval parameters they yield the best results in the classification process. This is also in accordance with the findings of Struijk and co-workers [10] who did similar investigations on resting ecg recordings obtained from adults.

As is known, time series constructed from RR or QT intervals exhibit circadian properties. This issue was targeted during our work by the examination of three segments from the recodings covering the morning, the evening and the night. But our results indicate that circadian effects have only little influence on the discriminative power of the used features as table 3 and 5 demonstrate.

Although our findings indicate good separative properties for the time interval, area and morphology measurements they have to be interpreted with care as they are based on a relatively small dataset. The reason for this small dataset is the fact that although most of the identified patients underwent genetical testing no established mutations were found on the genes.

The results achieved so far and the aims to find ecg features that can be used in addition to the Schwartz score future investigations are needed. In this study all patients had definite results for the Schwartz score. But in cases with borderline or negative results the experience of the investigating physician is needed. So the next task is to analyze recordings from this special group of patients to validate our findings.

Conclusion

Using a collection of parameters from the categories of time interval, area and morphology measurements a good separation of patients with LQTS from healthy control persons could be achieved. The parameters were selected complementary to those mentioned in the Schwartz score and introduce new aspects in the process of diagnosing LQTS. Future work should focus especially on the question whether patients who are not identified by the Schwartz score can be recognized using the additional ecg features. Furthermore ecg parameters were evaluated that might be used to discriminate the most common LQT subtypes LQT1, LQT2 and LQT3. Although only a small dataset was available the achieved results motivate for further investigations.

Acknowledgments

This work is supported by a grant from the Ministry of Science, Research and the Arts Baden-Württemberg, Germany

References

[1] Schwartz PJ, Priori SG and Napolitano C. The Long QT Syndrome. In: Zipes DP and Jalife J. eds. Cardiac Electrophysiology – From Cell to Bedside. W.B. Saunders Company, 2000; pp 597-615

[2] Schwartz PJ, Moss AJ, Vincent GM, Crampton RS. Diagnostic criteria for the long QT syndrome. An update; Circulation. 1993 Aug; 88(2):782-4.

[3] Zareba W. Genotype-specific ECG patterns in long QT syndrome. J Electrocardiol. 39 (2006) S101-6

[4] Moss AJ, Zareba W, Benhorin J, Locati EH, Hall WJ, Robinson JL, Schwartz PJ, Towbin JA, Vincent GM and Lehmann MH. ECG T-wave patterns in genetically distinct forms of the hereditary long QT syndrome. Circulation. 1995; 92:2929-34.

[5] Zhang L, Timothy KW, Vincent GM, Lehmann MH, Fox J, Giuli LC, Shen J, Splawski I, Priori SG, Compton SJ, Yanowitz F, Benhorin J, Moss AJ, Schwartz PJ, Robinson JL, Wang Q, Zareba W, Keating MT, Towbin JA, Napolitano C and Medina A. Spectrum of ST-T-wave patterns and repolarization parameters in congenital long-QT syndrome: ECG findings identify genotypes. Circulation. 2000;102:2849-55.

[6] Dickhaus H, Maier C, Bauch M, Rengshausen U, Khalil M, Hessling G and Ulmer HE. A System for Automated Analysis of RR-Series Applied to Identification of Patients with Neurocardiogenic Syncope, Proceedings of the Second Joint EMBS/BMES Conference Houston, TX, USA, 2002, 1620-1.

[7] Siegler A. Genotyp-Phänotyp Relation bei Patienten mit Long QT-Syndrom durch das 24-Stunden EKG. Master Thesis. Heilbronn University 2006.

[8] Zareba W, Moss AJ and Konecki J. TU Wave Area-Derived Measures of Repolarization Dispersion in the Long QT Syndrome. J Electrocardiol. 1998; 30 Suppl:191-5.

[9] Acar B, Yi G, Hnatkova K and Malik M. Spatial, temporal and wavefront direction characteristics of 12-lead T wave morphology. Med Biol Eng Comput. 1999; 37(5): 574–584.

[10] Struijk JJ, Kanters JK, Andersen MP, Hardahl T, Graff C, Christiansen M and Toft E. Classification of the long QT syndrome based on discriminant analysis of T-wave morphology. Computers in Cardiology. 2005; 25-28: 511- 514

Address for correspondence

Dipl.-Inform. Med. Matthias Bauch
Department of Medical Informatics
Heidelberg University
Im Neuenheimer Feld 400
D-69120 Heidelberg
Email: mbauch@hs-heilbronn.de

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Temporal Abstraction and Data Mining with Visualization of Laboratory Data

Katsuhiko Takabayashi [a], Tu Bao Ho [b], Hideto Yokoi [c], Trong Dung Nguyen [b], Saori Kawasaki [b], Si Quang Le [b], Takahiro Suzuki [a], Osamu Yokosuka [d],

[a] *Division for Medical Informatics and Management, Chiba University Hospital, Inohana, Chuou-ku, Chiba, 260-8677 Japan*
[b] *Japan Advanced Institute of Science and Technology,Tatsunokuchi, Ishikawa, 923-1292 Japan*
[c] *Division for Medical Informatics, Kagawa University Hospital, Ikedo, Miki, Kida-gun, Kagawa, 761-0793 Japan*
[d] *Department of Gastroenteology, Chiba University Hospital, Inohana, Chuo-ku, Chiba, 260-8677 Japan*

Abstract

To analyze the laboratory data by data mining, user-centered universal tools have not been available in medicine. We analyzed 1,565,877 laboratory data of 771 patients with viral hepatitis in order to find the difference of the temporal changes in laboratory test data between Hepatitis B and Hepatitis C by the combination of temporal abstraction and data mining. The data for one patient is temporal for more than 5 years. After pretreatment the data was converted to abstract patterns and then selected into sets of data combination and rules to identify Hepatitis B or C by D2MS and LUPC which were originally produced by ourselves. Not only data pattern, but also temporal relations were considered as a part of the rules. In the course of evaluating the results by domain experts, even though there were not so remarkable hypotheses, visualization tools made it easier for them to understand the relations of the complicated rules.

Keywords:

databases, liver function test, hepatitis

Introduction

Data mining is beginning in medical application. Even though there are many data mining techniques to analyze the database, most of them are still experimental and in the hands of computer scientists. Medical doctors cannot use the tools for their own data analysis individually. User-centered universal tools should be applied for medical researchers to analyze their own data. In Japan we started the national project of active mining to analyze the same database of viral hepatitis and evaluate the availability and validity to implement it in medicine[1]. In this approach from the medical point of view, not only providing data, but we performed pretreatment of medical data and attended evaluation with mathematical researchers as active mining. One of the keys of laboratory data analysis in medicine is how to treat temporal data. By using temporal abstraction, we aimed to solve this problem.

The goals

One of the concrete goals of this project was to discover the differences in the temporal patterns of hepatitis B (HBV)and C (HCV) which has not been clearly defined, and, more importantly, to examine whether the methods we applied here can work well and be applied to other fields.

Materials and methods

Database

The hepatitis data is located in two databases, one is the laboratory database of the hospital, and another is the biopsy and clinical data in the department of Medical School, Chiba University. The contents are as follows:

- Basic information of patients (total 771 records)
- Results of biopsy (total 960 records)
- Information about measurements in in-hospital tests (total 459 records)
- Results of out-hospital tests (total 30,243 records) and those of in-hospital tests (total 1,565,877 records)

The data in the hospital are more than 20 years. Those patients were performed liver biopsy once at least.

Preprocessing

Our preprocessing of hepatitis data includes data cleaning, data integration, data reduction, deidentification and data transformation. We removed only redundant and not unsuitable suffix data for further processing and we eliminate noisy data in the next temporal abstraction step. For the purpose of temporal abstraction, we have to integrate original relational data tables into one data table, where each column represents laboratory examination. By combining the expert guidance and the frequencies of attributes presented from 983 examinations, we selected 15 most significant examinations. The numbers of examinations for each patient are different, and the examination periods are irregular.

Temporal abstraction

Temporal abstraction (TA) is one approach to deal with time-related data in medical research. TA methods are

those able to derive an abstract description of temporal data by

extracting their most relevant features[2] .The key idea is to transform time-stamp points by abstraction into an interval-based representation of data by extracting their most relevant features [3]. TA task can be defined as follows.

The input includes a set of time-stamped data points (events). The output includes a set of interval-based, context-specific unified values or patterns (usually qualitative) at a higher level of abstraction. TA can be generally considered in two phases: *basic* TA for abstracting time-stamped data from given episodes (which are significant intervals for the investigation purpose) and *complex* TA for investigating specific temporal relationships between episodes that can be generated from a basic TA or from other complex TAs.

Basic temporal relations

We started by a separation of two groups of tests, one with values that change rapidly in the short term such as GOT, GPT, TTT and ZTT and the other with values that change slowly in the long term such as T-CHO, CHE, ALB.

Basic temporal abstractions typically extract *states* (e.g., low, normal, high), and/or *trends* (e.g., increase, stable, decrease) from a uni-dimensional temporal sequence.

The essential ideas of our temporal abstraction methods here is to deal with long and irregular time-stamp sequences, and doing abstraction in efficient. We introduce the notion of "changes of state" to characterize the slowly changing tests, and the notions of "base state" and "peaks" to characterize the rapidly changing tests .

Temporal abstraction primitives

From observation and analysis, we defined the following temporal abstraction primitives:

1. State primitives: N (normal), L (low), VL (very low), XL (extreme low), H (high), VH (very high), XH (extreme high).
2. Trend primitives: S (stable), I (increasing), FI (fast increasing), D (decreasing), and FD (fast decreasing).
3. Peak primitives: P (peaks occurred).
4. Relations: ">" (change state to), "&" (and), "–" (and then), "/" ("X/Y" means majority of points are in state X and minority of points are in state Y).

The thresholds to distinguish the state primitives of tests are given by medical doctors, for example, those of the test GOT are 40, 100, 200, respectively. We define structures of abstraction patterns as follows:

<pattern> ::= <state primitive>

<pattern> ::= <state primitive> <relation>

<pattern> ::= <state primitive> <relation> <peak>

<pattern> ::= <state primitive> <relation> <state primitive>

Examples of abstracted patterns in a given episode are like follows:

GOT = H (all the values of GOT in one case are above the normal region as shown in the upper left in Figure 1),

"GPT = H&P" (The values of GPT in one case are always high and with peaks like lower left of Fig 1),

"I-BIL = N>L>N" (I-BIL first is normal, then changed to the low region, and finally changed to the normal region in one case like the right bottom in Fig 2) etc.

Figure 1 shows typical possible patterns (8 and undetermined) for rapidly changing tests, and *Figure 2* shows typical possible patterns (21 and undetermined) for slowly changing tests [4].

Abstraction of rapidly changing test results

From our observation and analysis, especially GPT and GOT were defined as rapidly changing attributes, which can go up in a very short period and go back to a stable" state. Thus two most representative characteristics of these tests are a "stable" base state (BS), and the position and value of peaks, where the attributes suddenly go up. Based on this, we formulated the following algorithm to find the base state and peaks of a test. Rapidly changing tests applied also to TTT and ZTT and they showed 9 patterns.

Algorithm 1 (for rapidly changing tests)

Input: A sequence of patient's values of a test with length N denoted as S_{00} = {s1, s2, …, sN} in a given episode.

Output: Base state and peaks, and an abstraction of the sequence derived from them.

Parameters: NU, HU, VHU, XHU: upper thresholds of normal, high, very high, extreme high regions of a test, a (real).

Notation:

- Mi: Set of local maximum points of S
- BS: base state of S
- PEi: set of peaks of S

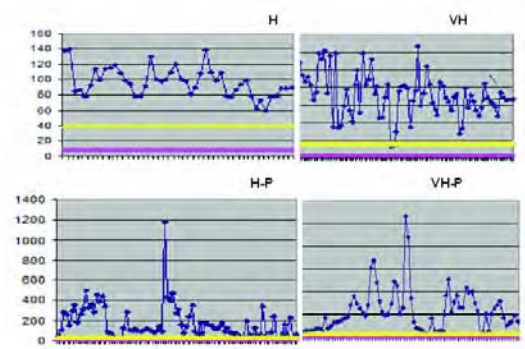


Figure 1 - rapidly changing test patterns

Abstraction of slowly changing test results

The key idea is to use the "change of state" as the main feature to characterize sequences of the tests. It can be seen that the "change of state" characterize information of both state and trend of the sequences.

From the beginning of a sequence, the first data points can be at one of the three states “N”, “H”, or “L”. It will happen that:

Either the sequence changes from one state to another state, smoothly or variably (at boundaries), or the sequence remains in its state without changing.

We provided 22patterns for slowly changing data.

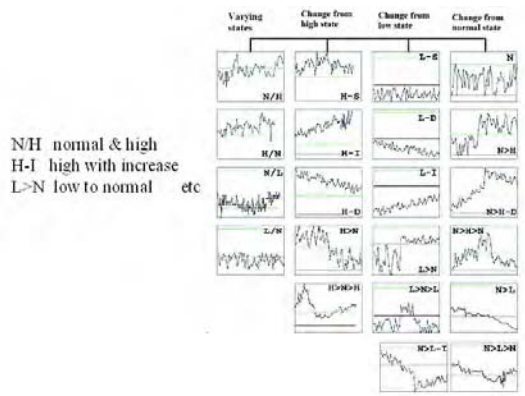


Figure 2 - Slowly changing test patterns

Temporal relationships

The temporal relations between the abstracted events of laboratory data were also treated here as phenomena and a part of the rules by comparing the period of the state. They are classified into seven relations by Allen[8].

$$Z = \frac{conf(R) - p(C)}{\sqrt{p(C)(1-p(C))/n_A}}$$

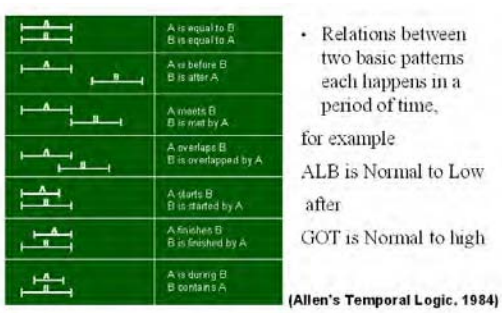


Figure 3 - Temporal relations

Complex temporal abstraction

Mining abstracted hepatitis data with system D2MS

The authors developed an interactive visualization tool in decision tree construction called D2MS (data mining for Model Selection) for supporting an effective cooperation of the user and the computer in classification. D2MS shares many features with WinViz [5] and Cviz that both use parallel coordinates. WinViz allows the user to visually examine a tabular database and to formulate query interactively and visually. Cviz is an attempt to integrate visualization into the knowledge discovery process.

D2MS facilitates the trials of various alternatives of algorithm combinations and their settings. The data mining methods in D2MS consists of programs CABRO for tree learning and LUPC for rule learning [6]. CABRO produces decision trees using R-measure and graphically represents them in particular with T2.5D tool (trees 2.5 dimension). As shown in *Figure 4*, visualization made us easily recognized the different pattern of HBV and HCV.

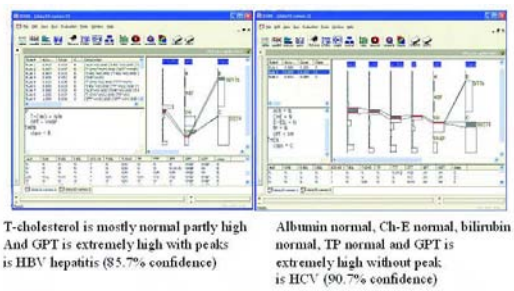


Figure 4 - Rules for HBV and HCV by D2MS

We examined statistical significance of the consequence according to the method of [7], which prunes discovered rules statistically as follows.

Assume a rule R: A C (or R: A ¬C) with confidence conf(R). If conf (R) = p(C) then R is eliminated. To test whether conf (R) = p(C), we use the following test statistic where n_A is the number of cases satisfying C

Mining abstracted hepatitis data with LUPC

LUPC is a separate-and-conquer algorithm that controls the induction process by several parameters. The parameters allow obtaining different results and this ability allows the user to play a central role in the mining process [6].

LUPC is developed to learn prediction rules from supervised data. Each rule found by LUPC is a conjunction of attribute-value pairs that may present an interesting pattern. The main features of LUPC are (1) its ability of finding rules with associate domain knowledge (such as finding rules containing or not containing specified attribute-value pairs), as well as finding rules for minority classes; (2) it is integrated with D2MS’s rule visualizer and thus supports the user in selecting the appropriate rules which result from different possible settings of parameters. The performance of LUPC depends on several parameter specified by the user: for min accuracy of rules, for min coverage of rules, for maximal number of candidate

rules in the beam search, and for maximal number of attribute-value pairs to be consider. By varying these parameters we can find different sets of rules [6]. When using the setting with default parameters of = 80%, = 3, = 200, and = 100, we found 119 rules characterizing the hepatitis B and 152 rules characterizing hepatitis C.

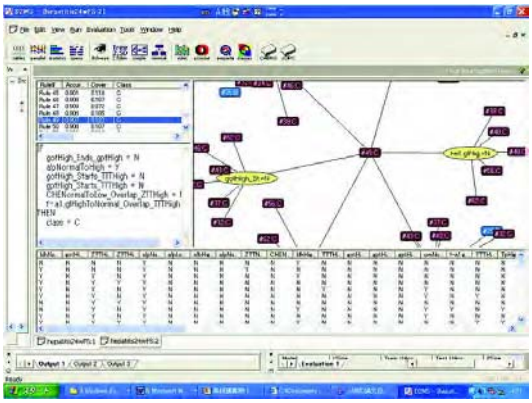


Figure 5 - LUPC Rules can be illustrated in a left figure

Evaluation

The produced rules were evaluated by three experts.

Results

By using D2MS, we discovered 33,477 rules for type B and C difference. These rules are complicated and sometimes contradictory to each other. For example, there is a rule if T-Bilirubin(Bil) is N(normal) , ZTT is N, and GOT is N with P(peak), then HBV, while there is another rule if D-Bil is N, TTT is N, GOT is N with P, then HCV, which are almost the same but the results are completely different.

After pruning by statistical aspects between HBV and HCV, there are only 27 rules (0.08%) left (*Table1*). However, these rules seemed not attractive for medical doctors even if they are statistically significant. For example, T-cholesterol is normal is HCV in 171/260 cases (66%), or GPT is high with a peak and ZTT is mostly high partly normal is HBV. They are too simple or vague and must be carefully assessed.

Different datasets were found by using LUPC with various parameters including temporal relations between laboratory tests. *Table2* presents the top five rules by LUPC from the point of coverage and confidence. From this table, especially rule3 and 5 are similar and could be merged as a rule that if TP decreasing high to normal and both ZTT and TTT is high, then it is HCV. In the evaluation of medical doctors, though most of them seemed not crucial or not useful in clinical medicine even the discovered rules covering many cases with high accuracy. However, some of the rules could be reasonable from the different clinical course of two types of hepatitis, especially when the experts checked and integrated the rules in the illustration.

Table 1 - 27 pruned rules produced by D2MS and satisfying chi-square test

No.	class	T-CHO	CHE	GOT	GPT	TTT	ZTT	D-BIL	T-BIL	I-BIL	TP	acc	ratio
1	C (95%)	N										0.66	171/260
2	C (95%)		N									0.72	183/256
3	C (95%)							N				0.73	180/248
4	C (95%)			H								0.76	89/117
5	C (95%)				H							0.76	78/103
6	C (95%)		N					N				0.82	142/173
7	C (95%)					N	H/N					0.92	11/12
8	C (95%)				H				N/H			0.93	14/15
9	B (95%)				H&P		N/H					0.92	11/12
10	B (90%)						N					0.68	83/93
11	B (90%)		H/N									0.7	14/20
12	B (90%)		N/H									0.74	23/31
13	B (90%)			N&P	H&P							0.7	16/23
14	B (90%)		N				N/H	N				0.88	7/8
15	C (90%)		N			N						0.8	67/84
16	C (90%)						H-I	N				0.95	63/66
17	C (90%)			H	XH							0.92	11/12
18	C (90%)		N	N					N			0.93	26/28
19	C (90%)										N/H	0.81	35/43
20	C (90%)					H-I				N		0.93	25/27
21	C (85%)				N&P	N						0.69	33/40
22	C (85%)		N	N								0.84	41/49
23	C (85%)		N	H								0.87	58/67
24	C (85%)					H/N				N		0.79	23/29
25	C (85%)				XH							0.72	18/25
26	C (85%)	N/L										0.8	28/35
27	C (85%)						H-D					0.83	33/40

Table 2 - Top 5 rules selected by LUPC

Rule 1(Coverage ; 4.098% confidence ; 100% coverage ; 25 cases)

creNormalToLow = Y — Creatinine decreasing from normal to low
gptHigh_Start_gotHigh = Y — and GPThigh start with GOT high
-> Class = C — is HCV

Rule 2(Coverage 3.443% confidence ;100% coverage;21 cases)
Bilirubin decreasing from normal to low before TTT elevates
is HCV

Rule 3 (Coverage 3.443% confidence ;100% coverage;21 cases)
TP decreasing high to normal and ZTT goes up to high after TTT up to high
is HCV

Rule 4 (coverage 3.279% confidence 100% coverage; 20 cases)
Creatinine decresing normal to low and bilirubin increasing normal to high
is HCV

Rule 5 (coverage 2.951% confidence 100 % coverage; 18 cases)
TP decreasing high to normal and ZTT goes up to high before TTT up to high
is HCV

In the viewer of LUPC we can see the accuracy and coverage on the upper left, rule itself in the middle and the relation of the rules and attribute value pairs in the figure to the right which can be manipulated by users (*Figure 6*). By handling it users can see the relations of each item. In this figure doctors could easily recognize if LDH is low to normal is false, then all the rules are related to HCV, while if creatinine is normal to low is false, then it is related to the rules of HBV except rule #48(center), so that the doctors could understand from a more comprehensive point. This technique was highly evaluated by medical doctors and some rules such as the top 5 were considered as meaningful.

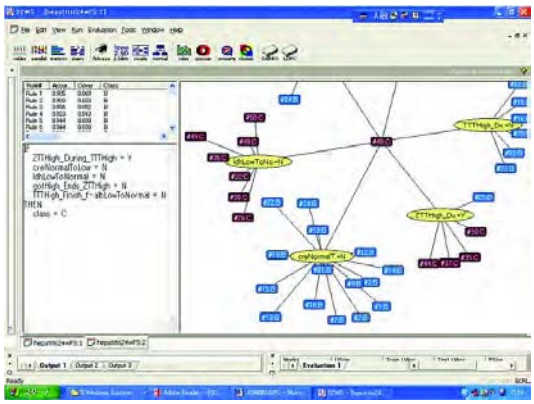


Figure 6 - LUPC makes it easier for the users to understand the relations comprehensively

Discussion

We have presented a temporal abstraction approach to data mining the temporal hepatitis data. Most doctors do not believe that they can distinguish HBV and HCV from the laboratory data change, so it might be true when we cannot obtain any new findings from this data mining. However, we in fact obtained many rules to identify HBV or HCV with statistical significance. Some of them look very simple. To confirm, we need to analyze them by stratified analysis, removing the modification of other factors such as treatment and then compare again. By LUPC, we can estimate the patient data comprehensively and expect new findings in many diseases because it is difficult for human beings to find data changes over a long period of time. some experts of liver diseases mentioned that the cases with HBV and HCV were apt to show different clinical changes, and our results would reflect these changes.

One of the major problems in rule based data mining is that there are too many rules deduced for us to evaluate. To select most important rules we introduced the chi-square test which was effective to decrease the number of rules as well as statistical reasoning. Another is by LUPC, not only selecting minority classes from large unbalanced datasets but visualization, it is not difficult to separate the important ones from many rules for medical doctors.

Other studies of data mining in medicine are mostly in the field of genomics and epidemiology and the analysis of laboratory data is quite limited. We provided the data of anti-phospholipid antibody syndrome to PKDD 1999 as model data in order to establish a new technique as well as hepatitis data.

Current medical study is deeply inclined to use a prospective way or Cohort as a scientific study design, which implies a carefully planed experiment. However, when we think of a long-term experiment lasting over 10 years, it is not realistic to study prospectively. There is a great possibility of new paradigms appearing before the study is completed. Retrospective studies are expected for these long term studies and data mining techniques will play a major role in this filed by creating high potential hypotheses.

Even though we did not discover crucial rules to show the difference of laboratory data change between HBV and HCV, we proved to show that this combination of TA and data mining with visualization is useful and effective. Furthermore, it could be applied to other fields of medicine and would be a basic model for the universal analysis of data mining for temporal data analysis in medicine.

Conclusion

The rules that show the difference of the laboratory changes in the long clinical course between the HBV and HCV could be deduced by D2MS. Pruning by statistical significance could decrease the number of rules but obtained rules were not interesting in individuals. Visualization made it easier for doctors to find the relations and led them to find reasonable results. This combination technique of temporal abstraction and data mining with visualization could be applied universally.

References

[1] Motoda, H., Active Mining: New Directions of Data Mining, IOS Press, 2002.

[2] Horn, W., Miksch, S., Egghart, G., Popow, C., Paky, F., "Effective Data Validation of High-Frequency Data: Time-Point-, Time-Interval-, and Trend-Based Methods", *Computer in Biology and Medicine, Special Issue:Time-Oriented Systems in Medicine*, 27(5), 389-409, 1997.

[3] Bellazzi, R., Larizza, C., Magni, P., Monntani, S., and Stefanelli,M. Intelligent Analysis of Clinic Time Series: An Application in the Diabetes Mellitus Domain *Artificial Intelligence in Medicine*, 20, 37-57, 2000.

[4] Ho, T.B., Nguyen, T.D., Kawasaki, S., Le, S.Q., Nguyen, D.D., Yokoi, H., Takabayashi, K. Mining Hepatitis Data with Temporal Abstraction. *ACM Inter. Conf. on Knowledge* Discovery and Data Mining KDD-03, 369-377,2003.

[5] Lee, H.Y., Ong, H.L., and Quek, L.H. Exploiting Visualization in Knowledge Discovery. *First Inter. Conf. on Knowledge Discovery and Data Mining,* 1995, pp. 198–203.

[6] Ho T.B., Nguyen T.D., Nguyen D.D., Kawasaki S., Visualization Support for User-Centered Model Selection in Knowledge Discovery and Data Mining. *Int. Journal Artificial Intelligence Tools,* 2001;10(4): 691-713.

[7] Bruzzese, D. and Davino, C. Statistical Pruning of Discovered Association Rules. *Computational Statistics,* 16 (3), pp. 387 –398, 2001.

[8] Allen E and Jai S. A decidable temporal logic to reason about many processes. Proceedings of the ninth annual ACM symposium on Principles of distributed computing table of contents. Quebec City , Canada 233-246,1990.

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Automated Interpretation of Optic Nerve Images: A Data Mining Framework for Glaucoma Diagnostic Support

Syed SR. Abidi[a], Paul H. Artes[b], Sanjan Yun[a], Jin Yu[a]

[a] *NICHE Research Group, Faculty of Computer Science, Dalhousie University, Halifax, Canada*
[b] *Department of Ophthalmology and Visual Sciences, Dalhousie University, Halifax, Canada*

Abstract

Confocal Scanning Laser Tomography (CSLT) techniques capture high-quality images of the optic disc (the retinal region where the optic nerve exits the eye) that are used in the diagnosis and monitoring of glaucoma. We present a hybrid framework, combining image processing and data mining methods, to support the interpretation of CSLT optic nerve images. Our framework features (a) Zernike moment methods to derive shape information from optic disc images; (b) classification of optic disc images, based on shape information, to distinguish between healthy and glaucomatous optic discs. We apply Multi Layer Perceptrons, Support Vector Machines and Bayesian Networks for feature sub-set selection and image classification; and (c) clustering of optic disc images, based on shape information, using Self-Organizing Maps to visualize sub-types of glaucomatous optic disc damage. Our framework offers an automated and objective analysis of optic nerve images that can potentially support both diagnosis and monitoring of glaucoma.

Keywords:

glaucoma, data mining, feature selection, clustering, Confocal Scanning Laser Tomography, Support Vector Machines

Introduction

Glaucoma is an eye disease that is characterized by slow progressive damage to the optic disc and corresponding deterioration of the patient's vision [1]. At present, there is a gap in the understanding of the cause, the types and the natural course of glaucoma. The use of sophisticated imaging technologies, such as Confocal Scanning Laser Tomography (CSLT), capture 3-dimensional images of the optic disc that are used for diagnostic purposes [2]. However, the interpretation of CSLT images is a manual and subjective process—a trained professional has to manually define the margins of the optic nerve based on his/her training and expertise and then classify the optic nerve as normal or glaucomatous. The current process allows for misjudgments/errors in the interpretation of the CSLT image, inability to distinguish between actual and noisy images and variance in the diagnostic recommendations over a set of practitioners. The challenge, therefore, is to automate the analysis of CSLT images of the optic disc, in an objective and quantifiable manner, to support practitioners in the diagnosis and therapeutic management of glaucoma.

Researchers have analyzed optic nerve data and CSLT based images with varying results. Bowd et al [3], working with retinal tomography images applied forward and backward feature selection methods for training Multi Layer Perceptron (MLP), Support Vector Machine (SVM) and linear discriminant functions; Park et al [4] used correlation analysis and forward wrapper model to select features from optic disc data for training SVM classifiers; Swindale et al [5] used a wrapper model for feature selection to train SVM classifiers.

We have developed a data-driven Glaucoma Diagnostic Support (GDS) system that features the automatic interpretation of CSLT topography images of the optic nerve to support (a) the classification of the optic disc images to distinguish between healthy and diseased optic discs; (b) the identification of the sub-types of glaucomatous optic disc damage. This is to help further sub-classify the glaucoma patients in order to provide treatments in line with the specific morphological patterns of damage [6]; and (c) the visualization of the temporal progression of the disease for a patient over a period of time.

In this paper we present an automated approach to CSLT topography image analysis to support glaucoma diagnosis. Our multi-stage approach is a hybrid of image processing and data mining methods. In Stage 1, we apply image-processing techniques to CSLT images to derive image-defining features. In Stage 2, we apply data classification methods to the image's shape-defining features to develop classifiers that can discriminate between healthy and glaucomatous optic discs. An important task at this stage is feature selection whereby we select an optimal subset of image features that exhibit high image classification capabilities. In Stage 3, we apply data clustering techniques to the optimal subset of image-defining features in order to identify the different sub-types of glaucomatous images in the image data-set. The emergent image clusters are subsequently used to both visualize the progression of the disease and the identification of noisy optic nerve images. In Stage 4, we apply rule-induction techniques to the optimal subset of features to induce classification rules (not discussed here). These symbolic rules provide practitioners with a justification of the diagnostic recommendations by our image classifiers. For our experiments, we worked with 1257 tomography images taken at different time intervals from 136 subjects (51 healthy subjects and 85 glaucoma patients).

Methods

Figure 1 illustrates the functional design of our GDS system. We explain the methods developed for each processing stage.

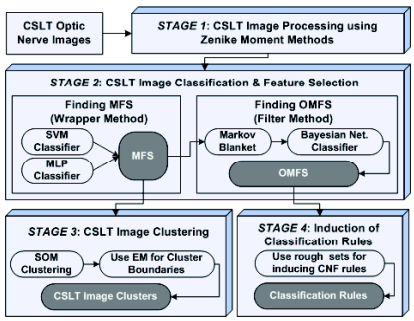


Figure 1 – Functional design of our GDS system

Stage 1: CSLT image processing

This stage involves the extraction of shape-defining features from CSLT images. These features are used to develop the image classification and clustering models. We use an image processing technique referred to as Moment Methods that describes the properties of connected regions in binary images as *Moment* features. We use Zernike moments [7] which use a set of complex polynomials to describe the image's properties by their order (n) and repetition (m) with respect to a digital image—the low order moments capture gross shape information and high order moments incrementally resolve high frequency information (representing detail) of the digital image. Two attractive features of Zernike moments for our purpose is that (a) moments can be made invariant to shifts, rotations and magnification changes; and (b) the optic nerve is centered in the image, thus avoiding the requirement for an independent segmentation stage in which the object is explicitly identified.

For each CSLT image we generated 254 Zernike moments, grouped in an incremental order ranging from 1 to 29—each group comprises a set of ordered moments. Low order moments capture fundamental geometric properties and high order moments represent detailed information of the image [7].

For efficient classification of CSLT images, it was important to select an optimal number of lower order moments. This is a non-trivial task because: (a) there is no objective measure to determine the exact number of (low order) moments needed to achieve high classification accuracy; and (b) there is no discernable relationship between the moments that can be utilized. Given these challenges, next we pursue feature subset selection in conjunction with image classification.

Stage 2: Classification of CSLT images

In the previous stage, we derived a 254 moment based representation for each CSLT image. In this stage, we pursue the classification of CSLT images based on a sub-set of low order moments. This stage therefore involves two tasks—i.e. firstly feature (sub-set) selection and secondly image classification. We have developed a two-pass image classification method that incorporates feature sub-set selection as an integral element (see Figure 1). In the first pass, MLP and SVM based wrapper models are simultaneously used to generate a *Moment Feature Subset* (MFS) consisting of low order moment features. In the second pass, we apply a Markov blanket filter method [8] based on an inferred Bayesian network to select the highly relevant moments from the MFS—i.e. the *Optimal Moment Feature Subset* (OMFS)—that offers reasonably high image classification despite using a small number of moments.

Pass I: Using MLP and SVM

In the absence of any guiding principle to determine the size of the MFS, we devised an accumulative feature subset selection strategy as follows: (a) Generate training-set by incrementally adding the next higher order moments to an existing training set. We exploited the intrinsic partitioning of the 254 moments in terms of their order ranging from 1 to 29. Therefore, feature subset1 included moments with *order2*, feature subset2 includes moments with order 2 and 3, and so on. In total 29 different training sets were generated, where each training set covered all the images based on the moment orders chosen to represent it; (b) Train both a MLP and a SVM classifier separately using the 29 training sets. In total, 29 different MLP and SVM classifiers were trained. For training the MLP and SVM, we partitioned the images so that 75% images were used for training and 25% images were used for testing. For training the SVM, based on the training data a 5-fold cross validation was performed to find the optimal hyper parameters: C and λ; and (c) Determine the classification accuracy of both classifiers, using the test images that are represented by the same number of moments as used to train the classifier.

The next step was to determine the size of the MFS and based on it to select the most efficient MLP and SVM classifier. Our objective was to select the largest possible number of moments without compromising the classification accuracy. To do so, we plotted the classification accuracy of both classifiers and then identified the highest accuracy point on the plot (i.e. with respect to n moment groups) just prior to a downward trend in the classification accuracy as a result of the inclusion of the next higher moment group. The most low order moment groups that achieved the highest classification accuracy were selected as the MFS. And, the MLP and SVM classifiers trained using the MFS were deemed as the most efficient.

A comparison of the classification accuracy trends for both the MLP and the SVM classifiers (see figure 2) shows that both classifiers exhibited a similar classification accuracy trend—i.e. they both start with a relatively high accuracy with the first moment group and then the accuracy drops with the addition of the next few moment groups. But later the accuracy starts to pick up again such that for the MLP it peaks when the feature subset constitutes the first 8 moment groups, whereas for the SVM the accuracy peaks for the first 11 moment groups. It is interesting to note that the classification accuracy with higher order moment groups is relatively low as compared to the peak achieved with just the lower order moments.

Based on classification accuracy trend for both classifiers (shown in figure 2), we determined the MFS to contain the first 11 moment groups—i.e. the first 47 moments. With 11 moment groups the SVM exhibited the highest accuracy and the MLP produced its second highest accuracy level.

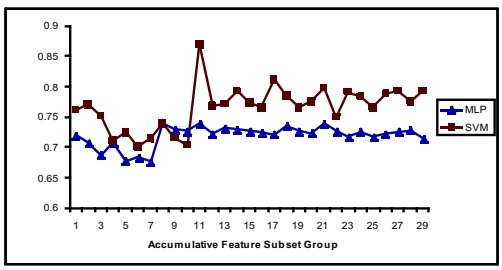


Figure 2 - Classification accuracy for both MLP and SVM

Pass II: Using Markov blanket and Bayesian network

In the second pass, we attempt to further reduce the size of the MFS in order to generate the OMFS that comprises only the highly salient moments. We use a filter model based on a Bayesian Network (BN) and the Markov blanket of the class label [8]. The choice of Markov blanket is guided by the observation that the correlation between most of the moments and their class label is weak, and the same is true for correlation between different moments. Hence, correlation based feature selection methods are not suitable here. We decided to use the Markov blanket approach as it considers every feature's probability dependence relationship during the learning procedure of the Bayesian network's structure.

In a BN where C_A is the set of children of node A, and Q_A is the set of parents of node A, the subset of nodes containing Q_A, C_A and the parents of C_A is called Markov blanket of A [8]. The Markov blanket of a specific feature is a subset of nodes in the BN; it comprises the feature's parent nodes, child nodes and all parent nodes of the child nodes. If we consider the class label node as the root node to learn a BN from data, then all nodes within the Markov blanket of the class node have probabilistic dependence relationship with it.

The steps to generate the Markov blanket were as follows: *Step 1*: We used the *K*2 algorithm to learn the BN. Initially, the 47 moments in the MFS were discretized using an entropy-based method, resulting in 29 moments to be discretized into a single value. These moments were removed from the MFS. Thus we were left with only 18 moments for training the BN. The following moments were retained: moments {1, 2, 5, 6, 7, 12, 16, 21, 23, 25, 27, 33, 36, 37, 43, 44, 45, 46}. *Step 2*: A BN was trained using the18 moments in their original order. Five-fold stratified cross validation was used to evaluate the classification accuracy (see table 2 for results). *Step 3*: The 18 moments were ordered based on the chi squared statistical test score x^2 between the moments and their class labels. The moments with the highest x^2 were: {1, 43, 16, 25, 21, 23, 6, 5, 36, 2, 27, 33, 37, 7, 46, 45, 44, 12}. A BN was learnt using the ordered moments (see table 2 for results). *Step 4*: From the BN learnt in step 3, we inferred the Markov blanket of the class label and found that only six (6) moments {1, 6, 16, 21, 37 46} were within the Markov blanket of the class label. These six moments were selected to form the OMFS. *Step 5*: In order to determine the classification capability of the selected OMFS we used them to train a BN. Next, 5-folds cross validation's classification accuracy was calculated (see table 2 for results) and it was noted that the OMFS offers quite high classification accuracy.

Stage 3: Clustering of CSLT images

In this stage we pursued the clustering of the CSLT optic nerve images, represented using the 47 moments in the MFS, to differentiate between the different subtypes of healthy and glaucomatous optic nerves. It may be noted that an important theme in glaucoma research is to develop an understanding of the large variation in the appearance of the optic nerve, both within groups of healthy subjects and in patients with glaucoma. It is therefore important, from a clinical standpoint, to recognise and differentiate between such patterns. However, the problem with the sub-classification of patterns of optic nerve damage is that it is a subjective task, giving rise to considerable levels of disagreement between trained experts. In this context, our aim was to develop an objective and automated method to characterize optic nerve images.

Our two phase clustering strategy was to: (a) partition the images into distinct clusters using Self Organizing Maps (SOM); and (b) draw clear and distinct boundaries around the clusters using the Expectation Maximization (EM) algorithm [9].

Phase A: Data clustering using SOM

We used a SOM to cluster the CSLT images based on the similarities between image shapes, where each cluster may represent a different subtype of healthy and glaucomatous optic nerves. The training of the SOM was conducted as follows: (i) we determined the topology of the SOM to be hexagonal lattice comprising 192 units that were arranged as 16 rows and 12 columns; (ii) The units were linearly initialized along the two greatest eigenvectors of the covariance matrix of the training data—i.e. images represented using the 47 moments in the MFS; (iii) The SOM was trained using a sequential training algorithm by first running a rough training phase comprising 100 epochs starting with a large neighbourhood radius of 12 that was linearly reduced to 3 with a learning rate of 0.5. This was followed by a second fine-tuning phase comprising 1000 epochs with a small initial neighbourhood radius 3 that was reduced to 1 with learning rate of 0.1. In both cases a Gaussian neighbourhood function was used and the learning rate function was set to be inversely proportional to the training epochs; (iv) Finally, we achieved a trained SOM that placed similar images into close proximity, thus leading to the image clusters. We applied principal component projection to the learnt SOM to determine its projection. This was followed by developing a U-matrix representation of the learnt SOM by spreading a colour map on the projection. Based on the visualization offered by the SOM, it was noted that the data was partitioned into discernable clusters.

Phase B: Defining the cluster boundaries

After determining broad clusters of CSLT images, in this phase we objectively determine the cluster boundaries. The processing was guided by our assumption that the dis-

tribution of the clusters within the learnt SOM is Gaussian. Therefore, we used the EM algorithm [9] as it is suitable to find distinct components in the case of Gaussian mixtures. Functionally, the EM algorithm initiates with an estimate of the number of components and their parameters. Our strategy was to maximize the likelihood of the optic nerve images into distinct clusters given the parameters and a maximum likelihood measure that indicated how well the Gaussian mixtures fit the data into clusters. We used a Bayesian Information Criterion (BIC) [9], where the best estimate (e.g., number of clusters) was chosen based on the highest BIC value.

To achieve the cluster boundaries, using the EM method with BIC, we initialized the EM using 10 random re-starts method, and then selected a parameter setting to maximize the log-likelihood of our clusters from the SOM. EM clustering was performed for different number of clusters. Table 1 shows that the maximum BIC is achieved when K = 4. Hence, we determined that given the learnt SOM there are 4 clusters—one cluster represents health images and the three clusters for sub-types of glaucomatous images—in it that best fit the data (see Table 1). To finalize the cluster boundaries for the 4 clusters, we calculated the assignment probabilities of each CSLT image to all the cluster labels, the cluster label with the highest probability value was assigned to the CSLT image. Figure 2a shows the SOM with the emergent clusters, the clusters are coded using grey scale for visualization purposes.

Table 1 – Number of clusters vs. BIC values

K	2	3	4	5	6	7	8
BIC	29100	30409	31354	30516	29125	27456	25486

Evaluation and discussion

In this section we present the evaluation results for the various methods developed for stages 2 and 3 of our GDS system.

Evaluation 1: Evaluating CSLT image classification

Table 2 presents the CSCLT image classification accuracy for the different classifiers trained in phase 2, using test images not previously seen by the classifiers. It is interesting to note that both the MLP and the SVM classifiers offered higher accuracy with the MFS as compared to the original 254 moments. This vindicates our hybrid feature sub-set selection strategy, and also confirmed the theoretical assumption that low order moments contain more shape information that is relevant for classification as compared to the information content of high order moments. In the second pass, we determined that the MFS could be further reduced to just 6 moments—i.e. the OMFS—without compromising the classification accuracy. The highest accuracy for MFS was offered by the SVM—i.e. 86.96%. The highest accuracy for the OMFS was 83.82% offered by a BN. Therefore, the compromise in classification accuracy is just 3 %, yet the gain in computational efficacy is quite significant. Note that the BN (with Markov Blanket) offers the most optimal classification results when compared with both MLP and SVM trained on the OMFS. We therefore selected the BN classifier trained with the OMFS to distinguish between healthy and glaucomatous optic nerves.

Table 2 - Classification accuracy for different classifiers

Feature Subset Size	**Classifier**	**Accuracy**
Pass I		
254 moments	MLP	72.88%
254 moments	SVM	77.50%
47 moments in MFS	MLP	74.00%
47 moments in MFS	SVM	86.96%
Pass II		
18 moments (original order)	BN	77.21%
18 moments (chi^2 order)	BN	80.88%
6 moments in OMFS	**BN**	**83.82%**
6 moments in OMFS	SVM	80.26%
6 moments in OMFS	MLP	72.84%

Evaluation 2: Examining the CSLT image clusters

Evaluation of the clustering stage involved mapping a series of optic nerve images for individual patients (i.e. test cases with explanations provided by experts) onto the SOM and noting the Compactness Factor (CF) between the activated units. The CF measures how close the images are with respect to each other in terms of the average distance between the centroid of all active units. The CF is an objective measure for evaluating the clustering goodness based on our initial observation that for a patient the series of optic nerve images are quite similar over a period of time; over time the differences are quite minute and should not lead to large variations between consecutive images. This implies that when visualizing the optic nerve images for a subject, the active units should be in close proximity and therefore yield a low CF.

Figure 2a show that the results for patient 4209643, and it maybe noted that the 7 optic nerve images, taken over a period of time, map on to a single SOM unit resulting in a compactness factor of to 0. The numeral within the active unit shows the number of images mapping on to that unit. This demonstrates the best possible clustering outcome as the learnt SOM recognizes the similarity between all the 'healthy' optic nerve images for this patient. Figures 2b shows the 11 optic nerve images of patient 112455 being mapping on to 4 neighboring SOM units within one cluster, with a compactness factor of 0.20808. This result again implies the close proximity of the images for this patient. These results are in line with the visual observations of these images by experts, who also concurred that the images for these patients are quite similar in shape.

We use the learnt SOM to visualize the disease progression for a patient over a period of time. Images taken over time for a patient were mapped onto the SOM. The pattern of the active units indicated the potential progression of the disease from one cluster to another, where each cluster may represent images of a specific glaucoma sub-type. In Figure 3a the images fall into two adjoining clusters, and the path across the clusters suggests the progression of the

disease from one sub-type to another. Figure 3b shows the progression over time.

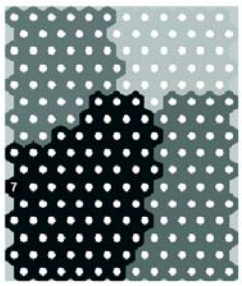


Figure 2a – SOM showing all images mapped to a unit

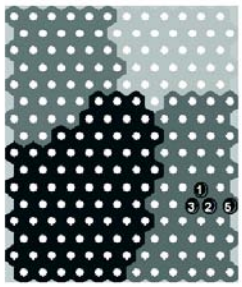


Figure 2b – SOM showing all images mapped in one cluster

Evaluation 3: Visualizing disease progression over time

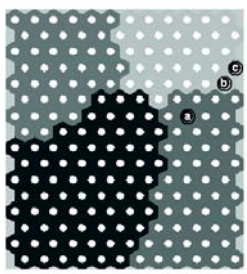


Figure 3a - SOM showing the dispersion of images over two adjoining clusters

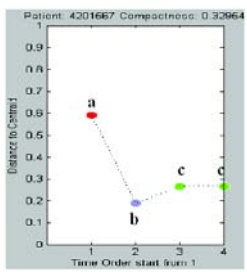


Figure 3b – The disease progression path. Note the high CF between the images

Figure 4a - SOM showing a noisy image that is distant from the other images.

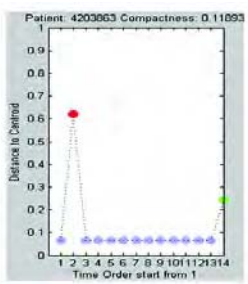


Figure 4b – The spike shows that the 2nd image is noisy, as it does not follow the pattern.

Evaluation 4: Identifying noisy CSLT images

We used the learnt SOM to identify noisy CSLT images that typically occur due to various factors related to the capture of the optic nerve image. With the knowledge that consecutive images do not manifest drastic changes, if an image is noted to be significantly dissimilar from its neighbors it can be regarded as a noisy image. At present there are no objective means to identify noisy CSLT images. Figure 4 (a-b) shows 14 images for a patient, where the 2nd image is identified as a single noisy image because it is in a different cluster, whereas the remaining images all map onto just two other units that are very close to each other.

Concluding remarks

We presented a data mining framework to objectively analyze medical images, and applied it to investigate glaucoma. The novel features of our approach are that: (a) we process CSLT images to derive shape information by using image processing techniques. This is in contrast to the traditional approach of using morphological features to analyze CSLT images; (b) we have developed a feature selection strategy that identifies the most salient image-defining features without compromising the classification accuracy; and (c) we are able to visualize the CSLT images in terms of clusters of similar images. These clusters provide an opportunity to visualize the dispersion of multiple observations for a subject, and we show how this information can help to (i) determine a potential progression of the disease due to changes in the optic disc over time; and (ii) identify noisy CSLT images. We believe that our framework takes a step towards the automated and objective analysis of optic nerve images to support glaucoma diagnostics.

Acknowledgments

This project is supported by a research grant from Nova Scotia Health Research Foundation, Canada.

References

[1] Hoskins H, Kass M. Primary open-angle glaucoma, Becker-Shaffer's Diagnosis and Therapy of the Glaucomas. The C.V. Mosby Company, St.Louis, 1989: 277-307.

[2] Zinser G, Wijnaendts-van-Resand RW, Dreher AW. Confocal laser tomographic scanning of the eye. Proc. SPIE 1161, 1980: 337–344,

[3] Bowd C, Cban K, Zangwill LM, Goldbaum MH, Lee T, Sejnowski TJ. Comparing neural networks and linear discriminant functions for glaucoma detection using confocal scanning laser opthalmoscopy of the optic disc. Investigative Opthalmology & Visual Science, Vol. 43 (11), 2002.

[4] Park J, Reed J, Zhou Q. Active feature selection in optic disc nerve data using support vector machine. IEEE World Congress on Computational Intelligence, 2002.

[5] Swindale NV, Stjepanovic G, Cbin A, Mikelberg F. Automated analysis of normal and glaucomatous optic nerve head topography images. Investigative Opthalmology & Visual Science, Vol. 41 (7), 2000.

[6] Broadway DC, Nicolela MT, Drance SM. Optic disc morphology on presentation of chronic glaucoma. *Eye* 17 (6), 2003: 798-799.

[7] The CH, Chin RT. On Image Analysis by the methods of moments. IEEE Trans. Pattern Analysis by Machine Intelligence 10(4), July 1998: 96-513.

[8] Hruschka Jr. ER, Hruschka ER, Ebecken NEF. Feature selection by Bayesian network. Canadian AI Conference, LNAI 3060, Springer Verlag, 2004:370-279.

[9] McLachlan G, Krishnan T. The EM algorithm and extensions. New York: John Wiley & Sons, 1997.

Address for correspondence

Syed Sibte Raza Abidi,
Faculty of Computer Science,
Dalhousie University,
Halifax, B3H 1W5, Canada.
Email: sraza@cs.dal.ca

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Intelligent Querying and Exploration of Multiple Time-Oriented Medical Records

Denis Klimov[a], Yuval Shahar[a]

[a] *Medical Informatics Research Center, Ben Gurion University of the Negev, Israel*

Abstract

Querying and analyzing multiple time-oriented patient data is a key task during medical research, clinical trials or the assessment of the quality of therapy. In this paper, we present several aspects of the VISITORS system, which includes knowledge-based tools for graphical querying and exploration of multiple longitudinal patient records. We focus on the syntax and semantics of the knowledge-based aggregation query language for multiple time-oriented patient records, and on the graphical query-construction interface. The query language assumes an underlying computational method for deriving meaningful abstractions from single and multiple patient records, such as we had previously developed. The aggregation query language enables population querying using an expressive set of constraints. By using our underlying temporal mediator architecture, the time needed to answer typical temporal-abstraction aggregation queries on databases of 1000 to 10000 patients was reasonable.

Keywords:

intelligent visualization, temporal abstraction, multiple patients, medical informatics, Human-Computer Interfaces

Introduction: Knowledge-based exploration of multiple time-oriented records

A key task facing clinicians and medical researchers is the analysis of time-stamped, longitudinal medical records, in particular records of multiple patients. This capability is necessary to support, for example, quality assessment tasks, analysis of clinical trials, or the discovery of new clinical knowledge. Although the task of assessing patient data has been mostly solved through the increasing use of Electronic Medical Record (EMR) systems, there still remains the task of intelligent processing of multiple time-oriented patient records, including the capability for interactive exploration of the results. Standard means, such as tables, temporal statistical tools, or more advanced temporal data mining techniques, are often insufficient or can help only in particular cases.

To solve the computational aspect of this problem, we have been using the *knowledge-based temporal abstraction* (KBTA) method [1] for automated derivation of meaningful context-specific interpretations and conclusions, called *temporal abstractions*, from raw time-oriented patient data, using a domain-specific knowledge-base (KB). In general, the KBTA method is defined as follows: The input includes a set of time-stamped parameters (e.g., platelet, red blood-cell (RBC), and white blood-cell (WBC) counts) and events (e.g., bone-marrow transplantation (BMT)), which create the necessary interpretation context (e.g., the therapy protocol used). The output includes a set of interval-based, context-specific parameters at the same or a higher level of abstraction and their respective values (e.g., a period of nearly 3 months of grade 0 bone-marrow toxicity in the context of that therapy protocol).

Furthermore, the output temporal abstractions can be efficiently visualized. The KNAVE-II system, which we developed previously [2], supports the visualization and exploration of raw data and derived temporal abstractions for an individual patient or small number of patients. Evaluation of the KNAVE-II system in the oncology domain [3] has demonstrated that, by using KNAVE-II and its underlying temporal abstraction computational architecture, physicians can more quickly and more accurately answer clinical queries about patients.

However, to analyze clinical trials, or for quality assessment purposes, an aggregated view of a group of patients is more effective than exploration of each individual record separately. In addition, certain patterns can only be discovered through the analysis of multiple patients. Therefore, we have extended the KNAVE-II system into a system called VISualizatIon of Time-Oriented RecordS (VISITORS) [4] which supports the visualization of a group of time-oriented records at different levels of abstraction.

The following three important features distinguish the VISITORS framework from other exploring data tools :

1. Time-oriented data are graphically displayed and explored for both individual and multiple patients.
2. The temporal dimension is a first class citizen. It can be explored in various granularities, such as hour, day, and month. We also support a calendar timeline and a time-line relative to special events (e.g., the six months following a particular intervention).
3. The computational reasoning supports not only a view of raw time-oriented data and their statistics but also a meaningful summarization of the raw data, based on the temporal-abstraction domain ontology and the KBTA computational mechanisms. The exploration interface is also based on that ontology, which supports a semantic exploration of the data and enables navigation of semantically related raw and abstract concepts. For example, the user can explore graphically an instance of a pattern that was derived by the KBTA pattern-detection inference mechanism and view all the

abstract components and raw data from which the pattern was derived.

In this paper, we explain in detail the syntax and semantics of the VISITORS query model, which directly affect the semantics of the computational and display modules. We also introduce several of the graphic modules which we have implemented to assist users in the interactive definition of temporal aggregated queries and exploration of multiple time-oriented records. Finally, we provide results of a preliminary functional evaluation and discuss the implications of the VISITORS framework.

Methods

General architecture

The VISITORS system is an intelligent interface to a distributed architecture specific to the tasks of querying, knowledge-based visualization, and interactive exploration of time-oriented data. We assume that the necessary elements of the temporal abstraction framework (shown in Figure 1 by striped lines) are available. Figure 1 describes the overall architecture: End users (clinicians) interact with *Query Builder* of VISITORS to submit time-oriented queries regarding patients. The Temporal Abstraction Mediator, for example our previously described systems, the goal-directed IDAN Mediator [5], integrates the relevant data and knowledge from the appropriate sources, indicated by the user, to answer queries regarding raw data or to derive using a Temporal Abstraction Service the abstract time-oriented concepts. The resultant data can be visualized and explored by the user.

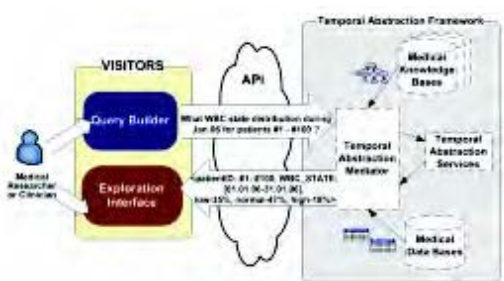


Figure 1- The distributed VISITORS architecture

Semantics of the temporal aggregation queries

We assume that the clinical time-oriented databases (DBs) have the following virtual patient record structure: the patient identification data (e.g., patient ID), name of entity (e.g., measured parameter, medications, interventions, etc.), time of laboratory test or medication and value. For example:

<John Smith,WBC,11.01.06 13:45:00,6.7*103 cells/ml>,

In our work we have designed and developed a formal query language, which is ontology-based, i.e., we use domain knowledge to formulate and display the queries. We distinguish three types of aggregated queries.

Select patients query

This query retrieves a list of patients from selected database who satisfy a set of constraints, defined by user:

GetPatients (KB, DB, <PatientsConstraints>) $\Rightarrow$ *<patients>*,

where *GetPatients* is an external procedure that queries the selected *DB*, *KB* is an appropriate knowledge base that includes the parameters definition used in *<PatientConstraints>* - a set of complex conditions defining criteria for patient selection, and *<patients>* is a resultant list of identification data of patients who satisfy the set of *<PatientConstraints>*.

The *<PatientConstraints>* aspect of the query includes the list of Boolean and temporal conditions of three types:

- **Demographic** or **non-temporal constraints** (i.e., only the last value is relevant) such as patient's ID, age, sex, physician, etc. The user can define Boolean constraints among the attribute's values (i.e., OR/AND logical operators). However, we do not recommend using complex Boolean expressions, since they make understanding why a patient was included in the output less intuitive. The logical operator NOT was omitted for a similar reason.
- **Time and value constraints** for both raw and derived concepts. To construct the query, the user can define constraints on the value of a concept, its duration, and its start/end points (See Figure 2). Both absolute (i.e., calendaric) and relative (i.e., measured from a reference event) timelines are supported. Defining pair-wise temporal relations between interrelated concepts is supported using Allen's temporal logic relation as well as relations among specific time points (i.e., start/end).

  We use the conjunction logical operator between the time and value constraints. Note that the computational functions and procedures that enabled us to derive the abstract parameters are defined in the knowledge base as part of the domain ontology. The name of the KB is specified in the query.

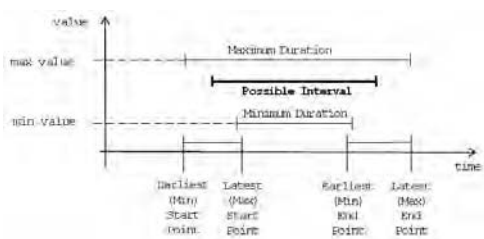


Figure 2-Time and value constraints for concept

- **Statistical Constraints** enable the user to aggregate and filter the patient's data on the basis of a specific statistical function. Using such constraints, the user is able to investigate who are the patients in the database who have specific values (or a range of values) within a given statistical range of threshold values. For example: "Select all patients whose state of WBC was derived as "low" or "very low" during more than 25% of the period Jun 1 – Oct 31 2006".

Figures 3 & 4 show an example of constructing a *Select Patients Query* with both demographic and knowledge-based constraints, whose informal definition is "Select all male patients, either younger than 20 or older than 70, whose hemoglobin (HGB) state was abstracted at least as

"moderate low" or higher, during at least seven days, starting at least two weeks after the allogenic bone marrow transplantation(BMT),and whose WBC counts were increasing during the same period." That is, either young or old patients whose bone marrows have been recovering following the BMT procedure.

The bottom part of the interfaces shown in Figures 3 and 4 displays the query that is automatically and incrementally being created from the user's graphical specification (displayed in the top part). The graphical interface used for query construction has a similar structure to the main exploration interface of the VISITORS system, which we discuss in the next section (e.g., the ontology browser is displayed on the left side; the panel display used for query definition is similar to the one used for data display, etc.). The highlighted rectangular area denotes the ranges of the time and value constraints on the patient data or abstractions relevant to the query.

Select time intervals query

Given a set of time-oriented patient data, this query returns a list of time intervals that satisfy the constraints defined by the user. In other words, the goal of this query is to find *when* a certain portion of the patients has a specific value or value in a predefined value range.

Formally, the Select Time Interval Query has the following structure:

```
GetTemporalIntervals (KB,DB,
<Constraints>, [refPoint]) ?
     relation*(<start_time, end_time>),
```

where *GetTemporalIntervals* is an external procedure that queries the selected *DB* and *KB* is an appropriate knowledge base that includes the parameters definition used in *<Constraints>*. The output *relation*(<start_time, end_time>)* is a data structure that includes a list of temporal intervals distinguished by *start_time* and *end_time* time points. The optional parameter *refPoint* requires that the output be calculated from the denoted reference point (e.g., a key clinical procedure).

<Constraints> is a logical expression without nested parentheses of the following atomic constraints related by conjunction logical operator:

```
<Concept,    min_thresh,    max_thresh,
min_value, max_value>,
```

where *Concept* is a concept name, the *min_thresh* and *max_thresh* are minimal and maximum thresholds values in percents of patients, and *min_value* and *max_value* is a range of values of the selected *Concept*.

Figure 3-Definition of a query's demographic constraints

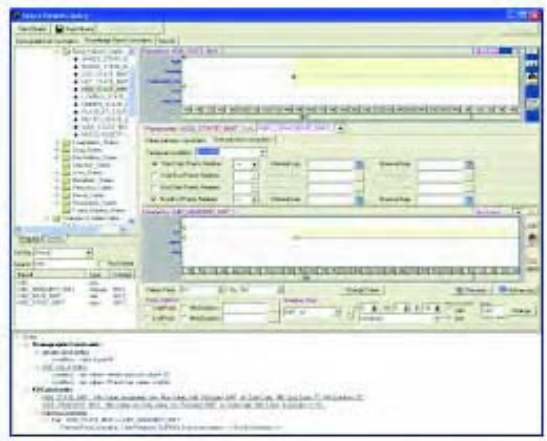


Figure 4-Definition of a query's knowledge-based constraints

For example, a typical *Select Time Interval Query* is: "Find [relative] time intervals following the BMT, during which the WBC count was increasing, and the state of the Platelet count was "normal" or higher, for more than 30% of the patients." Its formal XML based definition would be:

```
<SelectTimeIntervalsQuery>

  <Concept name="Platelet state"
min_value="normal"

           max_value="very high"

           min_theshold="30"
max_theshold="100"/>

  <Concept name="WBC
gradient"min_value="increasing"

           max_value="increasing"

           min_theshold="30"
max_theshold="100"/>

  <ReferencePoint name="BMT"
event="last"/>

</SelectTimeIntervalsQuery>
```

Get patients data query

Given a list of patients ID's and, optionally, a list of time intervals, the query retrieves the patients' raw data, or the derived temporal abstractions of one selected concept within the selected time intervals for the selected patients. The default patient list is all of the patients in the DB, and by default there are no time-interval constraints. The formal expression of the query is the following:

```
GetDataConcept(KB,DB,Concept,<patients>,[<t
ime intervals>])?
       relation*(<patient_n,
start_time_n,m, end_time_n,m,value_n,m>),
             1  n  N, 1  m  M,
```

where *GetDataConcept* is an external procedure that queries the selected DB, *KB* is an appropriate knowledge base that includes the *Concept* definition, *<patients>* is a list of patients identification data, and *<time intervals>* is an optional parameter that constrains the time of the returned

data. The output *relation*(<patient$_n$, start_time$_{n,m}$, end_time$_{n,m}$, value$_{n,m}$>)* is a data structure that includes the time-oriented data of *N* patients in <*patients*> list with *M* data records for each patient.

In contrast to the previous two queries, the *Get Patients Data Query* does not have a separate interface. This query is constructed automatically during the users' exploration of the patient's data in the main interface of VISITORS.

The main interface in VISITORS

We described previously several visualization tasks in VISITORS [4]; here we intend to explain only the main design principles of the interface.

The main interface in the VISITORS system is divided into three logical parts (See Figure 5):

The top panel (A) is used for the patients selection tasks. The user can select previously retrieved groups from the table, choose patients from a DB, input the patient ID by himself, or construct a new *Select Patients Query.*

The middle small panel (B) is used for time interval selection. The user can use the previously returned intervals, define a new interval, or construct a new *Select Time Intervals Query.* Note that both calendaric and relative timelines are supported. The main part of the interface (C) is used to explore the patients' time-oriented data. The left side includes a browser to the clinical domain ontology, retrieved on the fly from the relevant domain KB. Clicking on a concept node in the ontology tree displays the data of that concept for the selected group of patients. In this case, the user explores the data of a group of patients named *My patients* (58 patients), previously retrieved by the *Select Patients* query. The 1st panel from the top displays all of the WBC laboratory test values during March 95. The top (red) line represents the daily maximal value of the WBC count. The 2nd panel shows the daily mean value of HGB for each patient during 1995. The top (red) line and bottom (blue) line represent, respectively, the monthly maximal and minimal patient HGB values within the group. The 3rd panel displays the distribution of the WBC state-abstraction values for each month of 1995. For example, in Aug 95, 29 % of patients in the group have had a *moderately_low* value. The bottom panel displays statistical and temporal associations within the specific time period among the selected raw-data or abstracted concepts. In this case, WBC and RBC values are displayed over March 1995, and the Platelet and HGB values over 1995. Only 25 patients in the group have data in the selected time intervals. Values of all parameters for each patient are connected by lines.

The functional evaluation

We are currently performing a functional and usability evaluation of the *Query Builder* module in an oncology domain. Evaluators were asked to construct eight *Select Patients* and two *Select Time Intervals* queries at different difficulty levels (Examples of queries are showed in Table 1).

Figure 5-The main VISITORS interface

Table 1 – Examples of aggregated queries

Complexity	Examples of Queries
Select Patients Query	
Easy	Find male patients older 50 who have had during Sep 95 WBC laboratory test value less than *4000 cells/ml*.
Moderate	Find all patients who had the HGB state value in the value range "*low*" to "*normal,*", within the following time constraints: the episode of HGB starts between 3 to 4 days after allogenic BMT and its duration is at least 20 hours but no more than 60 hours.
Hard	Select male patients, whose bone-marrow toxicity grades have been decreasing for at least seven days during the period starting from 2 weeks after the allogenic BMT procedure, and whose liver toxicity grades were also decreasing during the same period.
Select Time Intervals Query	
Easy	Find time intervals when the state of WBC was considered as "normal" or higher for less than 30% of the patients
Hard	Find time intervals following the BMT procedure during which more than 20% of the patients had "*moderate*" anemia, and more than 30% of the patients had "*low*" or "*very low*" WBC counts

In all tests we used a retrospective DB of more than 1000 oncology patients after bone-marrow transplantations (BMT), who were followed for two to four years. In addition, we assessed the performance of the overall architecture when answering these queries. Results show that performance times using IDAN architecture were reasonable: several seconds for answering aggregated queries with demographic and time and value constraints for raw and abstracted parameters. We intend also to evaluate the system in the diabetic domain.

Discussion

In this paper we have presented the novel ontology-based multiple records query language, which enables users to construct a new query and retrieve a set of relevant patients or time intervals, using a broad set of constraints. Such aggregated queries, and the graphical *Query Builder* tool used to construct it, are part of the VISITORS system that enables clinicians to query, visualize and explore both raw time-oriented medical data and meaningful interpretations (including complex temporal patterns), derived from the these data, based on a domain knowledge base. The main advantage of our system is successful integration of different methods, such as information visualization and knowledge-based temporal reasoning.

In previous studies the time-oriented aspect of querying was commonly addressed by adding a temporally-oriented extension to standard SQL tools [6, 7]. However, these systems are limited in the temporal analysis by SQL capabilities. Moreover, they do not include the temporal reasoning mechanisms. In the area of visual querying, several tools have been proposed. The TimeFinder system [8] is a visual exploration and query system for exploring time-series data sets, based on a direct manipulation metaphor. Chittaro and Combi [9] provide visual representations framework of temporal intervals and relations. However, these techniques focus on query and exploration only of raw longitudinal data. Attempts to support intelligent query and retrieval are provided in the TimeLine system [10], and in knowledge-based spatial temporal query language (KSTL) [11] which support, however, only the medical imaging domain. Thus, we aimed to develop a domain-independent visual query and exploration system that enables clinicians to explore multiple longitudinal patient records, in an intelligent manner that supplies querying both raw data and their *knowledge-based meaningful interpretations*.

Acknowledgments

This research was supported by Deutsche Telekom Company and Israel Ministry of Defense award no.89357628-01

References

[1] Shahar Y. A framework for knowledge-based temporal abstraction. *Artificial Intelligence*, 90(1-2), 1997.

[2] Shahar Y, Goren-Bar D, Boaz D, Tahan G. Distributed, intelligent, interactive visualization and exploration of time-oriented clinical data and their abstractions. *Artificial Intelligence in Medicine* 38 (2), 2006, pp. 115-35.

[3] Martins SB, Shahar Y. Evaluation of KNAVE-II: A tool for intelligent query and exploration of patient data. *Proc. of Medinfo 2004*.

[4] Klimov D, Shahar Y.A framework for intelligent visualization of multiple time-oriented medical records. *AMIA Annu Symp Proc. 2005*, pp. 405-9.

[5] Boaz D, Shahar Y. A distributed temporal-abstraction mediation architecture for medical databases. *Artificial Intelligence in Medicine* 34 (1), 2005, pp. 3-24.

[6] Nigrin D, Kohane IS. Temporal expressiveness in querying a time-stamp–based clinical database. *J Am Med Informatics Assoc*, vol. 7, 2000, pp.152-63.

[7] Combi C, Cucchi G. GCH-OSQL: a temporally-oriented object-oriented query language based on a three-valued logic. *Proc of the 4th International Workshop on Temporal Representation and Reasoning*, 1997

[8] Hochheiser H, Shneiderman B. Visual Specification of Queries for Finding Patterns in Time-Series Data. *Proc. of Discovery Science 2001*, 2001, pp. 441-46.

[9] Chittaro L, Combi C. Visualizing queries on databases of temporal histories: new metaphors and their evaluation. *Data and Knowledge Engineering*, 44, 2003, pp. 239-264.

[10] Aoyama D, et al. TimeLine and visualization of multiple-data sets and the visualization querying challenge. *Journal of Visual Languages & Computing* 18 (1), 2007, pp.1-21.

[11] Chu W, Alfonso F, and Ricky K. Knowledge-based image retrieval with spatial and temporal constructs. *IEEE Transactions on Knowledge and Data Engineering* 10(6).

Address for correspondence

Mr. Denis Klimov, Tel: 972-8-6477160, email: klimov@bgu.ac.il
Department of Information Systems Engineering, Ben-Gurion University of the Negev, P.O.Box 653, Beer Sheva, 84105, Israel.

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Analyzing Web Log Files of the Health On the Net HONmedia Search Engine to Define Typical Image Search Tasks for Image Retrieval Evaluation

Henning Müller [a], Célia Boyer [b], Arnaud Gaudinat [b], William Hersh [c], Antoine Geissbuhler [a]

[a] *Medical Informatics Service, University and Hospitals of Geneva, Geneva, Switzerland*
[b] *Health On the Net (HON), Geneva, Switzerland*
[c] *Oregon Health and Science University (OHSU), Portland, OR, USA*

Abstract

Medical institutions produce ever-increasing amount of diverse information. The digital form makes these data available for the use on more than a single patient. Images are no exception to this. However, less is known about how medical professionals search for visual medical information and how they want to use it outside of the context of a single patient. This article analyzes ten months of usage log files of the Health on the Net (HON) medical media search engine. Key words were extracted from all queries and the most frequent terms and subjects were identified. The dataset required much pre-treatment. Problems included national character sets, spelling errors and the use of terms in several languages.

The results show that media search, particularly for images, was frequently used. The most common queries were for general concepts (e.g., heart, lung). To define realistic information needs for the ImageCLEFmed challenge evaluation (Cross Language Evaluation Forum medical image retrieval), we used frequent queries that were still specific enough to at least cover two of the three axes on modality, anatomic region, and pathology. Several research groups evaluated their image retrieval algorithms based on these defined topics.

Keywords:

log files analysis, image retrieval evaluation.

Introduction

An increasing amount of medical information is being produced digitally, making it available for further processing and use, i.e., for teaching and research. Much of the produced data and experiences from past cases can be used to create tools for diagnostic decision aid. A great deal of medical information is also available on the Internet, as there are increasing requests for medical information by patients and professionals [1]. MedLinePlus is one example of a repository created to inform non-professionals, patients searching for information. Another example is Health On the Net (HON), which develops quality criteria for medical web pages and has an accreditation service for pages adhering to several quality criteria. HON also runs web search engines for medical web content aimed at patients and medical professionals with a multilingual search interface[1] [2]. Much research has been done on the searching of medical texts [3] but less on how images are used and searched for, although the amount of image data being produced is rising [4]. Many medical image databases are available within institutions, mainly for teaching, but some are also made available on the Internet. These include Casimage, HEAL (Health Education Assets Library), MedPix, and the Pathopic datasets. MIRC[2] (Medical Image Resource Center) is an initiative of the Radiological Society of North America (RSNA) to unite teaching files under a single interface. These databases contain thousands of annotated images. Unfortunately, the images are only rarely indexed in search engines such as Google as they are usually only available through the search in database fields. Another problem is that the annotation is often incomplete and information on the image modality is not always given. A search for "lung CT" with Google image search in October 2005 brought 160 results, about half of them lung CTs. The abovementioned databases alone contain several thousand lung CTs.

Outside of medicine, visual information retrieval has been an extremely active research domain for more than 15 years [5]. Studies on domain-specific user requirements have been performed, for example for journalists searching for images [6] or in the cultural heritage domain [7]. In the medical field, visual information retrieval has been proposed many times as extremely useful [8, 9]. Still, most research has a limited focus on retrieval for one particular group of images [9]. Although this might be a domain with high potential impact, teaching and research are more likely to profit first from possibilities to browse very large and diverse image collections by visual properties. In the context of ImageCLEFmed [10], a challenge evaluation for medical image retrieval, two surveys were performed among medical image users [11, 12] to find out more about typical information needs and search tasks. CLEF (Cross-Language Evaluation Forum) is a challenge evaluation for retrieval of multilingual information. ImageCLEFmed in particular evaluates the quality of retrieval from multilingual medical image retrieval available on the Internet. The

1 http://www.wrapin.org/ & http://www.hon.ch/HONselect/
2 http://mirc.rsna.org/

Selected for best paper award.

surveys include five user groups: medical professionals for diagnosis, teaching, and research as well as medical students and librarians. The goal of the work descried in this paper was to create realistic search tasks for ImageCLEFmed[3] based on information needs of web users. The analysis resulted in 30 search tasks used by participating research groups. Among the techniques used was analysis of log files, an active research domain [13], mainly to analyze web page design.

Materials and methods

Used data sets

The data used for this study were log files containing query terms of the HONmedia[4] search engine. The examined period of queries included ten months, from January 1, 2005 to October 31, 2005. This period was sufficient for a representative evaluation of search terms. Variations of search frequency or quality over the months were not part of our analysis. The original data set contained 53'970 queries. With each automatically extracted query term, the date and time of the query was stored. It was also stored whether the query was directly done via the HONmedia interface or referred to from Google towards HONmedia search. Many queries were in French, as the French-speaking medical community frequently uses the HON query engine. It was not possible to perform an automatic translation of the topics, as language detection is hard with only very few words. Other languages identified for the queries were English, German, Spanish, and Italian.

Pre-treatment of the data and evaluation techniques

The analysis of the data was done on a Linux computer using Perl to analyze the text files. The original data sets were transferred to pure text and the information on time and date of the query were discarded. Perl was used mainly to pre-treat the data. As data were extracted automatically and as robots perform queries on web interfaces there are many different formats for queries (sometimes broken), plus a variety of international character containing umlauts and accented characters sets that need to be combined.

Results

The data contained two groups of queries, queries directly asked via HON and queries forwarded via Google. These groups were treated separately. A total of 37'293 queries were directly performed via HONmedia and 16'677 were forwarded via Google.

Text normalization

First, normalization was necessary for the text to remove differences in coding of the strings, parameter options transmitted and for broken queries containing graphical symbols. We did not treat the word order in the queries. The steps were mainly based on a manual analysis of the data:

- Unify coding issues, to remove accents, Umlauts, national symbols, and any sort of non-text: –"()+–.
- Remove commands and options send by web robots or search engines.
- Remove URLs or fragments of URLs.
- Convert all characters to lower case.
- Change plural of frequent terms such as "images".
- Remove frequent terms to define the target media: image(s) (5'796), media (512), video(s) (334).

Over 100 rules for normalization and removal were defined and applied to clean the data. Even after the removal steps, it was apparent that an extremely large number of different queries remained. In total, there remained 5'365 different unique queries (of 16'677) for the Google queries and 17'643 different HON queries (of 37'293). This meant that almost half the queries were unique being asked only once, which made a systematic evaluation of the entire dataset hard. The number of words per query was small. Google queries contained an average of 2.01 words in our study and HON queries 1.50 words, after removing the words image, video and media. This resulted in 191 empty queries for Google and 150 for HON. The same number of queries contained only a single character.

Removal of unclear queries

After term normalization, it became clear that there are queries unimportant for further analysis. First, a group of queries concerned sexually explicit queries: In the Google queries, the following frequent terms were removed: xxx (334 times). For HON the following terms were removed: penis (114), vagina (108), breast (102), sex (65), clitoris (32), gynecology (24). Another group of queries implicitly contained similar ideas; for Google these were: accouchement (childbirth, 143), cesarienne (33). For HON: home childbirth (239), nurse (130), birth (69). Third, another group of queries were processed to remove those not containing a precise information need, some of them, such as the term "search," were simply placed by web robots trying to access information stored in web-accessible databases. For Google this included the following terms: medical images (508 times), HON (116), health (62), medical illustrations (32), repository (30). For HON, these terms included: search (1493), medical images (79), doctor (70), anatomy (65).

Most frequent queries and terms

After normalization and removal of queries, we analyzed the most frequent remaining terms. Table 1 shows the most frequent remaining terms forwarded from Google. This list contains very specific medical search requests, from specialists rather than patients. Most of the terms are in French, actually all of the most frequent 20. The specialized nature of the terms and the fact that they are in French can be explained with the fact that only these technical queries link towards HONmedia.

3 http://ir.ohsu.edu/image/

4 http://www.hon.ch/cgi-bin/HONmedia/

Table 1 - Most frequent terms forwarded from Google

Term	*Frequency*
Nerf sciatique	154
Kyste pilonidal	76
Leucemie aigue myeloblastique	72
Glossite exfoliatrice marginee	67
Fracture humerus	66
Grenouillette sublinguale	60
Hematome sous dural	57
Polype nez	56
Appendice xiphoide	53
Leucomalacie periventriculaire	51
Leucemie	46
Purpura rhumatoide	46
Scarlatine	44
Hematome retroplacentaire	40
Kyste thyreoglosse	39
Leucemie myelomonocytaire chronique	39
Leucoplasie	38
Apophyse odontoide	37
Hidradenite	37
Scoliose	34

Table 2 shows the most frequent terms directly queried with HONmedia. These terms are more likely to be from patients than specialists. The first 20 contain only a single word. More terms are in English than in French, actually all top 20, whereas a large number of the less frequent terms are in French. Most terms are of two groups: Terms describing an anatomic region or a disease. Only other terms found in the most frequent 20 are concerning symptoms or a treatment in the largest sense, such as *injection, bacteria and pain.*

Table 2 - Most frequent terms from the HONmedia search

Term	*Frequency*
Heart	381
Asthma	242
Brain	211
Diabetes	160
Liver	101
Cancer	98
Marfan	93
Kidney	77
Lung	69
Knee	69
Injection	67
Bacteria	64
Eye	60
Foot	58
Pain	58
Ear	58
Pancreas	57
Aids	57
Blood	55
HIV	54

Classified term occurrences important for us

This section analyzes only queries directly from HON as they correspond better to our needs concerning patient information search. We particularly note the most frequent terms for *anatomic region, pathology, imaging modality, symptom* and *treatment*, as these are axes to model search tasks along.

Table 3 - Frequent terms regarding modality

Term	*Frequency*
Ultrasound	47
Ecg/ekg	34/32
MRI	33
X-ray	21
Endoscopy	18

Table 3 shows modalities searched for. Interestingly, a commonly used modality (CT) is not mentioned often, whereas ECG, often discarded in medical image databases, is frequently used as it corresponds to the information needs.

Table 4 - Frequent terms regarding symptoms

Term	*Frequency*
Bacteria	64
Pain	58
Burns	42
Stress	37
Blood pressure	30

Table 4 shows symptoms searched for, where symptom is taken in a broad sense. Bacteria is not a symptom but might be interpreted from patients with flu-like symptoms looking for more information on a particular situation.

Table 5 - Frequent terms regarding treatments

Term	*Frequency*
Injection	67
Surgery	46
Stethoscope	36
Anesthesia	24
Vaccination	22

Table 5 lists terms concerning treatments, taken in a wide sense, as stethoscope is not a treatment.

Table 6 - Frequent terms regarding anatomic region

Term	*Frequency*
Heart	381
Brain	211
Liver	101
Kidney	77
Lung	69

In Table 6, frequent anatomic regions are listed that correspond well to the most frequent causes of death [14]. Also the search terms regarding pathology correspond well to diseases mentioned in [14]. Only Marfan is surprisingly frequent.

The 500 most frequent terms were analyzed accounting for almost half the search terms in total. Besides the identified five axes, some other terms are frequently queried, which are hard to classify: Human body (41), smoking (38), CPR (computerized patient record, 33), cardiology (26). It is hard to know what images or videos the users were searching for.

Table 7 - Frequent terms regarding pathology

Term	*Frequency*
Asthma	242
Diabetes	160
Cancer	98
Marfan	93
Aids/HIV	57/54

Constraints to define search tasks based on the results

From the most frequent concepts and the average number of query terms it becomes clear that users express fuzzy information needs and describe them with few terms. As the information in the HON queries corresponded better to our goal, we only used these. It is clear that information needs are often broad and it seems to aim at general illustrations (CPR, human body, AIDS …) than towards precise images of a particular modality and anatomic region. Illustrations also need to be taken into account as

frequent query words such as doctor, nurse, injection or bacteria show.

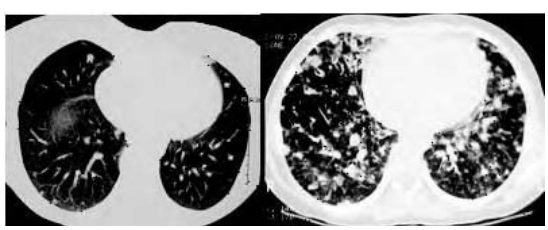
Table 8 - A collection of longer queries

Term	*Frequency*
Autonomic nervous system	16
Heart conduction system	10
Artrite reumatoide juvenil	9
Lupus vasculitis central nervous system	8
Fetal alcohol syndrome	7
Sickle cell anemia	7
Epilepsy frontal lobe	6
Respiratory distress syndrome adult	6
Spinal cord compression	6
Shoulder impingement syndrome	6

Other queries contained expected concepts but not as detailed as desired. If looking for images of the heart, all modalities, views and pathologies combined produce an extremely large number of images to be found. Such tasks are not suited to find out more about the quality of a retrieval system. For this reason, we evaluated the most frequent queries with at least three words. Table 8 lists these frequent search terms. The table shows that several terms still contain a single concept (autonomic nervous system). Most queries contain two distinct concepts, either pathology and anatomic information (epilepsy frontal lobe) or a disease and a patient group (respiratory distress syndrome adult). Still, few of these queries can be taken as query tasks for a benchmark directly.

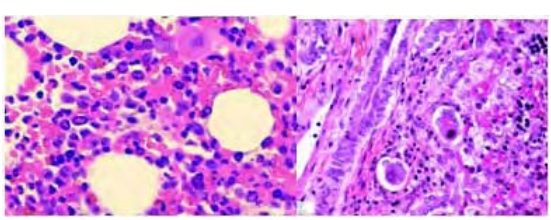
Example Query tasks of ImageCLEFmed

Finally, it was decided to use 30 real but rarer queries of the log files that cover at least two of the axes modality, anatomic region, and pathology. Example topics with example query images can be seen in Figures 1 and 2.



Show me chest CT images with nodules.
Zeige mir CT Bilder der Lunge mit Knötchen.
Montre-moi des CTs du thorax avec nodules.

Figure 1 - A visual query of ImageCLEFmed 2006



Show me microscopic images showing parvovirus infection.
Zeige mir Mikroskopien mit einer Parvovirusinfektion.
Montre-moi des images microscopiques qui montrent une infection parvovirale

Figure 2 - A semantic query for ImageCLEFmed 2006

The 30 query topics generated in this way were sent to all 40 participating research groups together with an image database. After retrieval experiments by participating groups and pooling of results, a group of physicians performed relevance judgments to compare the retrieval results of the participating retrieval systems. More about the results can be read in [15].

Discussion and conclusions

The normalization of query terms that we applied is not completely sufficient for a system that is used in several languages. A translation of the terms towards a single language or terminology would be best but with most queries being single words, this is difficult. At least 10 languages were identified. Spelling errors and abbreviations were other problems. Part of this was corrected with manual analysis but a large number of queries for the same terms could not be combined.

It can be seen that many queries for visual medical content are being performed with HONmedia search. About 52'000 queries in ten months is a large number for a small specialized search engine. Some queries are not for medical content but erotic, which is a phenomenon known by all search engines, particularly searches for images. Many queries are for illustrations of broad concepts, where the users seem to be willing to browse through a large number of varying results without a clear idea in mind and rather to illustrate an article or a presentation. Most queries are for a particular anatomic region or a certain disease. Users of the search engine do not seem to be used to supplying precise information needs concerning images. They follow the behavior of textual Internet search using broad concepts. Most image databases on the web are not well annotated and much of the information is incomplete resulting possibly in poor results.

Compared to text analysis and retrieval, medical visual information retrieval is still in its infancy. Currently, large data sets are being created and made available. Still, the applied search methods are mostly based on text, only. Techniques for visual retrieval do exist [9] and if we want to apply them in real clinical settings we need to build prototypes and make users familiar with the techniques, the possibilities and the limitations. In this sense, ImageCLEFmed is an important initiative for bringing image retrieval systems closer to routine use, through evaluating their quality. To do so, the common image databases need to be shared and realistic visual information needs have to be defined. For this, resources such as the HONmedia log files are important for us as only few medical visual search engines exist in routine use. It is also important to educate users to define their information needs more precisely using text as well as visual means and also relevance feedback.

An interesting future research topic is the analysis of query terms over short time frames. How does this behavior change with respect to events in the world (such as the bird flu)? Could the beginning of a flu outbreak be detected through keyword changes for related terms? Medical

image search on the Internet and in institutional databases has a high potential but more research is needed and particularly prototypes that can be made available to the users for testing to find out more about concrete information needs.

Acknowledgements

This work was partially supported by the Swiss National Science foundation with grant 205321-109304 and by Grant 0325160 of the US National Science Foundation..

References

[1] Rice RE, Influences, usage, and outcome of Internet health information searching: Multivariate results from the Pew surveys, International Journal of Medical Informatics, 75:8-28, 2006.

[2] Gaudinat A, Boyer C, WRAPIN (Worldwide online Reliable Advice to Patients and Individuals). In MEDNET 2003, The 8th Annual World Congress on the Internet and Medicine, Geneva, Switzerland.

[3] Hersh WR, Hickam DH, How well do physicians use electronic information retrieval systems? A framework for investigation and systematic review, Journal of the American Medical Association, 1998, 280: 1347-1352

[4] Gould P, The rise and rise of medical imaging, physicsweb, 16(**8**), August 2003.

[5] Smeulders AWM, Worring M, Santini S, Gupta A, Jain R, Content-Based Image Retrieval at the End of the Early Years, IEEE Transactions on Pattern Analysis and Machine Intelligence 22(**12**) pp 1349-1380, 2000.

[6] Markkula M, Sorminen E, Searching for photos - Journalists' practices in pictorial IR. In JP Eakins, DJ Harper, J Jose (Eds.), The Challenge of Image Retrieval Newcastle upon Tyne, UK, 1998.

[7] Choi Y, Rasmussen EM, Users' relevance criteria in image retrieval in American history, Information Processing and Management 38: 695-726, 2002.

[8] Tagare HD, Jaffe C, Duncan J, Medical Image Databases: A Content-Based Retrieval Approach, Journal of the American Medical Informatics Association (JAMIA), 4(**3**):184-198, 1997.

[9] Müller H, Michoux N, Bandon D, Geissbuhler A, A review of content-based image retrieval systems in medicine – clinical benefits and future directions, International Journal of Medical Informatics, 73, pp 1-23, 2004.

[10] Clough P, Müller H, Deselaers T, Grubinger M, Lehmann T, Jensen J, Hersh W, The CLEF 2005 Cross-Language Image Retrieval Track, Springer Lecture Notes in Computer Science LNCS **4022**, pp 535-557, 2005.

[11] Hersh W, Müller H, Gorman P, Jensen J, Task Analysis for Evaluating Image Retrieval Systems in the ImageCLEF Biomedical Image Retrieval Task, Slice of Life, multimedia in medical education, Portland, OR, USA, 2005.

[12] Müller H, Despond-Gros C, Hersh W, Jensen J, Lovis C, Geissbuhler A, Medical professionals' image search and use behavior, Medical Informatics Europe, 2006.

[13] Paliouras G, Papatheodorou C, Karkaletsis V, Spyropoulos CD, Tzitziras P, "From Web Usage Statistics to Web Usage Analysis," IEEE Conference on Systems Man and Cybernetics, 1999.

[14] Anderson RN, Smith BL, Deaths: Leading causes for 2001, National Vital Statistics Report 52(**9**):1-86, 2003.

[15] Müller H, Deselaers T, Lehmann T, Clough P, Kim E, Hersh W, Overview of the ImageCLEFmed 2006 Medical Retrieval and Medical Annotation Tasks, Springer Lecture Notes in Computer Science (LNCS), 2007 – to appear.

Address for correspondence

Henning Müller (PhD)
University & Hospitals of Geneva, Medical Informatics Service
24, rue Micheli-du-Crest, 1211 Geneva 14, Switzerland
Tel +41 22 372-6175, Fax +41 22 372-8680
henning.mueller@sim.hcuge.

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Improving Computer Aided Disease Detection Using Knowledge of Disease Appearance

Tatjana Zrimec[a,b] James S. Wong[b,]

[a] *Centre for Heath Informatics, University of New South Wales, Australia*
[b] *School of Computer Science & Engineering, University of New South Wales, Australia*

Abstract

The accurate quantification of disease patterns in medical images allows radiologists to track the progress of a disease. Various computer vision techniques are able to automatically detect different patterns that appear on images. However, classical pattern detection approaches do not perform satisfactorily on medical images. The problem is that texture descriptors, alone, do not capture information that is pertinent to medical images, i.e. the disease appearance and distribution. We present a method that uses knowledge of anatomy and specialised knowledge about disease appearance to improve computer-aided detection. The system has been tested on detecting honeycombing - a diffuse lung disease pattern in HRCT images of the lung. The results show that the proposed knowledge guided approach improves the accuracy of honeycombing detection. A paired t-test, shows the improvement in accuracy to be statistically significant ($p < 0.0001$).

Keywords:

lung HRCT, lung diagnosis, computer aided diagnosis, medical imaging, honeycombing detection

Introduction

Medical imaging systems are constantly improving in image quality because of increased image resolution. This results in a growing number of images that have to be inspected for diagnosis. For example, high resolution CT (HRCT) imaging protocols of the lungs can generate from 40 up to 600 images per study. These high-resolution axial images provide anatomic detail similar to that available from gross pathology specimens of lung slices [1]. Now radiologists can clearly see the alterations in lung anatomy caused by a disease process. Unfortunately, image analysis is still performed manually, which is often a difficult and time-consuming task. Consequently, there is an increasing need for computerised image analysis to facilitate image-based diagnosis.

We are developing a system for computer-aided detection of diffuse lung diseases, a large group of disorders that primarily affects the lung parenchyma. They are characterised by specific abnormal findings, mostly texture-like in appearance. Consequently, most of the automated detection algorithms, being developed to analyse CT scans are texture based. The classical approach is to use a set of image features to describe the image content and to use some classification scheme to distinguish between different patterns. For example, Uppaluri et al. [2] used twenty-two independent texture features to characterise a tissue pattern in the overlapping square regions of the lung. A Bayesian classifier was trained to discriminate between six different patterns. Uchiyama et al. [3] proposed a similar texture based technique. They trained an Artificial Neural Network with twelve features, calculated on regions with different sizes, to classify new regions. The system was trained to distinguishing between seven different patterns, which included normals and six patterns associated with diffuse lung diseases.

Our system, developed to automatically detect lung disease patterns, adopts a similar approach. However, we use a much bigger set of image attributes to describe the content of the image. We experimented with different attributes subsets and different learning schemes to improve the system's performance. The results reveal that classical pattern detection approaches do not perform satisfactorily on medical images. The problem is that texture descriptors, alone, do not capture information that is pertinent to medical images, i.e. the disease appearance and distribution. Therefore we incorporated domain knowledge of lung anatomy and lung structure to help and improve image analysis.

In this paper we focus on detecting honeycombing, an important diffuse lung disease pattern, in HRCT images of the lung. As the goal of the system is to provide radiologists with a second opinion on a lung diagnosis, it is important to achieve high accuracy. In this paper we present a new method developed to improve computer-aided detection. The method uses specialised knowledge of disease appearance in axial images. It also uses information about lung regions that are used in radiology reporting [9]. To determine if using knowledge can significantly improve the system's performance, we incorporated the knowledge-guided approach into two classification methods, one based on decision tree learning and the other using Naïve Bayes.

In the remainder of the paper, we present a computer-aided detection system using a classical pattern detection

approach. We then present the improved system and compare their performance.

Materials and methods

Lung diseases – honeycombing pattern

Honeycombing is one of the main indicators of diffuse lung diseases. It can be seen in many diseases leading to end-stage pulmonary fibrosis. Honeycombing is characterised by small, uniform (2-10mm) cystic air spaces with well-defined thick walls, (See Figure 1 left). Honeycomb cysts usually form clusters that have the characteristic appearance of "honeycombing" on HRCT images. The visual appearance of honeycombing in cross-section scans is a combination of dark and light patches. It is one of the more difficult disease patterns to detect because honeycombing can often be mistaken for other normal structures in the lung, for example, bronchi and pulmonary vessels (see Figure 1 right).

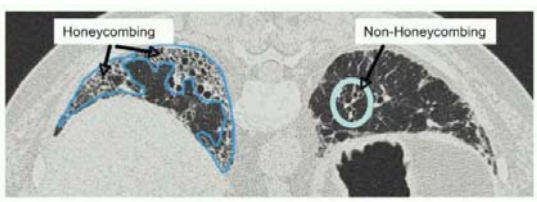


Figure 1 - Left lung- outlined region with honeycombing. Right lung –outlined example of bronchovascular structures, which has similar appearance as honeycombing.

The HRCT images

The HRCT images used for developing the computer aided disease detection system were obtained from a radiology practice. They were acquired using a SIEMENS scanner and a standard imaging protocol with 750ms exposure time. The HRCT generates volume data with spatial resolution 512x512 and 1.0mm slice thickness. For diffuse lung disease detection, radiologists usually use data with 10mm or 15mm slice gap, The data are stored in DICOM format as 16-bit greyscale images.

Describing the disease

There are specialised texts that describe how to interpret HRCT images, accompanied by illustrative examples [1]. Although this highly specialised knowledge is very useful for humans, computers cannot make direct use of it.

For computers to recognise a disease in medical images, the content of the image is represented by a set of image attributes, calculated for regions of interest (ROI) in an image. The values of these attributes depend on the characteristics of the regions. We can use these characteristic values to distinguish between normal and pathological regions as well as between different pathologies.

The method

The easiest way for a radiologist to communicate expert knowledge about how diseases appear in HRCT images is to provide examples. Using a specially developed image mark up system [4], we collected a set of images with marked and labelled regions of honeycombing and other lung diseases patterns. Figure 2 (top) shows example of an HRCT image with marked regions. Having examples of image regions with and without a disease, we were able to use supervised machine learning to generate rules for recognising different patterns in HRCT images.

The method consists of there main steps:

1. Data preparation:
   a) Image pre-processing and segmentation
   b) Feature extraction - calculating attributes for regions of interest
2. Knowledge generation - training:
   a) Feature selection – finding informative attributes for a particular disease pattern
   b) Generating rules via machine learning
3. Knowledge verification -testing the quality of the learned rules:
   a) On part of the training data
   b) On new data

Data pre-processing and segmentation

As we are interested in detecting patterns in the lungs, we first pre-process the images and segment the lungs. We have developed a lung segmentation technique based on adaptive thresholding, morphological operators and active contour snakes. Adaptive thresholding is applied to segment the darker regions in the image that represent the air-filled lung. Morphological operators are then used to include structures within the lung that have a high attenuation. Active contour snakes [5] are used to generate the lung contours (see Figure 2 bottom row).

Feature extraction

Having segmented the lung, we proceed to extract features from the image that best represent the underlying texture. A set of attributes was calculated for each pixel and it's surrounding area. We used a ROI with two window sizes, 7x7 and 15x15 pixels, to capture the characteristics of small and larger honeycombing cysts.

We calculate first and second order texture attributes and grey-level difference for each ROI [6]. The first order texture attributes measure the grey-level distribution within the ROI. Those attributes include: the mean HU[1], variance, skewness, kurtosis, energy and entropy. The second order features describe the spatial distribution of the grey-levels within these ROIs. To do this, a co-occurrence matrix is calculated that specifies the frequency of a particular grey-level occurring near another grey-level. Each pixel, with it's surrounding area, is represented by 63 attributes per window, resulting in a feature vector with 126 attributes, (63 for ROI_{7x7} and 63 for ROI_{15X15}).

1 Hounsfield unit (HU) a unit used in medical imaging (CT or MRI scanning) to describe the amount of x-ray attenuation of each "voxel" (volume element) in the three-dimensional image.

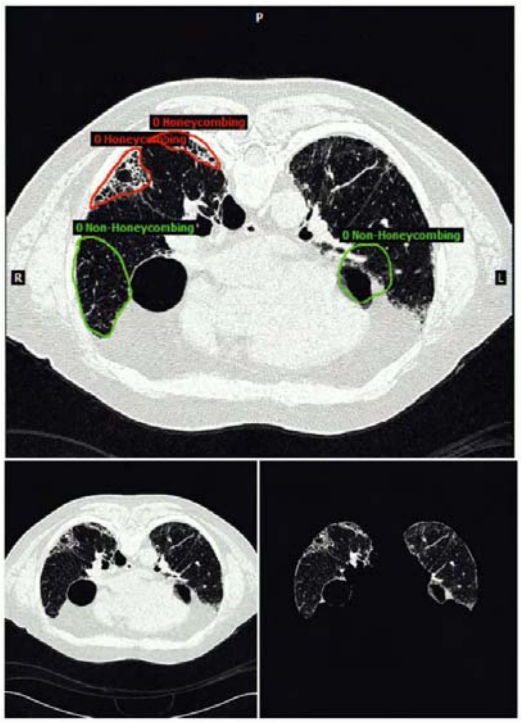


Figure 2 - Top: an image with marked and labelled regions by a radiologist showing disease patterns; bottom left: original HRCT image, bottom right: segmented lungs.

Knowledge generation process - feature selection

With a feature vector of 126 dimensions, the classifier generated would be computationally intractable. We reduce the dimensionality by selecting a subset of features that best discriminates honeycombed and non-honeycombed regions. In this study, we use Correlation-based Feature Selection (CFS) [7]. CFS selects subsets of attributes that are highly correlated with the class and that have low intercorrelation.

Generating rules via machine learning

Even with the reduced set of attributes, it is difficult to manually determine the attribute values that are characteristic for a particular pattern. We used supervised machine learning to automatically generate rules for discriminating between different patterns. In supervised learning, a set of pre-classified training examples is used to generate classification rules. In our case, the training examples consist of a set of attribute values representing a region with or without honeycombing pattern. The training set was prepared from the images with labelled regions provided by radiologists (see Figure 2 top).

We experimented with two machine learning algorithms to build the classifier: Naive Bayes and the decision tree learner, J48, both implemented in Weka data mining environment [8]. The Naive Bayes algorithm is based on a probability model. The probability of a class given a feature vector is determined using Bayes' rule:

$$P(c \mid F) = \frac{P(F \mid c)P(c)}{P(F)} \quad (1)$$

where c is the class and F is the vector of features. The class with the highest probability is assigned to the ROI. Although the Naive Bayes approach is optimal when the features are independent, in reality it still works well without this assumption. The decision tree learner generates a tree in which nodes represent tests on attributes, branches represent attribute values and leaf nodes represent classes, i.e. decisions. More informative attributes appear higher in the tree.

To train the system to detect honeycombing pattern, we used 42 images from 8 different patients that showed patterns representative of honeycombing and non-honeycombing tissue. After prepossessing and segmentation, feature extraction and selection were performed as described previously. A set of 18,407 labelled region of interest were used for training the machine learning algorithms from which 9,467 ROIs contained honeycombing and 8,940 ROIs did not. Two classifiers were built, one using Naïve Bayes and the second using decision tree induction.

Knowledge verification

Ten-fold cross validation was used to estimate the accuracy of the learned classifiers. In this validation scheme, 90% of the training data are randomly sampled for learning and 10% for testing. This is repeated 10 times and the results averaged.

The performance of the classification during training and testing was evaluated by calculating *accuracy*, *sensitivity, and specificity*. In our case *accuracy* measures the proportion of the lung that is classified correctly. *Sensitivity* determines the proportion of actual honeycombing that has been detected as honeycombing. *Specificity* measures the amount of non-honeycombing that has been classified as non-honeycombing.

Although the *accuracy*, *sensitivity* and *specificity* were comparable with the results published in the literature, we were not satisfied with the system's performance. It produced some spurious honeycombing classifications in regions where honeycombing cannot appear.

Improvements based on domain knowledge

Instead of developing post-processing methods for handling misclassifications, we decided to make use of domain knowledge about the lung structure as well as expert knowledge about the appearance of diseases. For example, Web [1] pp 91, states that "Honeycombing results in cysts …which have a peripheral predominance". This simple statement is not simple to implement. We first had to develop a model of the human lung. Next we had to develop algorithms that use anatomical knowledge to automatically generate lung regions, such as, peripheral, central, apical and basal, which are frequently used in disease reporting [9], [10]. These enabled us to determine the

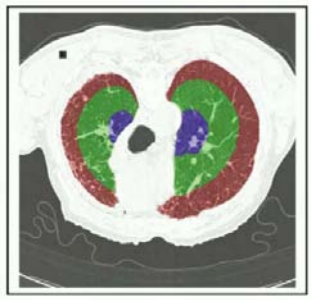


Figure 3 - An HRCT image with lung regions: blue-central, red - peripheral.

lung periphery on each axial scan, which helped in desease classification.

Knowledge guided classification

In many systems (e.g. [2, 3]), all regions within the lung are classified starting from the top of the image. However, for diseases that show honeycombing, the pattern spreads from the periphery of the lung. We developed a knowledge-guided strategy for classification. This strategy uses seeded region-growing [11] and works as follows:

- The algorithm initially only classifies peripheral regions. Peripheral regions are determined using the lung regions masks. ROIs in the periphery of the lung with honeycombing are set as the "seed points" for the algorithm.
- The algorithm only classifies a region of interest if it is near other ROIs already classified as honeycombing.
- The algorithm will stop when there are no more ROIs to consider.

In summary, the knowledge guides the system to classify all ROIs that are either in the periphery of the lung or in close proximity to other ROIs classified as honeycombing.

Results

In order to test the clinical viability of the system, we evaluated the performance of our system on part of the training data and on new, previously unseen data.

Testing on the training data **-** We used ten-fold cross validation. The number of ROIs used for testing varies for each fold, as the size of the lung in each slice affects the number of ROIs that we extract. On average 9,337 ROIs were used for testing (876 ROIs for honeycombing and 8,461 ROIs for non-honeycombing). The number of ROIs containing non-honeycombing was significantly larger as most of the lung region does not show honeycombing. The average of the results are shown in Table 1.

From the results presented in Table 1, it can be seen that the knowledge-guided approach improved the accuracy of the honeycombing detection. The improvement in accuracy is attributed to a decrease in false positive classifications (sometimes by over 2%). The increase in accuracy shows that the technique is well suited for honeycombing detection. A paired t-test, shows the improvement in accuracy to be statistically significant ($p < 0.0001$).

Table 1 – Results of the two classifies: Decision Tree Induction J48 - (DTI-J48) and Naïve Bayes used with classical and knowledge-guided approach.

Classifier	DTI-J48	DTI-J48	Naïve Bayes	Naïve Bayes
	Classical	Knowledge-guided	Classical	Knowledge-guided
Accuracy	88.20%	89.70%	85.50%	87.20%
Sensitivity	96.70%	96.60%	97.50%	97.40%
Specificity	86.80%	88.60%	83.50%	85.50%

Testing on new, previously unseen data **-** Images from 8 patients, 4 patients with homecoming present and 4 patients with different disease patterns were used to test the performance of the detection system. The evaluation was performed on six images with honeycombing present and six images with out honeycombing. In total, there were 3150 regions with honeycombing and 60318 regions without honeycombing. Sensitivity of the algorithm dropped to 85%.

Conclusion

The accurate quantification of disease patterns in medical images allows radiologists to track the progress of a disease. We have developed a system that uses machine learning to automatically detect honeycombing patterns in HRCT images of the lungs. Applying a classical, texture-based, approach resulted in over detection of honeycombing. It was not possible with simple post-processing to remove the false positive regions. We improved the perfor-

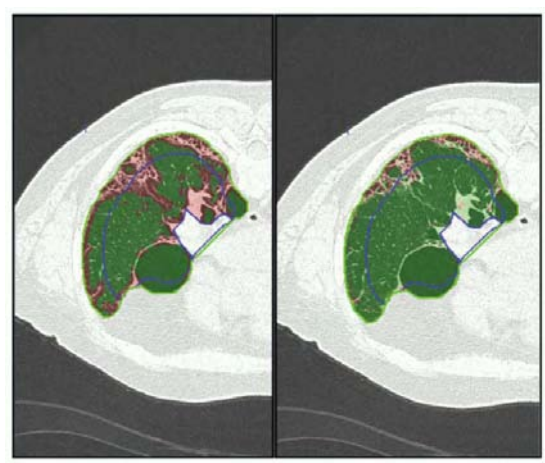


Figure 4 - Left: classification without using knowledge; right: knowledge-guided classification; red regions are classified as honeycombing; the blue line is the mask defining the periphery.

mance of the system by using knowledge-guided classification.

We tested the system with part of the training data and with new, previously unseen data. The results showed a high degree of accuracy (89.7%) and sensitivity (96.6%) on the training data. The accuracy sensitivity, however drop from 97 to 71% when testing on new data.

Experiments in building classifiers with different machine learning algorithms, Naïve Bayes and J48 decision tree learner, showed that the knowledge-guided classification performs better in both cases. The results showed that using knowledge-guided classification using the texture based Naïve Bayes classification lead to significant improvement according to pared t-test.

The domain knowledge not only improved the results of the classification, but it also improved the representation of the results. For example, a computer aided system without knowledge of lung structure reports that 5% of the fifth image contains honeycombing and that 20 % of the images 9, 10 and 11, also contain homecoming. Our system reports that 5% of honeycombing was detected in the apical area and 20% in the basal area, predominantly in the lower lobe of the left lung. Being able to use knowledge of lung anatomy in image analyses will significantly improve the detection and quantification of other lung diseases.

Acknowledgments

We thank Dr. Peter Wilson for providing medical knowledge and guidance and Medical Imaging Australia for providing data. We thank Claude Sammut for his comments. This research was supported by the Australian Research Council.

References

[1] Webb WR, Müller NL, Naidich DP. High-Resolution CT of the Lung. 3rd ed. Lippincott Williams & Wilkins, Philadelphia, 2001.

[2] Uppaluri R, Hoffman EA, Sonka M, Hartley PG. Computer Recognition of Regional Lung Disease Patterns. American Journal of Respiratory and Critical Care Medicine 1999:160(2): 648–654.

[3] Uchiyama Y, Katsuragawa S, Abe H, Shiraishi J, Li F, Li Q, Zhang CT, Suzuki K, Doi K. Quantitative computerized analysis of diffuse lung disease in high-resolution computed tomography. Medical Physics 2003:30(9): 2440–2454.

[4] Rudrapatna M, Sowmya A, Zrimec T, Wilson P, Kossoff G, Wong J, Busayarat S, Misra A, Lucas P, LMIK – Learning Medical Image Knowledge: An Internet-based medical image knowledge acquisition framework, Electronic Imaging Science and Technology - IS&T/SPIE's 16th Annual Symposium, San Jose, CA. 2004.

[5] Papasoulis J, LMIK - Anatomy and Lung Measurements using Active Contour Snakes, Undergraduate thesis, supervisor: T. Zrimec, Computer Science and Engineering, University of New South Wales, Sydney, Australia, 2003.

[6] Haralick RM, "Statistical and Structural Approaches to Texture," Proceedings of the IEEE 67, pp. 786– 804, 1979.

[7] Hall, MA. Correlation-based feature selection for discrete and numeric class machine learning. In: ICML '00: Proceedings of the Seventeenth International Conference on Machine Learning, San Francisco, CA, USA, Morgan Kaufmann Publishers Inc. 2000; pp. 359–366

[8] Witten, IH, Frank E. Data Mining: Practical machine learning tools and techniques. 2nd edition, Morgan Kaufmann, San Francisco, 2005.

[9] Zrimec T, Busayarat S, Wilson P. A 3D model of the human lung with lung regions characterization. ICIP 2004 Proc. IEEE Int. Conf. on Image Processing, 2004, pp. 1149–1152.

[10] Zrimec T, Busayarat S, 3D modelling and visualisation of the human lung, Proc. 2nd Int. Symp. on 3D Data Processing, Visualization and Transmission, Thessaloniki, Greece, September 2004.

[11] Adams R, Bischof L. Seeded region growing. Pattern Analysis and Machine Intelligence, IEEE Transactions on 1994: 16(6): 641–647.

Address for correspondence

Tatjana Zrimec
Centre for Health Informatics
University of New South Wales
Sydney, 2052
tatjana@cse.unsw.edu.au
Tel: +61 2 9385 9034
Fax: +61 2 9385 9006

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

MR Atlas for Articular Cartilage Morphology: Potential to Detect Shape Differences

Hussain Z. Tameem[a,b], Usha S. Sinha[b]

[a] *Department of Biomedical Engineering, University of California, Los Angeles, United States*
[b] *Medical Imaging Infomatics Group, University of California, Los Angeles, United States*

Abstract

An atlas of the cartilage was created using free form transformation of MR images of the cartilage from 20 subjects. The deformation required to move each voxel to its corresponding location in the atlas is used to determine the differences in shape between cartilages of subjects in a population. Based on these active shape models, it is possible to localize regions of high morphological variance in population cohorts. The atlas, reported here, is based on 20 male subjects; ten symptomatic of arthritis and ten asymptomatic. The active shape models based on this atlas show regions of high morphological variance corresponding to cartilage thinning in the arthritic group. This method has the potential to differentiate between normal and arthritic population groups by detecting subtle morphological changes in articular cartilage.

Keywords:

active shape models, articular cartilage, image registration, magnetic resonance imaging, morphometry

Introduction

Osteoarthritis (OA) is a complex, progressive disease of the joint characterized by degenerative and regenerative morphological and structural changes in the articular cartilage and subchondral bone [1]. OA is a slowly progressing disease characterized clinically by pain, enlargement and deformity of the joints, and limitation of the motion. OA is the most prevalent form of arthritis and leading cause of disability and work limitation among adults resulting in enormous cost to society [2, 3]. Approximately 21 million American adults have physician diagnosed OA, [4] a diagnosis usually based on the combination of joint symptoms and radiographic changes. The prevalence of OA in a population is difficult to determine because: 1) the degree of radiological changes in symptomatic individuals varies greatly and 2) many individuals with radiographic evidence of OA have no symptoms. By age 60 nearly half of the population has radiographic evidence of OA in one or more joints, and by age 80 these findings are universal [5, 6].

Among the various sites being affected in OA, knee is the major source of reported disability and loss of function. About 40% of the adult population age 55 and older has frequent knee pain or definite x-ray evidence of knee OA [7-9]. Advanced OA accounts for majority of knee replacements surgeries among Medicare recipients. Well over 200,000 knee replacement procedures for OA are performed every year in the United States [2].

As of today, there are no reports of any disease modifying therapies for knee OA and all treatments are predominantly designed to relieve pain [10]. Approaches to prevent knee OA development, progression, or related disability are also very limited, in large part due to incomplete knowledge of potentially modifiable factors responsible for these outcomes.

In this paper, we report a method based on free-form transformation to generate an average shape atlas of the femoral cartilage and apply it to study shape differences in a population cohort. This method has potential applications in the detection of subtle shape differences in normal and diseased population groups.

Background

Previous research on cartilage morphological assessment includes its volume and thickness measurements and the impact of various factors on normal knee, such as sex, body weight and height, maturity and age, body mass index, leg length and foot size, knee bone size, bone mineral density, muscle mass, level of physical exercise and genetics [11,12,13,14]. But none of the research to date focuses on studying the shape changes in cartilage between normal and diseased state population. Also there are studies which indicate the use of mathematical frameworks such as principal component analysis (PCA) to describe general shape variations. Marcus [15] used PCA to study the variation in the skull measurements of rodent and bird species. The resulting principal modes were interpreted as size and gross shape components. Cootes et al. [16] applied the theory of PCA to build statistical shape models of organs based on manually chosen landmarks. This model provided the average positions of the points and the principal modes of variations were computed from the dataset. The ability of the method to locate structures in medical images was demonstrated in a set of experiments with echocardiograms, brain ventricle tracking and prostate segmentation. Le Briquer and Gee [17] applied PCA to analyze the displacement fields obtained from registering a reference image volume of the brain to a set of subjects, based on the elastic matching framework. The

analysis provided the inference of morphological variability within a population and was the basis for the construction of a statistical model for brain shape, which could be used as prior information to guide the registration process. Duchesne et al. proposed shape models for segmentation of the medial temporal brain structures [18].

Materials and methods

Image acquisition

We obtained the images from a pilot study conducted for the National Institute of Health OA initiative (OAI) version 0.A.1. MR images were acquired using a water-excitation double echo steady-state (DESS) imaging protocol with sagittal slices at 3.0T (Magnetom Trio®, Siemens). The imaging parameters for the sequence were: TR/TE: 16.3/4.7 ms, matrix: 384∞384, FOV: 140 mm, slice thickness: 0.7 mm, x/y resolution: 0.365/0.365 mm. Figure 1 shows a sagittal slice of the magnetic resonance image obtained using DESS sequence.

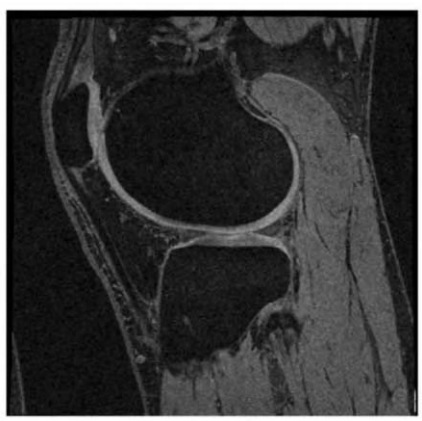


Figure 1 - MR image acquired with a DESS sequence with water-excitation at 3.0 T. The images are obtained from National Institutes of Health (NIH) OA initiative

The OAI data consists of a stratified random sample of 200 participants based on gender, sub-cohort assignment (progression and incidence) and clinic (four recruitment centers). The progression sub-cohort contains participants with symptomatic knee OA at baseline where symptoms are pain, aching or stiffness in or around the knee on most days for at least one month during the past 12 months. The incidence sub-cohort contains participants with no symptomatic knee OA at baseline, but has characteristics that place them at the risk for developing symptomatic knee OA during the study. For this study we randomly chose 20 male participants, 10 each from progression and incidence cohort group.

Atlas creation

We manually segmented the cartilage for all 20 subjects and then interpolated the raw data to a pixel resolution of 0.365x0.365x0.365 mm^3. We created the atlas in the following steps, which are illustrated in figure 2.

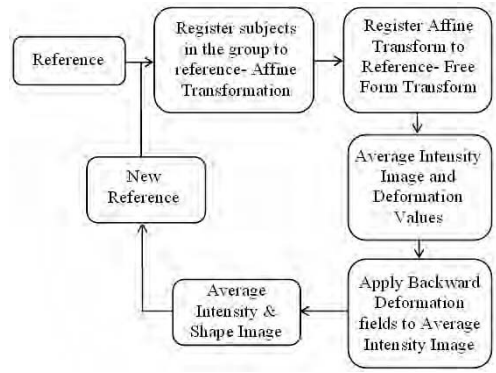


Figure 2 - Schematics of steps involved in atlas creation

Step I: We randomly chose one subject from the pool of 20 to serve as a reference to which the rest of the images were aligned utilizing the mutual information based affine transformation. Affine transform corrects for subject positioning and global size differences.

Step II: An elastic registration based on demons algorithm [19] was employed to locally map all the images in the group of subjects to the reference image using the affine transformation parameters as an initial estimation. This provides 3D deformation fields that can map the spatial locations on an individual in the group to the coordinate system of the reference. The registration algorithm computes the transformation iteratively using equation 1.

$$v_{n+1} = G_\sigma \otimes \left(v_n + G_\sigma \otimes \frac{1}{2}\left[\frac{C(T-S)\|\nabla S\|\nabla T}{(\|\nabla T\|^2 + \|\nabla S\|^2)\|\nabla S\|^2 + \|\nabla T\|^2 + 2(T-S)^2)} \nabla S \right] \right) \quad (1)$$

v_{n+1} is the correction vector field at iteration n+1, G_σ is the Gaussian filter with variance σ2, ⊗ denotes the convolution, C is the scaling factor and T and S are the target and transformed images respectively. The algorithm estimates the displacement which maps a voxel at location (*x,y,z*) in *T* to the corresponding anatomical location in *S.* The algorithm is implemented hierarchically and to preserve the morphology, deformation vector fields were computed utilizing both the forward and backward transformation.

***Step III*:** A mean intensity image with the shape of the reference image is created by averaging the globally and locally transformed images of the group.

Step IV: A mean deformation field that encodes the shape variation between the reference image and average shape of the elements in the subject group is created by averaging over 3D deformation vector fields of the individual subjects of the group.

Step V: Inverse average deformation field is applied to the average intensity image to generate and average intensity and deformation image template for the group under study.

Step VI: Steps 1-5 are iterated until no significant change in the deformation field is observed relative to the previous computation. At the end of each iteration the original reference image is replaced by the average template

Figure 3 - Accuracy of registration. Left to Right: reference, test, result of affine transform and free form deformation image. Top and bottom row showing slices at different locations. Outline from reference image shown superposed on test & aligned images.

constructed at *Step V* generating both average shape (morphometric) and intensity atlases that represent the centroid of the population data set.

Active shape models

Active shape models are used to represent the variance in cartilage shape within a given population. Active shape models based on principal component analysis of the deformation fields were created using the data from the last iteration of the atlas creation procedure [17]. With reference to the atlas creation, each iterative process results in an average atlas and deformation field d_{ir}. The deformation field, $d_{ir,}$ is the amount required to move the voxel from its original position (after global affine transformation) to the corresponding location in the atlas. Here *r* refers to the voxel, *i* represent the subjects (1-20 used to create the atlas) and *d* represents the deformation vector. The following analysis based on principal component analysis for data reduction was performed on all n voxel of the cartilage and consists of the following steps.

Step I: Calculation of the mean deformation for N subjects at each voxel as shown in equation 2 where dmean is the mean deformation at any voxel over all subjects.

$$d_{mean} = (1/N)\left(\Sigma\, d_i\right) \tag{2}$$

Step II: Computation of the deviation from the mean value as shown in equation 3

$$\Delta d_i = d_i - d_{mean} \tag{3}$$

Step III: Calculation of n∞n covariance matrix, C, to find the basis for the space as shown in equation 4

$$C = (1/N)\left(\Sigma \Delta d_i d_i^T\right) \tag{4}$$

Step IV: Diagonalization of the covariance matrix to obtain the eigenvectors, ν_k and the eigenvalues, λ_k.

Step V: Construction of the linear model as shown in equation 5, where $\nu = (\nu_1, \nu_2, \ldots, \nu_k)$ is the matrix of the first eigenvectors, and W_s is a vector of weights, also called the shape coefficient. This results in a shape model.

$$d = d_{mean} + \nu W_s \tag{5}$$

The shape variations by ±2 SD from the mean shape along the first two principal modes were generated using eigenvalues and eigenvectors derived from this analysis. These images were synthesized by setting the weights of the first or the second modes at $\pm\sqrt{\lambda_i}$ and $\pm 2\sqrt{\lambda_i}$ (where i = 1, 2 for the first and second eigen mode respectively) and all other weights to zero. The synthesized images provide a visual representation of the possible variance in shape of the cartilage based on 20 image sets to create the atlas.

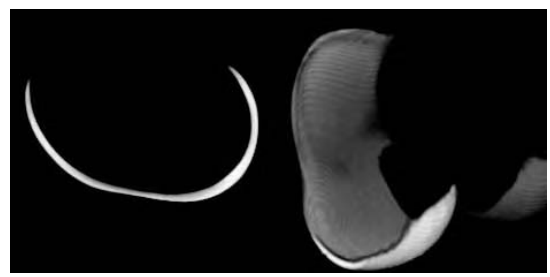


Figure 4 - Left: Shows the sharp edges on the atlas. Right: 3D volume reconstruction of the cartilage from the atlas

Results

The alignment of test image to the reference image using affine and freeform transformation is shown in figure 3. A region of interest is selected on the reference image and overlaid on the test, affine and the image obtained after free form transformation to show the accuracy of registration. It can be clearly seen that the affine transform corrects for positional and global scale changes and the local deformation corrects for physiological changes. The accuracy of the alignment can be visually evaluated by the sharp edges of the cartilage in the 2D images of the atlas as

Figure 5 - 3D active shape models. Average shape shown in the middle top row with standard deviation (SD) variations along the first and second mode shown in rows 2 and 3: Left to right - 2SD, -1 SD, +1 SD and +2 SD. Arrows indicate regions with changes from average bone shape

shown in figure 4. At the end of each iteration the reference image is replaced by the average image which moves closer to the centroid of the images in the group. There is no significant difference in the atlas after the third iteration confirming its convergence. Figure 4 shows 3 dimensional rendering of the cartilage generated from the atlas volume.

Figure 5 shows the active shape models and the variations in shape for ±2SD along the two leading eigenmodes. The outline of the cartilage is overlaid on the variations along the first and second modes that show the variations seen in the current set of 20 subjects. The first mode shows the larger changes since it captures the largest variation in the data as compared to the second mode.

Discussion

We successfully developed an atlas for the articular cartilage derived from the MR images at 3T. This is the first report of creating a cartilage atlas from images acquired at high resolution and isotropic resolution (0.365x0.365x0.365). The results show very accurate alignment which could be used for clinical purposes. We are currently working on creating a 3D Active contour without edges segmentation algorithm proposed by Chang T and Vesse L to extract the cartilage from the MR images.

We hypothesize that it is possible to automatically determine the cartilage location by analyzing its overall shape variation. If this hypothesis holds, then we can reverse the process and use the information obtained from the shape analysis to automatically segment cartilage from rest of the structures. In future we intend to use the active shape models and show that unsupervised learning can be used to explore the anatomy and facilitate segmentation. This methodology could potentially be used to classify different population groups. Structural shape characterization using PCA has been used to study gender and disease-related morphological differences in the corpus callosum, putamen, ventricles and hippocampus [20, 21]. It should be noted that within the scope of this paper we demonstrate the feasibility of generating the active shape models and that the application for classification will require far more image volumes to be included in the training set from the population cohorts of subjects under investigation.

Conclusion

We have developed an atlas for the cartilage and active shape models which when combined can be used to detect the subtle shape changes in cartilages. We see significant changes using this technique within the group of 20 subjects we selected. These models have a potential to be used in the future to discriminate normal and diseased states with larger databases.

Acknowledgments

We would like to thank The OAI study sponsored by NIH for providing the data used in this study

References

[1] Aubray JH, Jr, Sokoloff L 1989 Pathology of osteoarthritis. In: McCarty DJ (ed.) Arthritis and Allied Conditions, 11th Ed. Lea & Febiger, Philadelphia, PA, USA., pp. 1571-1580.

[2] Dunlop DD, Manheim LM, Yelin EH, Song J, Chang RW. The cost of arthritis. Arthritis Rheum 2003; 49(1):101-13.

[3] Guccione AA, Felson DT, Anderson JJ, Anthony JM, Zhang Y, Wilson PW, Kelly-Hayes M, Wolf PA, Kreger BE, Kannel WB. The effects of specific medical conditions on the functional limitations of elders in the Framingham study. Am J Public Health 1994;84:351-8.

[4] Lawrence RC, Helmick CJ, Arnett FC, Deyo RA, Felson DT, Giannini EH. Estimates of the prevalence of arthritis

and selected musculoskeletal disorders in the United States. Arthritis Rheum 1998; 41:778-99.

[5] Bagge E, Bjelle E, Valkenburg HA, Svanborg A. Prevalence of radiographic osteoarthritis in two elderly european populations. Rheumatol Int. 1992;12:33-8.

[6] Van Sasse JLCM, Van Romunde LKJ, Cats A, Vandenbroucke JP, Valkenburg HA. Epidimealogy of osteartritis: Zoetermeer survey. Comparison of radiological osteoarthritis in a Dutch population with that in 10 other population. Ann Rheum Dis 1989;48:271-80.

[7] Urwin M, Symmons D, Allison T, Brammah T, Busby H, Roxby M, Simmons A, Williams G. Estimating the burden of musculoskeletal disorders in the community: the comparative prevalence of symptoms at different anatomical sites, and the relation to social deprivation. Annals of Rheumatic Diseases 1998; 57:649-55.

[8] Peat G, McCarney R, Croft P. Knee pain and Osteoarthritis in older adults: a review of community burden and current use of primary health care. Annals of Rheumatic Diseases 2001; 60:91-7.

[9] Felson DT, Zhang Y. An update on Epidemiology of knee and hip osteoarthritis with a view to prevention. Arthritis Rheum 1998;41(8):1343-55

[10] Felson DT, Lawrence RC, Hochberg MC, McAlindon T, Dieppe PA, Minor MA, Blair SN, Berman BM, Fries JF, Weinberger M, Lorig KR, Jacobs JJ, Goldberg V. Osteoarthritis: New Insights. Part 2: Treatment Approaches. Ann Internal Med 2000; 133(9): 726-37

[11] Cicuttini F, Forbes A, Morris K, Darling S, Bailey M, Stcukey S. Gender differences in knee joint cartilage volume as measured by magnetic resonance imaging. Osteoarthritis Cartilage 1999; 7:265-71.

[12] Ding C, Cicuttini F Scott F, Glisson M, Jones G. Sex differences in knee cartilage volume in adults: role of body and bone size, age and physical activity. Rheumatology (Oxford) 2003.

[13] Cicuttini FM, Teichtahl AJ, Wluka AE, Davis S, Strauss BJ, Ebeling PR. The relationship between body composition and knee cartilage volume in middle-aged subjects. Arthritis Rheum 2005;52:461-7.

[14] Jones G, Ding C, Scot F, Cicuttini F. Genetic Mechanism of knee osteoarthritis: a population based case-control study. Ann Rheum Dis 2004; 63:1255-9.

[15] Marcus L. Traditional Morphometrics In: Rohlf FJ, Bookstein FL, editors. Proceedings of the Michigan Morphometrics Workshop, The University of Michigan Museum of Zoology, Ann Arbor, 1990. p. 77-122.

[16] Cootes T, Hill A, Taylor CJ, Haslam J. Use of active shape models for locating structures in medical images. Image Vision Comput 1994; 12548:126-32.

[17] Le Briquer L, Gee JC. Design of a statistical model of brain shape. Lect Notes Comput Sci 1997; 1230:477-82.

[18] Duchesne S, Pruessner JC, Collins DL. Appearance-based segmentation of medial temporal lobe structures. NeuroImage; 2002,17:515-532

[19] Thirion JP. Image Matching as a diffusion process: An analogy with Maxwell's demons. Medical Image Analysis, 1998; Vol 2 (3), 243-260

[20] Kim SH, Lee JM, Kim HP, Jang DP, Shin YW, Ha TH, Kim JJ, Kim IY, Kwon JS, Kim SI. Asymmetry analysis of deformable hippocampal model using the principal component in schizophrenia. Hum Brain Mapp. 2005 Aug; 25(4):361-9

[21] Shenton ME, Gerig G, McCarley RW, Szekely G, Kikinis R. Amygdalahippocampal shape difference in schizophrenia: the application of 3D shape models to volumetric MR data. Pyschiatry res. 2002 Aug 20; 115(1-2):15-35

Address for correspondence

Hussain Tameem
UCLA Medical Imaging Informatics
924 Westwood Blvd, Suite 420
Los Angeles, CA 90024 USA
Email: htameem@mii.ucla.edu

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Automatic Image Modality Based Classification and Annotation to Improve Medical Image Retrieval

Jayashree Kalpathy-Cramer[a], William Hersh[a]

[a] *Department of Medical Informatics and Clinical Epidemiology, Oregon Health and Science University, Portland, Oregon, USA*

Abstract

Medical image retrieval can play an important role for diagnostic and teaching purposes in medicine. Image modality is an important visual characteristic that can be used to improve retrieval performance. Many test and on-line collections do not contain information about the image modality. We have created an automatic image classifier for both grey-scale and colour medical images. We evaluated the performance of the two modality classifiers, one for grey-scale images and the other for colour images on the CISMeF and the ImageCLEFmed 2006 databases. Both classifiers were created using a neural network architecture for learning. Low level colour and texture based feature vectors were extracted to train the network. Both classifiers achieved an accuracy of > 95% on the test collections that they were tested on. We also evaluated the performance of these classifiers on a selection of queries from the ImageCLEFmed 2006. The precision of the results was improved by using the modality classifier to re-sort the results of a textual query.

Keywords:

medical imaging, neural networks, image annotation, content-based image retrieval

Introduction

Medical images form a vital component of a patient's health record. Effective medical image retrieval systems can play an important role in aiding in diagnosis and treatment; they can also be effective in the education domain for healthcare students, instructors and patients alike. As a result of advances in digital imaging technologies, there has been a large growth in the number of digital images stored in recent years. In addition to the Picture Archival and Communication Systems (PACS) that are becoming omnipresent in hospital and clinics, there are numerous on-line collections of medical images. On-line atlases of images can be found for many medical domains including dermatology, radiology and gastroenterology. The sheer volume of medical image data provides for numerous challenges and opportunities in the arena of medical image retrieval.

Historically, the task of indexing and cataloging these collections has been performed manually. This is an arduous and painstaking task, and is prone to errors. Consequently, there is a desire to be able to automate the task of indexing these collections with a goal to improve the ability to search and retrieve relevant documents.

Medical image retrieval systems have traditionally been text- based, relying on the annotation or captions associated with the images as the input to the retrieval system. The last few decades have offered advancements in the area of content-based image retrieval (CBIR) [1]. CBIR systems have had some success in fairly constrained medical domains, including pathology, head MRIs, lung CTs, and mammograms [2]. However, purely content-based image retrieval systems currently have limitations in more general medical image retrieval situations, especially when the query includes information about pathology [3, 4]. Mixed systems (using both textual and visual techniques) have demonstrated improved retrieval performance, especially with regards to precision at the top of the list [4].

Medical image databases used for image retrieval or for teaching purposes often contain images of many different modalities, taken under varied conditions with variable accuracy of annotation. This can be true for images found in various on-line resources, including those that access the on-line content of journals[1].

Image modality is an important, fundamental visual characteristic of an image that can be used to aid in the retrieval process. However, the annotations or captions associated with images often do not capture information about the modality. Images that may have had modality associated with them as part of the DICOM header can lose that information when the image is compressed to become a part of a teaching or on-line collection. There have also been reported errors in the accuracy of DICOM headings [5].

The medical image retrieval task within ImageCLEF has provided both a forum as well as test collections to benchmark image retrieval techniques. The ImageCLEF campaign has been a part of the Cross Language Evaluation Forum since 2003 [3]. CLEF itself is an offshoot from the Text REtrieval Conference (TREC). In 2004, ImageCLEFmed, a domain–specific task, was added to evaluate medical image retrieval algorithms and techniques.

1 http://gm.arrs.org/ (accessed 3/26/2007)

Approaches combining both visual and textual techniques for retrieval have shown some promise at medical image retrieval tasks [3]. In 2005, a medical image annotation task was added to ImageCLEF. The goal of this task was to correctly classify 1000 test images into 116 classes given a set of 10,000 training images. The classes differed primarily in anatomy and view of the image. It should be noted, however, that these images were primarily of a single modality (X-rays).The goal of the ImageCLEF medical image retrieval task of 2006 was to retrieve relevant images for thirty topics from a test collection of about 50,000 annotated images of different modalities. These tasks were divided by the organizers into those expected to be amenable to textual, visual, or mixed retrieval techniques.

We participated in both the medical image retrieval and the automatic medical image annotation tasks at ImageCLEF 2006 [6, 7]. The techniques developed for those tasks have been extended for the more general task of medical image modality classification and annotation.

Using medical image modality for image annotation and retrieval has recently been studied. Florea et al [8] have compared the efficacy of two different systems (MedIC and MedGIFT) in classifying the modality of a database with six standard modalities for radiology and nuclear medicine images.

In this paper, we compare the results obtained on our system with those described in previous publications [8] for the six modalities of the CISMeF database. We will also extend this technique to classify colour images from the ImageCLEF medical retrieval task collection [6] into six categories. We will finally report on the improvement in precision that we observed for a selected number of tasks of the ImageCLEF medical retrieval task for 2006 by incorporating the modality classifier in series with a text-based retrieval system.

Methods

We employed a supervised machine learning approach to problem of medical image modality classification using a hierarchical classification scheme as seen in figure 1. There were two primary databases that were used to create and test the classifiers. We worked with a small subset of the CISMeF database as the primary target for our grey-scale (radiographic and nuclear medicine) image classifier [9]. This database had a set of 1332 images classified into one of six classes based on modality. These include angiography, computerized tomography scans (CT), X-ray, Magnetic resonance (MRI), ultrasound, and scintigraphy. The images in this database had been acquired under differing conditions over a long period of time. Consequently, there was considerable intra-class variation in quality, size, contrast, illumination and background.

The imageCLEFmed database contains 50,000 images of differing modalities, including radiography and nuclear medicine, as well as microscopic and histopathological images, photographs and gross pathology images, power point slides, electroencephalograhical images (EEGs) and electrocardiograms (ECGs), as well as a few miscellaneous images.

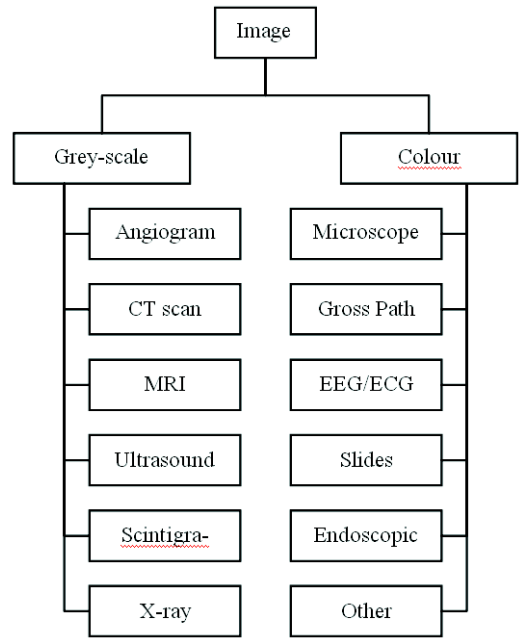


Figure 1 - Hierarchical classification scheme for images

A neural network-based scheme using a variety of low level, primarily global image features was used to create a six-class classification system for the grayscale images. The multilayer perceptron architecture used a hidden layer of approximately 50-150 nodes. The classification system was created in MATLAB[2], in part using several routines modified from the Netlab toolbox[3].

We experimented with a variety of feature vectors as inputs to the network. A combination of texture and intensity histogram features provided the best classification [10, 11]. All images were first resized while maintaining the aspect ratio such that the smaller dimension was 256 pixels. The image was divided into five overlapping blocks. Grey level correlation matrices were computed for each block using four angles and an offset of 1 pixel. Contrast, correlation, energy, homogeneity and entropy were calculated for each matrix. A quantized grey scale histogram was then appended resulting in a 132-dimension feature vector for each image for the texture. All inputs to the neural network (the image feature vectors) were normalized using the training set to have a mean of zero and variance of 1.

The 1332 images in the database were randomly split into a training set of 1000 images and a test set of 332 images. A small random subset of the training images was initially used to create the classifier (200 images). The classifier

2 www.mathworks.com (accessed 3/26/2007)
3 http://www.ncrg.aston.ac.uk/netlab/index.php (accessed 3/26/2007)

was then applied to the entire training set and images that were misclassified were then added to the images used to refine the classifier. The classifier was finally tested on the test images.

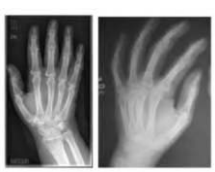
A similar scheme was used to create the classifier for colour images. We believe this novel idea can improve the retrieval performance of purely textual systems or for images for which the associated modalities are not known. Although modality detectors for grey-scale medical images have been reported [9], we are unaware of a similar effort for classification of other categories of medical images like those produced by histo-pathology and endoscopy. The images used for this classification task were taken from the test collection of images used in the ImageCLEFmed retrieval task. 2250 colour images in this collection were broadly categorized into six categories as microscopic, gross pathology, EEG/ECG or other charts, powerpoint slides, endoscopic images and other. There was considerable intra-class variability in this dataset. These 2250 images were again split randomly into training (1750) and test images (500). A similar training methodology to that described above was used to incrementally improve the classifier, starting with a smaller subset of the training database.

A two-layer architecture with 25-150 hidden nodes was used for the neural network. The feature vector in this case consisted of colour histogram features, as well as texture features obtained using the grey level correlation matrices. The image was split into 9 uneven blocks. Colour histogram properties of image after conversion into the L*A*B* colour space were calculated, while texture features were calculated after converting the image to grey-scale

These neural network classifiers can be created to further classify images within a given modality. For instance, x-ray images could now be classified to account for anatomy. Anatomical classifiers were used in the automatic annotation task at ImageCLEFmed.

The tasks had been stated in English, German and French, and had and provided example images. All but three of the tasks stated the desired modality of the image to be retrieved. Two examples of the tasks are shown in figure 2.

Show me images of a hand x-ray.
Zeige mir Röntgenbilder einer Hand.
Montre-moi des radiographies de la main.



Show me blood smears that include polymorphonuclear neutrophils. Zeige mir Blutabstriche mit polymorphonuklearer Neutrophils. Montre-moi des échantillons de sang incluant des neutrophiles polymorphonucléaires.

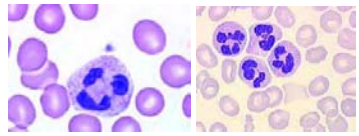


Figure 2 - Sample textual and visual queries at ImageCLEFmed 2006

Once our classifiers had been trained to achieve >95% classification accuracy, they were tested on a random subset of the ImageCLEFmed topics.

The schematic of our modified retrieval system is shown below. The query was initially fed to our Lucene[4] based text retrieval system. The queries were manually edited by one of the authors. The resulting images were subsequently classified by the hierarchical classifier for modality. Images of the desired modality (as stated in the query or as discerned by the automatic classifier based on the sample images) were moved to the top of the list while maintaining the ranking of the textual system within a class.

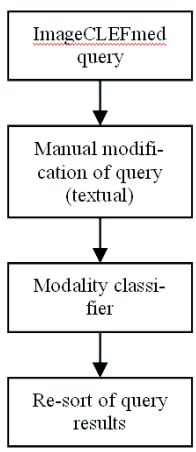


Figure 3 - Image retrieval system used for the ImageCLEFmed 2006 test collection

We compared the results of our purely textual system with that including the addition of the modality classifier.

Results

A classification accuracy of 96.4% was achieved on the CISMeF database. The confusion matrix suggests that the primary misclassification occur between the MRI and CT scan classes. This is not surprising as these classes are visually quite similar. Florea et al [8] have reported similar results both in terms of accuracy and inter-class misclassification patterns. The classification of grey-scale medical images into commonly occurring modalities using low level image features and machine learning techniques appears to be a tractable task. We expect to achieve over 98% accuracy with further refinement of our machine

4 http://lucene.apache.org/ (accessed 3/26/2007)

learning approach by the use of more advanced cross-validation, bootstrapping, boosting or bagging techniques.

Preliminary testing of the classifiers on 2250 colour images of the imageCLEFmed test collection resulted in a modality classification accuracy of 98.6%. Most of the misclassifications involved the "other" class with contained a set of miscellaneous images not belonging to the other five specific categories

The colour modality classifier was tested on a small random subset of the ImageCLEFmed 2006 topics. The topics for imageCLEFmed 2006 fell into three categories (visual, mixed, semantic) consisting of 10 tasks each. Although visual techniques had, in general, performed extremely poorly at the semantic tasks, use of some visual information (primarily modality) was shown to increase the precision [4].

Analysis of our textual results indicated that in many queries, especially those of a visual or mixed nature, up to 75% of the top 1000 results were not of the correct modality. A compelling example is given in figure 4 and table 1. Only 90 of the top 2000 images returned by the textual query were of the desired modality.

Task 1 - Show me images of the oral cavity including teeth and gum tissue

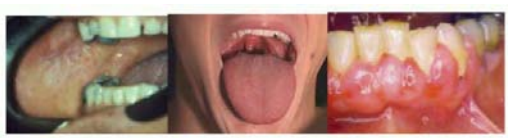


Image type	Number of images
Total returned by textual query	2000
Grey-scale	1831
Photograph/gross pathology	90
Microscope	71
Other	8

Figure 4- Sample query suitable for visual retrieval at ImageCLEFmed 2006

These images were then classified using our modality classifier. The ranked list of retrieved images was resorted taking into account the desired modality based on the query.

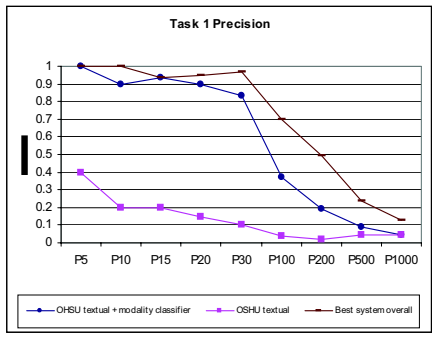


Figure 5 - Improvement in precision resulting from modality classification.

Figure 5 plots the precision for varying number of documents retrieved for the purely textual system, the improvement with the use of the modality classifier and the overall best system (mixed visual and textual based) that participated in ImageCLEFmed 2006. This increased the precision of the query as seen in figure 5. The improvement in precision at the top of the ranked list (P5 – P200) is better with the use of the modality detector compared to a purely textual search. We should note that a perfect modality classifier will only improve the precision of the search and not the recall if it is applied in the serial manner described above. The mean average precision (MAP) would still be limited by the number of relevant images that are retrieved by the textual search (recall of the textual search).

Even in searches that are expected to be semantic, we see an improvement in precision by using the modality classifier as seen in figure 6 and 7.

Task 2 - Show me microscopic images of tissue from the cerebellum (semantic query)

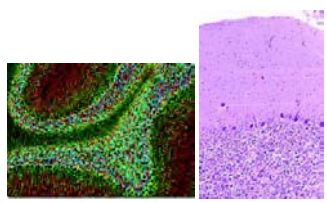


Image type	Number of images
Total returned by textual query	2000
Greyscale	1476
Photograph/gross pathology	408
Microscope	116

Figure 6 - Sample query suitable for visual retrieval at ImageCLEFmed 2006

The precision of this search was similarly improved by the use of the modality detector as seen in figure 7.

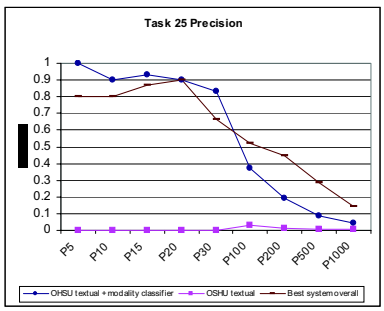


Figure 7 - Improvement in precision resulting from modality classification

Four of the six tasks tested showed improvement in precision by the use of the modality detector for colour images. There were two tasks amenable to textual methods for which there very little change in precision with the addition of the modality information.

We plan on testing the performance of the modality detector on the complete set of tasks for ImageCLEFmed 2005 and 2006. We also intend to index the entire collection of 50,000 images used in the ImageCLEFmed test collection using the modality classifier. Information about the class membership of an image will be added to the metadata. This should improve the performance of the retrieval in two ways. Clustering of the data by modality and perhaps anatomy will speed up the search process as fewer documents will have to be compared to the query image/text. Secondly, we expect that the overall precision of the search will improve by considering the modality of the image that is desired by the user. However, we can expect a small degradation in the recall due to potentially misclassified images not being searched.

Conclusion

We have developed a neural network based, hierarchical classifier for the modality classification of medical images. This system can classify colour images including histo-pathological and endoscopic images, and photographs as well as grey-scale (radiological and nuclear medicine). The classifier uses a histogram and texture properties as inputs to the two level neural network. This classifier results in a classification accuracy of greater than 95% for the grey-scale images of the CISMeF database as well as a selection of colour and grey-scale images from the ImageCLEFmed database. The use of this classifier increases the precision of retrieval of our primarily text based retrieval system by moving images of the desired modality to the top of the ranked list.

Acknowledgments

The authors would like to thank Henning Müller for his valuable input in discussions as well as providing access to the CISMeF images. We would also like to thank Jean-Nicholas Dacher at the Rouen Hospital, France, for proving the CISMeF images used in this paper. We would like to thank Dr. T. Lehmann at RWTH Aachen University, Germany, for access to the IRMA image database. We acknowledge the support of NLM Training grant 5T15 LM07088-15 and NSF Grant ITR-0325160.

References

[1] Smeulders AWM, Worring M, Santini S, Gupta A, Jain R. Content-based image retrieval at the end of the early years. IEEE Transactions on Pattern Analysis and Machine Intelligence 2000;22(12), 1349-80.

[2] Müller H, Michoux N, Bandon D, Geissbuhler A. A review of content-based image retrieval systems in medicine – clinical benefits and future directions. International Journal of Medical Informatics 2004; **73**, 1-23.

[3] Müller H, Deselaers T, Lehmann T, Clough P, Hersh W. Overview of the ImageCLEFmed 2006 medical retrieval annotation tasks, Evaluation of Multilingual and Multi-modal Information Retrieval, Seventh Workshop of the Cross-Language Evaluation Forum, CLEF 2006, editors: Peters C, Clough P, Gey F, Karlgren J, Magnini B, Oard DW, de Rijke M, Stempfhuber M, LNCS 2006,Alicante, Spain, to appear.

[4] Hersh W, Kalpathy-Cramer J, Jensen, J. Medical Image Retrieval and Automated Annotation: OHSU at ImageCLEF 2006. Working Notes for the CLEF 2006 Workshop, Alicante, Spain. http://www.clef-campaign.org/2006/

[5] Güld MO, Kohnen M, Keysers D, Schubert H, Wein BB, Bredno J, Lehmann TM, Quality of DICOM header information for image categorization, Proceedings SPIE, 4685, 280-287, 2002.

[6] Hersh W, Müller H, Jensen J, Yang J, Gorman P, Ruch P. Advancing biomedical image retrieval: development and analysis of a test collection. J Amer Med Inform Assoc 2006; 13(5).

[7] Lehmann TM, Güld MO, Thies C, Fischer B, Keysers D, Kohnen M, Schubert H, Wein BB. Content-based image retrieval in medical applications for picture archiving and communication systems, in Medical Imaging, SPIE Proceedings, 5033, 440-451, San Diego, California, 2003.

[8] Florea F, Müller H, Rogozan A, Geissbühler A, Darmoni S, Medical image categorization with MedIC andMedGIFT. Medical Informatics Europe (MIE 2006).

[9] Douyere M, Soualmia LF, Neveol A, Rogozan A, Dahamna B, Leroy JP, Thirion B, Darmoni SJ. Enhancing the MeSH thesaurus to retrieve French online health resources in a quality-controlled gateway. Health Info Libr J 2004; 21(4):253-61.

[10] Haralick R. Statistical and structural approaches to texture. Proceedings of the IEEE 67**,** 786-804, 1979.

[11] Howarth P, Ruger S. Evaluation of texture features for content-based imageretrieval. In: Proceedings of the International Conference on Image and Video Retrieval, Springer-Verlag, pp. 326-324, 2004.

Address for correspondence

Jayashree Kalpathy-Cramer
5th Floor, Biomedical Information Communication Center
3181 S.W. Sam Jackson Park Rd.
Portland, Oregon 97239-3098

Email: kalpathy@ohsu.edu

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Quantification of Myocardial Perfusion for CAD Diagnosis

Hartmut Dickhaus[a], Markus Erbacher[a], Helmut Kücherer[b]

[a] *Department of Med. Informatics, University of Heidelberg, Heidelberg, Germany*
[b] *Department of Cardiology, University Hospital, Heidelberg, Germany*

Abstract

We introduce a computer based algorithm for objective quantification of myocardial perfusion to support the diagnosis of cad patients. This new method is based on conventional cine angiographic films. In order to achieve maximal quality of the digital subtraction angiography images, the sequence is synchronized with the ECG. Optionally, the digital images can be motion compensated by a two step matching method. The spatio-temporal spread of blood, or the so-called blush, through the microvasculature to the myocardium - indicated by dye injection – represents a characteristic pattern for the myocardial perfusion. This dynamic temporal pattern is characterized by typical features as the maximal value of blush intensity, of increase and of decrease velocity which correspond with the different phases of flooding in and washout. On the basis of 100 different temporal blush profiles, an algorithm is established which classifies the acquired blush patterns into 4 different grades.

Keywords:

angiography, blush grade, CAD diagnosis, myocardial perfusion

Introduction

Currently, the coronary angiography is still the gold standard for coronary artery disease (CAD) diagnosis, although other methods have been proposed and discussed [1]. In almost all cases, the required invasive procedure of inserting a catheter for dye injection is combined with a percutaneous transluminal coronary angioplasty (PTCA). Therefore, the angiographic procedure allows an immediate estimation of the success of the therapeutic intervention.

For quantifying complete and sustained reperfusion of the infarcted myocardium and prognostic statements as for identifying patients at high risk, myocardial blood flow, expressed in so-called blush grades, is much more appropriate than any other angiography related measure. This is demonstrated by Stone et al. [2].

However, so far only qualitative descriptions for different extends of blush or myocardial reperfusion exist. Gibson et al. [3] presented 4 different grades which classify the perfusion in relation to its temporal dynamic and intensity. For example: grade 1 is defined as: "dye slowly enters but fails to exit the microvasculature. There is no blush or opacification on the myocardium in the distribution of the culprit lesion that fails to clear from the microvasculature and dye staining is present on the next injection 30 sec later."

Despite the principal advantages of blush characterization, qualitative descriptions like these are difficult to apply in an objective and reproducible stratification. For that reason, we tried to establish a computer assisted procedure to quantify blush grades corresponding with those of Gibson, however, in an objective and formal description. Furthermore, we developed a computer aided tool to visualize the spatial and temporal spread of the adelomorphic blush, of the myocardium from cineangiographic films after dye injection into the arteries. Under the control of the cardiologist, a specific blush grade is assigned for a specific myocardial region related to the three main supplying coronary arteries.

Materials and methods

In this study, 100 films from patients with various extent of CAD recorded in different projections, are quantitatively examined before and after PTCA. Because the heart is supplied by three different arteries: right coronary artery (RCA), left artery descending (LAD) and left circumflex (LCX), we look for blush occurrence in correspondent myocardial areas supplied from these vessels.

In order to enhance the contrast between blush and the arteries and surrounding tissue, images after dye injection have to be subtracted from that prior to injection, respectively in standard position. The initial image without dye is called a mask and is mostly established by averaging two or three consecutive frames in order to smooth small artifacts and noise. The difference of images shows the highest contrast between filled vessels, microvasculature and the surrounding, if we take the logarithmically transformed angiograms. In order to consider the motion of the heart and its vessels as well as the inserted catheter during the heart cycle, one has to choose images for subtraction which correspond to identical heart geometry and position. This can be achieved by synchronization of the image acquisition to the R peak of the simultaneously recorded ECG. Fortunately, the time interval between two cycles,

about 1 sec, is relatively small, so that the resulting time course of dye spreading is still sufficiently documented.

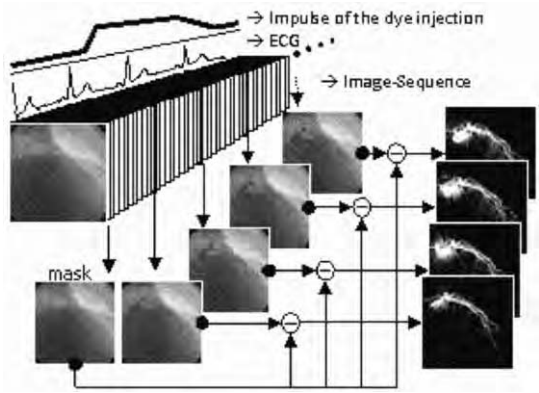


Figure 1 - Acquisition of ECG synchronized frames from the cine angiographic film. The black marked frames related to the R peak of every cardiac cycle are used for creating the sequence of difference images, displaying the spreading blush across the heart.

In figure 1 the acquisition procedure is schematically demonstrated. With increasing time we get a sequence of difference images displaying the temporal progression of dye and its spatial spreading through the arteries and small vessels into the microvasculature and the myocardium. The quality of these images demonstrating the blush is highly dependent to any change of body position or motion of inner organs, like the diaphragm, during breathing. For this reason, the patient is asked to stop breathing for about 20 sec, and the camera as well as the operating table have to be fixed for that period. However, in clinical routine it is difficult to accomplish these conditions completely. Therefore, we apply additional compensation procedures in two consecutive steps.

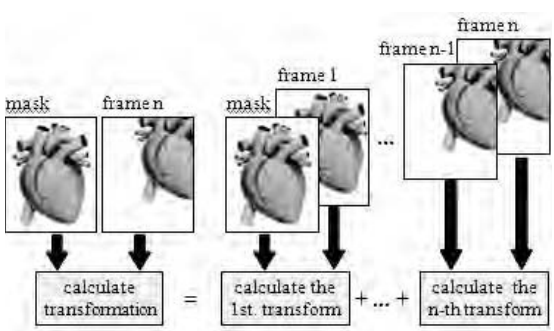


Figure 2 - Schema of the registration procedure between frame n and the mask: the registration is split into n single consecutive transformations and finally the sum of all transformations is performed.

Movements of the operation table or of the camera adjusted by the cardiologist for optimal viewing conditions are primarily compensated by a rigid matching procedure. Secondly motion of the inner organs particular caused by breathing needs an elastic compensation. In most cases the table or camera movements are larger but can be corrected more easily than the breathing effects. This two step procedure minimizes the extent and time for the elastic registration [4].

One general problem characterizes both registration steps, even the continuous spread of dye during the digital acquisition procedure. This means that besides the mentioned motion artefacts we have to consider differences between succeeding frames resulting from the dye distribution. In order to minimize this problem we apply the matching procedures always between two consecutive image frames. That means that the registration is performed every time relative to the previous frame. Finally all single transformations are summed up. Figure 2 demonstrates this process schematically. To calculate each transformation step, the rigid one as well as the elastic one, primarily the vessels have to be segmented for each frame. The pixels identified as vessels are then skipped during the calculation, in order to minimize the problems caused by the spreading dye.

The spatial distribution of the blush is represented by the gray value averaged from myocardium pixels excluding the arteries. That means we have to identify all myocardium pixels belonging to the blush. This task is very hard even for an experienced cardiologist. We perform this segmentation by an interactive procedure. For all pixels, we trace the intensity profile, respectively the gray value, over time. The incoming wave of dye after injection follows primarily the main coronary arteries, flows through the smaller vessels into the microvasculature, and reaches the myocardium. The washout is collected in the venous vessels. This time dependent process can be represented in its spatial distribution by detecting the maximal values of the intensity profiles of all image pixels. All pixels identified by their maximal gray values at a certain time instant represent more or less a specific structure of the pathway to the venous system. As it is demonstrated in Figure 3, it is easy to choose a color map representing the arteries map c) or d) or the perfused myocardium, pixel map f). For all segmented pixels of map f) we calculate the average gray value, cycle for cycle and plot the corresponding time course of blush intensity.

The cardiologist has to position a specific ROI at a region of the myocardium besides the arteries where a blush would be expected under normal conditions. We also calculate for this specific region the average gray values for different instants of time as characteristic time courses for a restricted perfusion. The different time dependent profiles have to be characterized by a few parameters which can be used to distinguish between different grades of blush. Furthermore, we would like to define parameters which can be easily interpreted in context with the existing qualitative characterizations of Gibson, and finally they should be related to the pathophysiological background.

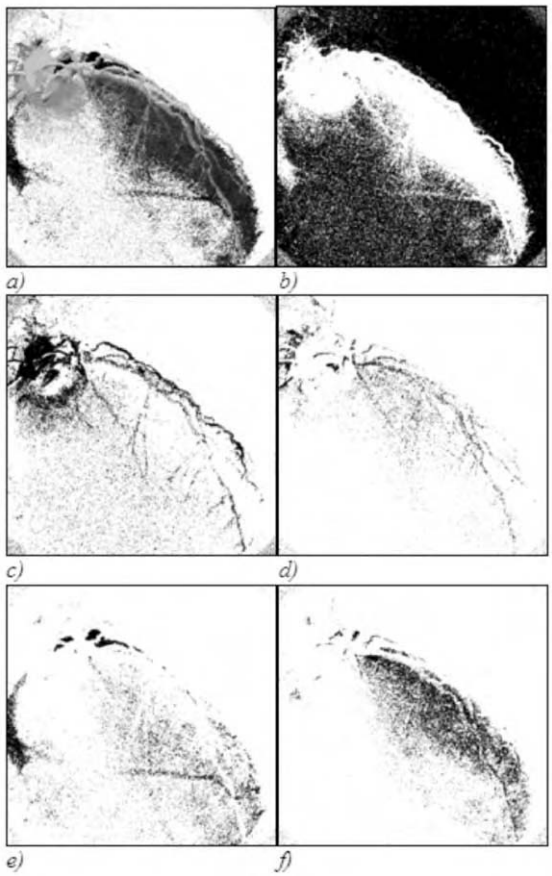


Figure 3 - Color maps for characterizing of specific structures during blood perfusion. The temporal incidence of the maximal dye intensity is mapped on to the pixel matrix. For different instances of time, corresponding with map b) to f), we display different anatomic structures which are perfused at this time; a) shows all maps superimposed: background, artery (LAD), branching, myocardium, vein; b) - f) the separated structures. Map f) is of particular interest because it represents the pixels of the myocardium.

We defined 4 different parameters for each intensity profile which are displayed in Figure 4: the maximal intensity Gmax; the time of Gmax; the maximal slope Imax of the rising profile, which corresponds with the maximal rising velocity of the perfusion; and the maximal slope Dmax of the falling profile, which corresponds with the maximal outwash velocity. For all calculated blush profiles, these four parameters are extracted as a typical characterizing feature set.

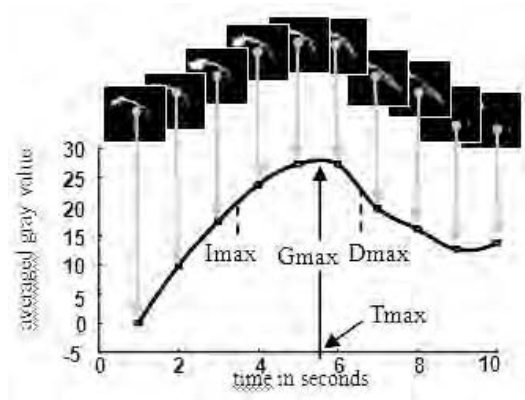


Figure 4 - The time course of averaged gray values from corresponding areas of the perfused myocardium. 4 specific parameters: Imax: Increase-slope, Dmax: Decrease-slope, Gmax: maximal gray-value, Tmax: time of Gmax are indicated

Results

For all 100 films the corresponding time courses are established and automatically characterized by Gmax, Tmax, Dmax and Imax. It is of particular interest to demonstrate the difference of blush spreading before and after a PTCA intervention by different temporal patterns.

Different degrees of stenosis are related with different extensions of blush and different patterns of dynamic profiles. The patient of Figure 5 had an occlusion of RCA. After PTCA the artery looks absolutely normal (Figure 5, lower trace). By positioning of a ROI in the supply area of this vessel we calculate the corresponding blush profile with its characterizing parameters. However, for the same region before PTCA we found a very flat profile (Figure 5 upper trace), which shows us that there is not any perfusion of the myocardium. Nevertheless, the reperfusion after PTCA is obvious. In this case we have consistent results with the re-opened artery and the blush profile. However, sometimes the situation demonstrated on the angiograms is not so clear. For these cases we get much more information by the quantified blush profiles.

In order to relate the various blush profiles to blush grades a non-parametric classification procedure based on the four specific parameters is under development. First promising results are in good agreement with the subjective grading of clinical experts.

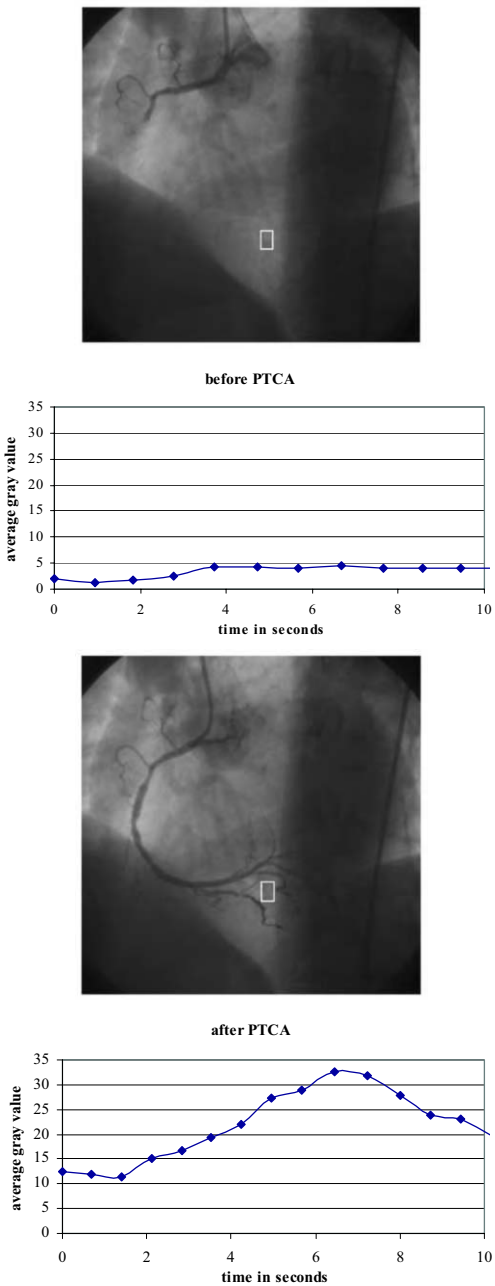


Figure 5 - Angiograms before and after PTCA of a patient with a complete stenosis of 100% for RCA. The corresponding temporal blush profiles document obviously the improvement of myocardial reperfusion

From the 100 digitized angiographic film sequences we could evaluate only 60 without considerable motion compensation. The introduced two step registration method allowed an additional evaluation of 30 more series. The remaining 10 films show such considerable motion artefacts of the diaphragm that an evaluation was not possible. In all of these cases the diaphragm moved over the whole hearth and modified its grayvalues significantly. Figure 6 shows the effect of motion compensation.

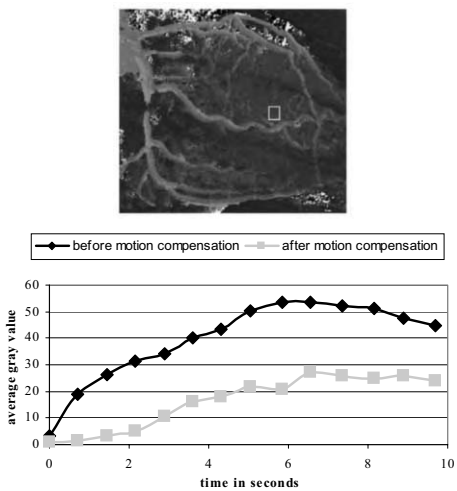


Figure 6 - The left image shows the extent of usually recorded motion artefacts whereas the right image demonstrates the benefit of motion compensation. The time courses before and after motion compensation are plotted in the same diagram below

Discussion

Our presented approach on the basis of digital subtracted angiographic images is an interactive procedure which helps the cardiologist to assign blush grades in strong relation to what he is used to do. He gets an intuitive impression of the spreading blood after the dye injection by displaying the processed colormaps. It is easy to confirm a quantitative hypothesis by looking precisely to selected areas of interest of the myocardium, and its temporal gray value profile or to integrate the average dye activity of the whole area.

The advantages of the developed system are given in a highly adaptable degree of automation. Dependent on the quality of angiograms, the more or less experienced user can interfere in different kinds. For example, simple adjustment of the contrast or corrections of motion artefacts lead to an improved quality of the digitally subtracted image sequence. The automated classification procedure is also easily adaptable to other types of blush patterns. Up to now it was our aim to implement the descriptions for blush grading of experienced cardiologists as it was proposed by Gibson. However, if we wish to characterize any other specifity of the perfusion, we can as well introduce more quantitative features related to the blush patterns, for example temporal measures. Furthermore, other approaches of motion compensation could be incorporated and evaluated for their practicability in respect of typical

angiographic artefacts. Of course, we have to look for more sequences to incorporate them in our learning sample. Therefore, we developed an intuitive user interface which helps the cardiologist to follow the different steps of the acquisition procedure and which offers him the various facilities to derive a reliable diagnostic statement. During the current evaluation of our system we are discussing the usefulness of a more automated selection of the ROI for blush profile quantification. However, up to now the experienced cardiologists are not convinced that this feature increases the clinical benefit significantly.

References

[1] Bibra H. Assessment of myocardial perfusion - role of contrast echocardiography. Z. Kardiol 2001: 90: pp. 848-859

[2] Stone GW, Lansky AJ, Mehran R, et al. Beyond TIMI 3 Flow: The Importance of Restored Myocardial Perfusion for Survival in high risk Patients undergoing primary or rescue PTCA. J Am Coll Cardiol 2000: 35: 403A

[3] Gibson CM, et al. Relationship of TIMI Myocardial Perfusion Grade to Mortality after Administration of Thrombolytic Drugs. Circulation, 2000: 101(2): pp. 125-130

[4] Modersitzki J. Numerical Methods for Image Registration. Oxford University Press, 2004

Address for correspondence

Hartmut.Dickhaus@med.uni-heidelberg.de
Prof. Dr. H. Dickhaus
Institute for Medical Biometry and Informatics
Department of Medical Informatics
University of Heidelberg
Im Neuenheimer Feld 400
D-69120 Heidelberg Germany
Tel.: 0049 6221 567483
Fax: 0049 6221 564997

This page intentionally left blank

Chapter 11.

Education and Training

This page intentionally left blank

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Enabling the Safe and Effective Implementation of Health Informatics Systems – Validating and Rolling Out the ECDL/ICDL Health Supplement

Michael J. Rigby[a], Carol Hulm[b], Don Detmer[c], Luca Buccoliero[d]

[a] *Centre for Health Planning and Management, Keele University, United Kingdom*
[b] *British Computer Society, Swindon, United Kingdom*
[c] *American Medical Informatics Association, Bethesda, MD, United States of America*
[d] *SDA Bocconi and CeRGAS Centre for Research on Health Care Management, Bocconi University, Milan, Italy*

Abstract

Sound understanding by end-users (health professionals and support staff) of key e-health principles and challenges is essential for the safe, effective, and sustainable use of health informatics systems. This is separate from, and ideally a precursor to, training on a specific system. However, hitherto this aspect has been little understood or addressed. Over the last few years, the concept of a customised Health Supplement to the well-established European/International Computer Driving Licence has progressed from idea to reality, through initial support for consultation by the then UK NHS Information Authority, followed by establishment of an international Expert Group by the global ECDL/ICDL Foundation. As a result, the ECDL/ICDL Foundation has developed a formal syllabus. This paper reports successful piloting, and progress in the development of local training and testing materials and national implementation plans, in three countries on two continents.

Keywords:

health informatics education; end-user; ECDL; ICDL; health informatics sustainability

Introduction

As defined by the theme for Medinfo 2007, there is rapidly growing recognition and dependence in all countries on the future major role of health informatics systems in enabling national health systems to be increasingly effective, efficient, and evidence based. Whilst each country starts from a different base line, all are moving forward at their local pace in increasing the use of informatics systems.

Rightly, the priority is increasingly for sustainability, which must include the reliable and efficient use of such systems, in line with intention and with capability. However, the fundamental oversight is to focus solely on systems design and implementation, important though they are. What is so often overlooked is the fact that systems are only as good as their users' understanding and capacity to use them appropriately.

Health informatics systems are a vehicle for collecting, processing, and making available relevant health information at the point of need. Therefore such support systems should not dominate or hamper health practitioners' activities except so far as they can make them more effective or efficient. However, all too often end users – health practitioners and their support staff – can find such systems threatening and restricting, and in turn this means that systems are used inefficiently and ineffectively, and either perform badly or fall into disuse.

Resistant mature staff

Though often a surprise to many, this picture of end-user poor compliance is well documented over a period of time, and thus should be anticipated [1]. Health professionals have in round terms a professional lifespan of approximately forty years. In nearly every country those practicing for more than ten years will have undertaken their basic professional education and early practice in an environment based on paper-based records. In many countries of the world electronic systems will only have been brought in even more recently. Furthermore, the most senior and experienced health practitioners – usually the professional champions and leaders – will certainly have been educated ahead of the e-health revolution. It is into this workforce environment that governments, health policy makers, and health informatics system advocates are seeking to introduce radical and comprehensive electronic systems.

An atypical IT system user population

A further complication is that not only are health information system end users more well versed in paper-based systems, in many other ways too they differ from the end users of almost all other organisation-wide modern electronics records systems, such as those in banking or the travel industry [2,3]. Not only are health users senior and mature staff rather than new entrants; they are totally dependent on information systems in order to carry out their principal daily business. However, this core work is not itself focussed on the information system - in general the use of the information system only forms between 10% and 20% of their duties - and to them this is a low-skill element of their work. Thus senior doctors, nurses, and other

Selected for best paper award.

health professionals will not give the same level of importance to being trained in new-generation information systems as they will to new clinical or healthcare techniques. Yet, even though being only part-time users, they are senior and often autonomous employees, and they have professional accountability over and above their duty to their immediate employer. Thus they will only be effective, reliable and regular users if they both understand the way to harness a system, and trust in its integrity and that of all partner users (who are largely unknown to them individually). The deeper analysis undertaken of these differences of health informatics system end user profiles compared to others system user profiles has been published [2, 3].

Change-inducing systems

Added to these challenges, health informatics systems in themselves do not support clinical practice in the simplistic ways that telephones support and replace meetings and written correspondence, or dictation systems replace the need for the physical presence of a secretary. Instead, health information systems require a very different pattern of working, ranging from new data recording processes, through to how to search a past record (which is navigated very differently if it is held in computer files compared with if it is a large collection of paper charts in a physical folder).

Informatics system sustainability but practice challenge

Therefore sustainability is highly dependent as much upon the pattern of use as it is on the pattern of design, yet is affected also by the user population's characteristics. Thus it is clearly naive to expect the complete cohort of the most senior and experienced practitioners in their country universally to welcome and endorse new systems which require radically different physical and cognitive skills, and which require immediate changes to patterns of data recording and assimilation honed over a life time. Thus, however good the system introduced, it will not be sustainable if the principal stakeholders are not comfortable with issues.

Counteracting the principal risks of health information systems

Apart from recognising the challenge to traditional practice of new health informatics systems, it is also important to recognise the inherent risks these bring if introduced without due preparation. Moreover, professionals will be aware in general terms of these risks, and be likely to militate against introduction of such systems with good intent unless they feel that these risks have been addressed and controlled. There are three types of such risk, as shown below.

Risks generated by the need for new skills

The use of a health informatics system requires a radical range of new skills. These commence with the basics of operating any computer system, through the skills required to record data electronically as apposed to by hand writing or filing a chart, to the skills needed to navigate a record which is stored in a highly structured and efficient way but which needs a new mind set in order to negotiate it effectively to find key and relevant information items. It would be unreasonable to expect a surgeon to use a new type of instrument or a radically new surgical procedure without adequate training, yet governments and policy makers worldwide are inappropriately labelling as "obstructive" health practitioners who are reluctant to change information management approaches with which they are familiar, for ones which they find unknown and intimidating. Further, 'smart' systems may make good evidence-based calculations, recommend particular treatment patterns or warn against particular prescribing intentions. These are safe provided the end user understands the rationale in both the clinical and computing logic inbuilt, but carry risks if the end user does not understand and know how to ascertain that logic.

The risks of new constraints

A key aspect of most health informatics systems is the fact that they require a standard approach to the description of histories, investigations, results, diagnoses, and interventions - in other words, the benefits of standard terminologies and taxonomies should reduce ambiguity and render observations and findings interpretable accurately by all. However, the converse of that is that an individual's well-developed means of indicating valuable subjective information such as uncertainty, provisional views, or feelings as apposed to hard evidence are rendered impossible. This may either exclude uncertain information, or result in it being recorded with a spurious impression of certainty. Users may know the clinical approach and specific skills of colleagues who hand-record narrative information, but this authorship and personalisation may be lost with electronically captured and stored data.

Risks of misuse

The very strength of health informatic systems – that they can search and present information from very large databases extremely quickly – is a potential risk as well as being a core purpose. Files can become increasingly comprehensive, and information can be obtained about many people. Careless use of such information can lead to inappropriate divulgence of confidential information, and without safeguards there are clearly risks of an ethical or malicious misuse. Thus all end users need to be educated to avoid these risks, and to adhere to robust corporate policies to control usage.

For all these reasons it is therefore important for end users to be adequately educated as to how to use systems soundly and effectively. This is a key part of the sustainability of systems. Evidence (or even suspicion) of misuse of systems, or of poor clinical decision making because of inability to use systems, will provide a rapid means of ensuring their demise. Such evidence or suspicions may come either through professional sources, press reports, or collective patient anxieties.

The move to a health end-user qualification

The need to recognise end users

The educational needs of end users are very different from those of technical health informatics staff. Over a decade ago a European Commission Concerted Action entitled EDUCTRA identified the informatics educational needs of health professionals as being different from those of IT staff [4]. This work suggested a range of necessary learning outcomes for each group – though for health professional users they focussed primarily on basic curricula for new entrants. Subsequently the International Medical Informatics Association (IMIA) addressed this topic, and produced recommendations of what should form the basis of health informatics education globally for each of the two staff communities. Both before and since these recommendations, the prime focus has been on informatics education, with development of many formal courses, and on the introduction of some informatics training into basic health professional education. Neither of these groups, however, form the general body of the practice community to whom new organisation-wide informatics systems are introduced or imposed.

The concept of an end user qualification

To overcome these anxieties and risks it seems self-evident that an appropriate end-user educational programme, and related qualification, for health informatic systems users should be developed, but this was not being addressed. Meanwhile, virtually all countries in the world have a qualification requirement for drivers of motor vehicles or pilots of aeroplanes, as such equipment is seen as extremely beneficial yet extremely risky if misused through ignorance or lack of skills. It is seen as a societal responsibility to provide a qualifications framework and regulation, and a citizen responsibility to ensure qualification before becoming a user.

This approach has already been taken with the more general use of computers, with the development of the European Computer Driving Licence (ECDL) to a standard international curriculum, as most of the issues of using computers safely and effectively are generic and universal. This has now developed into the global International Computer Driving License (ICDL), available in virtually all countries of the world [6].

It therefore seemed logical to develop a specific supplement or module for the ECDL/ICDL, given the risks and responsibilities inherent in using such systems. This concept was first promoted in 1999 in a European context [7]. Subsequently, the idea was developed at conceptual level in more detail, and support gradually developed [8]. Of significance for Medinfo 2007, details were sought by a principal Australian health informatics journal [9].

Practical steps to development

Following these moves towards the development a health supplement to the ECDL/ICDL as the best means of meeting this need, and thereby ensuring sustainable and safe implementation of health informatics systems through education, assurance, and empowerment of end users, many practical steps have been made towards achieving this reality.

In 2004 the NHS Information Authority, the then lead body in this field for the National Health Service in England, agreed to support two consultation workshops – one for key opinion leaders in health informatics from eight European countries, and one for a range of delegates from the National Health Service across the United Kingdom. As a result of the strong enthusiasm at both these meetings, the European Computer Driving Licence Foundation (the global regulatory and licensing body for the ECDL/ICDL) agreed to consider formalising the development process. The ECDL Board endorsed this, and in 2005 an Expert Group was set up comprising representatives of six European countries and of the United States of America. The resultant recommended syllabus was signed off by the Expert Group in early 2006.

The ECDL/ICDL health supplement content

The final ECDL/ICDL Health Supplement consists of a competencies framework defining knowledge and skills the candidate needs to possess in order to operate a health information system safely. It excludes generic issues covered in health professional training or staff induction (such as basic principles of confidentiality). Regarding computer recording, it focuses on those aspects which are different, or have different emphasis or importance, in health applications.

The core contents of the syllabus are copyright to the ECDL/ICDL Foundation, and comprise the following topics:

- Concepts
  - Health Information Systems
  - HIS Types
- Due Care
  - Confidentiality
  - Access Control
  - Security
- User Skills
  - Navigation
  - Decision Support
  - Output Reports
- Policy and Procedure

For each topic a number of defined knowledge areas or competencies are specified. The content is designed to accommodate specific national language and terminologies, organisations, and legal and professional frameworks. The normal pattern of assessment will be electronic, through a testing framework available on line or by other electronic means. It is based on the assumption that the candidate will already be competent in basic computer user skills.

International trials and validation

Since the specification phase, rapid progress has been made in significantly different countries, with very differ-

ent health systems, different languages, and also different terminology and nomenclature within the same language group.

United Kingdom

In the United Kingdom, the British Computer Society as national licensee for the ECDL, and with a strong relationship with the National Health Service, organised piloting of the syllabus utilising an interim training manual and testing framework in six very different sites. These encompassed very different localities, and different healthcare environments ranging from primary care through secondary care to mental health, and different health professions from research staff to medical consultants, and also health informatics and health data experts. This pilot involved 84 persons, who were all very positive on the value of the knowledge and competencies covered in the syllabus. The only significant comments received were about the interim testing framework, which was only ever intended to be temporary in order to facilitate consideration of the syllabus. More detailed reporting of these results is in press [10].

Consequent upon these successful pilots, a full electronic training resource has been developed from the interim one, and a definitive electronic testing framework built. The ECDL Health Supplement was launched to English NHS staff by NHS Connecting for Health in April 2007, with the on-line tutorial and testing available to staff free of charge.

United States of America

The American Medical Informatics Association (AMIA) and the national ICDL licensee, ICDL-US, worked closely during 2006 to create a US-version curriculum for 'anyone in a health-related entity who touches a keyboard containing person-specific health information' as well as an examination to certify mastery of this content. This will be entitled the Digital Patient Record Certification or DPRC. The curriculum group adapted the syllabus developed by the ECDL Foundation group; the US version was then reviewed by the ECDL group.

The test will be piloted in early 2007 with the expectation that the program will be functional late in 2007. ICDL-US and AMIA found the partnership to be mutually beneficial and there is a desire to work together on other products for the North American region.

AMIA has two major educational initiatives underway and this initiative is part of its "Got EHR?" campaign. AMIA also seeks to educate the general public about electronic health records and especially integrated personal health records, particularly as an integral part of the electronic medical record. The campaign also strives to increase the use of EHRs in small practice environments. The second initiative is the "10x10" Program, which is an effort to educate 10,000 applied clinical informaticians by 2010. This program now involves two universities and will involve at least five by the end of 2007.

Italy

The Italian Association for Computing and Automated Calculation (AICA), the Italian national ECDL licensee [11], instructed CERGAS Bocconi to work out the Italian health syllabus, based on the core syllabus, and the test structure. The related "ECDL Health Manual" has also been developed and printed. Between January and March 2007, Italy has implemented two pilot editions of the course, delivered to about 60 medical doctors and nurses of the Local Healthcare Units of Milan and Dolo (Venice). Participants have been offered four courses of ECDL Start (24 teaching hours), plus a specific course of ECDL Health (8 teaching hours). At the end of the pilot courses, in May 2007, final examinations will be held and skill cards issued accordingly (including, for the first time ever in Italy, those relating to ECDL Health). The examination to obtain the ECDL Health skill card will include practical exercises simulating the use of patient records management software. The courses will be included in national and regional programmes aimed at the continuous education of the NHS medical personnel and will enable participating medical doctors to obtain some compulsory education credits. These courses are expected to become a key element of the education and training programmes nationwide.

The Italian experience stands out for its special focus on the preliminary planning of the initiative, also being based on a scientific research project conducted in 2004 aimed at measuring the potential benefits of information education and training in the healthcare sector [12]. This research project has analysed and evaluated the "cost of IT ignorance" in the Italian healthcare sector through a sample survey, empirical measurement tests, and experience of a similar research project conducted on private businesses. Ignorance in the information field has proved to be a notable hidden cost for the Italian healthcare sector and the potential value of information education and training of the NHS personnel amounts to about 2 billion Euros per year. After being published and officially presented, the results enhanced the institutional awareness regarding development of targeted educational programmes.

Discussion

Development of any new qualification takes a considerable period of time, commencing with identification of the need. This initiative has sought to achieve this in a way that matches the differing needs of countries globally, and to pilot and validate it in differing countries in two continents. Though the piloting, and the development of the educational and testing frameworks and systems, was grounded on the needs of the individual countries involved, the resultant tools and products are likely to yield wider benefits and use by other countries wishing to follow suit.

As well as being a key contribution towards sustainable health informatics systems, this initiative also marks a policy development for the ECDL/ICDL Foundation. It began the development of the Health module as a global

concept with the help of a group of subject matter experts. However, it became clear that the application of Health informatics systems required a certification that was closely tailored to national requirements including patterns of practice, culture, language, and legislative frameworks. As a result the ECDL Foundation adopted a modem of product endorsement. The core syllabus for the Health supplement is specified by the Foundation, but individual national license holders propose the format in which the national certification is developed and assessed within a specific country, ensuring this has local relevance. Each national certification is then endorsed by the ECDL Foundation. The Health certification is thus tailored to each individual country's health system. It not only uses the languages of the country, but the terminologies and taxonomies for health care practice, as well as professional and legal codes. The ECDL/ICDL Health Supplement is the first successful implementation of this Endorsed Product concept.

Conclusion

With the increasing importance of health informatics systems, and the need to ensure their effective and safe usage, there has been a steady and increasing recognition of the importance of end user competence as a contribution to effective sustainable implementation and development of such systems. Moreover, given that the issues and risks are basically generic, coupled with increasing mobility of health staff, the advantage of devising a generic solution has become self evident. From this position, the progress in the last two years through an expert committee identifying and confirming the core syllabus, and three different countries undertaking trials and detailed implementation plans, is significant and encouraging.

Like motorcars, aeroplanes, or other items of advanced technology, health informatics systems are only as good as the competence (and confidence) of their users. Hitherto this has gone largely unnoticed, except for possible training in a particular system's operational instructions. The ECDL/ICDL Health Supplement has broken new ground, by recognising the high importance of the education and empowerment of the end user, whatever their level or profession. It thus makes a vital contribution to the sustainability of health informatics systems.

Moreover, this is a global solution, linking common generic requirements with local need through the ECDL Endorsed Product concept. Having been developed by an international expert group, endorsed by the ECDL/ICDL Foundation, and now validated simultaneously in three very different countries, this product is now available for use in any nation.

References

[1] Anderson JG, Aydin CE, Jay SJ (eds.). Evaluating Health Care Information Systems: Methods and Applications; Sage, Thousand Oaks CA, 1994.

[2] Rigby M. Improving Governance and Reducing Risk in Electronic Patient Record Systems: Ensuring Appropriate Competencies for Support and End-User Staff. In M Fieschi, E Coiera, Y-C J Li (editors). Medinfo 2004 - Proceedings of the 11th. World Congress on Medical Informatics, IOS Press, Amsterdam, ISBN 1 58603 444 8, 2004, 926-930.

[3] Rigby M. Protecting the Patient by Promoting End-user Competence in Health Informatics Systems – Moves towards a Generic Health Computer User "Driving License"; International Journal of Medical Informatics, 73, 151-156, 2004.

[4] Staff of EDUCTRA Concerted Action. Education and Training in Health Informatics: Guidelines for European Curricula; in Hasman A *et al.* Education and Training in Health Informatics in Europe; IOS Press, Amsterdam, 1995.

[5] International Medical Informatics Association. Recommendations of the International Medical Informatics Association (IMIA) on Education on Health and Medical Informatics; Methods of Information in Medicine, 39, 267-277, 2000. (also available on www.imia.org/wg1/rec.pdf)

[6] www.ecdl.com/main/index.php

[7] Rigby M, Draper R, Sheridan A. Confidentiality, Competence, and Confidence - Ensuring Ethics whilst Recognising Reality; in Proceedings, EuroRec '99 Third European Conference on Electronic Health Care Records, 1999, Sevilla, España; Sadiel, Sevilla, 1999.

[8] Rigby M. Protecting the Patient by Ensuring End-User Competence in Health Informatics Systems – Moves Towards a Generic Health Computer User "Driving Licence"; in Proceedings of IMIA Working Group on Education Conference, Portland, Oregon, 2003.

[9] Rigby M. The Health IT roll out: supporting the end users; Health Information Management, 33, 3, 93-94; 2004.

[10] Rigby M, Hulm C. Piloting in England of a Health Supplement for the European Computer Driving License (paper presented to HC 2007 and in press).

[11] www.aicanet.it

[12] Buccoliero L, Caccia C. Un approccio aziendale ai sistemi informativi in sanità. [A corporate approach to information systems in the healthcare sector.] In Borgonovi E, Camussone PF, Occhini G, (Eds.) L'Ignoranza informatica: il costo nella sanità. [IT ignorance: how much it weighs on healthcare] Milan, McGraw-Hill., 2004: 31-69.

Address of correspondence

Professor Michael Rigby, Professor of Health Information Strategy,
Centre for Health Planning and Management, Darwin Building,
Keele University, Newcastle-under-Lyme,
Staffordshire, ST5 5BG, UK.
Telephone +44 1782 583191, Fax +44 1782 711737,
email: m.j.rigby@hmp.keele.ac.uk

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

A Multi-method Approach to Assessing Health Information Systems End Users' Training Needs

Yiyu Qiu[a], Ping Yu[b] Dr. Peter Hyland[c]

[a,b] *Health Informatics Research Centre, University of Wollongong, Australia*
[c] *School of Information and Technology and Computer Science, University of Wollongong, Australia*

Abstract

Many existing training programs for Health Information Systems (HIS) do not adequately satisfy end users' needs nor meet training objectives. This is because they do not envisage the problems that users may encounter when performing specific tasks. Therefore the first priority for the development of an effective training program is to precisely assess the end users' training needs, a process called Training Needs Assessment (TNA). Applying traditional approaches for TNA, such as interviews or surveys alone, however, may be insufficient because they are limited in their capacity to reveal the cognitive processes of end users. Usability testing, with its ability to gather data about human computer interaction, overcomes the deficiencies of these traditional approaches. This paper proposes a multi-method approach, which combines usability testing together with traditional methods, such as interviews or questionnaire surveys to assess HIS end users' training needs. This innovative method is able to precisely reveal the training needs for different levels of HIS users. A case study, which applied this method to assess the training needs for users of a nursing information system demonstrates its feasibility.

Keywords:

health information systems, end users, training needs assessment, usability testing

Introduction

A well designed Health Information System (HIS) can not only save staff time in entering and retrieving client data, but can also increase the accuracy and completeness of such data. The adoption of HISs, however, is not common in the current Australian healthcare sector [1, 2]. One of the reasons is that shifting from the traditional paper-based documentation to electronic documentation requires the users of an HIS to have basic computer skills, to be familiar with the HIS, and to change their practices of information management [3]. Managers are more and more aware that end user training is an essential strategy to accomplish this [4, 5] and that the failure to implement appropriate end user training strategy will leave staff feeling frustrated and threatened by the new system [6]. In the worst case, they may even reject the new system or resign, which in turn will lead to loss of organizational resources like skilled healthcare workers. It is important, therefore, that significant investment be made into end user training and support in order to ensure that the introduced HIS will be accepted and used by the healthcare workers.

An effective end user training program should have the capacity to deliver a timely, effective, efficient and enjoyable learning experience to the end users [4, 7]. In other words, it plays the role of closing the gap between the complexity of an HIS and the users' cognitive capacity to master it. The majority of existing training programs, however, are not as effective as they promise [8, 9]. The most common problems of these training programs are that they are ill-directed and inadequately focused [10]. For example, some training programs provide healthcare professionals with huge amounts of unnecessary information because they have been developed as a "one size fits all" solution [8, 9]. The fundamental flaw is that training needs assessment, namely, the process of assessing the training objectives [11], is not properly conducted for these programs so the training program designer can not accurately envisage the problems end users may encounter when performing specific tasks using the HIS. Thorough TNA is required to improve end users' learning outcomes and to enable them to become familiar with a new HIS efficiently.

In order to conduct a thorough and accurate TNA, the method of analysis should be scientifically designed, and this is the topic to be discussed in this paper. Firstly this paper will critique the traditional methods for TNA. Afterwards it will propose a novel method for TNA, followed by a case study demonstrating how to integrate the new approach with traditional methods.

Need for new approach in training needs assessment

The primary purpose of a TNA is to identify what knowledge and skills end users should have in order to enable them to effectively interact with an HIS. Through identifying the usage problems, the gap between the necessary and the actual knowledge and skills that a user has for effectively interacting with this IS can be inferred [12]. In other words, experimental Human Computer Interaction (HCI) data such as usage problems, mistakes or inefficient behaviors, are effective indicators of what they do not

know [13], which, in turn, suggests what they need to know. Traditional methods of TNA, however, lack the ability to collect such detailed information about the cognitive process end users follow in their interaction with a new IS.

The common methods of assessing computer users' training needs are self-reported questionnaire surveys and interviews with end-users [13-15]. Questionnaire surveys have a number of distinct advantages, including the ease of distributing questionnaires to a large number of users and the automated analysis of the results with statistical packages. The typical process of quantitative assessment consists of managers setting the required level of skills for a particular task, then a staff member is requested to rate his/her skill level against this standard [16]. A comparison between these two sets of data suggests this staff member's skill gap in accomplishing the task, however, this method can only identify difficulties of which the designers or skilled external consultants are already aware [15]. It can not detect all the challenges or mistakes that a user may face or make in using a particular HIS.

Similar problems underlie other commonly used methods like interviews or focus group discussion, where end users are asked to reflect on their prior experience with an IS. Such qualitative assessments provide opportunities for users to express their perceived difficulties in using an IS but these conventional methods are not adequate to assess the learning needs of users, particularly for users with different levels of experience with the system. For example, novice users may not have sufficient knowledge about this new IS to enable them to realize the problems they may encounter in using it [17, 18]. Even experienced users may not be able to clearly recall their problems [19]. Verbalizing the process that a person follows to complete a task is also problematic, as it involves the expression of sequences of psychomotor movement in interaction with an HIS [20]. In addition, users' perceptions of the same problem may be different because of differences in their educational or technical levels [21].

Thus the use of traditional TNA methods alone may not reveal detailed information about the cognitive process end users follow in their interaction with a new IS [21, 22] so data gathered from these methods are not adequate to identify end users' training needs. In other words, basing the selection of training strategy on the analysis of such incomplete data may lead to ineffective programs because there is a significant gap between what is perceived to be a problem and the actual problems that a user encounters. Traditional methods for TNA need to be complemented by more effective new approaches. Careful observation of how users encounter and react to problems in interacting with a particular HIS is necessary. The next section will explain our proposed methods and suggest strategies for dealing with the problems discussed above.

Cognitive usability testing method

There is a growing role for the cognitive and behavioral sciences in health informatics, particularly as it pertains to human factors, and other areas such as information retrieval and educational research [22, 23]. From the perspective of informatics, cognitive science can provide a framework for the analysis and modeling of complex human performance in IS [23]. Theories and methods from the cognitive sciences can illuminate different facets of the design and implementation of IS [22, 23]. They can also play an important role in understanding and enhancing human performance on a wide range of tasks. These tasks may include developing training programs to reduce errors and increase efficiency for healthcare [23].

Usability testing encompasses a range of methods for identifying how users actually interact with a complete software product. Empirical testing is a form of usability inspection that observes actual user interactions with an interface (Neilson, 1994). Given that many HISs fail due to usability problems, organizations are starting to show interest in usability testing. Some preliminary studies have been performed testing clinical information systems. For example, Kushniruk [24] introduces a laboratory-based usability testing method to evaluate the effectiveness of HISs. According to Kushniruk [24], usability testing refers to "the evaluation of an IS that involves testing of participants who are *representative* of the target user population, as they perform *representative* tasks using an IS". During the testing, all interactions a user has with an IS are recorded (e.g. video recordings made of all computer screens or user activities and actions). In addition, this technique generally includes the collection of "think aloud" reports, involving the recording of users as they verbalize their thoughts while performing particular tasks [24].

In brief, this approach focuses on classifying users' cognitive ability and then identifying the problems they encounter during their interaction with the IS. With its ability to gather rich empirical HCI data, this method provides an excellent opportunity to complement the weaknesses of interviews or questionnaires for assessing training needs. Although this method was originally designed for testing the usability of an IS, it is also useful for identifying training needs. In the early '90s, Simpson proposed a framework to describe how testing methods could be used in the planning phase of designing online documents [25]. A recent case study has used novice users' interaction with a search engine to reveal the knowledge and skills that such users need [18]. Our preliminary study which employed the usability testing method to assess the training needs of nursing students also demonstrates the feasibility of this method [26].

A multi-method approach for training needs assessment

We have demonstrated the viability of cognitive usability testing methods in capturing process data on how an end user interacts with a particular IS. We propose that a multi-method approach combining usability testing with conventional methods like interview or questionnaires can precisely and thoroughly understand the process end users follow in processing information in a particular HIS. It can also suggest the knowledge and skills that these users need to learn in order to use the HIS.

The process of our proposed approach is to:

- observe how novice, intermediate and veteran users use an HIS to complete representative tasks.
- interview or survey these users before or after conducting cognitive usability testing, in order to ascertain their level of the knowledge and skills that are relevant to the HIS that they are trained to use
- extract patterns of strategies used by the novice, intermediate and veteran users to complete various tasks with this HIS.
- identify the key knowledge gaps (learning needs) of different levels of learners based on the different interaction patterns that they display when using the HIS.

In the following sections we describe three approaches which can be utilized for an innovative TNA.

Usability testing approach

The following three issues need to be considered when performing usability testing.

Participants

Participants for the experiment are potential or actual end users. Based on their experiences with the HIS (e.g. measured by the time and frequency of their usage of the system), they can be categorized as novice, intermediate, and experienced users. The recommended sample size varies between usability experts. As Nielsen and Mack suggest [27], usability testing can be carried out using only five participants, and the results will demonstrate 85% of the usability problems. Kushniruk *et al.* [21] suggest that up to 80% of usage problems can be detected from 8 to12 participants evaluating a system.

Outcome measurement

Kushniruk *et al.* [21] suggest that the usability testing should involve setting up recording equipment that allows for continuous recording of computer screens during the process of human-computer interaction. To achieve this goal, Camtasia Studio, screen recording software, can be used to record each participant's mouse movement and keyboard strokes. In addition, participants' "think aloud" reports can be audio-taped. The data analyzed for usability analysis included both the video and the audio file.

Data analysis technique

Prior to analyzing the video data, a coding scheme should be defined for use in identifying specific occurrences of users' problems and aspects of cognitive processes from transcripts of a participant's "think aloud" comments. A coding taxonomy developed by Kushniruk can be used for analyzing human computer interaction data [22, 24].

- Navigation: used when participants comment that they are navigating, or indicate that they are incapable of moving through the interface to find the relevant information or accomplish what they are supposed to do.
- Understanding: used when participants comment on understanding the meaning of labels, instructions or errors.
- Ease of use: used when participants comment on the level of "ease of use" of the system (from easy to hard) or any confusion or frustration experienced.

A case study

The NIS described in this case was the Care Planning Assessment Tool (CPAT), owned by the Hammond Care Group. The CPAT was introduced to help nurses carry out systematic assessments for nursing home residents. The "Clients" menu enables users to perform the most crucial functions of the CPAT, i.e., doing data entry and assessments for clients (i.e. residents), see Fig 1 for a screen shot of the assessment screen. In the assessment screen, users can enter detailed assessment results for a resident. The program can then generate various assessment related reports.

This particular study aimed to assess the training needs for different levels of the CPAT users, so as to develop the "right" training materials for the "right" groups of users. The training materials for the CPAT are:

- User Manual: this is usually for novice or first time users and should be very detailed.
- Online Help: this type of training is traditionally for the relatively experienced users who require help while using the product; it usually contains information on how to conduct a task.

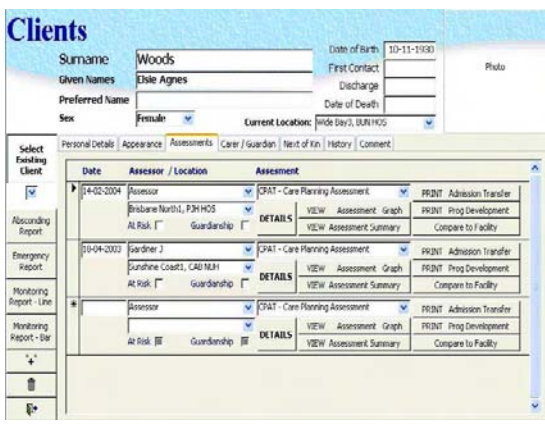


Fig 1 - The assessment screen of the CPAT

Procedure

Based on users' knowledge and experience with the CPAT, they are classified into two categories: novice and experienced users. The process of the training needs assessment in this case study was to:

- conduct laboratory-based usability testing to observe how novice CPAT users complete a series of data entry tasks using the CPAT;
- interview a cohort of experienced CPAT users to understand the problems they often encounter in using the CPAT software
- analyze two sets of data to identify knowledge gaps;

- integrate the findings into the design of training materials.

Usability testing experiment

Eight participants for the experiment were recruited from third-year nursing undergraduate students from the University of Wollongong. All the participants majored in geriatrics. They were potential users of the CPAT but had no previous usage experience with the software.

The participants were asked to perform the following three representative tasks supported by the software:

- entering data for a resident;
- doing an assessment for a resident;
- generating a change monitoring report.

The participants were encouraged to "think aloud" or verbalize their thoughts if they were uncertain about how to conduct the above documentation tasks using the software.

Findings

There were eight video and audio data collected from the experiment. Analyzing the triangulation of audio and video data identifies a series of problems that novice CPAT users encountered in this training session. An excerpt of a coded section of such triangulation is given below to show how users' interaction with the software was coded.

> 20:40 – user finished scoring "communication problems" and intended to answer the next group of assessment questions.
>
> "How should I go to the next group of questions?"
>
> *Navigation - having problems navigating between assessment criteria.*

By coding all of the participants' usage problems, three groups of problems that users encountered were identified:

- basic computer skills, e.g., users do not recognize the drop-down icon in the selection fields.
- knowledge about the software, e.g., users do not know where to score questions about assessment criteria
- domain knowledge about nursing documentation, e.g., users do not understand some assessment questions

Focus group discussion

A semi-structured focus group discussion involving nine veteran users from the Hammond Care Group was conducted to explore their learning and work experience with the CPAT. The participants were the actual users of the CPAT. Their roles included facility manager, trainer, consultant and dementia care worker. Most of them have more than three year's CPAT usage experience.

Participants were prompted to provide their answers for the following questions:

- How did you learn to use the CPAT?
- Could you recall any problems encountered when using the CPAT?
- What kind of help do you expect when you encounter problems?

Findings from the focus group discussion

The normal method of learning the CPAT was labor-intensive, one-on-one coaching, followed by self-directed practice. If the user encountered any problems, they could either approach the trainer or try to solve the problem by themselves. A list of frequent usage problems was identified through focus group discussion. It ranged from system-related problems to computer-related problems.

Discussion

The findings from the TNA identified both procedural and conceptual usage problems in the format of HCI (usability testing) and verbal expression (focus group), which provided valuable input into the design of both the user manual and online help for the CPAT software. For the user manual, usability testing vividly revealed three types of novice-user knowledge gaps, which would become the main focus of the user manual. In addition, frequently asked questions gathered from focus group discussion can be effective contents in the user manual as these problems are also common for novice users. For the online help, problems that were identified in usability testing and were summarized in focus group became the core contents of online documentations. In addition, feedback from experienced users suggested that the learners like functions like video demonstrations. Therefore, this new approach for TNA has proved its capacity to precisely and thoroughly identify the training objectives for both novice and experienced users.

Conclusion

In this paper we have firstly presented the strengths and weakness of a number of traditional methods (interviews and questionnaires) that are used for TNA for novice users of an HIS. Next we described the capacity of cognitive usability testing to capture the cognitive process of HIS users in their interaction with the system. We argue that using the conventional methods alone has limitations and that they could be complemented through combining cognitive usability testing with the conventional methods. This proposed new approach has been explained in detail, particularly through demonstration of a case study, which involved assessing the training needs for users of a nursing information system through two approaches: cognitive usability testing and focus group discussion.

Acknowledgement

We would like to acknowledge the contributions by:

All the participants in the laboratory and focus group.

Mr. Richard Fleming, owner of the CPAT application, for his support in providing introduction to the CPAT software for the participants of this experiment.

Dr. Patrick Crooks and Ms. Joanne Joyce and staff in Nursing school in University of Wollongong for encouraging and recruiting their students to participate in the experiment.

Dr. Madeleine Cincotta, for her generous help in editing and proof-reading the final manuscript.

References

[1] Gosbee J, Ritchie E. Human-Computer Interaction and Medical Software Development. Interactions. 1997;4(4): pg 13--8.

[2] Hunt EC, Sproat SB, Kitzmiller RR. The nursing informatics implementation guide. New York: Springer 2004.

[3] Yu P, Comensolia N. An exploration of the barriers to the adoption of information technology in Australia aged care industry. Health Informatics Conference 2004; 2004; Brisbane, Australia: HISA; 2004.

[4] Gjestland C, Van Slyke C, Collins R, Cheney P. End User Training: A Decade of Research; 1997.

[5] Huang AH. End-User Training Management Systems. Journal of Educational Technology Systems. 2002;31(1):19-33.

[6] Coiera E. Guide to health informatics: Arnold 2003.

[7] RadhaKanta M, Vincent SL. Evaluating end-user training programs. Association for Computing Machinery Communications of the ACM. 2005;48(1):n/a.

[8] Kidd MR, McPhee W. The 'lost generation': IT education for healthcare professionals. The Medical Journal of Australia. 1999;171(10):p510-1.

[9] Ward JPT, Gordon J, Field MJ, Lehmann HP. Communication and information technology in medical education. The Lancet. 2001;357(9258):p792.

[10] Washer P. Barriers to the use of web-based learning in nurse education. Nurse education today. 2001;21:455-60.

[11] Ostroff C, Ford JK. Assessing training needs: critical levels of analysis. In: Goldstein IL, ed. *Training and Development in Organizations*. San Francisco: Jossey-Bass 1989.

[12] Gupta S, Bostrom RP. End-user training methods: what we know, need to know. Proceedings of the 2006 ACM SIGMIS CPR conference on computer personnel research; 2006; Claremont, California, USA: ACM; 2006.

[13] Howard S. User interface design and HCI: identifying the training needs of practitioners. SIGCHI bulletin; 1995; 1995. p. 17-22.

[14] Nelson RR, Whitener EM, Philcox HH. The assessment of end-user training needs. Communications of the ACM. 1995;38(7):27-39.

[15] O'Brien E, Hall T. Training needs analysis: the first step in authoring e-learning content. 2004 ACM symposium on applied computing; 2004; Nicosia, Cyprus: ACM Press; 2004.

[16] Boydell T, Leary M. Identifying Training Needs. London: Institute of personnel and development 1998.

[17] Sein MK, Bostrom RP, Oflman L. Individual differences and conceptual models in training novice end-users. Human computer interaction. 1989;4(3):197-229.

[18] Lazonder AW. Exploring novice users' training needs in searching information on the WWW. Journal of computer assisted learning. 2000;16:326-35.

[19] Smith SM. Frustrated feelings of imminent recall. In: metcalfe J, Shimamura AP, eds. *Metacognition: knowing about knowing* 1994.

[20] Norton B. Selecting learning experiences to achieve curriculum outcomes. In: Billings DM, Halstead JA, eds. *Teaching in nursing : a guide for faculty*. St. Louis, Miss: Elsevier Saunders 2005.

[21] Kushniruk AW, Patel VL. Cognitive and usability engineering methods for the evaluation of clinical information systems. Journal of Biomedical Informatics. 2004;37:56-76.

[22] Kushniruk AW, Patel VL. Cognitive approaches to the evaluation of healthcare information systems. In: Anderson JG, Aydin CE, eds. *Evaluating the organizational impact of healthcare information system: second edition*: Springer 2005.

[23] Patel VL, Kaufman DR. Cognitive Science and Biomedical Informatics. In: Cimino EHSJJ, ed. *Biomedical Informatics: Computer Applications in Health Care and Biomedicine (Third Edition)*. New York: Springer-Verlag 2006.

[24] Kushniruk AW. Evaluation in the design of health information systems: application of approaches emerging from usability engineering. Computers in biology and medicine. 2002;32:141-9.

[25] Simpson M. How usability testing can aid the development of online documentation. Proceedings of the 8th annual international conference on Systems documentation; 1990; Little Rock, Arkansas, United States: ACM; 1990.

[26] Qiu Y, Yu P. Nursing Informaion Systems: applying usability testing to assess the training needs for nursing students. Methods of Information in Medicine. 2006;to be appear.

[27] Nielsen J, Mack RL, eds. Usability inspection methods New York Wiley 1994.

[28] Courage C, Baxter K. Understanding your users: a practical guide to user requirements methods, tools and techniques 2005.

[29] Crowl TK. Fundamentals of educational research. Madison, Wisconsin: Brown & Benchmark 1996.

[30] Wilkinson D, Birmingham P. Using research instruments : a guide for researchers. New York ; London: RoutledgeFalme 2003.

[31] Morgan DL, ed. Successful focus groups: advancing the state of the art: Sage Publications 1993.

[32] Kirakowski J. Questionnaires in usability engineering. 2003.

[33] Kushniruk AW. Analysis of complex decision making processes in health care: cognitive approaches to health informatics. J Biomed Inform. 2001;34(5):365-76.

Address for correspondence

Yiyu Qiu
Email: yq518@uow.edu.au
69/39 Northfields Avenue, Wollongong, Australia 2500

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Making Health Informatics Competencies Useful: An Applied Health Informatics Competency Self-Assessment System

H.D. Covvey, S. Fenton, D. Mulholland, K. Young, B.A.E.

Waterloo Institute for Health Informatics Research, University of Waterloo, Waterloo, Ontario, Canada

Abstract

Years ago we undertook to define Health Informatics (HI) competencies. This effort resulted in the creation of a document that articulated HI roles, the challenges faced by HI professionals, the high-level tasks that they needed to undertake to address these challenges and the competencies (skills, knowledge, and experience) they needed to complete these tasks. Unfortunately, in so doing we created what is arguably the most boring book in history, shoes contents are very difficult to extract, use, maintain and improve. We report here the completion of a pilot of a system that we believe corrects this situation. It is a web-based tool that incorporates all of the material, from roles to detailed competencies, enabling them to be accessed and used for a variety of purposes, the most notable of which is professional self-assessment.

Keywords:

health informatics education, competency assessment

Introduction

Canada Health Infoway and other Canadian organizations involved in the deployment of ehealth infrastructure have pointed out that they face major challenges in accessing human resources with Health Informatics backgrounds. There have been estimates that as many as 2,000 positions in Canada go unfilled or are filled by less-than-fully qualified individuals.

Canada already has university and community college-based programs in HI, but these produce less than a hundred HI professionals per year. Even with a number of new education programs emerging, there is little hope of increasing this number significantly over the next 3-5 years.

In addition to this, the reality is that few students in our high schools or many of our post-secondary institutions have even heard of opportunities for careers in HI, or even what HI is!

This lack of awareness was echoed by academic experts and industry leaders from Canada, United States and Britain at the eHealth Conference 2005 in Toronto, Ontario who called for aggressive action to resolve the health informatics human resources gap. Rapid education deployment programs like AMIA's 10 x 10 program and the Waterloo Institute for Health Informatics Research Boot-camp program have been launched as attempts to correct this situation. However, there is still minimal awareness of the nature of HI, the roles that HI professionals perform, the education required to play a role in this field, and the opportunities for employment for graduates.

Moreover, many health-related public and private sector organizations that would benefit from HI expertise are only just becoming aware of the competencies required of health informaticians, or that education and training is available in our universities and colleges. One consequence is that recruitments are not properly informed regarding required competencies, when candidates are found, poor choices can easily be made in hiring.

The field of Applied Health Informatics (AHI), that most relevant to health organizations and the vendors that serve them, blends the informatics and health disciplines to find and deploy the best possible technical solutions to a wide range of healthcare information-related challenges. Thus, people in AHI need to have a solid technical background, a detailed knowledge of the healthcare industry and of the capabilities it has produced, and a wide range of personal, organizational and cultural, business, managerial and analytical skills.

AHI self-assessment system

We undertook several years ago to address the need for Health Informatics career awareness in part by beginning the development of a unique and innovative web-based Health Informatics competency self-assessment tool. This tool helps students, potential students, and the individuals who want to hire them: (1) to understand the types of roles that exist for individuals trained in Applied Health Informatics, (2) to understand the competencies required to fill these roles, (3) to self-assess their (or their candidate's) knowledge and skills against the competencies required in these AHI roles, (4) to take objective tests to validate their perceptions of their competencies and (5) to find educational resources that enable them to address deficiencies that are identified.

The primary purpose of this tool is to serve as a self-assessment system and to provide an index to learning resources. It is intended as a resource for students, teachers and employers related to understanding the skills, knowledge and experience expected of competent Applied Health

Informaticians. Furthermore, the competencies and competency categories embodied in the tool and supporting documentation[1] can assist curriculum developers in defining the educational content for Applied Health Informatics education programs. The system is also intended to be able to serve as a "front-end" or entry point to an educational program that can be used to document progress related to the competencies the student needs and has acquired.

It should be noted that our framework to define competencies and the tool we built to support self-assessment and access to learning resources are entirely general and can be applied to any discipline, not just to Health Informatics. This makes our framework and tool reusable and of broad potential interest.

Methodology

In earlier work [1], we led a process that defined HI competencies and supporting curricula using a team of approximately 100 health and Health Informatics professionals (including HI teachers/researchers, curriculum developers, human resources professionals in healthcare organizations and HI companies, vendor staff, government representatives, and current and candidate students). This work involved the comprehensive documentation of HI roles, the challenges faced by professionals in each of these roles, the high-level tasks (which we called "micro-roles") that professionals in these roles need to undertake to address the challenges, and the competencies required to accomplish the tasks/micro-roles [1,2]. This material was then used to define content that could be used as the basis for a number of programs world-wide [3]. One such program has been established at Conestoga College, in Kitchener, Waterloo, Ontario, which registered its first students in the fall of 2005. The competency definition project was funded by the Canadian Institutes for Health Research (CIHR), its products are frequently cited, and today it stands as one of the few definitions of HI competence that have been derived using a logical framework (a work breakdown structure) rather than being solely based on the preferences and opinions of teachers with their own programs.

Once having completed the documentation of competencies (there are on the order of 400 for AHI alone) and the other components of this project, we recognized that the form they were in needed to change. In particular, they were incredibly boring to read, making improving them an un-motivating task. In 2001, F. Lau at the University of Victoria made the suggestion that we consider some way of encapsulating or packaging the competencies and other content in a directly usable form. We considered this suggestion and realized that significant value could be derived by packaging the competencies in a software tool that made them accessible, available and integrated for review, understandable, and actually usable by individuals in the field.

We called this tool "WebSAT" and built it using the Web-based Informatics Development Environment (WIDE) developed by the Computer Systems Group at the University of Waterloo, led by D. Cowan [4]. WIDE is intended as a rapid pilot system definition environment, based on a decade of Software Engineering research, composed of software services and are customized using declarative techniques. The goal of WIDE is to reduce technological barriers to system design and development.

WIDE is primarily based on open source software technology and consists of a number of services and supporting frameworks. Applications can include input forms or reports containing extensive multimedia materials such as imaginative use of maps or any 2-dimensional diagram, websites, databases, indexing and searching methods, agents, and push technologies. WIDE also contains a knowledge management system that supports documentation of technical information and best practices. The structures underlying the services are usually expressed in an XML-based declarative language that uses metadata and XSL. In the WIDE metadata context, "programming" has effectively been replaced with a declarative methodology thus making it possible to provide a wizard or forms-based approach to building Web-based systems. Internally WIDE uses a bootstrap approach; its extensions are implemented using its own metadata technology. WIDE can support a rapid development paradigm and new applications can be quickly built and demonstrated. The components of WIDE are described next.

A mapping services framework. Interactive maps are delivered from a map-server, which supports zoom-in or zoom-out functionality and positioning over areas of interest. When connected to a database or other directory the maps can be used to display and interact with the location of a geo-referenced object. Map searches can be defined by a circle, or general polygon. The mapping service framework does not use traditional GIS software.

A diagram and chart services framework. This framework manages and delivers specified interactive diagram and chart types upon request for presentation of data on the Web or in other formats. The diagram and chart services are also based on the Scalable Vector Graphics (SVG) open W3C standard and so provide similar functionality to the mapping service framework.

An XML-based metadata framework. The structure of databases, websites, agents, and applications including reports and input forms with maps and diagrams are described using XML. They are transformed into operating applications through the use of XSL "programs." Any application can describe and subsequently access databases or Web sites reachable anywhere over the Internet.

A report services framework. This framework supports the management of interactive report and input form types including maps, charts and diagrams and delivers them on request for presentation on the Web or in other formats. The user indirectly specifies the form type and the data that that is to be presented or requested; the framework chooses the report or input form type and populates it with the requested data.

A content management services framework. This framework supports the management of text and multimedia information in a database where it can be viewed, searched, maintained and then published for use on the Web or in other formats.

An access control service framework. Access to any content such as a database, website or other text and multimedia content can be provided with multi-level access controls to determine who can read or change data.

A Web and database searching service framework. This service framework contains an indexing agent and search engine that will index known websites and databases and support searching. The results from Web searches are categorized based on different search criteria such as the proximity of words in a phrase. The results of combined database and Web searches can be presented together. The results of the two searches can be compared to see if new results have appeared in the intervening time interval.

A push/notification service framework. The general push/notification service framework allows developers to create systems that allow users to specify conditions under which they wish to be notified or have information pushed at them.

An agent service framework. The agent service framework supports the description of agents that will act autonomously to perform utility tasks within an application. Agents are often defined to manage redundancy. For example, agents could be defined to verify the content of "local" databases against authoritative sources or to allow a user to type information once while submitting the data to multiple databases or Web sites.

The academy – a knowledge management framework. The academy framework is used to support widespread dissemination of "documentation" and knowledge describing how applications can be built from the WIDE Toolkit.

Description of WebSAT

WebSAT (Web-based Self-Assessment Tool) is a Web-based AHI self-assessment system that enables individuals to review HI competencies, assess their own competencies and compare these to the competencies required for specific roles. The development of this system was funded using internal resources and volunteer labor and is now available as a demonstration website on request to the author. In return, we ask users to complete a brief assessment of the system for us.

Our competency definition work identified three types of HI professionals: (1) AHI (Applied Health Informatics) professionals, who define the requirements for, procure, deploy, implement, manage, guide the use of, and evaluate HI systems and methods in health enterprises and their supporting industries; (2) RDHI (Research and Development HI) professionals, who teach, do research, and develop innovative HI tools for the health system and are typically found in academia and private industry research labs; and (3) clinicians who need HI competencies to be good clinicians (called Clinician HI or CHI).

In order to make the initial version of the system of manageable scope, we limited ourselves to providing a tool for those interested in assessing their own competencies relative to those required by AHI professionals. Furthermore, we did not address many of the user interface issues. For example, we did not allow individuals to assess themselves first at a high-level (versus broad categories of competencies) and then at a more granular level within these categories.

This past year we extended the development of this tool through the able assistance of a student working in the Undergraduate Research Assistant (URA) program in the David Cheriton School of Computer Science (D. Chodos). This student added references to Web-based educational materials for sample set of competencies. This allows users to click on a link and access educational documents, on-line educational programs, and on-site courses that provide a means to correct competence deficiencies. This has turned out to be an excellent capability that supports lifelong learning in this field. This latter work was reported at the recent eHealth 2005 Conference in Toronto [5].

During the last 6 months we have made major improvements to the system:

- A new user interface has been developed that makes use of the system more intuitive.
- Users can now assess themselves at a high-level (compared to competency categories) or at a detailed level, reducing the effort if the user does not need to go to a deeper level.
- Improved graphics show the comparison of the user's stated competencies to those required by the selected role.
- There is a new query capability that allows recruiters to retrieve the competencies required for various roles, based on the importance of the competencies.
- The system supports the self-assessment of experience.
- A capability to take objective tests is now included so users can validate there perceptions of their competencies.

Finally, the overall system now has a better look and feel to improve the user experience.

Mode of operation

WebSAT operates as follows:

- Users who access the system are asked to register under a user identity and password that they create.
- The user can review the definitions and detail of any of the components used to define competencies, including possible roles, challenges, skills, etc. This is the part of the system that supports the understanding of AHI itself and the review of details that the content developer may add, delete or improve.
- The user can then select one of the potential AHI roles, after reviewing descriptions of each role. The users stated competencies will be compared with this role (this can be changed at any time).

- The user can then access approximately 20 "competency categories" (groups of similar competencies), each containing multiple specific competencies, and can assess him/herself as to the level of knowledge or skill he/she has. Competencies are assessable as one of seven levels from "no knowledge" to "expert"). These levels include: UNA=Unacquainted; ACQ=Acquainted; PAM=Passing Familiarity; GAM=General Familiarity; FAM= Working Familiarity; CAP=Capable; EXP=Expert)
- If the user wishes, she/he can drill down to detailed competencies within the competency categories.
- The system allows corrections and the saving of input for later completion.
- Once the user has responded to all competencies at whatever level of detail desired, the system compares the user's competencies to the selected role and provides graphical feedback (a bar chart of responses versus requirements) as well as textual feedback. Different roles can be tried, to see if there is a better fit with these.
- For areas requiring further work, the users have the opportunity to receive system guidance to educational resources.
- Users can access objective tests for each competency.

The competency categories included in the system are:

1. Personal Competencies for AHI
2. General Computing Competencies
3. Health Computing for AHI Professionals
4. Key IT Usage for AHI Professionals
5. General Health System Competencies
6. General Business/Management
7. General IS Department Management
8. Team and Human Resources Management
9. Re-Engineering and Management of Change
10. Strategic and Operational Planning
11. Assessment of the Value, Effects, and Cost of IT
12. General Technology/Systems Life-Cycle Management Competencies
13. Procurement Competencies
14. Systems Implementation and Integration
15. Systems Maintenance and Support System Customization/Ad Hoc Development
16. Project Management Competencies
17. Education and Training Competencies
18. Vendor/Service Provider Competencies
19. User and Process Observation and Assessment Competencies
20. Security Management Competencies
21. Information and Data Collection, Analysis and Management Competencies

A total of approximately 400 separate competencies allocated under these categories are addressed by the system.

Results

WebSAT has now been tested by both by graduate students in our HI program and by students in a new HI program at a nearby community college.

Students have reported a high degree of satisfaction with the system. They found the system useful and informative, and it gave them a clear view of what they still need to learn and where they stand relative to the requirements of various roles.

We have recognized, however, that other improvements are possible, and we continue to enhance the system further along the following lines:

1. We are undertaking further improvements to its user interface so that it can be more efficiently used by inexperienced users.
2. We are completing the addition of learning links to Web-based educational sources to be referenced by students with identified weaknesses.
3. We are adding additional objective tests and creating a new testing engine. The existing one does not support sufficiently complex multiple choice options.
4. We will extend the tool to incorporate Research and Development HI (academic-level) self-assessment to the same degree as the AHI assessment.
5. We are in the process of offering the system to other programs for use as a tool for students, and we are seeking recruiters to test its support for their activities.

Summary and Conclusions

We have developed a Web-based competency self- assessment system for Applied Health Informatics. This system allows individuals to assess themselves as to the congruence of their competencies with those required for specific roles. The system has been tested on students and is in the process of being enhanced and disseminated. Individuals interested in accessing the system personally or using it as a component of their programs are invited to contact the first author. The system is offered without charge.

Acknowledgements

This work was supported by NSERC under the Executive Industrial Research Chair program and by Canarie Inc. The authors are grateful for the contributions of D. Chodos, research assistant, at the University of Waterloo. The authors are grateful to F. Lau at the University of Victoria who first suggested the development of this system.

References

[1] Covvey HD, Zitner D, Bernstein RM. "Pointing the Way: Competencies and Curricula in Health Informatics", H/IT Advisors, March 2001 (Book: at http://hi.uwaterloo.ca, May 2007).

[2] Covvey HD, Zitner D, Bernstein R. "The Development of Model Curricula for Applied Health Informatics", Medinfo 2001, London, England, September 2-5, 2001.

[3] Covvey HD, Zitner, D Bernstein RM, MacNeill JE. "The Development of Model Curricula for Health Informatics",

Proceedings of the International Medical Informatics Association WG 4 Meeting, April 2003.

[4] Cowan D, Fenton S, Mulholland D. Lowering Technological Barriers to Developing, Deploying and Maintaining Web-based Interactive Information Systems, to appear in the International Journal of Technology, Knowledge and Society 2005.

[5] Chodos D. Accessing Applied Health Informatics Educational Resources via a Web-based Self-Assessment Tool, eHealth 2005 (Poster), May 2005.

Address for correspondence

WIHIR, DC3333, University of Waterloo, 200 University Ave. West, Waterloo, ON, Canada, N2L 3G1; dcovvey@uwaterloo.ca

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

E-learning for Students in their First Year: A French Experimentation at the Medical School of Grenoble

Jean-Marie Renard[a, d], Daniel Pagonis[b,c, d], Jean-Philippe Vuillez[c,d], Jean-Paul Romanet[c, d], Bernard Sele[c, d]

[a] *CERIM, CNRS-EA2694, Faculty of Medicine, Lille2 University, France*
[b] *SIIM, University Hospital, Grenoble, France*
[c] *TIMC, Faculty of Medicine, Grenoble University, France*
[d] *UMVF, French Virtual Medical University, http://www.umvf.org*

Abstract

A local study carried out in the Medical School of Grenoble shows that teaching in the first year in medicine studies satisfies neither the students, nor the teachers. The Faculty of Medicine of Grenoble decided to set up a reform in order to offer a high quality education. This reform leads to a complete reorganization of the curriculum and to the intensive use of new information and communication technologies of information, in particular, the use of multimedia documents. The communication and information technologies team of the Faculty of Medicine of Grenoble carried out an innovating and daring reform to start at the academic year 2006-2007. The new course is built on three activities: self learning on multi-media resources, meetings with teachers for questions-answers sessions and tutorials animated by older students. This article reports the first results for this successful project. In the academic year 2006-2007, are concerned 1290 students, 40 teachers and 8 disciplines.

Keywords:

e-learning; first year medical curriculum; multi-media; Grenoble, France.

Introduction

Observed difficulties during the first-year curriculum

A degradation of working conditions

This degradation was explained as a consequence of years of teaching without innovations or improvements. From a teaching point of view, the acquisition and transfer of knowledge remain extremely antiquated. The majority of learners remain passive. The student remains isolated before a task that cannot be managed without help. This may generate harmful consequences on his learning capacity..

A regular increase in the number of students

The increasing number of students is a factor which amplifies the degradation of the working conditions. Since 1975 the number of admission in second year has dramatically increased. The "numerus clausus" of Grenoble corresponds today to 12% of the number of registered students (year 2006, 1300 registered students against 166 students allowed to continue the second year course). The amphitheaters have a normal capacity of 650 places whilst 800 students actually attend the lectures. This number cannot possibly be increased for safety reasons.

A contest which no longer fulfils its role

The contest no longer fulfils its double role of regulating the number of students and selecting the best ones. Studies showed that the harder a contest is, the less effective the selection of the best students is. This strong selection has many direct and indirect consequences such as some influence on the content of the lessons. Difficult lessons only exist for their selective character and not for their relevance.

Inadmissible behaviours

The extremely selective aspect of the contest has generated an attitude of unfair competition behaviour, such as voluntary disturbance during the lectures by a few students, voluntary locking of the books to limit there availability, paper planes sent through the amphitheatres, going to the extreme: sale of already annotated lectures showing voluntarily added errors. These behaviours go against the university values of solidarity and equity amongst students.

Less equity between the students

Many a student registers in private courses which offer method and regularity in work in order to make up for the effects of these bad working conditions. This is also a means for the getting of annals and self assessment in relation to others. In the French system, the inscription to these courses generates a discrimination against money since the first access price exceeds 1500 € .

Effects over the forthcoming years

One of the indirect but considerable effects of this situation is that the motivation to work is getting slack. This is understandable considering the very great efforts provided for during the first year. Teachers estimate that students need three years to come back to a normal working attitude.

Solutions already tested by other universities were not very conclusive

Faculties of Medicine, facing the same difficulties, and drawing the same conclusions adopted various solutions:

- One of them is the repetition of the courses, which obliges the teachers to reproduce at various gaps the same course in many amphitheatres. That demands a great involvement from the teachers and availability of many extra amphitheatres (with a capacity exceeding 700 places). The risk is this might encourage the action of disturbing students, increase the loss of interaction between teacher and students and generate problems with regards to safety. For all of these reasons, this solution cannot be applied in all Faculties.
- Another very widespread solution adopted by various Faculties is the broadcasting of the course by Video-transmission in a secondary amphitheatre. The principle consists in diffusing on line and on large screen in a secondary amphitheatre the course done by the teacher lecturing in the first amphitheatre. This reduces the number of mobilized teachers. However, to be added further to the already listed problems, the high costs for expensive, fragile and not always reliable equipment which cause considerable discontent among students. Answering the student's questions is difficult and this increases the risk of disturbance.

The need for an important reform

Whatever the solutions considered by the universities, the quality of the teaching is unfortunately not improved, on the contrary it is damaged. To face the whole of these drawbacks, the Faculty of Grenoble team has proposed a reform which leads to an original solution.

Materials and methods

Elaboration of an original reform

Emergence of the project of a reform for the first year medical curriculum, all the actors at the medical school carried out a complete thinking on the modernization of teaching. Administrative people (Dean...), teachers, and students (elected representatives) endeavoured to find a solution which would improve the quality of the teaching. In there analyse, they took into consideration different dimension like behaviours [1,2], cognitive sciences [3,4] and possibilities offered by the communication and information technologies [5,6]. A reform combining a new organization of teaching associated with an intensive use of communication and information technologies appear to be the best alternative.

Specificities of the medical profession

Given the teacher's specific statute, the Faculty of Medicine offers more freedom. Most teachers of Medical School are not subject to a quota of hours before the students. A modification in manner of delivering of the courses will not be a problem due to their teacher statute.

Grenoble specificities

This project is in the continuity of previous reforms made for the years following the fourth year (2002-2003), and for the years between the second and the third year (2003-2004) of the medical curriculum. Another particularity is the strong motivation of all the teams involved in teaching for an intensive use of new communication and information technologies.

New methods of training

The reform does not change either the curriculum program or the organization of the contest.

The year (2006-2007) is divided into 2 six-month study periods. Each semester ends with a part of a contest (January and May 2007). One six-month period is divided into 12 training sequences, each sequence divided into 4 weeks.
Each week is devoted to a different training activity.

- The first week is devoted to the study of the courses on DVD-Rom: one or two different disciplines are studied over the week in the form of multi-media courses animated and wired for sound by the teacher's commentary.
- The second week is used for the formulation of all questions destined to teachers. These questions exclusively relate to the multi-media courses studied during the previous week. They are the base of the Question/Answer Meetings. The questions are formulated on the Faculty of Medicine of Grenoble's web site which is destined to the PCEM1 students (www.medatice-grenoble.fr). The on-line form is open each week, from Saturday to Tuesday exclusively. The access is secured by a login-password provided for at the time of final registration.
- The third week is devoted to the Question-Answer Meetings. Each discipline studied during the sequence is followed by a 2 hour Question/Answer Meeting. These meetings are held in small amphitheatres. Students are divided into eight groups. The meetings are ensured by the teachers in charge of the discipline. The answers are built from the questions collected on the on-line form (www.medatice-grenoble.fr).
- The fourth week is dedicated to the tutorials. The tutorials are directed by third year tutor medicine student. These meetings are intended for QCMs training. QCMs are validated by the teachers, corrected and explained during the meeting by the tutor students. These students are trained and supervised by the teachers in charge of the disciplines and the contest. The meetings are held the week following the questions-answer meeting.

Recording of the multi-media courses

Recordings are carried out by the Stendhal University Audio-Visual Service. The multimedia resources consist of animated slides commented and recorded by the teachers.

Evaluation of the reform

The evaluation of the project lies on statistical data describing the correct working of the project as well as on satisfaction surveys carried out weekly. At the end of each tutorial session, students are questioned. The information collected concerns the evaluation of the lesson and training methods.

Agenda for the project

The Table 1 reports the agenda for the project.

Table 1 - Agenda of the project

Date	Event
November 2005	Realization of the preliminary study
December 2005 to January 2006	Redaction of the performance specification
February 2006	Validation of the project by the Université Joseph Fourier's council
March 2006 to September 2006	Production of the multi-media records
July2006	Opening of the web designed for the students in their first year
August 2006	Production of the DVD-Rom number 1 (2000 pieces)
September 2006	start of the curriculum – distribution of the DVD-Rom N°1 to students

Results

Numbers of students and teachers

The teaching of the first year of the medical curriculum in Grenoble concerns 1290 students, 40 teachers and 8 disciplines for the year 2006-2007.

Self learning on DVD-ROM

There was 230 sessions corresponding to 460 hours of recordings. The slides are in flash-R, mp3 and xml format. The materials are equivalent to 220 hours of listening by the students.

Every six months, a new DVD-ROM is distributed to the students. The DVD-ROM contains 4, 7 Go of data (The mean size of one course is about 26Mo). DVD-ROMs are pressed in 2000 specimens. Each DVD-ROM cost 1,60 euro.

The courses are also available on the web site dedicated to students under the form of Podcast using Open source Xvid file format [7].

The user assistance for DVD-ROM usage only received 16 calls and all of them were solved by the assistance.

The hot-line was overloaded by calls coming from students registered in Universities other than Grenoble to get the DVD-ROM.

Students will fill an electronic form after each course. Every week 1050 students will fill the evaluation form. 36 000 forms will be delivered during the year.

The students' satisfaction rate ranges between 75% and 85%

Figure 1 – self-learning rich media

The questions-answers sessions

The students send their questions through an on-line form. The questions must be related to the courses delivered during the previous week and are collected from eight independent groups of students.

At the end of the first cycle (the first four weeks), 900 students have used the system and asked 7484 questions which were marked "pertinent" 38 879 times by other students. The satisfaction level for that system was 85%.

The session with teachers was organized in small amphitheatres. Every student followed two sessions per week. Each session lasts two hours The ratio questions-answers sessions over self-learning time is two hours of questions-answers for ten hours of self-learning time.

The number of questions per teacher varies between 200 and 300. The quality of the question is qualified as good by teachers. The satisfaction level of the student is between 55% and 75%.

Tutorial session

The 1290 students are divided into 40 small groups. Every student follows two sessions per week. Each session lasts two hours. 96 hours per student and per year are envisaged.

The tutors are older students (120 students in their third year of medical curriculum). A session is animated by two students (always the same two students). The two students are supervised by teachers and spend two hours for the preparation of a session which content is a set of QCM.

1 000 000 QCM are planned to be written over the year.

The mean satisfaction rate is 94%.

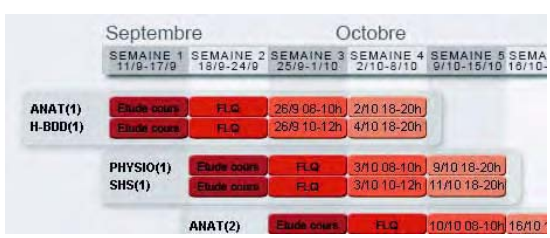


Figure 2 – The four activities of the learning cycle: self-learning on multimedia supports (Etude de cours), preparation of questions for teachers (FLQ), questions-answers session with teachers (ie 26/9 08h-10h) and tutorials (ie 2/10 18h-20h).

Logistics

A 700 places amphitheatre was converted into a multimedia room. It was adapted with 130 desktop computers, electrical connectors for personal laptop computers and with WIFI network connexion.

This room can accept 248 students who work in twos. It is open from 8h-20h and from Monday to Saturday

There is a low cost rental service for laptop computers and, in some cases, laptops can be lent.

Other rooms were adapted with computers and dedicated to questions-answers sessions as well as tutorials sessions.

During the first week, only 5 students used the 700 place amphitheatre. During the first month, the maximum number of students was 30.

The Web Site

It started in June 2006 and is reserved to first year students (www.medatice-grenoble.fr). It provides information on the organization of the curriculum. It gives access to the forms planned for their asking questions to the teachers. It is an entry point to access to the podcasts.

The average number of daily hits is 875 and the maximum 1200. In October, there has been 21 899 visits and 448 932 pages loaded.

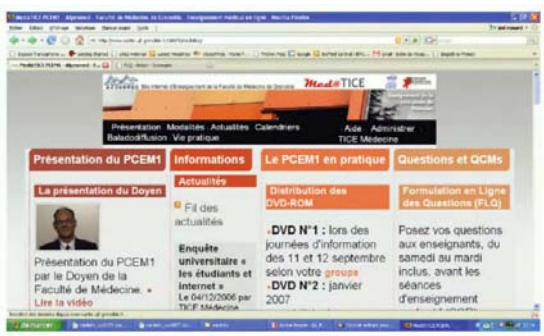


Figure 3 - The web site reserved to first year students

Discussion

The first result of satisfaction from students must be monitored to confirm the efficiency of this organization.

An element that needs to be confirmed is the uselessness of the multimedia amphitheatre.

A curriculum based on a four week cycle requires strong motivation from the teachers, how it will evolve in time is the question to ask.

The efficiency of a learning cycle lead on the sequence: self-learning on multi-média materials, questions-answers with teachers and tutorials with senior's student must be the subject for further analysis.

Conclusion

"The doctor of the 21th century must also have teaching competences in technologies which must be acquired at the time of his training" (Thierry KARSENTI).

The MedaTICE project showed that motivation and conviction of a team make it possible to carry out a joint and innovative project to it success.

Undoubtedly, the action taken by the Faculty of Medicine of Grenoble will be the starting point of a great scale disruption and will lead other faculties to become involved in similar projects.

It will be very important to follow on a monthly basis the statistics on the project and to do a complete analysis at the end of this first year.

Acknowledgements

Our warm thanks to Michèle Jacquemart for her assistance in preparing this paper.

References

[1] Apollon D. L'enseignement supérieur face au numérique en Europe : Tradition, innovation et résistance. Université de Bergen, 2006.

[2] Basque J. Influence de béhaviorisme, du cognitivisme et du constructivisme sur le design pédagogique.1999. www.teluq.uquebec.ca/~jbasque.montréal.

[3] Bernatchez PA. Vers une nouvelle typologie des activités d'encadrement et du rôle des tuteurs. Université de Montréal, 2001.

[4] Mayer, C. When learning is just a click away: Does simple interaction foster deeper understanding on multimedia messages?. Journal of educational psychology 2001: 390-397.

[5] Bernartchez PA, Marchand L. Encadrement médiatisé par ordinateur. Distances 2005 : Vol. 7. http://cqfd.teluq.uquebec.ca/distances/v7n1_b.html.

[6] Bellier S. Le E-learning. Editions Liaisons, Rueil-Malmaison, 2001.

[7] Xvid, Open source MPEG-4 Video Codec; Wikipedia Nov 2006. http://en.wikipedia.org/wiki/Xvid

Address for correspondence

Jean-Marie RENARD
CERIM, EA-CNRS-2694
1, place de Verdun
59045 Lille Cedex
mail: jean - marie . renard @ univ - lille2 . fr

Daniel Pagonis
Faculté de Médecine de Grenoble
Domaine de la Merci
BP 170 LA TRONCHE
38042 GRENOBLE CEDEX 9
mail: Daniel . Pagonis @ imag . fr

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

E-learning at Porto Faculty of Medicine. A Case Study for the Subject 'Introduction to Medicine'

Cristina Costa-Santos[ab], Ana Coutinho[c], Ricardo Cruz-Correia[ab], Ana Ferreira[bc], Altamiro Costa-Pereira[ab]

[a] *Department of Biostatistics and Medical Informatics, Faculty of Medicine, Univ. of Porto, Portugal*
[b] *Centre for Research in Health Technologies and Information Systems – CINTESIS, Faculty of Medicine, Univ. of Porto, Portugal*
[c] *Department of Informatics, Faculty of Medicine, Univ. of Porto, Portugal*

Abstract

The main objective of the Introduction to Medicine (IM) subject of the first year of the Medical Course at the Faculty of Medicine of the University of Porto is to provide students with a first contact with the areas of Biostatistics, Medical Informatics, Bioethics and the History of Medicine in the belief that they will be better prepared to learn, research, evaluate, share and decide within their practice. This paper presents a case study that describes how the subject IM is organized and how the b-learning tool (Moodle) is used to correct and grade the students' work. From the 239 students registered to attend the Introduction to Medicine subject 12% failed. The average grade among the successful students was 16 (out of 20). In the previous academic year only 2% of the students failed. However, among the successful students, the average grade was inferior (15 out of 20). The e-learning model that was described in this paper was successful because the results show that the students that made use of the Moodle got better grades.

Keywords:

education, medical faculty, e-learning

Introduction

The practice of Medicine requires the use of methods in order to acquire, store, process, analyse, transmit, evaluate and assess information as well as medical knowledge [1]. The study of scientific areas such as Biostatistics, Medical Informatics, Bioethics and the History of Medicine can contribute to improve the practice of Medicine and make it well supported by ethical principles and, therefore, socially effective. The main objective of the Introduction to Medicine subject of the first year of the Medical Course at the Faculty of Medicine of the University of Porto (FMUP) is to provide first year medical students with a first contact with the areas described above in the belief that, as future doctors, the students will be better prepared to learn, research, evaluate, share and decide within their practice [2]. Although there was some existing material on the Internet, it was only in October 1999 that an intranet for this subject was designed and implemented by the responsible professor and lecturers of the same subject [3]. The first interface, developed using PHP and HTML programming languages and a relational database, namely Postgres, in a LINUX server, was used to store and manage information. In the academic year of 2005/2006 an e-learning tool (i.e. Moodle) was introduced. This was done in order to use information systems' technology to provide for automatic students' work correction and evaluation.

This paper presents a case study that describes how the subject of Introduction to Medicine is organized and how the 'b-learning' tool is used to correct and grade the students' work in the academic year of 2005/2006.

Motivation for the use of an e-learning tool

A vast majority of first year Medical students are not interested in learning basic scientific subjects. They probably think that their content is not important for the clinical work ahead of them, and for which they have motivation. The introduction of technologies such as e-learning tools that bring new interactive methods for communication and simulation can help the lecturers to motivate the students and facilitate their learning process. Several case studies support the use of Web technologies in order to teach undergraduate Medical students [4].

Initial expectations

Although some of the material from previous academic years already included interactive means to communicate and learn (e.g. forums, chats, placard) the main objective to introduce Moodle was to use the specific module that allows students to make exercises and tests online, which are corrected automatically. With this module, the lecturers wanted the students and themselves to have a better understanding of the evolution of the students' learning results along the year with the main goal to improve the students' final results at the end of the year. This module eliminates the waiting time that was needed for the lecturers to correct the exercises and tests made by the students therefore simplifying the process of correcting and grading the weekly correction of 230 tests.

Objectives

The main objective of this case study was to create an e-learning platform that allows the students and lecturers to

follow the progress of students' work along the year. This helps the students to understand how they need to change or not their learning process with the real-time feedback they get.

Other objectives include facilitating students' access to the lectures' material as well as other interactive material and online communication means between students and lecturers that help and motivate the students to study and learn.

During the academic year, the lecturers noticed that the objectives were being achieved because the students were using regularly the material described above. This could be seen by consulting the usage statistics module within Moodle.

Materials and methods

Structure and grading system

Introduction to Medicine is a subject from the first year of the Medical Course of the Faculty of Medicine of the University of Porto that integrates four modules regarding the following scientific areas: Biostatistics, Medical Informatics [5], Bioethics and the History of Medicine. In the academic year of 2005/2006, 239 students were registered in order to attend this subject.

The Bioethics and History of Medicine modules are only taught with theoretical lectures and graded with a written examination. Each module weights 2 out of 20 for the final subject grade.

The other two modules, Biostatistics and Medical Informatics are taught with theoretical lectures, an e-learning component and practical lectures. Each one of these modules weights 5 out of 20 in the final classification (2 for the written examination and 3 for the work students do during the year at the practical lectures).

The 6 values necessary to complete the maximum grade of 20 come from a group project that is compulsory for all the students to complete at the end of the academic year. The group projects are supervised by the lecturers that help the students to understand and organize their work. In addition, students can attend seminars that give important information for the development and implementation of the group project.

Figure 1 - Moodle e-learning platform that is used within the Introduction to Medicine subject

B-Learning - integrating e-Learning components within the teaching process

The Biostatistics and Medical Informatics modules are taught with theoretical lessons (where the main concepts are presented to the students) and practical lessons (where the lectures support the students in their continuous self-learning process).

It is within the practical lessons that the integration of the theoretical concepts with the e-learning component available at the Moodle platform is made (Figure 1).

The practical lessons have the duration of 2 hours and a half every week and the students have access to their own personal computer. The lecturers start by discussing with the students the concepts given at the theoretical lessons and then these concepts are practised within the e-learning platform. The lecturers assess students' acquired knowledge (all the 239) individually with online mini-tests (Figure 2) at the end of every lesson. The mini-tests have a maximum duration of 5 minutes and include 2 or 3 multiple or numeric questions. These questions are selected randomly from a pool of questions that is created by the lecturers. The pool about a specific issue is big enough so that when students start the test they have most probably different questions to answer in their test or at least setup in a different order from all the other tests, so that copies among the students is almost impossible. All the information that is available within the Moodle can be accessed at anytime from any computer that has Internet connection. However, the mini-test can only be accessed during the practical lessons, only in some selected PCs where the lecturers need to insert a key so that the students can enter the right interface. This feature avoids problems with counterfeit of mini-tests because their results are taken into account for the students' final grade. When the 5 minutes expire the mini-test is corrected automatically and the final grade is showed to the student. If the student does not submit his answers before the mini-test finishes, the answers given until that moment are automatically submitted and the grade given accordingly.

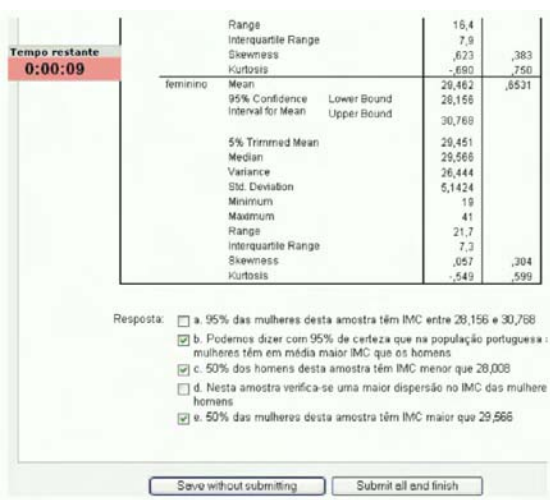


Figure 2 - Example of a weekly mini-test presented to the students, with automatic correction and timing

The average of all the results from the mini-tests can contribute to a maximum of 6 out of 20, which is the final grade.

Within the Moodle e-learning platform the content is organized on a weekly basis regarding each week of the academic year, in a total of 33 weeks.

Different types of content are included:

- SLIDES – the presentations made within the theoretical lessons;
- FORUM – used to exchange and share information about the subject between the students and the lecturers;
- EXERCISES – exercises to apply the theoretical concepts are made within the practical lessons and they allow the students to use real data from clinical studies. The tables with the real data that does not affect ethical issues related with the patients are also available in the Moodle.
- SELF-ASSESSMENT – the students can practice what they have learnt with self-assessment exercises with automatic correction;
- MINI-TESTS – these are exercises that count for the students' final grade. They are timed and corrected automatically by the Moodle;
- GROUP PROJECTS – the group projects are submitted by the students and commented by the lecturers within the Moodle, during 3 review phases;
- LINKS TO OTHER ONLINE MATERIAL:
  - MEDSTATWEB [6] - http://medstatweb.med.up.pt/ - the biostatistics interactive manual (figure 3) was developed by the lecturers and explains the basic biostatistics theoretical concepts with the use of practical examples and simulations (figure 4)
  - MEDICAL INFORMATICS - http://im.med.up.pt/ - the medical informatics manual was developed by the lecturers and explains theoretical concepts of Medical Informatics (see figure 5).

Advantages and disadvantages

Without an e-learning platform like Moodle it would be very difficult to make all the exercises and corrections that are done now because the number of students is too large compared with the number of lecturers.

This model allows students and lecturers to search and update the contents anytime, anywhere via Internet. It also permits the lecturers to monitor each student individually and the preliminary as well as final results of the subject and all the students in a generic way by allowing the analysis of the accesses that were made to its contents.

A disadvantage to consider can relate to the privilege that students with access to the internet at home may have in relation to the ones that do not have. To minimize this issue the students are encourage to use the computer laboratories within the Centre for Informatics at the FMUP that provide the Internet connection.

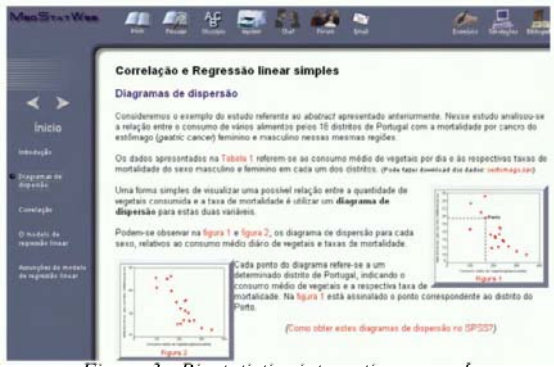


Figure 3 - Biostatistics interactive manual

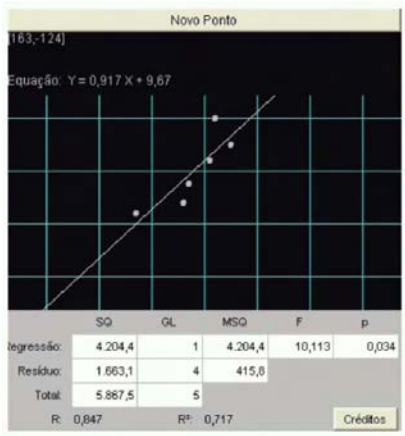


Figure 4 - Example of a simulation within the Biostatistics manual that explains the concept of regression. The user can move or insert points. The regression line is the ANOVA table and the relation coefficient automatically updated according to the points

Figure 5 - Medical informatics manual

Structure and implementation

The implementation and integration of the Moodle e-learning tool was done in collaboration with the Centre for Informatics at the FMUP. The contents have been developed, improved and updated by the lecturers during the past 10 years. In the academic year of 2005/2006 a group including 7 lecturers, the professor responsible for the sub-

ject and 2 other people from the Centre for Informatics at the FMUP integrated the contents of the subject within the Moodle platform, version 1.5.3 [7] along with a web server and a Mysql database server version 4. Both servers are shared with other services that the FMUP provides so the e-learning system does not have a specific server adjacent to it. Nevertheless, the performance of the system is not compromised. Both tools (Moodle and Mysql) are open source software so they do not bring additional costs to the FMUP. Furthermore, Moodle is a flexible tool, and being open-source software, it can be altered in order to adapt some of its functionalities in order to generate data and lists of relevant statistics to be used in this case study. The actual version of the Moodle does not include these functionalities.

The Moodle platform is available at http://moodle.med.up.pt.

The authentication platform to Moodle is synchronized with the authentication for the users of the FMUP allowing for its easier integration.

This platform was also used by integrate another subject form the Medical degree and 5 more subjects from a post-graduate summer course organized and lectured by the Biostatistics and Medical Informatics Department.

Results

Results for the use of the biostatistics module

The biostatistics module was lectured with theoretical and practical lessons during the first semester of the academic year of 2005/2006. During the course of this semester 227 students accessed the e-learning material available from Moodle. The students made an average of 1,4 accesses to each chapter of the interactive manual, the MedStatWeb, 1,5 accesses to each exercise of the practical component, 2,8 downloads of each set of slides from the seminars and 2,5 downloads of each set of slides from the theoretical lessons.

From the 227 students, 31 got an average grade inferior to 10 (out of 20) in the mini-tests, 149 got an average grade between 10 and 15 whilst 47 students got grades equal or superior to 16.

202 students wrote the final exam and the average grades of the Biostatistics module was 11.

We noticed that the students with better grades in the mini-tests were the ones that made more accesses to the MedStatWeb interactive manual and practical exercises within the Moodle. Although some of the students with better grades also made more downloads of the theoretical lessons' slides and seminars, these differences were not statistically significant (table 1).

Evaluation of papers

All papers will be reviewed by three (3) independent reviewers. Reviewers will use a standardized form for review, a sample of which appears on this Web site (click here for direct link to that form). Authors should be familiar with the criteria used in evaluation.

The students that got an average grade inferior to 10 in the mini-tests for this module got an average of 7 (standard deviation of 4) in the final exam for this same module. Those who got an average grade between 10 and 15 in the mini-tests got an average of 10 (standard deviation of 4) in the final exam. The students that got an average grade equal or superior to 15 in the mini-tests got an average of 14 (standard deviation of 4) in the final exam. These differences have a statistic significance (p<0,001 - One-Way ANOVA).

Table 1 - Median (minimum-maximum) of accesses made by the students for each type of content available in Moodle

	Average grades for the weekly mini-tests			
	<10 n=31	**10–15** n=149	**>15** n=47	**p***
Number of accesses during the first semester to:				
7 chapters from the interactive manual	8 (0-36)	10 (0-38)	13 (0-51)	**0,047**
10 exercises with real data	15 (6-25)	15 (4-29)	16 (5-33)	**0,033**
10 sets of slides – seminars	24 (1-136)	25 (0-125)	30 (2-66)	0,283
10 sets of slides – theoretical lessons	23 (2-91)	24 (0-73)	27 (5-56)	0,386

* kruskal Wallis test

Results for the use of the medical informatics module

The medical informatics module was lectured with theoretical and practical lessons during the second semester of the academic year of 2005/2006. During this semester 221 students accessed the e-learning material available in Moodle. The students made an average of 2,2 accesses to each exercise of the practical components, 2,5 downloads to each set of the theoretical lessons. From the 221 students, 37 got an average grade inferior to 10 (out of 20) on the mini-tests, 124 got an average grade between 10 and 15 and 60 students got an average grade equal or superior to 16. 202 students did the final exam for this module and the average grade was 12. The students with better average grades in the mini-tests made significantly more accesses to the practical exercises than the ones who did not. Although some of the students with better grades also made more downloads of the theoretical lessons' slides and seminars, these differences were not statistically significant (table 2).

The students with an average grade inferior to 10 in the mini-tests within this module got an average grade of 10 (standard deviation of 6) in the final exam on the same module.

Those who got an average grade between 10 and 15 in the mini-tests got an average of 13 (standard deviation of 5) in the final exam. The students that got an average grade equal or superior to 15 in the mini-tests got an average of 15 (standard deviation de 3) in the final exam. These differences have a statistical significance (p<0,001 - kruskal Wallis test).

Statistical data for the use of the platform

During the whole academic year the students made, within the Moodle, about 22 000 accesses to the slides used for

Table 2 - Median (minimum-maximum) of accesses made by the students for each type of content available in Moodle.

	Average grades for the weekly mini-tests			
	<10 n=37	**10–15** n=124	**>15** n=60	**p***
Number of accesses during the second semester to:				
6 exercises with real data	11 (2-23)	13 (0-38)	14 (6-45)	**0,026**
9 sets of slides – theoretical lessons	17 (0-39)	22 (0-77)	25 (1-60)	0,070

* kruskal Wallis test

theoretical lessons and seminars, 6 650 accesses to the practical exercises that relate with the theoretical concepts, 3685 downloads to real data databases, 10 383 submissions of the self-assessment exercises with automatic correction, 3 947 mini-tests that were included in the student final grade, 420 group project submissions and 94, 29 and 14 discussions made within respectively the forum shared between students and lecturers, the forum used only by the lecturers and the forum where news were announced to the students.

The e-learning model developed by the lecturers supports and motivates the study using other web material such as interactive manuals namely MedStatWeb and Medical Informatics manual.

There was an average of 53 daily visits was made to the MedStatWeb, with 2565 hits per day. Each student made an average of 6 visits per month on this same manual. There was an average of 38 daily visits to the Medical Informatics interactive manual with a total of 587 per day. Each student made an average of 5 visits per month on this same manual.

Anonymous survey applied to the students

At the end of the academic year of 2005/2006 anonymous surveys were completed by the students that attended the subject of Introduction to Medicine and came to write the final exam (n=170). 90% of the students agreed that there was a good relationship between students and lecturers, 82% that the subject was well organized, 79% that it was easy to access the available material, 75% believe that was motivated by the interaction of the practical lessons, 65% thought that was able to pay attention and be interested during the lectures, 59% said that they accessed the online material on a regular basis, 54% stated that they could study regularly the subject and 36% said they attended the theoretical lessons regularly.

Final results

From the 239 students registered to attend the Introduction to Medicine subject, 231 were assessed from which 204 (88%) were successful and 12% failed. The average grade among the successful students was 16 (out of 20). In the previous academic year, when another e-learning platform was in use only 2% of the students failed. However, among the successful students, the average grade was inferior (15 out of 20).

Conclusions

The e-learning model that was described in this paper was successful because the results show that the students that made use of the Moodle got better grades. Also, the results from the survey illustrate that a vast majority of students think that the subject I well organized and that was easy to access the available study material. They also agree they were motivated to participate more actively in the learning process.

It is important to refer that the significant statistic differences reflect the fact that students that access more the online exercises and simulations are the ones that got better grades, not the ones that made most downloads of the lectures content material.

However, it is surprising that the students made so many downloads from the slides available since one download and printout for each set of slides would be enough.

According to the results obtained the mini-tests are beneficial for the students learning process. The students that could not obtain a satisfactory grade in the mini-tests could not obtain a similar grade in the final exam. On the other hand, the students that got good grades in the mini-tests were able to get the best results in the final exam.

This e-learning platform facilitates and simplifies the lecturers work because it would be very hard for them to correct 200 mini-tests every week and give the students the updated feedback every time. The integration between the traditional teaching with the e-learning was successful as 90% of the students agree that there was a good support between them and the lecturers and this happened mostly during the practical lessons.

In conclusion, although the percentage of failed students was bigger than in previous years (reflection of a tougher final exam because lecturers thought the students were better prepared this time), the final grades obtained for the Introduction to Medicine subject were very satisfactory because the students that passed the final exam had generally better grades than the ones in previous years.

For the future use of the Moodle platform the layout needs to be improved to be better adapted to the FMUP characteristics and objectives. Also, the use of forums and discussion groups will be more encouraged as they can help the students to clear their doubts and discuss the lecturers' topics in a more informal way.

Acknowledgements

We would like to thank the Introduction to Medicine lecturers for the academic year 2005/2006 that are not co-authors for this paper but participated actively in this e-learning project, namely Luís Filipe Azevedo, Filipa Almeida, João Fonseca and Mário Ribeiro. Further we would like to thank other collaborators such as: Camila Dias, Clara Tavares and Pedro Marques, as well as Armando Teixeira Pinto and Alberto Freitas that did not lecture this specific academic year but have contributed in previous years to improve the use of the e-learning tools. Finally, special thanks to Miguel Ângelo from the Centre for Informatics at the FMUP that contributed with the implementation of the Moodle layout.

References

[1] Costa-Santos C, Ayres-de-Campos D, Costa-Pereira A, Bernardes J. Utilização do Web Site de um Projecto de Investigação na Educação Médica. Educação Médica 1998; 9 (2): 110-6.

[2] Costa-Santos C, Caso de Estudo de Introdução à Medicina. Elearning UPorto 2004-2005 : 331-340.

[3] Costa-Santos C, Teixeira-Pinto A, Costa-Pereira A. Web-based Learning at the Faculty of Medicine of Oporto: Description and Evaluation of a Student's Internet. Proceedings of the 8th International Conference of European University Information Systems 2002; 152-4.

[4] Bernardo V, Ramos MP, Plapler H, De Figueiredo LF, Nader HB, Ancao MS, Von Dietrich CP, Sigulem D. Web-based learning in undergraduate medical education: development and assessment of an online course on experimental surgery. Int J Med Inform. 2004 Sep;73(9-10):731-42.

[5] Freitas JA, Cruz-Correia R, Almeida F, Costa-Pereira A. Evolution of Medical Informatics teaching in a medical undergraduate course .Technology and health care - official journal of the European Society for Engineering and Medicine: Proceedings of MedNet 2003, Eighth World Congress on the Internet in Medicine 2003; 11 (5): 312-314.

[6] Teixeira Pinto, Parry G, Vieira P, Martins C, Costa-Pereira A. MedStatWeb: a Web-Based Course on Medical Statistics. Proceedings of the meeting Web-Based Learning Environments 2000:158

[7] Moodle - a free, Open source Course Management System for Online Learning. Available at http://moodle.org/

Address for correspondence

Ricardo João Cruz Correia (rcorreia@med.up.pt)

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Designing m-Learning for Junior Registrars – Activation of a Theoretical Model of Clinical Knowledge

Anne Marie Kanstrup[a], Niels Boye[b], Christian Nøhr[a]

[a] *Virtual Centre of Health Informatics, Dept. of Development and Planning, Aalborg University, Denmark*
[b] *Regional Hospital of Randers / Aarhus University Hospital, Denmark*

Abstract

The MINI-project aims at supporting junior registrars in the learning process of how to utilize their theoretical knowledge from Medical School in everyday clinical reasoning and practice. Due to the nature of the work – concurrent moving, learning and producing - we designed an m-learning application. This paper introduces the possibilities and challenges for design of the m-learning application based on a) analytical findings on learning and mobility as derived from the design case – an emergency medical ward b) theoretical perspectives on medical knowledge, and c) presentation of the design of an m-learning application. The design process was based on user-driven innovation and the paper discusses considerations on how to combine user-drive and generic models.

Keywords:

software design; user-computer interface; medical informatics applications; decisions support systems, clinical knowledge; physician practice patterns; clinical reasoning; diagnostic reasoning; handheld computers; guidelines; education; internship

Introduction

Handheld computers or personal digital assistants (PDAs) have become common in clinical settings, and are used by physicians as well as nurses [2; 6]. The most common functions for PDAs in clinical settings are to provide clinical reference guides for drug information, patient tracking and various clinical guidelines. Increasing wireless connectivity combined with more patient data in digital form, introduces new application areas such as electronic prescribing, real-time medical records access and point-of-care evidence-based literature searches [13]. This paper reports from the MINI-project which is experimenting with the use of handheld computers for mobile e-learning/ m-learning to support physicians' introduction to clinical work. The case is an emergency medical ward at a large regional hospital in Denmark.

The case: learning and mobility at a medical ward

During their first years of medical practice physicians need to operationalize their knowledge from medical school, in the terms of Dreyfus & Dreyfus going from "knowing that" to "knowing how" in stages from novice to competent [5]. The current practice for supporting physicians in this learning process is an apprenticeship process [10] where junior registrars make experiences in daily hospital production and 'reflections on action' with chief physicians. The 'reflection *on* action' [14] is organized formally in conversations taking place during shifts, x-ray and other conferences but can alternatively, if the problem justifies it, take place "on the fly". At these conferences physicians (junior registrars and chief physicians) discuss their experiences, diagnoses, the situation and status of patients at the ward, care plans, etc. When it comes to 'reflection *in* action' [14], the back-up of registrars and conversations between novice and experienced physicians is, however, more difficult especially due to time pressure and lack of resources. Photograph 1 displays the ward used as case and is taken during ward-rounds displaying a typical situation of contact between junior registrars (standing in a circle) and an experienced chief physician, here briefly passing by answering questions and giving suggestions and then on to the next patient. At a design-workshop focusing on learning resources when on duty, junior registrars expressed how "you are mostly alone when on duty and when making your decisions" and how "one of the difficult things to learn is to make decisions on your own". Their back-up for decision-making is carried in the pockets of the white-coat, which is stuffed with reference books, instructions and guidelines, personal notebooks, etc. as seen in photograph 2.

Photograph 1 - Chief physician giving brief advice to registrars on duty. Photograph 2 - The pockets/back-up of a registrar (2,6 kg).

The objective of the MINI-project is to experiment with design of m-learning [11] to support physicians especially in their first period of clinical work as junior registrars. The aim is to design ePockets to replace the paper pockets displayed in photograph 2 and more importantly to improve knowledge support for decision making by inexperienced but professional physicians: the right information, at the right time, at the right place, and right at hand via mobile technologies and m-applications.

Empirical work at the ward/with staff (observations, interviews, conversations, workshops, design-meetings) has pointed out the following important characteristics for learning and mobility in this specific design context [7]:

- Mobile means walking – the physicians walk while looking up information and interaction techniques for 'one the move-interaction' like one hand interaction, is important.
- Mobile means on the way to the next task – maximum time for looking up information or making notes is 3 minutes and navigation supporting 'easy to find' is important.
- Learning at work is filled with interruptions – the physicians are constantly interrupted and as a consequence working hard to keep track via personal notebooks on their patient. It is important not to design yet another interrupting artifact but rather a personal device supporting their tracking of their work.
- Learning takes place in situ when working. Consequently it is important that the learning content is related to this 'situ' (vs. general information from reference guides, etc.)
- These characteristics of mobility and learning challenge the dominant 'office-domain' and call for experimentation with interaction techniques and content development (e.g. from text towards AV or VR).

Methods: user-driven innovation and (generic) models

The primary outset for the design process has been user-driven innovation in order to root the design in its context [8]. The sustainability of the specific design is to be seen in this perspective – an information ecological perspective to design and use of information technologies as an interwoven part of practice, values, people and technologies [12]. User-driven innovation is a meeting-point for designers and users making it possible to swap roles and bring use-practice, values and users to the foreground of design-processes. Hence, user-driven innovation provides a nice point of departure for sustainable software development from an information ecological point of view.

To combine design and use, theory and practice, is, however, not an easy task solved by involving users in the design process. A significant but very difficult aspect of software development in a wider perspective than just one case (one information ecology) is the development of generic models which make it possible and easy to make, share, maintain, and not least understand applications. Experiences with development of Electronic Health Records (EHR) emphasize this difficulty where the concept of archetypes has been elaborated for many years and where the goal has been sustainable systems with an easy exchangeable knowledge component [3]. Currently, a similar picture is surfacing within the area of providing online instructions or guidelines to medical staff in Denmark. Several different projects are emerging most of them focus on transmitting information to the WWW or PDAs. In the MINI-project we have developed a theoretical model for clinical knowledge as the important basis for understanding the problem to be supported and not least as a basic model for the designed m-learning application. The following presents this model and how it has been activated in the design of our MINI-m-learning application.

Towards a model of medical knowledge

Our employed model of knowledge in the medical domain is inspired from Adolfsen's epistemological model of everyday problem solving, where knowledge is stratified in three layers [1]. We define medical knowledge as consisting of three interrelated layers.

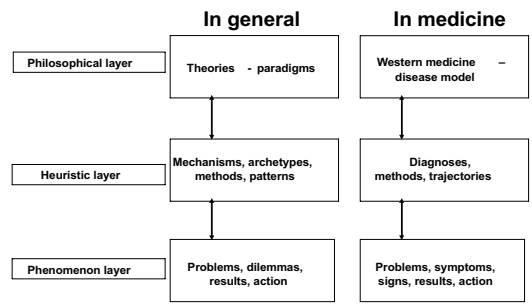


Figure 1 - The employed model of knowledge. The western medicine model is simple and says: "That to every disease or syndrome is a specific lesion, this can be functional, anatomical, physiological, biochemical, genetic, or social". Another model in the philosophical layer is Watson & Cricks model of DNA. For the other parts of the figure, see text for explanation.

The figure shows how the same dilemma can co-exist in three appearances of abstraction: in the bottom level the concrete everyday problem in contact with the everyday action and environment. In the middle heuristic layer, as a general problem, an archetype, a pattern, a diagnosis, or a method. In the upper model layer, as a theory or philosophy.

A medical example:

- bottom level – the patient's problem: palpitations corresponding to the doctors problem in this phenomenon layer: tachycardia (fast heart-rate)
- In the heuristic, archetypical layer the disease entities that can cause fast hearth-rate are located: e.g. thyrotoxicosis, fever or heart disease (there are many more).

- At the theoretical level are factors promoting a fast depolarization of the sinus node in the heart.

There are multiple other causes of fast hearth-rate than given in this example, this serves as an illustration of the different instantiations of the same problem and that "clinical reasoning" is translation and knowledge acquisition, activation, and operationalization (only possible in the phenomenon-level) in the different layers and this is not a strait forward algorithmic exercise. A recent review article "Educational strategies to promote clinical diagnostic reasoning" discusses this in detail mainly in the phenomenon layer [4].

Doctors need to master a problem in all three layers including the action relevant to diagnostic and therapeutic procedures. Patients have their primary interest and attention in the phenomenon layer, since the concrete problems are experienced in or on her body. Patient education efforts aim at bringing patient knowledge and reasoning to the heuristic layer, and thus provide them with "higher level tools and understanding" in cope and self-care of their disease problems.

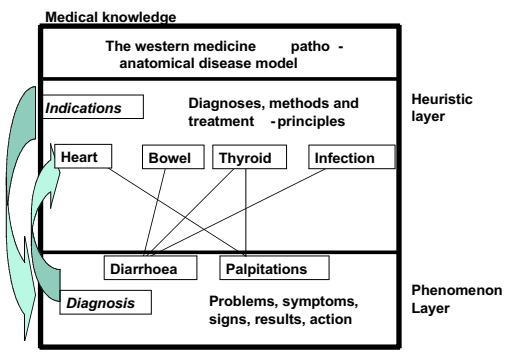


Figure 2: Example of the (non-algorithmic) dynamics of operationalization of medical knowledge. A patient presents diarrhoea and palpitations as problems. The doctor generalizes these problems to diagnoses in the diagnostic process. During this he activates knowledge from the upper theoretical layer. After the diagnostic process an indication of treatment is formulated employing both theoretical and heuristic knowledge about treatment principles. The principle is brought into action tailored to the specific problems and patient. The explanation for the choices made is "the indication" for the treatment.

Designing m-learning for medical knowledge

This perspective on medical knowledge put emphasis on support for the bridging between the different layers of medical knowledge – the relations – when designing m-learning for physicians in medical training i.e. focus on the heuristic layer as the meeting point for relations and support knowledge to and from this point. This perspective and focus is not new in the practice from the medical ward used as case. The heuristic layer is currently supported by the ward-instruction "the acute" organized in archetypes (heuristics for the ward e.g. 'the patient with fever, 'the arthritic patient', etc.) with relations to both the phenomenological layer (e.g. symptoms) and less prominent the philosophical layer (from University studies).

Consequently, based on both empirical analysis and the theoretical perspective on medical knowledge, the point of departure for the MINI-project has been to digitalize archetypes a) because it is learning content related to the specific use situation of the m-learning application b) because it supports the relations in medical knowledge grounded in the heuristic layer. It implies that the junior registrars are respected as professionals able to perform "clinical reasoning" and activate different layers of relevant knowledge regarding the specific patient – to fit a patient into a correct archetype and perform the necessary individual patient specific adjustments from this non-existing, average patient described in the archetype.

The transformation of existing paper archetypes into digital archetypes was carried out on the basis of a modified general work-flow model: with four "milestones" – all situated in the phenomenon layer: conclusion of the diagnostic interview (anamnesis, symptoms and signs) – often the basis for activating a specific archetype

1. Checklist of symptoms and signs contained specific ideas for further and alternative information relevant to a more precise positioning of the patient within the range of the archetype diagnoses.
2. Checklist of diagnostics, objective findings and test – contained relevant ideas for specific hospital procedures.
3. Checklist – treatment-plans
4. Checklist – monitoring and alternative actions.

In the technical dimension transformation of the paper archetype aimed at:

- A structure useful for a database
- Software useful as editor for producing archetypes within this structure
- A navigation design for MINI-archetypes/PDA-interaction

Microsoft Word™ was used as editor for production of the text for digital archetypes. Microsoft Word™ was used for producing the existing text for ward instructions and therefore well-known software to the archetype-producers (staff specialists). The archetype-producers were allowed to mark and prioritize the order in which they entered the text for later hyperlinking by means of a parser constructed by the programmer in the MINI-project.

The mini-mizing of the rather long text for each archetype (some were up to 25 regular pages in Microsoft Word™) was done by carefully designing the navigation for application. In order to work with the call for 'easy-to-find-information-while-walking-in-3-minutes' we have

- focused on the understanding of not only the screen of the PDA but also *the hardware as interface*. A short-

cut to the MINI-application was programmed and works by activation of a button at the front of the PDA.

- focused on gathering central information in *a MINI-frontpage* giving easy access to 1) instructions organized in archetypes with search function in the four areas presented in figure 3, 2) personal notes, 3) tables and 4) link to the Danish website for medical handbooks, drug catalogues etc. used very often – and carried in heavy books – by physicians. A click on the logo of Aalborg University always present in the top of the screen brings the user back to the MINI-frontpage (fig. 3a)
- focused on providing direct access to specific information via *search functions* facilitating free text search or providing access to information by means of lists of symptoms, diagnostics, treatments, and monitoring (fig. 3b)
- focused on providing overview of the rather long text for archetypes by giving a) an overview of headlines (designed in the editor and in co-operation with the archetype-writers/chief physicians) under which text can be unfolded by clicking the drop-down icon, b) giving 'at a glance' information on whether there are notes to the text by using a transparent icon of a document when no notes are available and making this icon clear when notes has been inserted. The same icon with the '+' is the icon used when inserting notes to the text (fig. 3c).

Fig. 3 - Screen-dumps from the MINI-application showing a) the frontpage of the application, b) search functions and c) a mini-mized digital archetype organized in headlines which are unfolded/folded by use of the drop-down icon and linked to notes by use of the note/paper icon.

A real challenge related to mobile interaction has not only been organizing of long texts but also mini-mization of rather large tables often used by physicians. Figure 4a shows an example of an often used table in 'real size' which, if minimized 1:1 would be impossible to read. An area in the table makes no sense if not related to other areas or at least to the columns and rows. Consequently we programmed the interaction in the table with fixed columns and rows and scroll within the areas (fig. 4b).

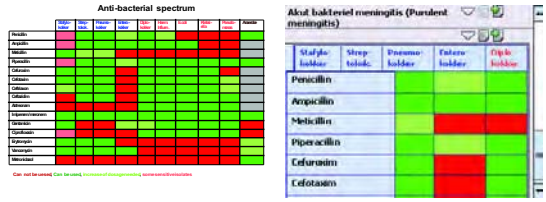


Figure 4 - a) a table in 'natural size', b) a screen-dump of the table in the MINI-application displaying the need for navigation here handled by keeping fixed columns and rows while scrolling in the table.

Ambitions of experimenting with voice interaction, primarily when making notes, failed due to technical problems but are on the top of a list for further research and development of the prototype.

Discussion: sustainability between user-drive and generic models

In the user-driven design process physicians were primary drivers behind the presentation and modeling of the clinical knowledge (their knowledge!) which was used in the design of the MINI-prototype by interaction designers. It is our impression that sustainable design cannot be identified at its source, in a (one!) practice (one information ecology) only. Likewise sustainable design cannot be validated on the generic potential of a model. Sustainability rather seems to lie in the interrelationship between the two – an interrelationship between user-drive and generic models which should not come as a surprise to software designers. However, it should emphasize (what cannot be emphasized enough) the importance of close collaboration between software developers and clinical personal in design processes. Further challenges and perspectives of the MINI-project are both the evaluation of the prototype but also inquiries into the use of PDAs at a national level. Both evaluations are taking place in the winter and spring of 2007.

Conclusions

The paper has presented results from the MINI-project on design of m-learning for physicians in hospital training. We have presented a) contextual findings on mobility and learning in the context of a medical ward, b) theoretical perspectives on medical knowledge, and c) description of our activation of this knowledge in the design of digital archetypes for PDAs.

It is argued that the user-driven innovation process of the project has provided a useful and valuable platform for sustainable development: it has both provided a platform for user-drive in the specific use-practice and provided the development of (steps towards) a generic model of clinical knowledge useful for further development and design projects. In other words, the user-driven innovation process has forced us to emphasise the combination of design and use and theory and practice – a combination which we have argued is essential to sustainable design.

Acknowledgments

We acknowledge the work of Pernille Bertelsen, Ellen Christiansen, and Thomas Christiansen and thank physicians at the ward for participating in observations, workshops, discussions, and pilot-test. The MINI-project is granted by the Danish Ministry of Science and Technology.

References

[1] Adolfsen, J. Problemer i videnskab, en erkendelsesteoretisk begrundelse for problemorientering (Problems in Science, an epistemological argumentation for problem-orientation), Aalborg: Aalborg Universitetsforlag, 1992

[2] Al-Ubaydli, M. Handheld computers. BMJ 2004: 328: 1181-1184

[3] Beale, T. Archetypes and the EHR. Stud Health Technol Inform 2003: 96: 238-44

[4] Bowen, J.L. Educational Strategies to Promote Clinical Diagnostic Reasoning. Medical Education 2006: 355: 2217-25.

[5] Dreyfus, H.L., Dreyfus S. Mind over Machine: the power of human intuition and expertise in the era of the computer. Oxford: Blackwell 1989.

[6] Fisher, S., Stewart, T.E., Mehta, S. Wax, R., Lapinsky, S.E. Handheld Computing in Medicine. Journal of the American Medical Informatics Association 2003: 10 (2): 139-49

[7] Kanstrup, A.M., Christiansen, E. Mobility at a Medical Ward: design challenges and decisions for an m-learning application. Proceedings of the Sixth Danish Human-Computer Interaction Resarch Symnposium (DHRS), Aarhus, Denmark, November 15 2006: 31-32

[8] Kanstrup, A.M., Christiansen, E. Selecting and Evoking Innovators: combining democracy and creativity. Proceedings of the 4th Nordic Conference on Human-Computer Interaction (NordiCHI), Oslo, Norway, October 14-18 2006: 321-330.

[9] Klass, D.J. Will e-learning improve clinical judgment? BMJ 2004: 328: 1147-1148.

[10] Lave, J., Wenger, E. Situated learning – Legitimate peripheral participation. Cambridge University Press 1991.

[11] Metcalf, D.S. mLearning – Mobile learning and performance in the palm of your hand. HRD Press, Inc. Amherst, Massachusetts, USA, 2006

[12] Nardi, B.A., O'Day, V.L. Information Ecologies – using technology with heart. MIT Press 1999.

[13] Rothschild, J.M. Fang, E. Liu, V., Litvak, I., Yoon, C, Bates, D.W. Use and Perceived Benefits of Handheld Computer-based Clinical References. Journal of the American Medical Informatics Association 2006: 13 (6): 619-626.

[14] Schön, D.A. The Reflective Practitioner – How professionals think in action. Basic Books 1983.

Address for correspondence

Anne Marie Kanstrup,
V-CHI/Dept. of Development and Planning,
Aalborg University,
Fibigerstraede 13, 9220 Aalborg, Denmark.
E-mail: amk@plan.aau.dk.

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Evaluation and Assessment of the Online Postgraduate Critical Care Nursing Course

Reena Patel

School of Nursing, University of Auckland, New Zealand

Abstract

During challenging times facing the health service, strategies for sustaining further education for nurses in highly specialised areas call for alternate means for learning. Nurses, who were accustomed to traditional methods of learning and had no formal computer training as part of their curriculum, are now being confronted with new methods of learning. Evaluation of the effectiveness of a newly developed postgraduate critical care course delivered online for nurses was examined. A pre test and post test of 16 participants were conducted. Participants found coursework intellectually stimulating and their preference to learn from websites demonstrated the effect size (tau-b = .677) to be significant at the 0.01 level. The pre and post test results from the knowledge assessment tool indicated an advancement of mean test scores and at a significant difference value of p = .055. Ninety four percent of the participants agreed that they were able to integrate their learning from the coursework towards their clinical practice. Improvement in nurses critical care knowledge impacts positively on delivery of safe and effective health care.

Keywords:

education, nursing, critical care, computer assisted instruction

Introduction

Development of knowledge in the field of health is synonymous with using computer technology. The existence of the Internet and the vast amount of information available at the fingertips is the reality for nurses today. Computer technology is used to provide further education to nurses employed in critical care environments.

Traditional classroom settings are being replaced by the use of web-based and resource supplemented packages for continued learning [1]. To meet the needs for access and convenience, postgraduate nursing courses are being offered via distance learning environment [2]. As new web-based nursing courses emerge, the emphasis to examine quality means that evaluation and assessment need to be considered.

The increasing reliance on technology has become more apparent within the health science disciplines and has been integrated into many courses. The use of technology in nursing education has been directly extended into both graduate and postgraduate programmes. Nurses, who were accustomed to traditional methods of learning and had no formal computer training as part of their curriculum, are now being confronted with new methods of learning.

These new methods of learning for students encouraged more autonomy in learning [3]. Previously, the use of chalk and blackboards were the primary modes of communication in classes. Later, the use of overhead projectors, whiteboards and markers came to use. Eventually computers were integrated into the teaching and learning process. Today, the convenience of a learner having access to a course material anytime and anywhere [4] has removed the barriers of time and distance previously imposed by traditional classes [5]. The increased emphasis on technology in health care and nursing education should facilitate the development of flexible, knowledgeable and technologically competent practitioners [4].

This paper discusses the effectiveness of a newly developed critical care course delivered online as part of a postgraduate certificate program for nurses. Research is required to evaluate the effectiveness of online instruction, in the development of new knowledge and skills [6]. Additionally, the study examined the critical care nursing course for enhancing knowledge attainment and for student satisfaction with web based learning.

Methods

The longitudinal study design involved one study group of sixteen students. All participants completed the Online Post Graduate Specialty Practice Adult/Paediatric Critical Care nursing course. The design involved data collection at two intervals, prior course commencement, and at course completion. The questionnaire examined student application to course material and evaluation on the effectiveness of the online course. Critical care nursing knowledge was assessed utilising a tool known as the Australian Basic Intensive Care Knowledge Test (ABICKT) [7]. ABICKT was used to test participants prior to commencing the online critical care course. On completion of coursework at the end of semester two the test was reapplied to determine any difference in scores.

Results

Learning with technology

Results were obtained from a series of questions related to learning with technology. It was important to ask questions to the participants pertaining to whether additional learning and training with technology was required prior to or during the course. The participants responded to questions related to the ease in navigating around the course and their satisfaction towards the technical access to the course.

Prior to commencement of coursework, an introductory seminar regarding the means of accessing coursework and assessments was demonstrated. Participants were asked in the post course questionnaire if they had found that they required additional technological learning when course content was commenced. Of the 16 participants, 10 (62.5%) of them agreed that additional learning was required. Only 4 (25%) participants strongly disagreed that no additional technological learning was required when commencing the course content.

The post course questionnaire sought data on responses from participants regarding the ease of navigation around critical care course materials. Course materials consisted of Internet web site addresses, Microsoft PowerPoint presentations, online journal articles from the university course library page and coursework set out in Microsoft Word documents.

Results from this question indicated that 11 (68.8%) of the 16 participants found it easy to navigate around course materials and 5 (31.3%) found it somewhat difficult. Participants were then asked if they were satisfied with the technical access to the coursework. Of the 16 participants, 13 (81.3%) were satisfied with the technical access compared to 3 (18.8%) participants who disagreed to this statement.

Correlations between data obtained from participant's experiences with technology and levels of support were considered. The most significant correlation seen at the 0.01 level was between the data related to participant's satisfaction with technical access to coursework and their ease of navigation around the coursework material. The correlation was investigated using Kendall's tau_b correlation. Expectedly, there was a positive association between these two variables, p= 0.01 and tau_b = 0.55 (see Table 1).

Table 1 - Correlation of the variables for online learning experiences

		Ease to navigate in course	**Satisfaction of technical access for coursework**
Adequately prepared for technology used in course	Correlation Coefficient	.532(*)	.512(*)
	Sig. (1-tailed)	.012	.014
	n	16	16
Ease to navigate in course	Correlation Coefficient	-	.549(**)
	Sig. (1-tailed)	-	.009
	n	-	16

* Correlation is significant at the 0.05 level (1-tailed).
** Correlation is significant at the 0.01 level (1-tailed).

Interestingly, when the strength of the relationship between ease of navigation in the course and the participant being adequately prepared for the technology used in the course were considered, there was a strong correlation at the 0.05 level, p= 0.12 and tau_b = 0.53 (see Table 1).

Correlation between a participant's satisfaction with technical access to the coursework and being adequately prepared for the technology used was also considered. Again the results from Kendall's tau_b analysis showed a significant positive association between these two variables p= 0.01 and tau_b = 0.51, at the 0.01 level (see Table 1).

Coursework satisfaction

On completion of coursework and assignments, the post course questionnaire sought the responses from participants in relation to their advancement in critical care nursing knowledge. Of the 16 participants, 4 (25%) strongly agreed with this statement and 12 (75%) somewhat agreed (see Figure 1). There were no participants who disagreed with this statement. Further, they were asked if they were able to integrate course learning with their clinical practice. Of the 16 participants, 15 (94%) responded by strongly agreeing and somewhat agreeing to the statement (see Figure 1).

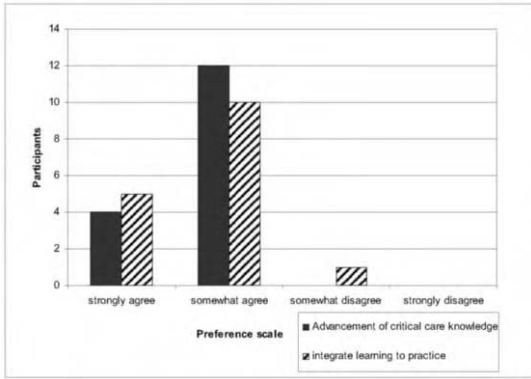


Figure 1 - Integration of advanced critical care knowledge to practice

The critical care course provided current world wide web (www) addresses as adjuncts to course learning. In the post course questionnaire participants were asked as to their preference for learning from these websites. Of the 16 participants, 11 (64.7%) preferred learning from websites and 5 (31.2%) indicated they did not prefer this method of learning (see Figure 2).

All participants were asked to indicate whether journal articles for each topic contributed to an increase in their understanding of the subject. Of 16 participants, 14 (87.6%) participants agreed that the journal readings deepened their understanding and 2 (12.5%) participants somewhat disagreed to this statement. There were no participants who strongly disagreed regarding the contribution of journal articles to the coursework learning and understanding (see Figure 2).

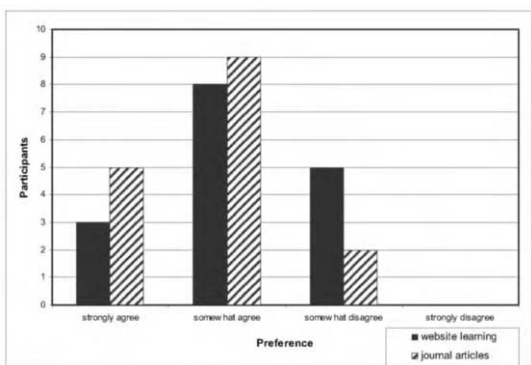


Figure 2 - Preferences of learning method

Importantly, there was a strong correlation between participants' use of online journal articles and their ability to integrate course learning to practice, p= 0.001 and tau_b = 0.726. Those participants who found coursework intellectually stimulating also reported a preference for learning from websites. This correlation was seen at a level of p= 0.002 and tau_b = 0.68, which was significant at the 0.01 level.

Critical care nursing knowledge test

Results from the pre course and post course critical care nursing knowledge test findings are presented. The post-test results revealed no changes in the number of higher scores (i.e. number of participants who achieved marks 80%). What it did reveal, however, was a positive shift of scores from the 60th percentile to the 70th percentile for several participants (see Figure 3).

The calculated median for the pre test results was 66, and the median for the post-test results was 71. This demonstrated an improvement of test scores post completion of coursework and assessments.

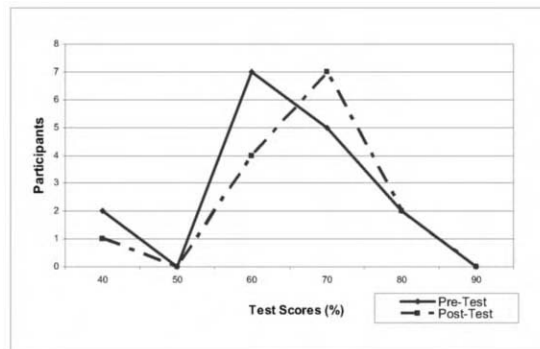


Figure 3 - Pre course and post course critical care knowledge test results

Wilcoxon Signed Rank test was used for analysis of the pre and post ABICKT results. It indicated that the two set of scores were approaching significant difference at p = 0.055. The median for the pre and post knowledge assessment test was 66 and 71, respectively. The median increased by 5% in the post-test. The majority of the participants achieved a score greater than 60% and below 90% in both the pre and post-test. There was a 10% increase in scores for some participants who scored 60% in the pre-test.

Discussion

The study examined the effectiveness of coursework learning and the application of the knowledge acquired to clinical practice. In nursing, the capacity to blend nursing education with clinical competency and problem solving skills has been a growing demand within the nursing profession. Dunn et al [8] and Morrison [9] have suggested that the ability to master the cognitive facts and apply the knowledge with critical problem solving and psychomotor skills in a clinical situation can be correlated with competency. The applicability of learning in the course to the clinical setting was relevant.

All participants in the course had indicated that they perceived their critical care learning had advanced. When questioned about how their learning was being applied to practice, 94% agreed that they were integrating their new knowledge to clinical settings. Nurses working in critical care particularly valued their training in physiology-based

decision-making. Saggs [10] highlighted this view and considered that physiology decision-making and practice skills were considered extremely relevant to postgraduate education in critical care. Equally significant, as discussed by Manias and Aitken [11], was the importance placed on nurses achieving technological competence in critical care.

In this investigation although there was a favourable outcome, one participant felt that their skills had not advanced in the clinical setting. In accordance to principles of Gestalt psychology, learning is perceived in different ways depending to the learner's experiences [12, 13]. The fact that this particular participant had not transferred new learning to clinical practice, suggests the level of exposure to clinical experiences need to be explored. Toth [14] concurred and proposed that increased knowledge occurs with more experience.

The relevance of content learned and its applicability to clinical practice was also examined in the study. Knowles et al [15] has pointed out that learning occurs not only when one is exposed to learning materials, but also occurs in the workplace. In the workplace, one is able to apply higher thinking processes leading to the interpretation and understanding of new learning. This has been well documented by learning theorists, who emphasised that adult learners need to know why they are learning something and then understand the meaning of it when applied to their real world [13, 16-18]. Content and intended learning should be related to the needs of the nurse in order for it to be applicable in practice.

Motivation to learn goes hand in hand with engaging the student with the learning material and ensuring the provisions of adequate intellectual stimulation. The results from this course evaluation showed that students perceived that the course had motivated them to learn, they had a clear idea of what was expected of them as they proceeded with coursework, which they were able to cope with. Online instruction engaged students in an interactive learning environment that extended beyond the classroom and increased their access to the wealth of educational resources online [6].

The online critical care course provided postgraduate study for nurses who were employed in critical care units. A variety of nurses were able to enrol into the course for example those who were new to practice, those seeking career development, or those who were re-entering into the critical care specialty. Students used technological competence to make sense of the critical care environment and to support and extend their existing practice. Wynd [19] proposed that critical care knowledge improved when nurses participated in specialised critical care courses which developed critical care experience. In this study, critical care nurses displayed an earnest desire to improve their knowledge and skills.

Interestingly, a strong association was seen between the variable of participant's use of journal articles for learning and the ability of participants to integrate course learning to clinical practice. A correlation of $p = .01$ and an effect size of tau_b - .726 was observed. Journal articles were all online and focussed on the topic objectives. Equally important was gaining information and learning from websites, yet only 65% of the participants preferred this mode of learning. This revealed differences in the manner students interacted with computer-mediated learning materials. Although learning involved printed material and computer mediated material, Benson [20] observed that learning via online instruction involved more than just obtaining lecture notes from technology mediated programs. It involved an integration of delivery options available for interactivity and learner engagement with the online content.

Interactivity has been considered to represent two-way pathway for the flow of information between the facilitator and the participant [1]. In this investigation, the high levels of interactivity have encouraged learners to participate in learning activities. As a consequence they have demonstrated enhanced retention, transfer of learning and satisfaction [20].

Conclusion

The use of technology in higher education has grown exponentially. In order to establish the worth and merit of these programs, careful evaluation needs to be conducted. Williams [21] has suggested that since there was no norm with which to compare how well learning with technology has worked, conclusions as to whether a program has in fact improved learning outcomes need to be examined critically.

The primary outcome of the research indicated that all participants perceived an increase of critical care nursing knowledge after completion of all coursework and assignments. Ninety four percent of the participants agreed that they were able to integrate their learning from the coursework towards their clinical practice.

The secondary outcome of the research indicated that the pre course and post course knowledge assessment test indicated an advancement of mean test scores and at a significant difference value of $p = .055$ in test scores.

As discussed by Williams [21], evaluation can prove to be a powerful partner for improving higher education, as long as relevant participants are involved in the process. It should be discussed in the context of its users and its purpose. To keep up with our changing society, nursing and nursing education should continue to develop innovative ways of teaching, reflecting the past but setting new directions for future knowledge and providing safe and effective health care.

Acknowledgements

This paper is based on research undertaken for Master of Health Science, Health Informatics, with thanks to Michelle Honey, fellow nurse and supervisor.

References

[1] Tegtmeyer K, Ibsen L, Goldstein B. Computer-assisted learning in critical care. Critical Care Medicine 2001; 29(8 suppl):N177-N182.

[2] Billings DM. A framework for assessing outcomes and practices in web-based courses in nursing. Journal of Nursing Education 2000;39(2):60-7.
[3] Jeffries PR. Development and testing of a hyperlearning model for design of an online critical care course. Journal of Nursing Education 2005;44(8):366-372.
[4] Magilvy JK, Smith MC. Chapter 9 Distance graduate education:The University of Colorado experience. In: Novotny J, editor. Distance Education in Nursing. New York: Springer Publishing Co; 2000. p. 136-152.
[5] Boyer SA. A comprehensive critical care course. Nursing Management 1996;27(6):32b-32f.
[6] Lashley M. Teaching health assessment in the virtual classroom. Journal of Nursing Education 2005;44(8): 348-350.
[7] Boyle M, Kenney C, Butcher R. Background information for users of the Australian basic intensive care knowledge test. Randwick, Australia: Prince of Wales Hospital; 1994.
[8] Dunn SV, Lawson D, Robertson S, Underwood M, Clark R, Valentine T, et al. The development of competency standards for specialist critical care nurses. Journal of Advanced Nursing 2000;31(2):339-346.
[9] Morrison S. Writing multiple-choice test items that promote and measure critical thinking. Journal of Nursing Education 2001;40(1):17-24.
[10] Saggs P. Developing postgraduate awards for critical care. Nurse Education Today 2003;23(4):307-314.
[11] Manias E, Aitken R. Achieving collaborative workplace learning in a university critical care course. Intensive and Critical Care Nursing 2003;19:50-61.
[12] O'Connor AB. Clinical instruction and evaluation: A teaching resource. Boston: Jones and Bartlett; 2001.
[13] Ally M. Chapter 1 Foundations of educational theory for online learning. In: Anderson T, Elloumi F, editors. Theory and Practice of Online Learning. Athabasca: Athabasca University; 2004. p. 3-31.
[14] Toth J. Basic knowledge assessment tool for critical care nursing version 4 (BKAT-4): Validity, reliability, and replication. Critical Care Nursing 1994;14(3):111-117.
[15] Knowles MS, Holton III EF, Swanson RA, editors. The Adult Learner. 5th ed. Houston,TX: Gulf; 1998.
[16] Knowles MS, Holton III EF, Swanson RA. The Adult Learner. 6th ed. London: Elsevier; 2005.
[17] Knox D. Learning Theories. In: McEwen M, Wills EM, editors. Theoretical Basis for Nursing. Philadelphia: Lippincott Williams & Wilkins; 2002. p. 321-344.
[18] Cutler LR. From ward-based critical care to educational curriculum 1: A literature review. Intensive and Critical Care Nursing 2002;18(3):162-170.
[19] Wynd C. Evidence-based education and the evaluation of a critical care course. The Journal of Continuing Education in Nursing 2002;33(3):119-125.
[20] Benson EP. Online learning: A means to enhance professional development. Critical Care Nurse 2004;24(1):60-62.
[21] Williams DD. Improving use of learning technologies in higher education through participant orientated evaluations. Educational Technology and Society 2002;5(3):11-17.

Address for correspondence

Reena Patel, School of Nursing, Faculty of Medical and Health Science, University of Auckland, Private Bag 92019, Auckland, New Zealand
reena.patel@auckland.ac.nz

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Development and Evaluation of a PDA-based Decision Support System for Pediatric Depression Screening

Ritamarie John[a], Penélope Buschman[a], Mosely Chaszar[a], Judy Honig[a], Eneida Mendonca[b], and Suzanne Bakken,[a,b]

[a] *Columbia University School of Nursing, New York, New York*
[b] *Department of Biomedical Informatics, Columbia University, New York, New York*

Abstract

Depression is under recognized in a variety of pediatric settings. The purpose of this paper is to describe the development and initial evaluation of a personal digital assistant (PDA)-based decision support system (DSS) for pediatric depression screening in ages 8 to 18 years of age by pediatric advanced practice nurse (APN) students. Three aspects are described: selection of depression screening instrument; integration of the instrument into the PDA; and quantitative (usage) and qualitative (focus group) evaluation. Only one third of eligible patients were screened. Twenty percent of those screened were identified as at risk for mood disorder. The barriers to screening identified through focus groups included a lack of time, knowledge, intervention protocol, referral resources, PDA usability issues, preceptor motivation and comfort, as well as perceived or real cultural barriers. Suggestions for educational, research, and interventions to integrate clinical based PDA-based screening are discussed.

Keywords:

depression, mood disorders, child, pediatrics, primary health care

Introduction

Childhood and adolescent depression is a serious public health problem [1]. Modern interest in child and adolescent depression arose in the late 1970 [2] and over the past 30 years, there has been an increased recognition of pediatric mood disorders [3]. It is estimated that depression occurs in 9 per 1000 preschoolers and 20 per 1000 school-age children [4]. The prevalence of a major depressive disorder (MDD) by late adolescence is ranges from 14 to 20% [5-8]. Depression is a risk factor for suicidal behavior [9], reduced social functioning [10], long term effects on neurocognitive functioning [11-12], poor school performance [13], serious conduct disorders [14], and drug use [15]. Untreated depressive disorders during childhood and adolescence can herald the onset of chronic and recurrent disorders during the person's lifetime [16-17]. Thus, early identification of children and adolescents at risk for a mood disorder is an important issue for clinicians.

While suicide can occur without previous depression in children and adolescents [18], depression is also sometimes a precursor to suicide During the last few decades, the suicide rate has steadily increased reaching a peak in the late 1990s [9]. Over the past few years the suicide rate has not changed substantially and at present, the rate of successful suicide is .6 per 100,000 in 5-14 year-olds and 9.7 per 100,000 in 15-19 year olds [19]. In addition, a history of prepubertal suicidal behavior predicts suicide attempts in adolescence [20-21]. Consequently, in screening for mood disorders, it is important to screen separately for suicidal ideation.

Health care providers in a variety of settings are in a unique position to identify, manage, and coordinate care for children with mental health disorders including depression [22]. Furthermore, the lack of child mental health specialists has placed an increased burden on primary providers to identify and treat children with behavioral health problems [23]. Multiple studies have confirmed that there is an under identification of mental health disorders in primary care settings [24-26]. If the provider does not use a standardized screening instrument, the detection of children with behavioral problems is lower [28-29].

At Columbia University School of Nursing, a PDA-DSS was designed to aid APN students to document and analyze patient encounters and to provide decision support for the screening and management of obesity, depression, and smoking cessation. A randomized controlled trial is in progress. The purpose of this paper is to describe the development and initial evaluation of the PDA-DSS for pediatric depression screening in ages 8 to 18 years of age by pediatric APN students. Three aspects are described: selection of depression screening instrument; integration of the instrument into the PDA; and quantitative (usage) and qualitative (focus group) evaluation.

Materials and methods

Selection of depression screening instrument

A Medline search was conducted combining the key words of mood disorders, pediatric depression screening, primary health care, and pediatric depression. In addition, an Internet search for pediatric depression screening was

conducted. The search yielded twenty different instruments with acceptable specificity and sensitivity statistics. Team members with expertise in pediatrics, psychiatry, and psychometrics evaluated the tools taking into account age range, length of questionnaire, and population that the instrument had been tested in.

Integration of the instrument into the PDA application

Questions from the screening instrument were added to the knowledge base that supports the PDA application. The content and algorithm were implemented for the Palm OS using the AppForge development environment.

Usage

SQL queries were developed to retrieve the data associated with depression screening from the central database. The following data were retrieved from the database for the time period of September 1, 2006 to November 26, 2006: the number of depression screening encounters versus the number of eligible depression screening encounters; medical diagnoses of those identified as at risk for a mood disorder and those not at risk for a mood disorder; characteristics of encounters in which pediatric depression screening occurred (patient demographics, family history of depression).

Focus groups

The objective of the two focus groups was to gain an understanding of the knowledge, attitudes, and beliefs of APN students regarding the use of a PDA-DSS. A purposive sample of 6 first and 6 fourth semester pediatric APN students were invited to participate in a 1-hour focus group. A 10-question, investigator-developed, interview guide was used to facilitate the discussion. The questions were developed from discussions with a panel of doctorally-prepared experts in informatics and qualitative research methods, as well as from a review of the qualitative literature on the attitudes of physicians [33-34] regarding PDA use in the clinical setting.

At the beginning of the group, each participant was given a copy of the interview guide. An experienced psychiatric clinical nurse specialist (PB) and an experienced pediatric APN (JCH) each conducted and facilitated one of the focus groups. An observer (RJ) took notes and audiotaped the discussion. During group discussion, the facilitator put forth a question and participants responded while interacting with one another. Probes were used to gain more detailed information and personal accounts. Data were content analyzed to extract major themes related to knowledge, attitudes, and beliefs.

Results

Depression screening instrument

After a careful review of the literature [30-32] and discussion among research team members, the team selected the Short Mood and Feeling Questionnaire (SMFQ) to measure risk for mood disorders rather than a longer diagnostic instrument. An addition four questions were added: two related to family history of depression [3] and two related to suicide [18].

Integration of the instrument into the PDA application

The SMFQ was displayed using 3-4 questions per screen (see screen shots below).

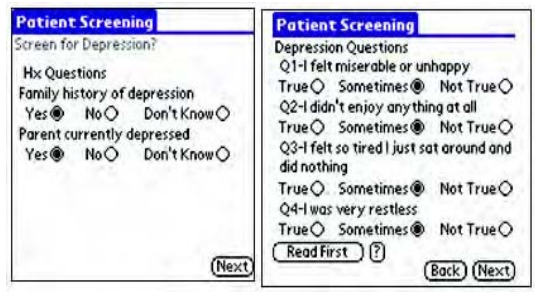


Figure 1 - Screen shots of application

The algorithm for the application is shown in Figure 2:

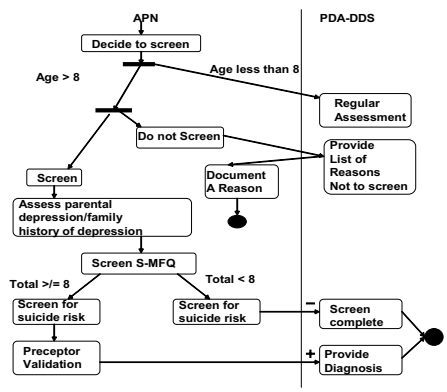


Figure 2 - Application algorithm

Usage

A total of 25 students were randomized into depression screening with 8 being fourth semester students and 17 being first semester student. In the fourth semester group, all students screened at least one child or adolescent for depression with an overall screening rate of 44.2%. The first semester students' overall screening rate was 37% with 23.5% doing no screening for depression. Students who failed to screen tended to be in specialty sites, ED follow-up clinic, or in private practices.

Of the 376 encounters eligible for pediatric depression screening, only one third (n=124) were screened for depression. If the student did not screen the child, they documented reasons for not screening. Over 81% cited other and about 8% noted a medical emergency as a reason for not screening. The remaining reasons included retardation (3.2%), learning disability (2.7%), patient refusal (2%), agitation (1.2%), other cognitive disorder (1.1%), and amnesic disorder (0.8%). Further explanation of the reasons for not screening in the other category is discussed in the qualitative data analysis.

Sociodemographic characteristics of patients in encounters in which screening versus no screening are shown in Table 1. Almost 80% of those screened were either Hispanic or Black.

Table 1 - Sociodemographic Characteristics in Screened vs. Not Screened (n=376)

	% Screened	% Not Screened
Ethnicity		
Hispanic	63.1	59.7
Black, not Hispanic origin	16.3	25.8
Asian or Pacific Islander	2.4	3.2
White, not of Hispanic origin	17.1	10.5
Other or unknown	1.1	0.8
American Indian or Alaska Native	0	0
Sex		
Female	51.6	52.4
Male	48.4	47.6
Age		
8-10	27.4	11.3
11-19	72.6	88.7

Overall, 20% of the patients screened were at risk for a mood disorder with 4% having a risk for a mood disorder and suicide. In the group at risk for a mood disorder, mean age was 13.1 years and 18.2% of patients were between 8-10 years. This was similar to the 13.4 years in the group at risk for suicide and a mood disorder. The latter group was totally female. No one screened was positive for suicide and negative for a mood disorder.

Within the group at risk for mood disorder, 35.3% had a family history of depression while 50% of those at risk for a mood disorder and suicide had a positive family history.

In terms of associated diagnoses, those identified as at risk of mood disorder had also had a high incidence of a psychiatric, behavioral or developmental disorders (n=44). The diagnoses with highest frequency were Attention Deficit Hyperactivity Disorder (ADHD), violence risk, and coping impairment. In the group at risk for a mood disorder, 68.2% had a behavioral or developmental disorder and only 31.8% had a diagnosis consistent with a physical disorder. However, of 467 diagnoses listed in the group not screened not screened for depression, only 6.6% had a psychiatric, behavioral, or developmental disorder diagnosis. The diagnosis with the highest frequency in the not screened group was well child (8.4%).

Focus group

Focus group discussions enabled exploration of students' knowledge, attitudes and beliefs about depression, depression screening, and the PDA-DSS. The major themes extracted for each of these areas are described in the following paragraphs.

Knowledge. Three main themes emerged during the focus groups: (1) limited, but some exposure to the problem of pediatric depression, (2) lack of a protocol within clinical site, and (3) lack of preceptor knowledge and support for depression screening. Many of the preceptors were not familiar or desirous of initiating screening.

Attitudes about depression screening. There were five related patterns centering on the theme of support that emerged during the focus groups. The students were concerned about a lack of time, knowledge, referral sources, comfort, and preceptor experience. The students expressed their need for more support from the site, preceptor, referral system, appointment system, and educational system to improve their knowledge base and comfort levels. The student felt it was inappropriate to do screening in emergency rooms or specialty clinics and wanted screening in primary care sites.

There was also a use and usability theme related explicitly to the PDA-DSS. Most students felt it interfered with the therapeutic relationship by creating a barrier between the patient and the student. One student commented, "The screening did not stimulate discussion. When I am done, there is not a transitional step." In terms of suicide screening, there was a similar theme of support. The students identified discomfort and apprehension regarding a positive screen. "You can ask the questions, but you must know what your next step is."

There were also cultural concerns raised since the students raised the issue of whether or not the patients felt comfortable with the idea of sharing their feeling. One student who was in a practice of recent Chinese immigrants commented, "Talking about feelings is not acceptable in this population."

Beliefs about barriers and benefits of depression screening. There was a discrepancy about the students' beliefs about depression screening and the barriers to depressions. Table 2 summarizes the themes identified.

Table 2 - Barriers and Benefits to PDA-Based Depression Screening

Benefits	*Barriers*
Prevent suicide *Enables sharing of feelings* *Opportunity to give holistic care* *Helps to pick up depression* *Improved quality of care*	*Time* *Perceived/real lack of referral resources* *Lack of preceptor knowledge and support* *Lack of knowledge of interventions* *PDA format* *Student discomfort with screening* *Cultural barriers*

Discussion

Clearly, there is a discrepancy between what the students believe are the benefits and their actual behaviors. Even though the students felt that screening should take place in a well visit, in 8.4% of the encounters in which no screening occurred, the diagnosis was a well child. In addition, only one third of eligible children or adolescents were screened. While recognizing that depression needed to be identified, the barriers overpowered the benefits. It is not clear whether the difference in the behavioral, develop-

mental, psychiatric diagnoses was a result of the greater awareness of the behavioral problems brought out by the depression screening or whether these children were presented with behavioral problems, cueing the students to screen.

This study has educational, clinical, and research implications. First, the students need to be empowered to intervene in children at risk for a mood disorder. Although teaching and counseling interventions are included in the PDA-DSS, students need further education about developing a therapeutic relationship and intervening with a single encounter. Educating preceptors about depression screening and interventions in a variety of settings may also promote screening.

The students need to improve their use of the PDA in clinical settings through strategies such as sharing the PDA screening with patients and entering initial information before entering the room.

Further research is needed to explore the patient's reactions to the PDA-based depression screening. It would also be important to compare different cultural group's reaction to the depression screening tools in paper versus PDA format. Further research is needed to assess whether the identification of depression helps students to identify other behavioral, developmental or psychiatric issues.

Conclusion

The data support that risk for mood disorder is fairly common in the population of primarily Hispanic and Black children and adolescents screened using the PDA-DSS and that there are missed opportunities for depression screening in a variety of settings. There are site, preceptor/student, and usability issues preventing depression screening. The organizational challenges are particularly difficult because APN students practice in different clinical sites with different resources and protocols. Since pediatric depression screening is relatively new, even experienced preceptors may lack the education to be effective primary contacts for the screening and identification of children at risk for a mood disorder or suicide.

Informatics innovations cannot be effectively implemented in practice without understanding the context of use. The triangulation of quantitative and qualitative methods was useful in helping the research team understand the screening behavior of students using the PDA-DSS and the associated organizational barriers in clinical settings. Moreover, these data provide direction for curricular changes and communication with preceptors regarding the use of the PDA-DSS.

Acknowledgments

This research is sponsored by a grant from the National Institute for Nursing Research (R01 NR008903). The authors thank the other members of the Mobile Decision Support for Advanced Practice Nursing for their advice and assistance. The authors also acknowledge the assistance of Dr. S. Ledlie in the design of the interview guide.

References

[1] Wells KB, Kataoka SH, Asarnow JR. Affective disorders in children and adolescents: addressing unmet need in primary care settings. Biol Psychiatry 2001;49(12):1111-20

[2] Pfeffer CR, Conte HR, Plutchick R, Jerrett I Suicidal behavior in latency age children: an empirical study. J Am Acad Child Adolesc Psychiatry 1979;18:670-692.

[3] Elliott GR Smiga S. Depression in the child and adolescent. Pediatr Clin N Am 2003;50:1093-1106.

[4] Besseghini VH. Depression and suicide in children and adolescents. Ann NY Acad Sce 1997;816:94-8.

[5] Lewinsohn PM, Rohde P, Seeley JR. Major depressive disorder in older adolescents: prevalence, risk factors and clinical implications. Clin Psychol Rev 1998;18(7):765-94.

[6] Demler O, Jin R Koretz D, Merikangas KR, Rush AJ, Walters EE, Wang PS; National Comorbidity Survey Replication The epidemiology of major depressive disorder: Results from the National Comorbidity Survey Replication (NCS-R). JAMA 2003;289:3095-105.

[7] Costello EJ, Mustillo S, Erkanli A, Angold. Prevalence and development of psychiatric disorders in childhood and adolescence. Arch Gen Psychiatry 2003;60(8):837-44

[8] Kessler RC, Berglund P, Demler O, Jim R, Merikangas KR, Walters EE. Lifetime prevalence and age of onset distributions of DSM-IV disorders in the National Comorbidity Survey Replications. Arch Gen Psychiatry 2005;62(6):593-602.

[9] Pompili M, Mancineli I, Girardi P, Ruberto A & Tatrelli r. Childhood suicide: a major issue in pediatric health care.Issues Compr Pediatr Nurs 2005;28(1):63-8.

[10] Lewinsohn PM, Rohde P, Seeley JR, Klein D N, Gotlib IH (2003) Psychosocial functioning of young adults who have experienced and recovered from major depressive disorder during adolescence. J Abnorm Psychol 2003;112:353-363.

[11] Arborelius L, Owens MJ, Plotsky PM, Nemeroff CB. The role of CRF in depression and anxiety disorders. J Endocrinol 1999;1600:1-12.

[12] Charmandari E, Kino T, Souvatzoglou E, Chrousos GP. Pediatric stress: hormonal mediators and human development. Horm Res 2003;59(4):161-79.

[13] Jellinek MS, Snyder JB. Depression and suicide in children and adolescents. Pediatr Rev 1998;19:255-64.

[14] Zalsman, G, Brent DA., Weersing VR. Depressive disorders in childhood and adolescence; an overview epidemiology, clinical manifestation and risk factors. Child Adolesc Psychiatric Clin N Am 2006;15:827-841.

[15] Greydanus DE, Patel DR. Substance abuse in adolescents: a complex conundrum for the clinician Pediatr Clin N Am 2003; 50:1179-1223.

[16] 2Lewinsohn PM, Rohde P, Seeley JR, Klein DN, Gotlib IH Psychosocial functioning of young adults who have experienced and recovered from major depressive disorder during adolescence. J Am Acad Child Adolesc Psychiatry 1999;38:56-63.

[17] Pine DS, Cohen P, Gurley D, Brook J, Ma Y. The risk for early adulthood anxiety and depressive disorders in adolescents with anxiety and depressive disorders 1998;34:566-578.

[18] Apter A King RA. Management of the epressed, suicidal child or adolescent. Child Adolesc Psychiatric Clin N Am 2006;15:999-1013.

[19] Eaton DK, Kann L, Kinchen S, Ross J, Hawkins J, Harris WA, Lowry, R, McManus R, Chyen D, Shanklin S, Lim C, Grunbaum J,Wechsler H. Youth risk behavior surveillance United States, 2005. MMWR. 2006 Jun 15; / 55(SS05):1-

108 Available from http://www.cdc.gov/mmwr/preview/mmwrhtml/ss5505a1.htm
[20] Pfeffer CR, Klerman GL, Hurt SW, Leser M, Peskin JR & Siefker CA. Suicidal children grow up: Demographic and clinical risk factors for adolescent suicide attempts. J Am Acad Child Adolesc Psychiatry 1991;30:606-616.
[21] Pfeffer CR, Klerman GL, Hurt SW, Kakuma T, Peskin JR, Siefker CA Suicidal children grow up: rates and psychosocial risk factors for suicide attempts during follow-up. J Am Acad Child Adolesc Psychiatry 1993;32:106-113.
[22] Kelleher KJ, Campo JV, Gardner WP. Management of pediatric mental disorders in primary care; where are we now and where are we going?
Curr Opin Pediatr 2006;18:649-653.
[23] Weitzman CC, Leventhal JM. Screening for behavioral health problems in primary care.
Curr Opin Pediatr 2006; 18:641-648.
[24] Williams J, Klinepeter K, Palmes G, Pulley A, Foy, JM. Diagnosis and treatment of behavioral health disorders in pediatric practice. Pediatrics. 2004; 114:601-606.
[25] Sayal K, Taylor E. Detection of child mental health disorders by general practitioners.
Br J Gen Pract 2004;54:348-352
[26] Brugman E, Reijneveld SA Verhulst f, Verloove-Vanhorick SP. Identification and management of psychosocial problems by preventive child healthcare. Arch Pediatr Adolesc Med 2001;155:462-469.
[27] Reijneveld SA, Brugman E, Verhulst F, Verloove-Vanhorick SP. Identification and management of psychosocial problems among toddlers in Dutch preventive childhealthcare. Arch Pediatr Adolesc Med 2004; 158:811-817.
[28] Simonian S, Tamowsi K. Utility of the Pediatric symptom checklist for behavioral screening of disadvaged children. Child Psychiatry Hum Dev 2001;31:269-278.
[29] Haughey MT, Claderon Y, Torres S., Nazario, S, Bijur P, Identification of depression in an inner city population using a simple screen. Acad Emerg Med 2005;12(12):1221-1226.
[30] Dahlberg LL, Toal SB, Swahn M Behrens CB. Measuring violence-rated attitudes, behaviors, and influences among youths: a compendium of assessment tools, 2nd ed., Atlanta, GA: Centers for Disease Control and Prevention, National Center for Injury Prevention and Control, 2005. Available from http://www.cdc.gov/ncipc/pub-res/pdf/YV/YV_Compendium.pdf
[31] Myer K, Winters N Ten-year review of rating scales. II: Scales for internalizing disorders. J Am Acad Child Adolesc Psychiatry. 2002;41(6):634-659.
[32] Angold A, Costello E, Messer S. Development of a short questionnaire for use in epidemiological studies of depression in children and adolescents. Int J Meth Psychiatric Res 1995;5:237-249.
[33] McAlearney AS, Schweikhart SB, Medow MA. Doctor's experience with handheld computers in clinical practice: qualitative study. BMJ. 2004; May 15;328(7449):1162
[34] McAlearney AS, Schweikhart SB, Medow MA. Organizational and physician perspectives about facilitating handheld computer use in clinical practice: results of a cross-site qualitative study. J Am Med Inform Assoc2005;12:568-575.

Address for correspondence

Ritamarie John, DrNP, CPNP
Columbia University
Mailbox 6
New York, NY 10032
Telephone: 1-212-305-5542
Fax: 1-212-305-6937
e-mail: rmj4@columbia.edu

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Data Mining Results from an Electronic Clinical Log for Nurse Practitioner Students

Patricia Trangenstein[a], Elizabeth Weiner[a], Jeffry Gordon[a], Ryan McNew[a]

[a] *Frist Nursing Informatics Center, School of Nursing, Vanderbilt University, Nashville, TN, USA*

Abstract

Traditional techniques for collecting data on clinical experiences have been greatly flawed. Data cannot be easily collected in real time to make programmatic or placement changes "on the fly". Furthermore, it is difficult to look at data across students, specialty areas, and years because the data is typically not in a digital format. In response to this problem, the Vanderbilt University School of Nursing has created a web/PDA based clinical log to document the kinds of clinical experiences the students are having. Since our initial report, three years ago, we have collected three years worth of data, over 220,000 different patient encounters. This past year the data has been very complete, giving a full picture of the types of experiences the students are having. Our faculty have begun to analyze the data in the clinical log to examine the kind of experiences the students are having and to make programmatic changes and placement adjustments in real time. In general, the results supported that students in the various specialties managed patients and performed services appropriate to their specialty. Patients varied in ages, ethnic groups, payment sources, and medical diagnoses. Students did progress from an observer role to a more independent role in either a linear fashion or in a more biphasic mode with an increase in the observer role at the start of a new semester

Keywords:

education, nursing, graduate; nurse practitioners professional practice; students, nursing; computers, handheld/utilization; education, nursing, graduate/ methods; nursing education research; preceptorship

Introduction

Traditional clinical education of nurse practitioner (NP) students has used preceptors for direct clinical supervision with faculty responsible for student progression, student evaluation and attainment of course objectives as well as curriculum revision. Nursing faculty collected data from nurse practitioner students to document their clinical experiences via on-site evaluation, discussions with the preceptor, paper logs and scan sheets completed at selected intervals during the course of study. These models led to a significant delay in receiving the information from students and providing appropriate feedback. In addition, it was very difficult to aggregate this data for a given student, or across students for a given semester or program because of the non-digital nature of the media Furthermore, it was difficult for students to examine their practice over time and for faculty to create reports and track student's progress.

A real-time method of data collection and analysis would allow for changes during, rather than after, the semester of clinical practicum. Faculty also thought that it was appropriate to map current evaluation methods against competencies being developed by the various specialty groups as well as the National Organization of Nurse Practitioner Faculty (NONPF). In short, the faculty wanted to re-design their clinical evaluation tool for the students in their specialty while preserving the ability to make comparisons across specialties in an aggregate format for purposes of accreditation.

A review of the literature revealed a scarcity of information with only one article describing the content and quality of a precepted clinical experience for nurse practitioner students. [1] As a result, the informatics team was invited to work with the specialty directors in the design, implementation, and evaluation of an online web based clinical log with PDA data input capabilities.

Methods

In 2004 VUSN began implementing an electronic (web/ PDA based) clinical log for nurse practitioner students. By 2006, the clinical log had been used by 200 students per year in seven different nurse practitioner programs for the entire length of their clinical experience. This article presents the results from the analysis of over 89,000 records collected during this past academic year. The implications for nurse practitioner education, curriculum revision, and student clinical portfolios for future employment are substantial. Not only was the log data used for program evaluation through data mining techniques, the students also used the log to create an electronic portfolio of their experiences and have, in some cases, secured employment as a direct result of their log entries.

Description of the system

The electronic clinical log was initially designed for data entry through the web. However, students raised the issue that they were entering data twice (once on paper patient side and again that evening as they transferred the data to

the log) a PDA interface was designed in response to their concerns. Students were instructed to synch their PDA with their desktops every night to avoid data loss. They were further instructed to upload their data once each week to the server based database. The records in the database were accessible via a web browser by students, faculty, and preceptors. The students could see only their own data, faculty and preceptors could see their students' data. Data collected included patient's age, gender and ethnicity; type of insurance, services rendered (professionally recognized criteria designed by the National Organization of Nurse Practitioner Faculty (NONPF)), ICD9 or DSM-IV codes and self assessment of the student's responsibility in patient management. Students and faculty could view the records and export logs to a spreadsheet for aggregation and graphing. Because of the unique nature of the Psychiatric/Mental Health Nurse Practitioner program, the PDA component for this specialty will be discontinued. Instead, next year, the PDA will be used to actually audio record the patient encounter, uploaded to the log, and made available for listening and comment by the faculty member or preceptor.

All electronic clinical logs (N = 114,206) collected during the entire year of study were subjected to data cleaning prior to analysis. If a student did not have data for the entire program of study all their records were eliminated in this analysis. After data cleaning a total of 89,401client encounters for 200 students remained for the purpose of this study. SPSS was used for all an analyses.

Results

Each time a student interacted with a patient it was treated as a single encounter. Therefore, if the student saw the same patient multiple times, as is common in the Psychiatric/Mental Health Nurse Practitioner specialty, each visit counted as one encounter.

The total number of patient encounters per nurse practitioner program was astounding and ranged from 2,072 for Pediatric Acute Care Nurse Practitioner students (PNPAC) to 34,700 for Family Nurse Practitioner (FNP). On average, the data showed there were 10 cases per week for each Acute Care Nurse Practitioner student and 21.4 cases per week for each Family Nurse Practitioner student. The mean number of encounters per student ranged from a low of 250 for the entire year for Acute Care Nurse Practitioner students to a high of 550 for Nurse Midwifery students. It should be noted that most of the specialties have an 8 month clinical rotation; however, the Nurse Midwifery program had a 13 month clinical experience. In general, Acute Care Nurse Practitioner, Pediatric Acute Care Nurse Practitioner and Psychiatric/Mental Health Nurse Practitioner students had fewer patients but longer encounters.

Acute Care Nurse Practitioner students and Adult Nurse Practitioner students (ANP) students saw older patients with a mean age of 55 while students in the Family Nurse Practitioner (FNP), Nurse Midwifery and Psychiatric/Mental Health Nurse Practitioner (PMHNP) programs saw middle aged patients (average age = mid 20's to early 30's). Students in all programs saw an almost equal percentage of males versus females except for Family Nurse Practitioner (FNP) and Psychiatric/Mental Health Nurse Practitioner (PMHNP) students who saw almost twice as many females as males, and of course the Nurse Midwifery students saw only female patients.

While the nurse practitioner students in all programs cared for many ethnic groups, the majority of the patients were Caucasian (49 – 84 percent across all programs), African Americans (11-24 percent across all programs) or Hispanic (1.5 – 24 percent across all programs). The majority of patient seen by Adult Nurse Practitioner (ANP) and Acute Care Nurse Practitioner (ACNP) students used either Medicare or private insurance as their source of payment, while the majority of patients seen by Family Nurse Practitioner (FNP), Nurse Midwifery, Pediatric Nurse Practitioner (PNP) and Pediatric Acute Care Nurse Practitioner (PNPAC) students had private insurance or TennCare (replacement for the state of Tennessee's Medicaid program).

The log allowed the students to enter up to four ICD9 codes per record. DSM codes were supported for the PMHNP students. Tables 1-3 describe the top ranking ICD9 or DSM-IV codes for student encounters across programs. While there were some similarities, the data revealed that students were caring for patients that were appropriate for their specialty. For example, Acute Care Nurse Practitioner (ACNP) students managed more CHF patients, and a significant portion of the Family Nurse Practitioner (FNP) students encounters were routine infant or child health checks.

Table 1 - Rank order of ICD-9 Codes for ACNP, ANP and FNP

ICD-9	ANP	ACNP	FNP
250, Diabetes	1	2	3
401, Essential HTN	2	5	
401.1, Benign essential HTN	3	1	1
272.4, Other hyperlipidemia,	4	3	
272, Disorders of lipoid metabolism	5		
428, Heart failure,		4	
401.9, Unspecified essential HTN			2
V20.2, Rout infant or child health check			4
477.9, 461.9 Allergic rhinitis or sinusitis, 465.9, Acute URI			5

Table 2 - Rank order of ICD-9 Codes for NMW and PNP

ICD-9 Codes	NMW	PNP
V22.1, Supervision of other normal pregnancy	1	
V22.2, Pregnant state, incidental	2	
V20.2, Routine infant or child health check		1
465.9, Acute upper respiratory infections of unspecified site		2
382.9, Unspecified otitis media		3
460, Acute nasopharyngitis [common cold]		4

Table 3: Rank order of DSM-IV codes for PMHNP students

DSM-IV Codes	Rank Order
296.9Mood disorder	1
309.81PTSD	2
314.01 ADHD combined subtype	3.5
296.33 Major depression without Psychotic features	3.5
296.8Bipolar	5
304.8 Polysubstance Dependence	6

The faculty members were also interested in the types of services their students provided to their patients. The services rendered were derived from the NONPF content competencies and were customized by specialty. Only those programs with some commonality with other specialties were reported here.

Acute Care students (ACNP and PNPAC) provided an average of 10 services per encounter while primary care student provided an average of 3 services per encounter. Tables 4 and 5 describe the highest ranking services provided as a percentage of the encounters. For example, Acute Care Nurse Practitioner

(ACNP) students identified and documented actual patient problems almost 72% of the time while Pediatric Acute Care Nurse Practitioner (PNPAC) students provided patient and family education nearly 80% of the time.

Table 4 – Types of service rendered as a percentage of encounters for acute care specialties

Services Rendered	ACNP	PNPAC
ID & documented actual pt prob	71.9	68.3
Analyzed all pharmacological agents	68.9	53.1
Collaborated with others	65.4	69.6
Obtained a H&P	63.8	58.9
ID & documented potential pt prob	57.7	63.9
Provided pt and family education	56.0	79.5
Developed an individualized Rx plan	54.4	69.5
Documented H&P	52.5	58.5
ID expected outcomes ind'l to the pt	51.2	59.2
Developed & documented list of differential diagnoses	48.5	57.6
Utilized evidence based practice	47.7	52.4
Documented ongoing eval of Rx plan		55.2

Table 5 – Types of service rendered as a percentage of encounters for primary care specialties

Services Rendered	ANP	FNP	PNP
Focused/episodic exam	84.6	85.7	68.3
Prescription	66.0	73.3	37.7
Health Education	47.6	71.9	55.7
Labs	45.0	35.6	18.5
Complete H&P			34.1
Developmental screening			25.2

The faculty members also wanted to evaluate the level of a student's responsibility in patient management, increasing from observational roles in early clinical experiences to more independent roles later in their clinical experiences. The student responsibility ranged from Observer, to Novice, to Beginner, to Advanced Beginner and had descriptors associated with each level for reliability purposes. Most of the students' clinical experiences transpired over two semesters. In order to determine if there was a natural progression from observer to more independent roles, the date of the encounter was converted to the month of the year indicating a progression in clinical experiences.

Correlations between students' responsibilities and the month were statistically significant for all of the programs at the .001 level. The correlations ranged from .192 for the Pediatric Nurse Practitioner (PNP) students to .396 for the Acute Care Nurse Practitioner (ACNP) students indicating that students did indeed progress from observer to advanced beginner during their clinical experiences.

While the correlations were statistically significant, they were low, indicating significant variability. In order to visualize where the variability existed histograms were constructed for each of the specialties. Figures 1 and 2 indicate two different paths from observer to advanced beginner that were noted.

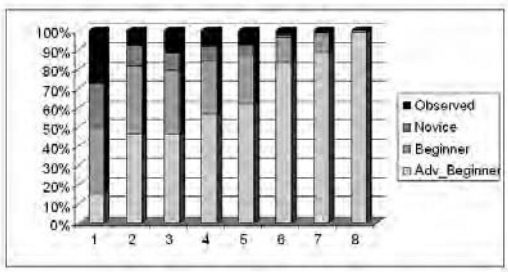


Figure 1 - Histogram of FNP students' responsibility by month of clinical experience

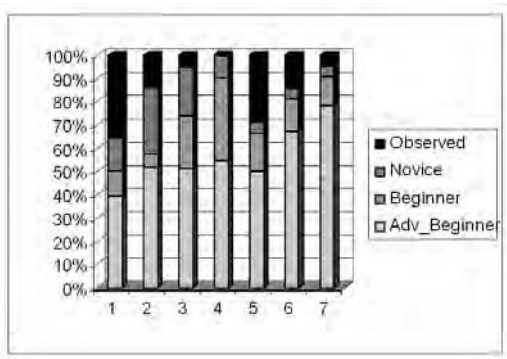


Figure 2 - Histogram of PNPAC students' responsibility by month of clinical experience

As can be seen from these two figures, the Family Nurse Practitioner (FNP) students showed a more orderly progression from observer to advanced beginner. On the other hand, the Pediatric Acute Care Nurse Practitioner students (PNPAC) students showed a more biphasic progression with an increase in the observer role in the fifth month. This corresponded to the beginning of a new semester and may reflect a change in preceptors, sites or performance expectations with the introduction of new skills. In any case, the faculty of the Pediatric Acute Care Nurse Practitioner students (PNPAC) program will look at the data to see if this trend continues and needs to be addressed or if it was an anomaly.

Discussion

NP faculty members have the ultimate responsibility for the supervision and evaluation of nurse practitioner students and for oversight of the clinical learning environment. [2] It is their responsibility to develop, evaluate and revise NP curricula. Students wish to secure the most favorable position possible and wish to make their case for such placement using data and evidence collect through the log. The NP faculty at the Vanderbilt University School of Nursing (VUSN) in conjunction with members of the Frist Nursing Informatics Center and with significant student input, developed electronic clinical logs (ECLS) for seven different nurse practitioner programs. Data presented represent a full program of clinical experiences for the students.

In general, the results supported that students in the various specialties managed patients and performed services appropriate to their specialty. Patients varied in ages, ethnic groups, payment sources, and medical diagnoses. Students did progress from an observer role to a more independent role in either a linear fashion or in a more biphasic mode with an increase in the observer role at the start of a new semester.

Conclusions

This article demonstrated the effectiveness of an electronic clinical log (ECL) for documenting the content and quality of a precepted clinical experience for NP students. Students entered data either by uploading from their PDA or by completing a web-based form. The records in the database were accessible via the internet by students, faculty, and preceptors. Data collected included patient's age, gender and ethnicity; type of insurance, services rendered (professionally recognized criteria designed by the National Organization of Nurse Practitioner Faculty (NONPF), ICD9 or DSM-IV codes and self assessment of the student's responsibility in patient management. Students and faculty could view the records and export logs to a spreadsheet for aggregation and graphing. Using informatics tools over 89,000 patient encounters for 200 NP students were analyzed.

Proposed revisions of the electronic clinical log (ECL) would include the creation of report dashboards to allow faculty members to determine "on the fly" whether the clinical experience was providing the right environment for the student. This would allow the faculty member to discuss with a preceptor the type of experiences the student was participating in as well as the quality of work the student was demonstrating. Currently a running total of the number of clinical hours has been well received by both students and faculty as a first piece of a dashboard. Given the volume of patient encounters the entire dashboard process would need to be automated with inadequacy flags automatically sent to the faculty member. We also plan on exploring the ICD9 data in detail to generate a culled-down list that would be more manageable for the students. We believe this shorter list will be far more useful to the

students who won't have to query through over 14,000 ICD-9 different codes.

Students have used the data to create a "clinical portfolio" listing their skills and activities across all of their clinical practice sites in their educational program and have analyzed their data using Excel. These advanced practice nurses will be participating in research and data collection/ analysis throughout their careers. A major benefit of our approach was that these students learned how to use Excel and demonstrated spreadsheet analysis techniques. The faculty were pleased that their students were as proficient with Excel as they are with Word.

An ECL will allow faculty to individualize learning based on the identification of gaps in the student's clinical experience. Faculty members can track the student's progress and types of patients seen. Data from the ECL can also be used to modify the curricula and provide documentation for grants and accreditation.

References

[1] Longworth, JCD, and Lesh, D. Development of a student nurse practitioner computerized clinical log. J. of the American Academy of Nurse Practitioners 2000:12 (4): 117-122.

[2] National Organization of Nurse Practitioner Faculties. Criteria for the evaluation of nurse practitioner programs: A report of the national task force on quality nurse practitioner education. National Organization of Nurse Practitioner Faculties, 2002.

Address for correspondence

Trish.Trangenstein@Vanderbilt.Edu

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Phase I Implementation of an Academic Medical Record for Integrating Information Management Competencies into a Nursing Curriculum

Carole A. Gassert, PhD, RN[a], Katherine A. Sward, MS, RN[a]

[a] *University of Utah School of Nursing, Salt Lake City, UT, U.S.A.*

Abstract

This paper is the report of the first phase of a case study from the University of Utah to help students and faculty integrate electronic information management into the nursing curriculum. Cerner AES, a live-production clinical information system with an academic overlay, has been implemented into the first semester of an undergraduate nursing program. A consortium of schools that use Cerner AES collaborate in the design and implementation of forms used by students. The consortium also allows members to share strategies for using the system. By using the system students are developing needed informatics competencies for beginning level nurses. The paper discusses the implementation strategies used and initial results of this project. Plans for expanding the project throughout the nursing curriculum are also presented.

Keywords:

electronic health record, nursing education, nursing informatics competencies, clinical information system, simulation

Introduction

Sustainable development can be defined as "development that meets the needs of the present without compromising the ability of future generations to meet their own needs" [1]. This concept has been adapted in the context of health care to define the capacity of a health system to continue over time without major interruptions [2].

Basic resources of sustainable health systems include data and knowledge, and the skills of the health workforce. Improving the safety, quality, and efficiency of health care systems requires ubiquitous access to complete patient information and decision support—that is, electronic health systems [3].

Information has become a capital good [4]. Nurses and other health professionals need skills that will allow them to more effectively manage information technologies. These skills must be learned from the beginning, during professional education [3, 5]. Staggers, Gassert & Curran [6, 7] identified information management competencies for nurses at multiple levels of practice. Despite recommendations by federal agencies that informatics content needs to be included in nursing curricula, information and information technology competencies have been slow to become part of the nursing curriculum [8] and many nurses may not be adequately prepared to manage information using technology [4, 9, 10]. A recent study of university students found that students' self-reports of information management activities were not an accurate predictor of their actual health information competencies [11]. Information management competencies should be assessed by observing actual use of an information system. Education regarding health information management should be conducted in the context of "real-world" applications and behaviors; that is, in the information environment where clinicians work, and should incorporate concepts such as confidentiality, systems thinking, and knowledge-resource evaluation [12].

One initiative that aims to increase the use of technology in nursing is the TIGER (Technology Informatics Guiding Education Reform) initiative [13]. This initiative, spearheaded by nursing and informatics leaders in the United States, has developed a 3-year plan to integrate informatics seamlessly into nursing.

Materials and methods

This paper describes the first phase of a case report at the University of Utah to incorporate information management competencies throughout the nursing curriculum by using a live-production application that simulates a clinical information system titled UCARE AES. The acronym stands for Utah Clinical Academic Record Excellence and is the name of the Cerner Academic Education Solution system installed for use in first semester undergraduate courses in May 2006. This on-going project is used to introduce curricular changes necessary to ensure that nursing students are adequately prepared to contribute to sustainable health systems. UCARE AES enables faculty to address many of the informatics competencies needed for beginning nurses identified by Staggers, Gassert, and Curran in their work. Examples of beginning level competencies that are taught in first semester undergraduate courses using UCARE AES are presented in Table 1.

Table 1 - Examples of beginning level competencies obtained through UCARE AES

Competency	Activity Example
Uses administrative applications for practice management	Searches for patient Retrieves demographics
Uses sources of data that relate to practice and care	Locate "patient data" in UCARE AES
Accesses, enters, and retrieves data for patient care	Charts on class activities in UCARE AES Creates plans of care in UCARE AES
Uses an application to document patient care	Uses UCARE AES to document results of class activities and simulated patients
Uses an application to plan care for patients to include discharge planning	Uses UCARE AES to document plan of care
Uses networks to navigate systems	Uses UCARE AES (hosted remotely)
Describes patients' rights as they pertain to computerized information management	Learns HIPAA regulations through UCARE AES
Identifies basic components of the current information system	Knows components of UCARE AES

Cerner academic education solution

An academic version of an electronic medical record is available through Cerner Academic Education Solution (AES), an application that simulates a clinical information system using Cerner's PowerChart. AES has an academic overlay that provides students with prompts and evidence-based practice information as they learn to document assessments and simulated patient events.

AES was first introduced at Kansas University [14, 15] and later at University of Missouri-Kansas City. With the introduction of UCARE AES at University of Utah, the three schools have formed a consortium to share design and management ideas and issues about AES. The consortium is a great forum for exploring new ways to use AES. Although the three schools share a domain server and databases that are hosted at Cerner headquarters in Kansas City, each school has a unique username and established structure. Many forms are shared as each school contributes their wisdom to the AES development process. All three schools are evaluating student outcomes. Monthly virtual meetings, held to discuss ideas and issues, are chaired by Cerner personnel.

UCARE AES implementation

It is interesting that each of the schools in the consortium has implemented the AES system differently. At the University of Utah planning for the phased implementation began in January 2006. Undergraduate faculty responsible for teaching the first level concepts of nursing, patient assessment and clinical care courses met, in spite of weekly snow storms, with the undergraduate program director and two informatics faculty in charge of the Cerner project to form an implementation team. The weekly meetings were held in a room that allowed access to and visualization of Cerner AES with an overhead projector. Minutes recorded decisions made by the team for later reference. The first task was for the entire implementation team to learn about AES and plan the structure for Utah. The team selected the name UCARE AES for the system and Swoopes Medical Center for our username. Units in the medical center were named for the focus of the semester, e.g., medical-surgical 1, medical-surgical 2, pediatrics, maternity, etc. and are created to accommodate the number of clinical groups assigned to the semester. For example, there are nine clinical groups in first semester so units Medical-surgical 1A through Medical-surgical 1I were developed. Students are then admitted as patients to their assigned "unit" and given a password to access the system.

To help students and faculty become more familiar with UCARE AES a number of "play patients" were admitted to the Utah Start Unit. Nicknamed the "Olympians" because their fictitious names reflect Olympic events, UCARE AES users were encouraged to chart anything on these patients, whether it made sense medically or not. Case studies for student use during the semester were created and admitted to a folder titled "UCARE 1." These patients were "off limits" for charting. When the semester began students located information needed for their assignments from the case studies, they charted the information requested on themselves as patients. To help students learn the principles of data security and confidentiality, they are warned that faculty could track where they have been on the system and if they are looking at fellow students charting.

Forms needed by students to complete their assignments were moved from the large repository of forms maintained by Cerner to specific folders designated by faculty in the first semester. They chose to have 5 folders – assessment, patient care, plan of care, clinical prep, and competencies. Most activities required forms from the first three folders. To make it easier for students to find needed forms, the Course Coordinator for the clinical course in first semester included instructions in her course syllabus. Examples are listed here:

Document the following on UCARE AES

- Vital signs: Located in Assessment Folder/Adult Vital Signs
- Results of incentive spirometer use: Located in Assessment Folder/Adult Respiratory Assessment/Incentive Spirometer (Left margin)
- Pulse oximeter results: Located in Patient Care Folder/Adult Vital Signs
- Teaching patient to TCDB: Located in Patient Care Folder/Adult Vital Sign/Oxygen Therapy/Document under "other."

- Use of O2: Located in Patient Care Folder/Adult Vital Signs

To allow the system built to be completed by Cerner in a timely manner, all design development was completed by mid-April for the May 15th go-live. First semester students and faculty were oriented to UCARE AES on the go-live day. Base-line data for first semester students' experience with and knowledge about information systems was collected at their orientation.

The system was implemented with accelerated baccalaureate students (students who hold a baccalaureate degree in another field and wish to become nurses) on their first day of clinical classes. Implementation team members and Cerner representatives were available during the first two days of clinical classes to help students sign onto the system and chart information required for class. The only problems encountered during the go-live were related to student IDs entered into the system, all problems were quickly resolved. Assistance was available in the Computer Lab to work with faculty and students as needed. During the planning phase, some faculty became champions for UCARE AES and were an essential part of the successful implementation of the system. It is also important to note that we had the full support of the Dean of the College of Nursing for implementing UCARE AES.

Since the implementation team is quite small, a decision was made to implement UCARE AES one semester at a time. As soon as the go-live was completed, second semester faculty responsible for clinical courses were asked to join the implementation team to plan for Phase II of the implementation. The plan is to have all four undergraduate semesters implemented by May 2007. Therefore, UCARE AES will have been implemented in the entire undergraduate curriculum in 13-14 months, an ambitious plan.

Results

Phase I student and faculty response

During the planning for implementing Phase I of UCARE AES, first semester faculty decided to tie the learning activities of the three courses more closely together. Timing of content was adjusted so students were learning the same concepts in all three classes. Faculty also began to standardize language used to teach concepts to students. During the design work, only one new form was created and one form was revised. Consortium members adopted the form that was developed for Utah. It seems that schools can use forms designed for others and such action will standardize some of the language used to teach concepts to undergraduate students across schools of nursing.

Data were collected at the beginning and end of the students' first semester. Data were also collected from the faculty at the end of the first semester. Specific data will be reported as aggregated data when more phases of the project have been completed. In general, faculty responses were mixed after the first semester of use; faculty reported feeling only moderately comfortable with how to use the UCARE system. Some faculty did not appear to grasp the purpose of learning to use an electronic record; with complaints about documenting on forms instead of free-text notes, and did not recognize the importance of teaching information management concepts, as reflected in comments such as "UCARE does not match the charting used at my clinical site". Other faculty recognized UCARE as a tool to teach information management, and commented that it was "easy to learn" and "an effective way to teach and learn terminology."

Our observations of nursing students concur with findings that student self-report of skills is likely not accurate. Incoming students rate their technology skills and knowledge as being low, and indicate they do not know how to find and manage information in an electronic health record, yet rate their information management skills as moderate. After one semester of use, the students reported higher technology skills and knowledge, and greater ability to find and manage information in an electronic record. However, these same first semester students also indicate that they want more guidance and direction with using the system.

The students were aware that their faculty were not yet comfortable with the system. Student comments were mixed, ranging from "busy work", "I don't get the point" and "it did not match the system at my clinical site"; to comments such as "Even though my clinical site has a different setup for their electronic records it still helped me" and "It was difficult at first to use not because of the system but because of my lack of nursing knowledge"; to comments such as "It was very cool to know about before starting clinical and to be able to say I know what to do - and how to get on and navigate the system."

Implementation status

The UCARE AES is an ongoing project that will eventually include undergraduate and graduate nursing, medicine and pharmacy students. Nursing is responsible for the initial design and has begun implementation at the undergraduate level. We recognize that repeated exposure is necessary to achieve a level of comfort with information systems and believe it will be crucial for both faculty and staff to continue regular interactions with the system; therefore we are incorporating use of the system into every semester of our undergraduate curriculum before moving into the graduate program with UCARE AES.

To date, it has been implemented into the curriculum for 3 (of 4) semesters of the undergraduate nursing program, although only Phase I is the focus of this paper. First semester students use UCARE AES to learn how to document individual "patient" assessments. Second semester students use the system to locate data in a medical record, synthesize information from multiple sources, and make clinical decisions based on the synthesized data. Third semester students currently use the system for pediatrics and maternity nursing. In pediatrics students assess scenarios on patient simulators (manikins) and obtain the "patient's" history from UCARE AES. In maternity students use UCARE AES to review a case study, prepare care plans, enter nursing orders, and practice reading fetal monitoring strips. All third semester students use a form

called SBAR (situation, background, assessment, and recommendation) to practice communication with nurse and physician providers. The implementation team is presently meeting with fourth semester faculty and mapping their curriculum to UCARE AES. As stated, all semesters of the undergraduate program will be live by May 2007.

Acute care nurse practitioner faculty are anxious to incorporate UCARE into their curriculum. They will be added as soon as the baccalaureate curriculum implementation is complete. For the nurse practitioner students, curricular content will be matched to the "experienced nurse" competencies of Staggers, Gassert, and Curran. The School of Pharmacy has already expressed a desire to be added to UCARE AES. The discussions with pharmacy are pointing to the use of collaborative and interdisciplinary case studies that focus on pharmacotherapeutics.

Faculty are becoming more comfortable using the system as their experience increases. In general, faculty are beginning to recognize the value of teaching electronic information management concepts and are continually verbalizing how they think they can expand the use of UCARE in their curriculum. To help faculty adopt more of their learning activities to UCARE AES, we anticipate adding an informatics student in the role of a teaching assistant (TA) to the implementation team to work with clinical teaching staff to assist them in using the system. The TA will also be available to work with students who may be uncomfortable using the system. Midway through the first semester we added a Systems Analysis position to the team. That individual helps to interface with Cerner to resolve any problems that arise and to request needed changes to the system.

Conclusions

Since UCARE AES has been added to the undergraduate curriculum it is clear that students are gaining competence in using information technology and in doing electronic data management. This will help with the development of sustainable health information systems. In addition undergraduate students are becoming aware of nursing informatics as a field and some have expressed interest in specialization in this area at the graduate level. We look forward to being able to report data that is being collected throughout the project as the remaining phases of the implementation are completed.

References

[1] WCED: UN World Commission on Environment and Development. Our Common Future. Oxford: Oxford University Press, 1987. Available online at http://www.are.admin.ch/are/en/nachhaltig/international_uno/unterseite02330/

[2] Madani S and Aronsky D. Factors Affecting the Sustainability of Information Technology Applications in Health Care. AMIA Annual Symposium Proceedings, 2003: 922. Available online at http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1479905

[3] Yasnoff WA, Humphreys BL, Overhage JM, Detmer DE, Brennan PF, Morris RW, Middleton B, Bates DW, and Fanning JP. Consensus Action Agenda for Achieving the National Health Information Infrastructure. *J Am Med Inform Assoc*. 2004;11:332–338.

[4] Gassert CA. The Challenge of Meeting Patients' Needs with a National Nursing Informatics Agenda. *J Am Med Inform Assoc.* 1998; 5(3): 263-268.

[5] Kinder G, Cashman SB, Seifer SD, Inouye A and Hagopian A. Integrating Healthy Communities concepts into health professions training. *Public Health Rep*. 2000 Mar–Jun; 115(2-3): 266–270.

[6] Staggers, N, Gassert, CA, and Curran, C. Informatics Competencies for Nurses at Four Levels of Practice. *J of Nurs Educ.* 2001; 40(7): 303-316.

[7] Staggers, N, Gassert, CA, and Curran, C. A Delphi Study to Determine Informatics Competencies for Nurses at Four Levels of Practice. *Nurs Research.* 2002; 51(6): 383-390.

[8] Ornes, LL and Gassert, CA. Computer Competencies in a BSN Program. *J of Nurs Educ.* 2007; 46(2): 75-78.

[9] Gugerty, B. Nursing at a Crossroads – Education, research, training, and informatics. *J of Healthcare Information Management,* 2007, 21(1): 12-14.

[10] McDowell, DE and Xiping, MA. Computer literacy in baccalaureate nursing students during the last 8 years. *CIN*, 2007; 25(1): 30-36.

[11] Ivanitskaya L, O'Boyle I, and Casey AM. Health Information Literacy and Competencies of Information Age Students: Results From the Interactive Online Research Readiness Self-Assessment (RRSA). Med Internet Res. 2006 Apr–Jun; 8(2): e6. Published online 2006 April 21. doi: 10.2196/jmir.8.2.e6 at http://www.pubmedcentral.nih.gov/articlerender.fcgi?tool=pmcentrez&rendertype=abstract&artid=1550696

[12] Staggers N, Gassert CA, and Skiba DJ. Health Professionals' Views of Informatics Education: Findings from the AMIA 1999 Spring Conference.J Am Med Inform Assoc. 2000; 7(6): 550-558.

[13] TIGER. The TIGER Initiative: Evidence and informatics transforming nursing: 3-year action steps toward a 10-year vision. Available online March 20, 2007 at http://www.tigersummit.com.

[14] Warren, JJ, Connors, HR, and Weaver, C. Clinical Information Systems and Critical Thinking: Outcomes of an Innovative Education Strategy. NI2003 Proceedings. 2003.

[15] Warren, JJ, Connors, HR, Weaver, C. and Simpson, R. Teaching Undergraduate Nursing Students Critical Thinking: An Innovative Informatics Strategy. NI 2006 Proceedings. 2006: 261-265.

Address for correspondence

Carole Gassert can be contacted at carole.gassert@nurs.utah.edu

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

ROC van Twente: Nursing Education in Care and Technology

William TF Goossen[a,c] Anneke TM Goossen-Baremans[b], Laura Hofte[c], Bert de Krey[c]

[a] *Results 4 Care, Amersfoort, the Netherlands*
[b] *Acquest Research, Development and Consulting, Koudekerk aan den Rijn, The Netherlands*
[c] *ROC van Twente, Enschede*

Abstract

The ROC van Twente offers nursing education at the diploma level (MBO), and is innovating the program to include a major /minor structure for education about care and technology. In order to achieve this, a new position was created: the Master Docent, Care and Technology. The task of the master docent includes development of education for nursing about technology, multidisciplinary cooperation, and service to health care institutions among others.

The first development concerns a module about electronic patient records, standards, and semantic interoperability for continuity of care. The module is delivered to nursing students and to students from the information technology department, who work jointly in 'development teams'.

This paper describes the background, the development of the educational material and program, and the core content of the module. The core content are the care information models that link clinical materials with health care information standards. The program has started end November 2006. At the Medinfo 2007 conference the results of the course for the first group of about 40 students will be presented.

Keywords:

health informatics education, HL7, vocabulary, standards, continuity of care, electronic patient records

Introduction

National infrastructures for health care information exchange present a challenge to clinicians to adopt many standards. However, clinicians usually did not get education about standards, and even in this age, the curricula of schools of health professionals usually do not include such topics in the overcrowded programs, despite the growing need for it.

Ageing populations and increasing numbers of people with chronic diseases change the focus of health care in such a way that the application of technology becomes essential. Example technologies applied include home care technology, video surveillance, smart homes, and information and communication technologies such as telehealth and electronic patient records (EPR). Increasingly, these technologies are integrated with each other, for instance based on internet standards. Thus, new kinds of healthcare emerge and for health care professionals there is an imperative to make sensible decisions about what technology to introduce in the care environment. The integration of technologies and the changes in health care delivery depend more and more on standardisation and quality assurance.

This paper discusses how this approach to development of electronic messages and EPR is chosen as the topic for nursing education and for the education of application developers. A new educational program is introduced in this area, currently offered by the school of nursing and the school of technology of the ROC van Twente (Regional Educational Centre of Twente).

Background

In the Netherlands, the activities of the National ICT institute for Health Care (www.nictiz.nl) [1] lead to the emergence of standards for electronic message exchange and development of electronic patient records [2]. The use of the EPR will be required by law in the near future. Thus there is a need for all health professions to learn using it, and to be able to support the development of EPR that address the clinician's needs.

The Netherlands have based their 'information for health care' strategy on the message paradigm, applying the international Health Level 7 version 3 (HL7 v3) standard for the safe exchange of patient information to authorized users via a national infrastructure. HL7 v3 has been used in about 20 projects now as a method to determine (clinical) user needs, modelling these needs, and implement clinical content in electronic messages. In addition, several vendors successfully base their electronic patient record systems on the HL7 v3 models.

A key part of the developments include the HL7 v3 messages for continuity of care: this is the care provision domain [3]. This care provision message is meant for referral, acceptance, record exchange, discharge summaries and so on. It is a generic structure covering a standard way to identify sender and receiver, the patient, the purpose of the message and the expression of supporting clinical data. Clinical details are expressed in the 'care statements' or 'clinical statements'. The care statements themselves can vary significantly, but the way they are

included in the message is consistent over different clinical domains. This has lead to the development of care information models [4, 5] that standardize clinical content in such a way it can be used and re-used in the HL7 v3 care provision messages [5]. The purpose of this approach is to realise semantic interoperability between health care information systems and technologies. Semantic interoperability is considered such that professionals receiving patient data electronically, clearly understand the meaning of the message and can adequately continue the required care.

The school of nursing in Twente is timely with this approach due to the fact that the Twente region is a national pilot for the electronic patient record [7]. Health care providers in this region have requested more education in the area of care and technology, specifically about the use of information and communication technology. The program described below has been developed with input from representatives of the health care providers. The program is considered a try out for both education within the school of nursing, and for continuing education of the existing nursing workforce. These students from the school of nursing of ROC van Twente will eventually work as nurses, but with additional knowledge and experiences with the EPR.

Design of the educational program

The master docent, care and technology

The ROC van Twente is positioned in the east of the Netherlands. Twente is the front runner for the national implementation of the Nictiz spearhead projects: the medication record and the general practitioner to general practitioner record exchange. Therefore there is a perfect situational context for education. Students will be confronted with the developments in their practical / clinical traineeships.

However, for a school of nursing it is difficult to gain immediate expertise to start participation. Therefore a new teacher role was established: the first master docent for diploma level education. The master docent has responsibilities to innovate the education, in this example about care and technology, to bring in knowledge and experience, in this example built up on many projects in health standards [7], to involve the teachers, in this example via a project team, and to deliver service to health care facilities in the Twente region. Service to health institutions is delivered for instance via participation in requirements gathering workshops and traineeships.

To start the developments, a choice was made to keep it simple in the beginning, but at the same time take an example technology that is innovative and that relates to the national and regional developments of information and communication technology in health care. Thus, semantic interoperability was chosen as the leading principle to start the developments. Health care agencies in the region agree that they require nurses with skills and knowledge to participate in development of patient record systems and messages.

The care and technology module

The student groups participating in the minor care and technology include about 25 students from the school of nursing and about 15 students of the application developers program of the school of technology of the ROC van Twente. The teacher team is a multidisciplinary team of nurse educators and information technology educators.

An overall goal for the module that started November 23 is that both student groups, each from their own perspective, understand the process of determination of information requirements, standardization and development and implementation of electronic patient record systems and electronic messages. The care information models are intended as a framework that bridges the often existing communication gap between system users and system developers.

The minor program serves as a differentiation within the nursing and within the technical education. Therefore it is assumed that students do know the basics of nursing care and have experiences in traineeships before entering the minor program. For the technical students, an equivalent background in systems life cycle and methods applied is expected.

Specific learning objectives / required competencies include:

- Multidisciplinary cooperation
- Communication and active participation
- Analysis of the need for care (nursing students)
- Analysis of information needs (application developers).
- Development of a care information model that includes purpose, description of variables, codes, HL7 v3 model, and technical data specification.
- Development of functional requirements for a electronic patient record system for continuity of care.

In total, the program consists of 12 weeks of education. A total of 4 contact hours per week is presented in small working groups with a mix of nursing and technology students. In addition, the students need an average of 4 hours a week for reading and preparation of the teamwork. The program is presented in Table 1.

The first five weeks are about the need for electronic patient records and messages, and the content of the messages. Then the application developers fall back on their normal program, and use time in between to work on the functional requirements and system design. In the meantime the nurse students have a clinical traineeship. For six of the nursing students, the traineeship involves participation in a nursing system development. In this particular setting, the ROC van Twente, a home care agency, and a vendor work together to create a new traineeship.

After about 16 weeks, both the nursing students and the application developers come back to school and continue another 7 weeks of education on this module. These 7 weeks deal with the subjects presented in Table 1. They include continuity of care and requirements for electronic patient records and messages. Further, the development of a care information model is a core element. The students

finish the program with a presentation of requirements, design and examples. Teaching materials include a module and reading materials, based on the work for the national information and communication infrastructure.

Software: stroke care record system

Another teaching tool is the electronic stroke care record system by Portavita [8]. Portavita is a vendor that agreed to have their software for stroke systems, still under development, made available for the students of ROC van Twente. This is a cost neutral arrangement, where both parties benefit: ROC van Twente gets access to a professional clinical information system for education, the vendor gets exposure and feedback on the system, and educational materials developed around the system become available for clients.

Table 1 – Overview of the program for Care and Technology

Week 1. The need for the electronic patient record (EPR) and standards.	
Getting acquainted with each other. Nurses: describe what nursing needs in the EPR. Technicians: support the nurses and apply methods for requirements gathering and functional design. Determine how to cooperate for 12 weeks.	
Week 2. Analyse information in care	
Nurses: explain which data are required in a stroke care record system. Technician: interview the nurses to get the requirements for a system.	
Week 3. Care information model 1	
Study the care information model structure and start with making one example. Nurses the clinical and terminology part, technicians the model and technical specification.	
Week 4. Care information model 2	
Develop a care information model, including clinical, terminology, HL7 v3 model and technical specifications.	
Traineeship intermezzo	
Nurse student	**ICT student**
Study the existing paper based methods for continuity of care during traineeship	Develop a functional design for a nursing record system
Week 5. Continuity of care for stroke	
Study needs for continuity of care and describe processes, professionals involved, roles, tasks and activities, **information** to be exchanged and apply UML modelling.	
Week 6. Continuity of care record	
Prepare a continuity of care record for stroke patients: contents, standards, sequence diagrams and functional design.	
Week 7. Review existing materials	
Select all relevant care information models from the repository www.zorginformatiemodel.nl	
Week 8. Classifications and codes	
Apply coding from standardised nursing and health vocabularies	
Week 9. Evaluate design against existing system	
Compare the functional design with existing system for stroke care	
Week 10. Functional Design	
Discuss the functional design Technology student presents to working group and makes final adjustments.	
Week 11. Preparation	
Prepare a presentation for the whole class and for teachers Criteria include: 1) Agenda and minutes, 2) Even distribution of the work in the small group, 3) The subject of the presentation has been negotiated with and approved by the teacher, 4) Apply presentation software, 5) Include the following: care information model, functional requirements, evaluation of group work, functional design	
Week 12. Final assignment	
Present as workgroup the results in public	
Evaluation	Evaluate the course and work

Care information models as a core topic

In order to have concrete materials available that are manageable for the students, the use of the care information models [5] is taken as the lead during the minor program. The care information models serve as a reusable building block within the framework of HL7 v3 Care Provision messages [2, 3, 5].

Care information models combine different standards materials and create valuable content for intelligent semantic interoperability [5]. They function as a communication bridge between clinicians and technicians and

facilitate inputs into the technical development of electronic messages and EPR systems.

The document structure for the care information models consists of meta-information, detailed description of the clinical instrument, and the reason for its application in practice [5]. It specifies clinical care using professional evidence, uses standardized terminology and coding, uses standard (HL7 v3) information models, and specify at the detailed level the technical requirements for the clinical content. Thus, the technical implementation according the HL7 v3 message and data specification are included via mapping tables, which are useful for EPR development as well. In most documents, one for every item of clinical activities, observation, or instruments, the current components include, in a recently revised format, the following components [5, 9]:

1. Version management and authorship
2. Explanatory introduction about the use of care information models
3. Aim of the instrument, index, scale, act, or observation
4. Scientific foundations / evidence base or other foundation such as guidelines
5. Description of variables / data items / values
6. Working instructions for practice
7. Interpretation guidelines for the results
8. Information on the topic relevant for care process
9. References / acknowledgements
10. An example of the instrument (when available)
11. HL7 v3 message model and description
12. Mapping table from domain to standardized terminology and to HL7 v3 domain message model
13. XML message example (extensible markup language)
14. Copyright issues, such as licensing of source materials, allowed use of care information models
15. Screen designs / screen shots for the instrument
16. Remarks, e.g. if a Dutch version is different from English version of an instrument
17. Contact information: how to contact the author(s)
18. Disclaimer

A current overview of the 90 care information models, in the earlier – less complete – format, is available at the website: www.zorginformatiemodel.nl. [4]

Students work together in small groups to complete one (draft version) of a clinical relevant topic and specify it according to the above format for the care information model and complete the HL7 v3 specification.

Exam

The exam of the module consists of a presentation by the students of the design of a system for continuity of care, based on the professional and information standards. Nursing students must underpin their requirements from a patient care perspective. Application development students must substantiate their part with the analysis, modelling and design of a system (component) that meets nurses' requirements.

Future plans

Currently the module is taught to the two student groups. However, once the module is delivered, more work for the master docent and the project group is waiting. The following new developments are on the agenda for 2007 and 2008.

1. ROC van Twente wants to integrate the module to other health care providers' educational programs.
2. The content and assignments will be put into an electronic learning environment, thus making it available to students wherever they are.
3. The minor program will be made available for continuing education for the health care agencies in the region Twente.
4. Other ROC's (diploma schools) have joined or will join and a wider spread of the materials and the education on these subjects is in preparation.

Discussion and conclusion

National and regional developments of information and communication technology in healthcare, such as electronic patient records and electronic messages, are emerging in order to deal with the changes in the health situation of the Dutch population. Due to results of standardisation efforts, useful materials become available for education. The ROC van Twente decided to develop a new role, the master – docent care and technology – to assist in establishing an innovative program for nursing students and technology students.

The innovation includes several challenging areas. The first is a new kind of content: the sensible use of technology in the care environment. Secondly, the application of clinical, vocabulary, message and technical standards for exchange of information for continuity of care, based on a well established format of care information models. Thirdly, the use of electronic record software for stroke patients currently under development. Fourth includes the multidisciplinary student groups: nurses and technical developers, handling the same problem each from their own perspective. Fifth, a minor program within the existing education, indeed with a traineeship for currently few nurse students with a focus on system development. Finally, a project group from teachers, representatives of health agencies, and experts working together to achieve this.

It is an exciting area of developments and we will be proud to present the results of the first course during the conference.

Acknowledgments

We want to thank the participation of all involved teachers and heads of departments of ROC van Twente, and the members of the project team from the health institutions in the Twente

Region. Nelleke Plaisir of Acquest for her contributions to the module and reading materials.

References

[1] NICTIZ Nationaal ICT instituut in de Zorg. NICTIZ website. Available at www.nictiz.nl. Accessed November 2006.

[2] Goossen WT, Jonker MJ, Heitmann KU, et al. Electronic Patient Records: Domain Message Information Model Perinatology. *Int J Med Inf.* 2003; 70 (2-3):265-276.

[3] Health Level 7. Message standards. WWW documents. http://www.hl7.org/ Accessed November 2006.

[4] NICTIZ, Acquest and Portavita. Care Information Models. www.zorginformatiemodel.nl. [Webdocuments, accessed November 2006].

[5] Goossen WTF. Intelligent semantic interoperability: integrating knowledge, terminology and information models to support stroke care. In: Park HA, Murray P, Delaney C (Eds). Consumer-Centered Computer-Supported Care for Healthy People. Proceedings of NI2006. Amsterdam etc. IOS Press, 2006: pp. 435-439. Studies in Health Technology and Informatics, volume 122.

[6] Information Technology in de Zorg in Twente (2007). www.izit.nl [Web documents accesed March 2007]

[7] Goossen W. Representing clinical information in EHR and message standards: analyzing, modelling, implementing. HINZ Primary Care and Beyond: Building the e-Bridge to Integrated Care. Eds: Christopher Peck & Jim Warren. Publisher: Health Informatics New Zealand (HINZ) & Health Informatics Society of Australia Ltd (HISA), 2006.

[8] Portavita (2006). www.portavita.nl [Webdocuments. Accessed November 2006].

[9] Van der Kooij J, Goossen WTF, Goossen-Baremans ATM, Plaisier N. Evaluation of Documents that Integrate Knowledge, Terminology and Information Models. In: Park HA, Murray P, Delaney C, (Eds). Consumer-Centered Computer-Supported Care for Healthy People. Proceedings of NI2006. Amsterdam etc. IOS Press, 2006: 519-522.

Address for correspondence

William Goossen, Results 4 Care, De Stinse 15 3823 VM Amersfoort, the Netherlands, williamtfgoossen@cs.com, results4care@cs.com

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Multiple Measures of Provider Participation in Internet Delivered Interventions

Thomas K. Houston[a,c], Ellen Funkhouser[b,d], Jeroan J. Allison[c,d], Deborah A. Levine[a,c], O. Dale Williams[c], Catarina I. Kiefe[c,d]

[a] Division of General Internal Medicine, University of Alabama at Birmingham
[b] School of Public Health, University of Alabama at Birmingham
[c] Division of Preventive Medicine, School of Medicine, [d] Birmingham VA Medical Center

Abstract

Evaluation of Internet-delivered continuing education for health care providers requires appropriate consideration of their level of participation. To fully assess participation requires multidimensional measures, including factors such as the volume of participation (page views), frequency (visits), variety (components accessed by each provider), and duration (months of activity). We calculated crude and refined (adjusted for study design) measures and then compared these measures across three longitudinal Internet-delivered continuing education to health care providers (N = 429). We found that participation varied across study, varied by factor and varied by specific measure. Correlation between crude and refined measures within a factor and across factors differed significantly. Participation assessment of internet-delivered interventions varies by the selection of measure and factor. Further research assessing the potential for these measures to predict intervention effectiveness is needed.

Keywords:

internet, continuing education, educational measurement, quality of health care

Introduction

The Internet has promise as a new tool to increase translation of research into clinical practice.[1] If proven effective, low-intensity Internet-delivered continuing education interventions have a high potential for dissemination. These interventions seek to improve provider knowledge, motivation, and behavior and are often multi-modal with access to case-based education, decision support, and patient education materials. Many such interventions have the ultimate goal of improving quality and enhancing safety. Although evaluations of Internet-delivered continuing education for healthcare providers are increasing[2-8], no standard method of evaluating these interventions exists. To more fully understand the impact of these interventions, investigators need to measure participation, including tracking of all provider "encounters" with the program.

Tracking provider-intervention encounters, however, is challenging. Web tracking software most frequently tracks page views and visits.[9, 10] However, all page views may not be of equal value. A frequent goal of website developers is to create "stickiness" - a measure of repeat usage of the intervention.[11] Thus, investigators might want to assess frequency of use per month enrolled in a longitudinal intervention. A combination of measures may best represent actual participation.

Few previous studies of Internet-delivered continuing education have measured differences in participation. Also, it is unknown how different measures of participation are correlated. As a first step toward developing a scientific approach to measuring participation, we propose a four-dimensional model. The general factors include volume, frequency, variety of components accessed, and duration of access. Within this model, we developed crude and refined measures of each factor and then compared these measures using data from three longitudinal Internet-delivered continuing education interventions. Our goal was to determine how the measures varied by study and to assess the correlation of crude and refined measures within factor and across factor.

Methods

Study design

Our group is conducting three separate group-randomized trials of Internet-delivered educational interventions to improve practice patterns of health care providers. Funded by the U.S. National Institutes of Health (NIH) and Veterans Affairs Health Services Research & Development (VA), these trials have similar designs and objectives but different target populations of providers and disease foci. The National Heart, Lung and Blood Institute (NIH) MI-Plus, conducted in two southern states, and VA MI-Plus, a national study, are parallel interventions targeting different provider populations: Medicare primary care providers in two Southern states and clinic-based primary care VA providers, respectively. Both MI-Plus studies seek to improve care for ambulatory post-myocardial infarction patients with multiple co-morbidities. Funded by the National Institute on Drug Abuse (NIH) DentalTobaccoControl.net (DTC) targets dental providers and seeks to improve tobacco cessation practice patterns in dentistry.

We prospectively tracked participation by 429 healthcare providers enrolled in the intervention arms of these trials. Users (private practice physicians, nurse practitioners, physician assistants, dentists, and hygienists) were recruited to the websites through mailings, phone calls, and emails and enrolled over multiple months ranging

from six to twelve, depending on the study. For this analysis, tracking data for each study was collected through a specific calendar month in 2006 (March for NHLBI MI-Plus, August for VA MI-Plus, and September for DTC). As enrollment was rolling, total months of enrollment varied for individual providers (mean months of enrollment was 14 (SD 5) for NHLBI MI-Plus, 11 (SD 4) for VA MI-Plus, and 12 (SD 4) for DTC). Once enrolled, all providers received scheduled email reminders, including notifications of new content, to encourage participation. All providers received continuing education credits specific to their specialty (medical or dental) and a certificate of appreciation for their participation. Additional incentives were provided in the NHLBI MI-Plus study (access to online journals and a textbook) and VA MI-Plus (subscription to the *Medical Letter*). Each study was approved by the appropriate Institutional Review Board.

Intervention descriptions

The core of all three interventions was case-based educational programs using interactive, web-based modules with tailored feedback based on responses to questions. During the tracking periods, the content and number of cases varied by study (6 for MI-Plus, 8 for VA MI-Plus, and 3 for DTC). All three studies had an accompanying "toolbox" with practice tools and patient educational materials that could be downloaded. The two MI-Plus studies included 1) a literature watch segment updated at intervals with reviews of the literature and 2) a guidelines component with summaries of current guidelines applicable to post-myocardial infarction patients. The literature watch and guidelines of the MI-Plus studies were analogous to the headlines and library components of the DTC study. Feedback of performance data with peer comparisons was provided to NHLBI MI-Plus providers. In the DTC study, testimonials of provider's success in encouraging smoking cessation were included. For this analysis, we focused on intervention components that were consistent across the three studies.

User authentication was required for all providers as they logged onto the interventions. Thus, we used server tracking logs linked to visit to calculate the measures of participation. The log included an individual user identification number and was tagged with date and time.

Measures of participation

As noted above, we propose a four-dimensional model to evaluate participation. The four factors are 1) volume, 2) frequency, 3) variety of components accessed, and 4) the duration of activity. For each factor, we developed crude and refined measures for each factor. Volume measures included total number of page views (crude) and the refined measures (number of page views per visit and number of pages per month). Frequency measures included total number of visits (crude) and number of visits per month (refined). Variety measures included number of components accessed (crude). Because the central component of the three interventions was the cases, we also created variety measures specific for the cases: number of case modules completed (crude), and mean percent of case modules completed (refined). Duration measures included 1) number of months actively participating (crude), defined as months from first to last logon, and 2) mean percent of enrolled months (refined), defined as the proportion of enrolled time known to be active (months active/months enrolled) because of the variation in potential enrollment across studies. We defined five categories of providers (private practice physicians, VA physicians, VA nurse practitioners/physician assistants, dentists and hygienists).

Other measures of participation could be calculated. Specifically, our measures focused mostly on counts, not session time, or time per webpage. Using session time is challenging as our providers rarely logged off, but would simply close the site so that session time continued indefinitely. Also, frequent outliers for session time existed (over five hours of activity in a single visit) suggesting that providers would just leave the page open and go to another task. Thus, we have not used session time in this analysis.

Statistical analysis

We calculated the means and standard deviations of each measure. We assessed differences in participation measures by type of provider using t-tests. Because of the multiple comparisons, we chose a significance level of = 0.01. Then, we evaluated participation in each intervention component by provider group and assessed differences between provider groups using chi-square tests. Finally, using participants in all three studies, we assessed pair-wise correlations between measures within and across factors using Spearman's rank correlation coefficient. We then repeated this analysis for each individual study again with = 0.01.

Results

We recruited 429 providers from 344 practices. These included 108 private practice primary care physicians in the NHLBI MI-Plus study, 193 VA primary care providers (125 physicians and 68 physician assistants or nurse practitioners) in the VA MI-Plus study, and 128 private practice dental providers (68 dentists and 60 hygienists) in the DTC study. Across studies, the mean number of months since enrollment was 12.2 months (SD 4.5).

Results by study

Overall, across measures of volume, frequency, variety, and duration of participation, values tended to be higher for the VA providers, both physician and non-physician, compared with private practice physicians and dental providers (Table 1). The mean visits per month among VA physicians, 0.62 (SD 0.52), was twice that of private practice physicians, 0.29 (SD 0.22), P = 0.001; the mean visits per month for VA Nurse-practitioners/Physician Assistants was even higher, 0.85 (SD 1.5), compared with private practice physicians, P = 0.0002. The point estimate for page views per month was highest among dentists in the DTC study (19.5) but the standard deviation was quite wide (106.6), thus this estimate was not significantly different than the estimates amongst the other providers.

Variety of access as measured by number of components was lowest among the dental providers [1.5 (SD 1.2) for dentists and 0.97 (SD 0.89) for hygienists] compared with the other studies, but as measured by proportion of cases completed, participation in the DTC study was similar to the VA MI-Plus study. Duration of participation ranged from a mean of 7.9 months (SD 4.1) in NHLBI MI-Plus (the longest running study) to 2.8 months (SD 4.2) for hygienists.

*Table 1 - Mean Measurements (Standard Deviations) of Participation for Providers in Three Internet-delivered Intervention Studies**

	NHLBI MI-Plus	VA MI-Plus		DentalTobaccoControl.Net	
	Private Practice Physicians (n = 108)	VA Physicians (n = 125)	VA NP/PA‡ (n = 68)	Dentists (n = 68)	Hygienists (n = 60)
Participation Measures*	**Mean (SD)**	**Mean (SD)**	**Mean (SD)**	**Mean (SD)**	**Mean (SD)**
A. Volume					
Total number of Page Views	60.4 (51.4) [a,b,c,d]	100.3 (74.5) [a,e,f]	89.2 (73.3) [b,h]	35.1 (33.0) [c,e]	24.5 (25.0) [d,,h]
Number of Page Views Per Visit	16.4 (8.8) [c,d]	15.6 (8.2) [e, f]	15.6 (9.6)	9.7 (7.11)[c, e]	9.5 (6.0) [d,f]
Number of Pages Per Month	4.5 (3.8) [a,b,d]	9.3 (8.1) [a]	13.6 (30.4) [b]	19.5 (106.6)	2.6 (3.8) [d]
B. Frequency					
Total number of Visits	3.8 (3.2) [a,b,d]	6.7 (5.5) [a,e,f]	5.9 (5.4)	3.0 (3.7) [e,i]	2.5 (2.4) [d,f,i]
Number of Visits Per Month	0.29 (0.22) [a,b]	0.62 (0.52) [a]	0.85 (1.5) [b]	1.00 (4.0)	0.25 (0.27)
C. Variety					
Number of Components	2.3 (1.3) [a,b,c,d]	2.8 (1.2) [a,e,f]	2.7 (1.3) [b,g,h]	1.5 (1.2) [c,e,h]	0.97(0.89) [d,f,g]
Number of Cases	1.7 (1.6)	3.5 (2.5)	3.2 (2.7)	1.5 (1.3)	1.3 (1.2)
Mean Percent of Cases completed	28%(27) [a,c,d]	44% (32%)[a]	40% (33%)	50%(44%)[c]	43%(41%) [d]
D. Duration					
Months from First to Last logon	7.9 (4.1) [c,d]	7.0 (5.4) [e,f]	6.2 (5.5)	4.8 (4.9)[c,e]	2.8 (4.2) [d,f]
Mean Percent of enrolled months	51%(38) [d]	56% (37) [e,f]	53 (40)%	39% (38) [e]	24% (35) [d,f]

* Participation measured in four dimensions – volume, frequency, variety, and duration of access. Cells in the same row with the same superscript are statistically significantly different, p<0.01
Crude measures, ‡ nurse practitioners and physician assistants
§ Percent of enrolled months = (Months for first to last logon/total months since enrollment)

When comparing use of individual intervention components, the VA providers had high participation, private practice physicians had moderate participation and dental providers had lower participation (Table 2). The VA physicians participated in a higher percentage of the cases (84%), compared with the private practice physicians (71%; P< 0.01) and the dentists (68%;P<0.01). Notably, a much larger proportion of VA physicians accessed literature watch than private practice physicians (82% vs 38%, P=0.01). Similar results were seen comparing the VA non-physicians with all other categories. The dental providers had the lowest rates of accessing each of the four factors.

Correlation of measures

The correlation between individual measures varied from strong (0.93) to weak (0.11) (Table 3). In general, "pages per visit" was the least correlated with the other measures. Also, the duration measures had a higher number of moderate or weak correlations with the volume, frequency, and variety measures. Within each of the four factors, correlations also varied. Some of the weakest within-factor correlations were again seen for measures of volume. The pattern of strength of the correlations seen in Table 3 was not qualitatively different when the analysis was repeated for each individual study (data not shown).

Discussion

To document the effectiveness of Internet-delivered educational interventions that seek to translate research into practice, we need more randomized trials. To fully evaluate these trials, investigators will need to describe the level of participation in the intervention as a surrogate for exposure. We evaluated a number of easily calculated measures of provider participation in Internet-delivered continuing education interventions for healthcare providers and noted three general patterns of variation – overall patterns by study, patterns across study by similar measures, and patterns of correlation among measures.

Table 2 - Participation in Individual Components by Provider Type in Three Internet-based Interventions

	Private Practice Physicians	VA Physicians	VA NP/PA*	Dentists	Hygienists
	N = 108	N = 125	N = 68	N = 68	N = 60
Any Interactive Case	71%	84% [e,f]	81%	68% [e]	63% [f]
Literature Watch or Headlines	38% [a,b,d]	82% [a,e,f]	75% [g,h]	32% [e,g,i]	6% [d,f,h,i]
Guidelines/Library	52% [c,d]	44% [e,f]	50% [h]	12% [c,e]	3% [f,h]
Toolbox	63% [d]	69% [e,f]	63% [g]	45% [e,i]	23% [d,f,i]

Cells in the same row with the same superscript are statistically significantly different, $p<0.01$
* nurse practitioners and physician assistants

Overall, the measures suggested varying rates of participation by study, with VA providers having the highest participation. This may be because VA providers have relatively greater computer access. It may also be related to "horizontal" use of technology. All VA providers in the community-based outpatient clinics use computers to access an electronic health record; thus, use of the computer for Internet-delivered continuing education during routine workflow may be more feasible and acceptable than in the private practice physician and dentist offices where use of the computer for other purposes likely has considerably wider variation.

We also found that the crude measures had more statistically significant differences across studies compared with the refined measures. Our refined measures included adjustment for variations at the study level to account for differences in exposure to the intervention. For example, number of months of enrollment varied across the studies. Thus, the total time available for participation varied considerably. Also, total number of pages was greater in the MI-Plus studies compared with the DTC study. Thus, the point estimate for total pages favored the physician groups but "pages per month" for dentists was higher although not statistically significantly so. Conceptually, we feel the refined measures are likely more appropriate when comparing participation across studies because they adjust for differences in the interventions and may provide a more accurate comparison across studies.

Third, we found that the correlation statistics varied comparing the various measures both within dimensions and across the dimensions of volume, frequency, variety, and duration of participation. "Pages per visit" had the weakest correlations with the other measures. Duration of participation showed only moderate correlation with the adjusted measure of visits per month and pages per visit.

These measures demonstrate both the successes of the three interventions and some difficulties. All three interventions successfully maintained provider participation over a number of months. However, participation did wane, with most providers only participating for around 50% (or less for the DTC study) of the total time available. In the DTC study, providers less frequently accessed the non-case intervention components. We have used this information to increase marketing emails encouraging current DTC study participants to access the toolbox and other intervention content. Our efforts to maintain participant

*Table 3 - Spearman's Correlation Coefficients for Measures of Participation**

	a	b	c	d	e	f	g	h	i
Volume									
a. Total Pages	1.0								
b. Pages Per Visit	0.58	1.0							
c. Pages Per Month	0.89	0.57	1.0						
Frequency									
d. Total Visits	0.85	0.13	0.74	1.0					
e. Visits per Month	0.72	0.11	0.84	0.82	1.0				
Variety									
f. Number of Components	0.80	0.54	0.73	0.67	0.56	1.0			
g. Number of Cases	0.92	0.55	0.82	0.78	0.65	0.71	1.0		
h. Percent of Cases	0.78	0.45	0.68	0.68	0.577	0.55	0.90		
Duration									
i. Months from First to Last logon	0.72	0.17	0.79	0.54	0.53	0.53	0.63	0.56	1.0
j. Percent of enrolled months	0.72	0.19	0.62	0.79	0.65	0.51	0.64	0.58	0.93

* Spearman's rank correlation coefficient, bolded cells represent those Not significant ($p > 0.01$)
Percent of enrolled months = (Months for first to last logon/total months since enrollment)

intervention activity had varying effects, with wide variations in participation of individual providers within studies. For some measures, the standard deviations were equal or greater than the point estimates. Future studies may require targeted marketing strategies to individual providers or smaller targeted groups to enhance participation. Also, working with individual practices to understand barriers to participation and to facilitate access may be required.

Each of our four factors has some face validity, the measures demonstrate within study variation, and we have demonstrated that the measures differ across studies. This suggests that our measures capture differences in level of participation, as intended. Still, further analyses are needed to fully assess the multi-factorial model of participation that we propose. Additional analyses evaluating factors that may be associated with participation (number of computers, acceptability of technology, computer literacy) should be conducted to assess whether participation can be predicted. Most important, analyses need to be conducted to assess whether these measures of participation predict outcomes including changes in knowledge and performance, and which measures are better predictors of outcomes.

Prior educational literature suggests that approaches that provide education in small content amounts separated by time ("spaced approach") are more effective than large content approaches. Thus, one might hypothesize that frequency and duration of participation are both critical. However, these factors cannot be considered in isolation, as one might have high frequency participation in the intervention with low volume and low variety. In prior research, a model for web surfing – the foraging model – suggests that there are discreet patterns of participation[12]. Thus, the four factors might be combined to categorize users into certain patterns of participation. We have not attempted to combine these measures into a single participation factor because we feel that further empiric validation of criterion validity is needed first.

A recent manuscript measuring participation in a patient website[13] noted that participation is but one measure of engagement and that participation should be considered in the context of satisfaction with content, assessments of knowledge, and process measures that may potentially lead to the desired endpoints. We feel this also holds for provider interventions. To fully predict endpoints, measures other than participation will also be needed.

Our study was limited in that data from only three interventions were available. As noted above, this is a first evaluation of these measures, and includes only correlational evidence, but additional assessment of construct and criterion validity are the object of ongoing research. Differences in participation across studies may be due to variations in interest, time availability, computer experience, and ease and type of computer access, none of which we could assess in this analysis.

Conclusion

In conclusion, participation of providers in Internet-delivered interventions varies widely across multiple dimensions of measurement. Primary and refined measures of volume, frequency, variety, and duration discriminated across studies. Within study variation of these measures was also high, as noted by the high standard deviations. In our studies, these measures have been used to track ongoing participation, suggest modifications to website design and marketing, and will be useful in understanding the outcomes of these interventions. Our preliminary analyses add to the prior literature by identifying categories of participation measures, and begin to approach the complexity of assessing participation. Much further research is needed to develop and validate the science of measuring participation for Internet-delivered continuing education.

Acknowledgement

This project was funded in part by the National Institute on Drug Abuse (R01 DA017971), the National Institute of Dental and Craniofacial Research (U01 DE16747) and the National Heart, Lung, and Blood Institute, (R01 HL70786).

References

[1] Weingarten S. Translating practice guidelines into patient care : guidelines at the bedside. *Chest.* Aug 2000; 118(2 Suppl):4S-7S.

[2] Allison JJ, Kiefe CI, Wall T, et al. Multicomponent Internet continuing medical education to promote chlamydia screening. *Am J Prev Med.* Apr 2005;28(3):285-290.

[3] Curran VR, Fleet L. A review of evaluation outcomes of web-based continuing medical education. *Med Educ.* Jun 2005;39(6):561-567.

[4] Wutoh R, Boren SA, Balas EA. eLearning: a review of Internet-based continuing medical education. *J Contin Educ Health Prof.* Winter 2004;24(1):20-30.

[5] Stewart M, Marshall JN, Ostbye T, et al. Effectiveness of case-based on-line learning of evidence-based practice guidelines. *Fam Med.* Feb 2005;37(2):131-138.

[6] Fordis M, King JE, Ballantyne CM, et al. Comparison of the instructional efficacy of Internet-based CME with live interactive CME workshops: a randomized controlled trial. *Jama.* Sep 7 2005;294(9):1043-1051.

[7] Butzlaff M, Vollmar HC, Floer B, Koneczny N, Isfort J, Lange S. Learning with computerized guidelines in general practice?: A randomized controlled trial. *Fam Pract.* Apr 2004;21(2):183-188.

[8] Chung S, Mandl KD, Shannon M, Fleisher GR. Efficacy of an educational Web site for educating physicians about bioterrorism. *Acad Emerg Med.* Feb 2004;11(2):143-148.

[9] Creese G, Burby J. *Web Analytics Key Metrics and KPIs (Version 1.0)*. Washington, DC: Web Analytics Association; 2005.

[10] Peterson E. *Web Analytics Demystified: A Marketer's Guide to Understanding How Your Web Site Affects Your Business*. Seattle: Celilo Group Media; 2004.

[11] Khalifa M, Limayem M, Liu V. Online consumer stickiness: a longitudinal study. *Journal of Global Information Management.* 2002;10(3):1-14.

[12] Pirolli P, Card S. Information foraging. *Psychological Review.* 1999;106(643-675).

[13] Danaher BG, Boles SM, Akers L, Gordon JS, Severson HH. Defining participant exposure measures in Web-based health behavior change programs. *J Med Internet Res.* 2006;8(3):e15.

Address correspondence and reprint requests to:

Thomas K. Houston, MD MPH
510 20th Street South, FOT 720B
University of Alabama at Birmingham
Birmingham, AL 35294-3407
thouston@uab.edu
205-934-7997, Fax: 205-934-7797

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Medical Students' Knowledge and Perceptions of e-Health: Results of a Study in Sri Lanka

Sisira Edirippulige[a], Rohana B. Marasinghe[b], Anthony C. Smith[a], Yoshikazu Fujisawa[c], Walisundara B. Herath[b], M. T. M. Jiffry[b], Richard Wootton[a]

[a] Centre for Online Health, University of Queensland, Australia
[b] Faculty of Medical Sciences', University of Sri Jayewardenepura, Sri Lanka
[c] Niigata University of Health and Welfare, Japan

Abstract

The present study investigates the knowledge, perceptions and attitudes of medical students in Sri Lanka in regard to e-health. We also examined the barriers which impede them to develop knowledge and skills in e-health within their medical curriculum. A questionnaire focusing on the knowledge, attitudes and expectations of medical students towards e-health was distributed to all final year students (n=136) at the Faculty of Medical Sciences, Sri Jayewardenepura University, Sri Lanka. Response rate was 74%. 43% of respondents stated that they were familiar with the term e-health. 51% rated their knowledge of e-health applications as minimal. 88% admitted that they had no e-health education or training of any kind. Over 80% of all respondents thought that e-health had an important role to play in the current and future health sector, particularly in developing countries. Our survey revealed that respondents had very poor access to computers and Internet use was rare. 77% of respondents admitted that they were not provided with systematic knowledge and skills in e-health through their medical curriculum and identified the absence of formal education in e-health as a serious shortcoming.

Keywords:

medical education, e-health, curriculum development, developing countries

Introduction

The healthcare industry in general is under growing pressure to provide services more efficiently and economically [1]. As far as developing countries are concerned, problems and issues associated with the health services are even more daunting [2]. Extreme poverty has prevented governments in developing countries from funding health adequately resulting poor state of health in the populations. Poor health infrastructure, inadequate facilities and the shortage of healthcare professionals are characteristic features of the health sector in developing countries. Rural and remote communities – where the majority of the population live in developing countries - are particularly deprived of health services. Addressing health issues in developing countries has become a global concern [3]. However, there is no quick remedy to health problems in developing countries as the improvement of health infrastructure and the increase of the number of health professionals require long-term investments and policy implementations [4].

In this context, e-health has been identified as one alternative to provide improved services and specialist care in developing countries [5]. The potential of e-health applications, i.e. the delivery of health services across a distance by using information and communication technologies (ICT) is being recognised for its potential, especially for the delivery of health services to rural and remote areas [6]. International organizations such as the United Nations (UN) and the World Health Organization (WHO) have acknowledged e-health as a potential alternative to address acute health needs in developing countries [7-8].

Among many other factors, the acceptance of e-health by health professionals is extremely important if this new modality of healthcare delivery is to become an integral part of mainstream healthcare. Knowledge, acceptance and enthusiasm to utilise e-health as an alternative way of service delivery by healthcare professionals, particularly by doctors would help facilitate the integration of e-health. Needless to say, developing countries are one of the most likely beneficiaries of e-health. Thus, knowledge and skills in e-health and keenness to use this tool by local healthcare professionals would undoubtedly help address at least some of the existing problems.

There is a growing body of literature showing that e-health has a role to play in contemporary healthcare [9-10]. Also literature is evident that e-health is useful for developing countries [11]. However, there is not many studies done to investigate the level of knowledge among health professionals [12]. A few studies show the level of IT knowledge in medical students [13].

Sri Lanka is a developing country according to number of indicators. Like other developing countries, health sector in Sri Lanka has been challenged by various problems. Total expenditure on health as a percentage of GDP in 2002 was 3.7% which is significantly lower than that is in

industrial countries. Overall access to health facilities for people in Sri Lanka is low. For example the number of hospital beds per 1000 population constitutes 2.9. The number physicians per 10,000 population is 4.11 [14].

Rural population in Sri Lanka constitutes 77% of the total population [15]. The most complete hospital facilities and highest concentration of physicians are in the urban areas, while many rural and remote areas suffer severe scarcity of health services. Thus the territorial disparity in health services is a characteristic feature in the island (Table 1). Emergency transport of patients especially in the countryside is still at a rudimentary level.

Table 1 - Number of medical specialists in urban and rural hospitals in Sri Lanka[1]

Speciality	Urban	Rural	Total
Cardiologists	18	0	18
Neurologists	9	0	9
Psychiatrists	15	0	15
Pathologists	23	9	32
Dermatologists	10	1	11
Radiologists	29	6	35
Microbiologists	16	0	16
Occupational therapists	46	1	47

[1] Source: Sri Lanka Government Health Web Portal, 2005, (http://www.health.gov.lk/)

Aims

The objective of this present study was to evaluate the knowledge, attitudes and perceptions of medical students towards e-health since their preparedness is a key to the success in implementing e-health in developing countries.

Methods

We designed and distributed a survey to assess the knowledge and attitudes of medical students towards the broad subject of e-health. The survey was distributed to all final year medical students (136) studying at the Faculty of Medicine, Sri Jayewardenepura University (SJU), Sri Lanka. Questions were divided into the following sections: demographic details, knowledge in e-health, relevance to future practice, the use of computers and the Internet and access to e-health education.

Results

Demographics

A total of 100 (74%) students completed the survey. 54% of respondents were female. The majority of respondents (about 91%) were between the age of 26-30 years and the remainder were between 23-26 years of age.

Knowledge of e-health

Nearly half of all respondents (43%) admitted that they were familiar with the term e-health. However, 51% of respondents described their knowledge and skills related to e-health as minimal while 22% were unsure. 86% of respondents had had no exposure to e-health education and/or training. 71% of respondents said they had never read any literature on e-health.

Relevance of e-health

Questions were asked to examine the perceptions of the students about e-health. About 86% of all respondents admitted that e-health will have an important role to play in the current and future health sector. Only 2% disagreed with that statement while 11% were not sure. Again 86% of respondents agreed with the fact that e-health will be useful in their future practice. Only a very small number admitted that e-health will have no use in their future practice. 78% of respondents admitted that e-health applications will improve their services. Majority of respondents (77%) believed that e-health would have particular relevance to developing countries and 85% agreed that e-health should be encouraged.

Use of computers and the internet

Several questions were asked to establish the knowledge and skills of the participants in computing and the level of the Internet use. The results of the survey showed that the availability of computers and the Internet for students was low. They admitted that the access to computers and the Internet was limited both at home and at the university. Only a very small number (3%) of students had frequent access to computers and the Internet. The majority of students (65%) used the Internet very rarely. Nonetheless a large number of students admitted that they were comfortable using computers and the Internet. Also 67% admitted that they had formal computer education and training. The majority of students expressed the desire to have better and more frequent access to computers and the Internet.

Access to e-health education

41% of respondents admitted that they had received no satisfactory knowledge of e-health through their medical program while 36% were not sure. While 79% of respondents suggested that e-health should be included in the medical curriculum and 56% thought that e-health must be offered as an elective. About 85% of survey participants suggested that e-health course must include a practical component to provide hands-on skills. More than half of respondents (64%) expressed their willingness to study e-health at post-graduate level.

Participants of the survey also identified the lack of appropriate educational programs, financial constraints, lack of

Table 2 - Information and Communication technology distribution

Countries	Main telephone lines per 100 persons	Residential main lines per 100 households	Monthly subscription as % of income per capita	Personal computers per 100 persons	Internet users per 10,000 persons	Internet hosts per 10,000 persons
Low income	2.9	11.4	14.1	0.6	62.2	1.0
Lower middle income	13.6	35.8	2.9	2.4	264.9	4.3
Upper middle income	22.7	59.8	2.0	8.2	992.6	78.7
High income	59.7	108.8	0.7	37.3	3992.9	1484.2
World	17.1	54.9	5.7	7.7	820.8	232.6
Africa	2.6	9.9	12.7	1.0	84.9	3.4
Americas	35.1	80.6	3.1	26.6	2164.3	1332.9
Asia	10.7	41.8	5.5	2.2	433.9	28.7
Europe	40.5	80.0	1.1	17.9	1804.5	191.5
Oceania	40.0	98.3	3.7	39.9	2771.6	885.2

Source: International Telecommunication Union, World Telecom Indicators 2002

sufficient access to technology and traditional methods of medical education as major barrier to develop systematic knowledge and skills in e-health.

Discussion

There are no quick solutions to the complex problems in the health sector in developing countries. Among others, e-health has been identified as one possible solution to address some of these problems. Under right circumstances new technologies can improve the quality of care and efficiency of services. Governments as well as private sector around the world have become aware of the potential of new technology. But the enthusiasm of policy makers and investments in infrastructure only cannot enable e-health to enhance health services. The expansion of knowledge and skills in e-health at grass-roots level (among health professionals) and their acceptance of these techniques are imperative factors for e-health to become sustainable.

Our survey revealed that although the majority of students were familiar with the term e-health their knowledge and skill to practice this modality was extremely limited. Indeed, the limited access to computers and the Internet is a serious barrier in developing countries. Unlike in industrialised countries, computers are still a luxury in the developing world. Presumably this barrier has limited the advantages they may gain from new technologies. This limitation also represents a significant factor preventing them from acquiring necessary knowledge and skills in e-health. The so called 'digital divide' is still a formidable barrier to be overcome [16].

Despite the fact that students have limited access to computers and the Internet, the majority of them are computer literate. In fact, there is a growing interest in computers and the Internet in the developing world [12]. The students found that one of the main barriers for them to develop appropriate knowledge and skills in e-health was the absence of formalised e-health educational components in the medical curriculum. There is a need to provide knowledge about the fundamentals of e-health, basic concepts and various applications with particular emphasis on low-cost e-health modalities. Such education must also include a practical component to provide medical students with necessary hands-on skills. This preparation would enable students to choose relevant applications in their own practice suitable for their own circumstances.

Conclusions

Health systems in developing countries can be a potential beneficiary of e-health applications. Not only local governments, but also the international organisations such as the UN and the WHO have identified the potential of ICT to address the health needs of developing countries. Efforts have been made to promote e-health in developing countries by investing funds, initiating projects and introducing technology and improving infrastructure.

However, in this effort, e-health education has been the least attended area. The knowledge, acceptance and enthusiasm of local health professionals, particularly doctors are vital if e-health is to be a significant component of mainstream healthcare. Cultivation of a positive attitude towards e-health requires systematic education. Students must be provided with a formalised e-health education within their medial and health curriculum to establish knowledge in basic concepts, terminology, various e-health applications, successes and failures in current practice. Such education must also include a practical component to provide hands-on skills.

Undoubtedly the impact of digital divide is still a serious problem for developing countries. Concerted efforts must

be made to enhance the access to technology in these countries. The potential of low-cost e-health applications in developing countries is still untapped. In order to use low-cost e-health modalities to their full capacity one must have an appropriate knowledge in e-health and understanding of local needs. Formalised e-health education embedded into medical and health curriculum is needed to enhance the knowledge and skills of local health professionals.

Acknowledgments

We thank the University of Queensland (New Staff Start-Up Research Fund) for funding this research. We also thank the medical students of Sri Jayewardenepura University for their participation in the study.

References

[1] Lee K, Fustukian S and Buse K. Health policy in a globalising world, Cambridge, Cambridge University Press, 2005.

[2] Carr D. Improving the health of the world's poorest people, Health Bulletin 1, Washington, DC: Population Reference Bureau, 2004.

[3] Wagstaff A, Claeson M. Millennium Development Goals for Health, Nutrition and population: Rising to the challenges, World Bank Publications, 2004.

[4] Akin JS, Birdsall N, De Ferranti DM. Financing health services in developing countries, World Bank Publications, 1987.

[5] Latifi R. Establishing telemedicine in developing countries: from inception to implementation, Amsterdam, IOS Press, 2004.

[6] Lakovidis I. E-Health: Current situation and examples of implemented and beneficial e-health applications, Amsterdam, IOS Press, 2004.

[7] United Nations, World Public Sector Report: E-Government at crossroads, United Nations Publications, 2003.

[8] World Health Organization, e-Health tools and services, needs of the member states, Report of the WHO Global Observatory for e-Health, 2006.Harrison JP, Lee A. The role of e-health in the changing health care environment. Nurs Econ 2006 Nov-Dec;24(6)283-8

[9] Reed K. Telemedicine: Benefits to advanced nursing practice nursing and the communities they serve. J Am Acad Nurse Pract 2005; 17: 176-80.

[10] Kirigia JM, Seddoh A, Gatwiri D, Muthuri LH, Seddoh J. E-health: determinants, opportunities, challenges and the way forward for counties in the WHO African Region. BMC Public Health, 2005 Dec 20;5:137

[11] Edirippulige S, Smith AC, Young J, Wootton R, Knowledge, perceptions and expectations of nurses in e-health: results of a survey in a children's hospital, Journal of Telemedicine and Telecare, 2006, 12 (Suppl. 3):S3:35-38.

[12] Gschwendtner A, Netzer T, Mairinger B, Mairinger T. What do students think about telemedicine? J Telemed Telecare, 1999;3(3):169-71

[13] Samuel M, Coombes JC, Mirinda J, Melvin R, Young EJW, & Azarmina P. Assessing computer skills in Tanzanian medical students: an elective experience, BMJ Public Health, 2004, available from http://www.biomedcentral.com/1471-2458/37.

[14] World Bank, Sri Lanka Home, Data and Statistics, available from http://www.worldbank.lk/WBSITE/EXTERNAL/COUNTRIES/SOUTHASIAEXT/SRILANKAEXTN/0,,menuPK:287056~pagePK:141132~piPK:141109~theSitePK:233047,00.html.

[15] Norris P. Digital Divide, Cambridge, Cambridge University Press, 2001.

[16] Smith R. Closing the digital divide, BMJ, 2003;326: 238.

Address for correspondence

Dr Sisira Edirippulige
Centre for Online Health
University of Queensland
Level 3 Foundation Building
Royal Children's Hospital
Herston Road HERSTON Q 4029
(Fax: +61 7 3346 4705)
Email: s.edirippulige@uq.edu.au

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Importance of Public Health Informatics: A Survey of Public Health Schools and Graduate Programs in the United States

Janise Richards, MS, MPH, PhD

*Centers for Disease Control and Prevention, Atlanta Georgia, USA**

Abstract

This paper examines the importance of data, information, and informatics to public health practice. Forty public health academicians from 40 schools and graduate programs of public health were interviewed. All agreed that informatics was important to public health practice. A qualitative analysis of their comments revealed their beliefs on the importance of informatics skills and knowledge to the practice of public health. The resulting comment groups varied from "some skills are more important than others" to "need all the skills." Eight "importance" comment groups were formed: 1) skills for all professionals; 2) some skills more than others; 3) yes, they need all the skills; 4) skills to become better practitioners; 5) usefulness to practitioners; 6) communication with public; 7) they're [the public] are depending on us; and 8) the future

***Keywords*:**

public health; informatics training; workforce development; needs assessment; education

Introduction

Unlike medical informatics, public health informatics is not a well established field. [1] Because medical informatics developed into a field first, much of the literature in American journals regarding informatics reflects the 'medical model.' There is little literature that reflects the 'public health model.'[2] Often medicine and public health are thought of as interrelated.

Medical professionals think of public health as a subspecialty of medicines and public health professionals view medicine as an arm of public health. [3] Each point of view has merit. Since managed care has encouraged clinicians to emphasize prevention and to examine the use of population-based outreach services, the differences between the two are becoming less evident. However, the two have different primary foci; the patient for medicine, and the community for public health. Also the role of information, core functions, data and information sources are more wide-ranging in public health practice. [1,2,3,4]

The role of information in public health

The foundation of public health is information. The 1996 World Health Report cites the continuing need to "disseminate health information widely, in the shape of epidemiological and statistical data, reports, guidelines, training modules and periodicals." [5] In practice, all the core public health disciplines – epidemiology, biostatistics, behavioral sciences, environmental health, and policy and administration – are supported by information and information technologies. Furthermore, according to the Institute of Medicine (IOM) the substance of public health is organized community efforts aimed the prevention of disease and promotion of health. It links many diverse disciplines and rests upon the scientific core of epidemiology. [2] Since epidemiology depends on the effective collection, analysis, interpretation, and dissemination of information, information is the foundation of public health.

Core functions of public health

In the IOM model, three core functions are necessary for sound public health practice: assessment, policy development, and assurance. A fourth core function, continuous evaluation, links the three core functions. [2] Each of these core functions is data and information driven.

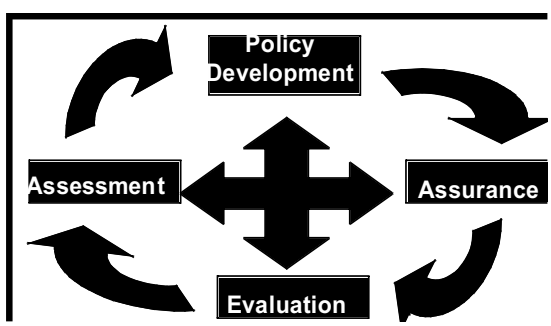


Figure 1 – Core Functions of Public Health

Assessment depends primarily on the core public health disciplines of epidemiology and biostatistics. [4] It consists of activities such as: surveillance; identifying needs; analyzing problems; collecting and interpreting data; case-finding; monitoring and forecasting trends; research and developing outcome indicators.

Policy development is a process by which public health professionals formulate, advocate and implement the appropriate action determined from the assessment of the situation. Policy development includes: decision-making;

mediating opposing views; negotiating common goals; developing programmatic goals and objectives; and allocating and mobilizing resources.

Assurance is the process that examines the quality of the situation, or impact of the policy or program implemented. Assurance activities include: determining the necessary services are provided to meet stated goals; regulating services and products; maintaining accountability and progress reports; and testing outcome indicators.

Evaluation is the ongoing assessment of the activities and processes. Evaluation data are gathered through a wide variety of sources.

Public health data sources

The public health core functions are performed in a data and information enriched environment. Data are collected through methods such as surveillance, surveys, and clinical assessments. The types of data generated include: demographic; mortality; morbidity; laboratory; physical assessments; health resources; health resource utilization; natality; disease occurrence; environmental assessments; and, public opinion data. There are as many sources for public health data as there are situations for assessment.

Public health information sources

Public health information sources also are unlimited. [6] Two common sources of categorized, meaningful data are Vital and Health Statistics and the National Health Interview Survey. Other more complete information sources include: the Morbidity and Mortality Weekly Report (MMWR); health department reports; health professions journals; newspapers; and magazines. Informal public health information sources such as phone calls, faxes, email, or meeting notes are a rich source of data and information about individuals and groups dealing with a specific public health situation.

Importance of public health information sources

Even with all the data and information sources available, public health professionals have difficulties obtaining the data and information they need. [5] Public health data and information sources lack standardization in organization, nomenclature and electronic transmission. Of the data are fragmented due to redundant collection or episodic data collection. Most notably, data and information is not always up-to-date. Timeliness of data is of importance. Nearly all public health actions depend on what is presently occurring in regards to a particular public health problem in a community.

If public health as a field is to become more effective, public health professionals need timely, quality information, better ways to communicate data and information, and betters tools to analyze new information. [1] Innovative methods of storing, organizing and disseminating the millions of pieces of data gathered during the core functions of public health need to be researched and tested. [6]

Because information is the major component of public health, and there is a need for public health professionals to know how to access, analyze and disseminate data and information. This study was designed to examine public health academician's belief in the importance of informatics for public health practitioners and the need to have informatics training in schools and graduate programs of public health.

Methods

This descriptive, qualitative study included 28 accredited schools of public health, 11 graduate programs in community health and 20 graduate programs in community health/preventive medicine which offer a MSPH or MPH degree within the United States.

A letter was sent to each school and program describing the study. A follow-up telephone call reached academicians at 21 public health schools, 14 graduate programs in community health/preventive medicine and 5 graduate programs in community health for a total population of 40 public health academicians. During this contact, one-on-one telephone interviews were scheduled.

Table 1. – Primary Contact in Study (N=40)

Number	Title
3	Dean
10	Associate Dean
1	Assistant Dean for Student Affairs
5	Professor
1	Department Head
4	Department Chair
5	Director of MPH Program
1	Associate Director of MPH Program
8	MPH Program Coordinators/Managers
1	Director of Student Services
1	Director of Distance Education

Data were analyzed from the interview surveys. The qualitative data analysis was conducted using a note-based analysis and open coding technique. [Note-based analysis involves immediately summarizing the notes take during the interview. Open coding describes the process of breaking down, examining, conceptualizing and categorizing the data.] Categories are the classification of concepts based on the comparison of one concept against another and those that appear to have similar characteristics are grouped together. [7]

Results

When asked "do you believe that informatics or informatics-related competencies are important for public health practitioners?" – 100% of the contacts said yes; several were emphatic. When asked why they believed informatics was important a variety of reasons emerged. Analysis of the comments revealed some interesting groupings.

Skills for all professionals

One subgroup of comments indicated these skills are necessary for any person to function in a professional capacity. This *skills for all professionals* subgroup had a global view of the need for informatics. The terms "computer age" and 'computerized society" were used to describe their belief; as one contact stated: "…you can't get along in the real world without these skills today."

Some skills more than others

Another subgroup of comments stated that some skills in informatics are needed by public health practitioners, but not all. This *some skills more than others* subgroup believe database use and development skills, information access skills, data and information analysis skills and an understanding of how communication technologies work is what should be taught. As one member of this subgroup succinctly stated: "…but not every practitioner needs to know all the details."

Yes, they do need all the skills

On the opposite end of the spectrum, the *yes, they do need all the skills* subgroup believes "there is a clear need to have expertise in informatics and the students need to understand the issues as well as the skills and knowledge." They believe the "shift to online based information is here to stay and [they] don't see any way to function in the public health arena without access to information technology sources." To teach these skills 'far outweigh the costs of the training and we need to support and train public health practitioners in this area."

However, "technology has gotten sophisticated so very quickly and those teaching need to understand the whole field of informatics and what it has to offer to public health. The faculty see only a narrow bit that is in their specialty area." In looking towards the future, they believe "we're not really preparing our public health students for the future" and "we think that it is criminal to graduate students without these skills."

Skills to become better practitioners

This subgroup believes informatics provides the tools to move the field of public health forward and there is a need for informatics *skills to become better practitioners*. They foresee that knowing when to use the "appropriate tools in each situation" will allow practitioners to be "more effective and cost-effective" in the practice of public health. Because this increased ability to be more effective, "informatics needs to become a basic professional skill." One person stated: "There is a clear need within public health to be able to use information technology."

Usefulness to practitioners

In this subgroup informatics *usefulness to public health practitioners* comments were categorized into different types: *information access*, *remote access,* and *cost-effectiveness*.

Widespread understanding of technology and information management concepts will "improve information access by practitioners" and allow practitioners to collaborate to "provide better services in their state or region." As one contact commented: "there is a lot of information to access and to be able to provide information in an intelligent way for others to find and use in practice, is a valuable need for public health."

Remote access was commented upon within this subgroup. In rural areas public health practitioners have limited access to information sources. The contacts believe information technology would greatly improve the practice of public health in these regions.

The cost effectiveness concept was stated by contacts in there belief that information technology has the ability to provide, "the latest developments in public health and has a lower cost and an easier accessibility than paper."

Communication with public

Another subgroup of comments characterizes the importance of informatics and information technology, the Web in particular, as methods to improve communication with the public. They believe it is "a vehicle by which we can communicate with the general public and with each other." They see expertise with information technologies necessary for public health practitioners to be able to "disseminate public health related information to the community and practitioners who want to implement intervention and prevention programs." A caveat to this improved communication is the frustration caused by lack of hardware in some populations and "there is a potential for creating a technology elite."

They're depending on us

An extension of the *communication with public* subgroup believes public health practitioners should take a leadership role in placing accurate health information on the web. The *they're depending on us* subgroup believes: "the world expects to find health information on the web"; the world expects that the health information on the Internet to be accurate"; and "public health has an obligation to provide accurate information [on the web]."

Future changes in practice

This subgroup believes the *future changes in practice* and information technology are inseparable. They believe graduates from schools and programs in public health "won't be employed in traditional public health programs, but in managed care organizations and the technology will become increasingly important in that arena," and "with managed care, from my occupational medicine point of view, data management is becoming more and more important to primary care." They also stated: "public health is moving away from the clinical services toward

the core areas of public health and within the core areas information is essential. For example, we need to maintain and access immunization registries to be able to accurately monitor health and well being of children and have the people with the skills to maintain them," and "in public health state departments [we're] moving from the delivery of services to the evaluation of programs which will emphasize the need for them [public health workforce] to have informatics skills and knowledge.´ This subgroups was very ardent in their beliefs. One contact summed up the thoughts persuasively by stating: *this is not a trend, but a modus operandi for the future*.

Conclusions

All forty contacts, regardless of their academic affiliation, stated they believed informatics and informatics related concepts were important to the practice of public health. They understand the importance of timely, accurate, and quality information. Yet, they seem to believe the information will automatically appear into information systems. To have the quality of data and information needed to practice public health effectively, the data must be gathered and organized in a standardized manner, and made available in an easily accessible form within information systems. The ability for public health practitioners to develop systems to ensure accessible, quality information should be provided through public health informatics training.

The emphasis on the information component is important, but there is value in understanding the technology. The distinct advantage of training in public health informatics is: The practitioner would not only have knowledge of what the data was needed for; they would be able to determine what technologies would be best suited to deliver the data and information expediently in the most useful format; and they would have the skills and knowledge to best organize the data or information to facilitate communication to other public health practitioners.

There is a need to develop information systems and technology to improve public health practice. Schools and graduate programs of public health are slowly developing courses in public health informatics to meet this need. [8] They believe informatics is important to the future of public health.

References

[1] Friede A, Blum HL, and McDonald M. Public health informatics: how information-age technology can strengthen health. Ann Rev of Pub Hlth 1995:16:239-52.

[2] Institute of Medicine (IOM). The future of public health. Washington DC: National Academy Press, 1988.

[3] Voelker R. Medicine and public health join forces. JAMA 1997: 277(20) : 1579.

[4] Afifi AA and Breslow L. The maturing paradigm of public health. Ann Rev Pub Hlth 1994: 15: 223-35.

[5] WHO. The world health report, 1996: fighting disease, fostering development. WHO 1997: 1-12.

[6] Friede A and O'Carroll PW. Public health informatics. In: Wallace, RB and Doebbeling BN eds. Maxcy-Rosenau-Last. Public Health and Preventive Medicine. 14th ed: Stamford Connecticut: Appleton & Lange 1998; pp 59-66.

[7] Strauss A and Corbin J. Basics of qualitative research: grounded theory procedures and techniques.

[8] Richards J. Informatics Training in Public Health. Masters of Public Health Thesis. The University of Texas School of Public Health. 1999.

Acknowledgments

Thank you to the generous support of the University of Texas system and the participants in this study.

*This study was completed at the University of Texas – this paper reflects the opinion of the author and not necessarily those of the Centers for Disease Control and Prevention.

Address for correspondence

Janise Richards, MS, MPH, PhD, National Center for Public Health Informatics, Centers for Disease Control and Prevention, 1600 Clifton Road, M/S E85, Atlanta, GA 30333

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Educating Medical Students as Competent Users of Health Information Technologies: The MSOP Data

Julie J. McGowan[a,b], Morgan Passiment[c], Helene M. Hoffman[d]

[a]*School of Medicine, Indiana University, Indiana, USA;* [b]*Regenstrief Institute, Inc., Indiana, USA*
[c]*Association of American Medical Colleges., Washington, D.C., USA*
[d]*School of Medicine, University of California San Diego, California, USA*

Abstract

As more health information technologies become part of the health care environment, the need for physicians with medical informatics competencies is growing. In 2006, a survey was created to determine the degree to which the Association of American Medical College's Medical School Objectives Project (MSOP) medical informatics competencies had been incorporated into medical school curricula in the United States. Methods: a web-based tool was used to create the survey; medical education deans or their designees were requested to complete the survey. Analysis focused on the clinician, researcher, and manager roles of physicians. Results: Seventy usable surveys were returned. Many of the objectives were stated in the schools' respective curricula and the competencies were being evaluated. However, only a few schools taught and assessed the medical informatics objectives that required interaction with health information. Conclusion: To insure that physicians have the knowledge, skills, and attitudes to effectively and efficiently interact with today's health information technologies, more medical informatics concepts need to be included and assessed in all undergraduate medical education curricula in the United States.

Keywords:

education, medical, undergraduate; medical informatics; hospital information systems; decision support systems, clinical

Introduction

Within the next decade, a large majority of hospitals and health care centers in developed nations and many in developing nations will have electronic health records and other forms of health information technology. Physicians will be expected to use these tools to improve patient safety, enhance the quality of care, and reduce costs. This expectation requires that physicians be trained, not as medical informaticians but as knowledgeable users of the health technology tools. However, most education in medical or health informatics has focused on the knowledge and skills needed by informaticians rather than health care professionals.

Recently in the United States, the President authorized the creation of the first Office of the National Coordinator for Health Information Technology. Several legislative initiatives were undertaken to promote the use of information technology within healthcare to improve process, quality and safety, thereby improving the health of our citizens. The promise of widespread adoption of electronic health records with the concomitant capabilities of provider order entry, decision support, and data mining for clinical research, as well as quality and safety evaluations, is about to become a reality. However, significant questions exist as to whether or not physicians will have the competencies necessary to effectively use these systems to achieve the goals outlined by the President and legislature.

Europe and Canada have long been leaders in the training of informatics-facile health care providers. The work of the European Centre for Medical Informatics, Statistics and Epidemiology (EuroMISE) has provided an early framework for such education in Europe.[1] The International Partnership for Health Informatics Education is in part an outgrowth of the earlier efforts and, in an environment of increasing globalization, emphasizes the need for international components in informatics education.[2]

Canada was also an early leader in medical informatics education and took a different but equally effective approach by integrating applied medical informatics into the undergraduate medical curricula.[3] However, such education must evolve with the changing technologies and the demand for more and more health care professionals to become information literate has resulted in an evaluation of current practices with more emphasis being given to emerging trends in both informatics and health.[4]

Other nations are beginning to recognize the need for more informatics training in the health professions.[5-7] In an attempt to address these very real issues, the International Medical Informatics Association developed recommendations on education in health and medical informatics.[8] These recommendations are the initial step in developing the educational framework necessary to insure that students possess appropriate qualifications to work in an information technology intensive health care environment.[9]

Leaders in medical informatics in these countries and others are calling for more targeted educational programs to insure that the systems being implemented will have physicians trained to use them.[10] However, the integration of such training into health professions curricula has been difficult at best and quite slow to develop.

Need for such training was beginning to be recognized in the United States in the 1980s with several calls from major organizations to prepare physicians for a future in an automated health care environment by integrating the necessary skills into the educational process.[11-13] However, little was realized in the form of concrete programs from these early inducements.

Understanding the potential impact of the growing interest in health information technology on the practice of medicine, and trying to take a more proactive stance in insuring that undergraduate medical students had a firm grounding in the knowledge, skills and attitudes necessary to become technologically savvy health care providers of the future, the Association of American Medical Colleges in 1998 convened an expert panel to develop educational objectives to satisfy this goal. The medical informatics panel of the Medical School Objectives Project (MSOP) II identified five medical informatics relevant roles played by physicians – lifelong learner, clinician, educator-communicator, researcher, and manager. The recommendations for educational content were developed within this framework and published in 1999.[14]

In part because of the increasing interest on the part of the government in facilitating widespread adoption of health information technology, in part because of the dearth of articles published about new educational programs in medical informatics in undergraduate medical curricula, and in part because of a growing need for information literate physicians, a small group of the educational leadership within the Group on Information Resources of the Association of American Medical Colleges surveyed and analyzed the responses of the 127 United States medical schools to determine whether or not they had implemented the MSOP medical informatics educational objectives and, if so, to what extent were the implemented.

Methods

An initial request to participate in the survey was sent to the respective deans of medical education at the 143 discrete medical schools in the United States and Canada. The deans were asked to either respond to the survey or refer it to someone who was knowledgeable about medical informatics content in the curriculum. The Web-based survey asked participants to respond to questions formulated directly from the MSOP II medical informatics educational objectives. These questions were grouped by the physician role with sub-groupings around concepts.

An initial question addressed whether or not the respondent was familiar with the MSOP Medical Informatics educational objectives. The subsequent questions asked the respondent whether or not each of the objective concepts was taught, had stated objectives, and was assessed. At the end of each of the five role divisions the respondent was asked to indicate who taught the concepts and how the concepts were assessed.

While virtually all of the respondents indicated they were familiar with the MSOP medical informatics educational objectives, the responses differed widely in regards to teaching, stated objectives, and assessment. In following up with a number of the participants about responses, it became apparent that many thought the medical informatics content was being taught as an integrated component of the clinical years. However others from the same institutions, many with long standing clinical information systems, stated that their medical students were exposed to these systems but did not have formal training or experiential learning with these systems.

Because of these discrepancies, a second survey was developed that limited responses to stated objectives and assessment because of the belief that having a stated objective would result in some educational action and would eliminate the possibility of someone assuming rather than knowing that the concepts were being taught.

The request to participate in the survey was again sent to the deans of medical education unless there was a different respondent on the first survey. The second survey was also Web-based and a request for participation was made in early 2006, almost a year after the first survey. Interestingly, individual school responses showed little change, however, several additional schools indicated establishing objectives.

Because the attributes for being facile with health information technology in the today's health care environment focused on three of the five physician roles, the responses for Life-long Learner and Educator-Communicator were not considered for this study. In addition, while data was collected on Canadian medical schools, because of their early embracing of the need to teach medical informatics in undergraduate medical education, only the responses from United States Medical Schools have been evaluated.

Results

Seventy usable surveys were "virtually" returned. Ninety-six percent of the respondents were familiar with the MSOP medical informatics educational objectives and eighty-eight percent indicated that there had been an overall strategy to integrate medical informatics objectives into the curriculum. However, the results of the specific competencies did not support this.

Clinician

Within the sub-group of effective use of clinical information systems, 60% of the respondents indicated that they had a stated objective on retrieving patient-specific information from a clinical information system and 49% assessed the competency. Forty-four percent had a stated objective on displaying selected subsets of information available about a given patient and 36% assess the competency. Forty-six percent had a stated objective about recording specific findings about a patient in a clinical

information system while 47% assessed the competency. Forty-six percent had a stated objective on recording orders (CPOE) directing the further care of the patient and 36% assessed the competency.

The sub-group of interpreting laboratory tests scored higher. Seventy percent of the respondents had a stated objective about recognizing the knowledge limitations of standard laboratory measurements and 66% assessed the competency. Seventy-seven percent had a stated objective about demonstrating the ability to integrate clinical and laboratory findings while 86% assessed the competency.

Within the sub-group of incorporating uncertainty explicitly into clinical decision making, fifty-seven percent of the respondents had a stated objective on demonstrating the ability to quantify and communicate the degree of certainty associated with specific items of scientific and clinical information and 50% assessed the competency. Forty-six percent had a stated competency on demonstrating the ability to identify and locate when possible the crucial pieces of missing clinical information and determine when it is appropriate to act on incomplete information and 40% assessed the competency. Sixty-three percent had a stated objective on demonstrating the ability to integrate verbal and statistical sources of medical knowledge with the facts of a specific clinical case and 61% assessed the competency.

Within the critical use of decision support tools sub-group, sixty-nine percent of the respondents had a stated objective on using textbooks and journal articles and 67% assessed the competency. Thirty percent had a stated objective on using diagnostic expert systems and fourteen percent assessed the competency. Twenty-three percent had a stated objective on using advisories or alerts issued from a computer based records and fourteen percent assessed the competency.

In responding to a student's ability to formulate a treatment plan, fifty-seven percent of the respondents had a stated objective that students should demonstrate the ability to express the relative certainties of a differential diagnosis while 69% assessed the competency. Sixty-one percent had a stated objective on expressing the relative risks and benefits of outcomes and treatment options while 66% assessed the competency. Forty-six percent had a stated objective on taking action by balancing risks and benefits while 53% assessed the competency.

Within the sub-group of respecting patient (and physician) confidentiality, 76% of the respondents had a stated objective on demonstrating the knowledge of the legal, ethical and medical issues surrounding patient documentation including confidentiality and data security while 79% assess the competency. Thirty-three percent had a stated objective on demonstrating the ability to use security-directed features of an information system while 27% assessed the competency.

Researcher

The first of the researcher group deals specifically with the use of clinical information systems. Twenty-four percent of the respondents had a stated objective on determining a practice's case mix and 20% assessed the competency. Twenty-nine percent had a stated objective on determining the incidences of diagnoses in a practice and 26% assessed the competency. Forty percent had a stated objective on testing the efficacy of a new treatment and 33% assessed the competency. Fifty-six percent had a stated objective on formulating testable hypotheses and 50% assessed the competency. Fifty-one percent had a stated objective on collecting, organizing, and interpreting data while 53% assessed the competencies.

Within the sub-group about determining what data exist relative to a clinical question or formal hypothesis, seventy-one percent of the respondents had a stated objective for demonstrating the ability to use information technology to locate existing data sources and 60% assessed the competency. Thirty-three percent had a stated object for demonstrating knowledge of data sources (including medical records claims and reimbursement information and online data) at one's own institution by identifying how these might be used to address a specific clinical question posed as research and 20% assessed the competency. Thirty-one percent of the respondents had a stated objective for demonstrating the ability to identify and locate existing data sets no maintained at one's own institution (e.g., national registry data) that might be used to address a specific clinical question posed as research and 16% assessed the competency.

For the sub-group executing a plan for data collection and organizing data for analysis, 24% of the respondents had a stated objective for selecting and appropriate computer database tool for collecting and organizing data and fourteen percent assessed the competency. Twenty-nine percent had a stated objective for properly representing data from a study in a form that is useful and supports computer-based analysis and sixteen percent assessed the competency.

Within the sub-group of analyzing, interpreting, and reporting findings, 23% of the respondents had a stated objective for selecting the appropriate computer software tools for analysis of data and ten percent assessed the competency. Thirty-one percent had a stated objective for using software to perform simple statistical analysis and portraying the results graphically and 23% assessed the competencies. Thirty-one percent had a stated objective for interpreting the reports of statistical software analysis and 27% assessed the competency.

Manager

There are three sub-groups within the Manager grouping. The first of these is the appreciation of the role of information technology in relation to managing the cost of medical care and its impact on individuals and society. Twenty-three percent of the respondents had a stated objective on using on-line sources of health care financing information and eleven percent assessed the competency. Thirty-nine percent had a stated objective on continuous quality improvement and process management and twenty percent assessed the competency. Twenty-four percent had a stated

objective on how information technology can be used to develop, implement and monitor compliance with clinical pathways and other forms of patient care protocols and eleven percent assessed the competencies. Thirty-three percent had a stated objective on how clinical information in the aggregated can be used to determine health care services planning for populations and 23% assessed the competency.

Within the sub-group of formulating and making decisions for individuals and groups, 55% of the respondents had a stated objective on demonstrating knowledge of cost/benefit issues in health care and 29% assessed the competency. Fourteen percent had a stated objective on using a decision-analysis package and seven percent assessed the competency. Thirteen percent had a stated objective on using software utilities assessing patients and six percent assessed the competency. Thirty-nine percent had a stated objective on incorporating economic and cost perspectives into decision making and 23% assessed the competency.

The last sub-group dealt with working effectively as an individual in inter-professional groups and as a member of a complex health care system. Nineteen percent of the respondents had a stated objective on using electronic personal and clinical scheduling systems and nine percent assessed the competency. Twenty-one percent had a stated objective on archiving and organizing digital information of personal and clinical import and fourteen percent assessed the competencies. Twenty-four percent had a stated objective on demonstrating knowledge of online resources for legislation, political advocacy, and local health care policy setting and six percent assessed the competency.

General questions

In all three of the physician role groupings, the content was taught generally through embedding it in core course. A few schools had an elective course in medical informatics and fewer still had a core course in medical informatics. Because the primary mode of teaching was through integration with other content, almost all of the assessment of competencies was done as part of a general educational evaluation schema. However, several schools had tests specific to medical informatics or used these in conjunction with the general assessment methodologies.

Discussion

The medical informatics educational objections presented by the MSOP expert panel were developed around the concept of information discovery and not predicated on computer literacy. For this reason, a number of the competencies can be taught without use of a computer. Examples of this are found in the interpretation of laboratory tests and the ability to formulate a treatment plan.

There were a total of 41 questions in the clinician, researcher, and manager role groups. Of those, 27 required interaction with a clinical information system or some ancillary system containing patient information. Eleven questions involved educational objectives that could be met without such interaction. Three questions related specifically to the competencies within the life-long learner role group but were also closely linked to clinician and researcher information management.

Of the five roles, the greatest number of medical school having stated objectives and competency assessments was found in the life-long learner role. This corresponds to the increase in teaching evidence-based medicine and the greater involvement of libraries for development of knowledge-based searching capabilities. For this reason, the life-long learning correlates, although requiring the use of computers to find information, were grouped separately.

In analyzing the responses by question type, less than a third (30.7%) of the medical school respondents had stated objectives for the 27 questions requiring use of computer systems and only slightly more than a fifth (21.1%) assessed competencies. There was one exception. Sixty percent of the respondents did have a stated objective about retrieval of patient-specific data from a clinical information system and 49% assessed the competency.

Of the three life-long learner correlated questions, approximately two thirds (67.7%) of the medical school respondents had stated objectives and slightly less (62.7%) assessed the competencies. Of those questions that did not require interaction with a computer system, over half (58.6% and 56.7% respectively) of the medical school respondents had both stated objectives and assessment of competencies.

In looking at the raw data and comparing the assessment to the stated objectives in all three of the physician role groups, there were 28 instances in which competencies were assessed within seven sub-groups without having stated objectives. These were virtually all in the clinician role and fell primarily under the non-computer based questions. A possible explanation is that the concept might have been considered too granular to include as a stated objective while it was included as part of a clinical evaluation schema.

Conclusion

Seventy of 127 surveys assessing the degree to which the MSOP medical informatics educational objectives have been incorporated into undergraduate medical curricula in the United Stated were completed. An analysis of these found that while many of the medical informatics concepts relevant to the clinician, research and manager roles were being addressed in the curricula, when broken down by those concepts that required health information technology interaction, only a few schools had stated objectives and fewer assessed the competencies.

The survey respondents were self-selected, and anecdotal information suggests that many who did not complete the surveys chose not to do so because they had little or no medical informatics in their curricula. Also, while these objectives are valid today, as HIT systems evolve and become more integrated into the health care system, the objectives also need to evolve. Some progress has been

made but much more needs to be accomplished to insure that physicians will be able to efficiently and effectively use the health information technology being installed in hospitals and health centers.

Acknowledgements

The authors gratefully acknowledge the support of the Group on Information Resources of the Association of American Medical Colleges for facilitating the survey distribution.

References

[1] Zvarova J, Engelbrecht R, and van Bemmel JH. Education and training in medical informatics, statistics and epidemiology in EuroMISE. Int J Med Inf 1997: 45: 3-8.

[2] Jaspers MW, Gardner RM, Gatewood LC, Haux R, Schmidt D and Wetter T. The International Partnership for Health Informatics Education: lessons learned from six years of experience. Methods Inf Med 2005: 44(1): 25-31.

[3] Kaufman DM and Jennett PA. Preparing our future physicians: integrating medical informatics into the undergraduate medical education curriculum. Studies Health Tech & Info 1997: 39: 543-6.

[4] Moehr JR. Where to in the next ten years of health informatics education? Methods Inf Med 2006: 45(3): 283-7.

[5] de la Calle G, Maojo V and Benito M. The INFOBIOMED Network of Excellence: facilitating training and mobility in biomedical informatics in Europe. Stud Health Technol Inform 2006: 124: 893-8.

[6] Sarbadhikari SN. Basic medical science education must include medical informatics. Indian J Physiol Pharmacol 2004: 48(4): 395-408.

[7] Hasman A and Haux R. Curricula in medical informatics. Stud Health Technol Inform 2004: 109:63-74.

[8] Haux R. Health information systems – past. Present, future. Int J Med Info. 2006: 75(3-4): 268-81.

[9] Muller S (Chair). Physicians for the twenty-first century: report on the project panel on the general professional education of physicians and college preparation for medicine. J Med Educ 1984: 59(11pt2): 1-208.

[10] Myers JD (Chair). Medical education in the information age. Proceedings of a symposium on medical informatics. 1986: Washington, DC: Association of American Medical Colleges.

[11] Association of American Medical Colleges. Advisory Panel on Assessing Change in Medical Education – The Road to Implementation. ACME – TRI Report: Educating Medical Students. Assessing Change in Medicine Education. The Road to Implementation. 1992: Washington, DC: Association of American Medical Colleges.

[12] Association of American Medical Colleges. Contemporary issues in medicine – medical informatics and population health: report II of the Medical School Objectives Project. Acad Med 1999: 74(2): 130-41.

Address for Correspondence

Julie J. McGowan, Ph.D., FACMI
Indiana University School of Medicine
975 W. Walnut Street IB-310
Indianapois, IN 46202-5121 USA

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Establishing A National Resource: A Health Informatics Collection To Maintain The Legacy of Health Informatics Development

Beverley Ellis[a], Jean Roberts[a], Helen Cooper[b]

[a] *Health InformaticsU nit, Lancashire School of Health and Post-Graduate Medicine, UCLAN, Preston UK*
[b] *Health Team, Library and Learning Resource Service, University of Central Lancashire, UK*

Abstract

This case study report of the establishment of a national repository of multi-media materials describes the creation process, the challenges faced in putting it into operation and the opportunities for the future. The initial resource has been incorporated under standard library and knowledge management practices. A collaborative action research method was used with active experts in the domain to determine the requirements and priorities for further development. The National Health Informatics Collection (NatHIC) is now accessible and the further issues are being addressed by inclusion in future University and NHS strategic plans. Ultimately the Collection will link with other facilities that contribute to the description and maintenance of effective informatics in support of health globally. The issues raised about the National Health Informatics Collection as established in the UK have resonance with the challenges of capturing the overall historic development of an emerging discipline in any country.

Keywords:

health Informatics, legacy, knowledge management, access

Introduction

Existing, lost and retained multi-media materials relating to topics in informatics to support health care represent a challenge in the context of the complexity of the current Health Informatics (HI) landscape. Absence of materials is most frequently felt by students and researchers, although the HI reference content also directly frames current operational developments. Looking at past projects and processes, and their longevity, success and failure may give pragmatic indicators for the level of acceptance or reluctance to embrace current development. If literature generated in the initial, basic and developmental phases of the projects are still available for reference, it will be possible to identify milestones and situations that can explain current entrenched positions and perhaps suggest catalysts to success that were missed on the first iteration.

The closure of the NHS in England Information Authority and the realignment of its roles and responsibilities, predominantly under the NHS Connecting for Health (NHSCFH) agency National Programme for IT (NPfIT) [1], made its reference resource library surplus to requirements. Members of the Health Informatics Unit at the University of Central Lancashire School of Health and Post-graduate Medicine (UCLAN) [2] had, with other academic colleagues, been considering how to sustain the legacy of HI. UCLAN was successful in bidding to house the residual resource; this is mainly due to its track record in managing and making available other specialist collections, such as the Livesey Collection relating to the Temperance Movement in the early 1800s, and the media collection of the Sport England organisation.

The National Health Informatics Collection (NatHIC) entity has been launched as a Collection rather than an Archive in order to confirm the commitment to maintain the heritage as a living and growing source. Personal donors who are, or were, key players in the domain have added to the initial materials, and continue to do so.

In order to keep a live Collection, the NatHIC objectives encompass making links to valuable knowledge facilities established elsewhere and to bodies who make available organisational resources on a rapidly changing basis. The linkages are in addition to managing physical resources on-site at UCLAN. The Collection is being made available for the benefit of all interested parties, not just those with close links to our university. NatHIC is intended to be available by reference, and where possible, loan of its content. NatHIC aims to act as a hub, maintaining physical multi-media resources, signposting other collections and materials, making linkages with other facilities and providers and providing guidance on identifying sources to those interested in informatics to support the health care domain. Users will not just be within the NHS but will also be from other care delivery agencies, their (commercial) solutions and service providers, and academia in all of the home countries and beyond. The desirable content will span both the historical legacy of HI and its contemporaneous / future materials, from a UK and wider perspective.

Method of development

The initial establishment of a repository of materials relating to the development of health informatics across the UK was first considered in the 1990s by members of CHIRAD,

Selected for best paper award.

a research institute in Winchester [4] whose members had had long and active participation in both teaching and learning and in the development of domain through the British Computer Society professional learned society and its subsidiary body, now called the Health Informatics Forum [5]. Despite various funding proposals the resources were not at that time available. However, the situation had changed sufficiently that in 2006 when substantial materials for the Collection became available, it was possible to put ideas into action at UCLAN. This Case Report paper describes the current situation, and discusses issues for future development in terms of content and accessibility for the domain internationally.

Access

Logistically, all users need to be recognized as authorized to use the Collection if they wish to loan any content. They must initially be registered as Distance (Associate Staff) Users of the UCLAN Library. This is a process already deployed in order to facilitate and enable generic library access by local NHS employees in the North West of England. It requires only completion of a form and provision of a photographic image to be used as authorisation if the holder ever wishes to physically enter the library on campus. UCLAN strives for equity of access for all through being cognisant of disability equity criteria including SENDA [6] Disability legislation.

NatHIC is initially intended to encompass any materials that impacts upon the HI developments across all the four UK home countries. Its principles and the lessons learnt from its establishment will perhaps be of use to other countries starting to identify similar omissions in their HI history. It is hoped that NatHIC will eventually interface to other resources elsewhere giving a global picture of HI developments over time.

Limitations

The current processing of donated will only capture legacy material that will set health informatics developments into context. UCLAN are putting into place mechanisms to include new works donated by publishers, authors and organisations creating informative resources over time.

Commentators have questioned the value of a back catalog rather than solely a comprehensive, contemporaneous resource. There are many significant instances where prior work has a distinct impact on current thinking. One particular instantiation of the continuing relevance of past concepts is a data quality principle of the early 1980s, where healthcare practitioners lead by the late Dame Edith Korner stated that a healthcare organisation should only (be required to) collect that information 'without which it is not feasible to be deemed to manage effectively' [7]. Thus data had to be collected at source, where any errors would be more readily identified and corrected; and that data had to be owned and have a purpose at that level in order to ensure the highest quality was maintained. At the time (1983) this reduced the number of resource wasting ad hoc studies that were previously carried out to answer one-off queries from government and strategic bodies. That principle can be seen reflected in the NPfIT 'Do once and Share' initiative but many younger HI professionals do not grasp that the underlying principles was defined over 25 years ago!

Others suggested that the current classification of cited publications was sufficient to afford a useful view of HI over time. However, as Machan [8] identified at the MIE2006 congress in Maastricht , analysis of the cited literature has a considerable omission in its knowledge capital – that of the negative findings and problem issues that are not formally published, but may have come to light in 'grey literature', trade papers and transient materials, such as newspapers, blogs and electronic commentaries.

As the Collection enlarges, there is a risk that space will become a premium. At that time, a usage review may be necessary to refine the most frequently requested sources to a particular geographic, technological, health or time focus, yet to be identified. Current library issue tracking systems can contribute to profiling what materials can/should be culled and can inform debate about what criteria should be used for content management.

Decisions relating to the scope, content, eligibility of materials for the Collection cannot be left to knowledge managers, library professionals or a domain experts with a close personal interest, (for example-ourselves) in maintaining the Collection, so an Editorial Management Committee with a range of areas of expertise is being established to reflect the extensive inclusive nature of the HI domain.

Determining ongoing functionality

Immediately after the launch of the Collection, a collaborative action research methodology was deployed in a Master Class facilitated by the NHS HI Faculty [9]. The session reflected on the basic Collection and its functionality and identified areas for further enhancement. These suggestions will be addressed over time in order to achieve as high quality resource as is feasible, given current technologies, funding and resources. The findings from the session were subsequently written up and validated through wider dissemination and feedback through the Faculty web area.

Various questions were identified, relating to establishing and maintaining resources for the benefit of all interested parties, they are explored further in this paper.

- What are the core resources that should be preserved
- What is the longevity of formal publications in practice
- How is it best to make available access to historic and future materials relating to HI
- Can a mechanism be developed that will accommodate self-input to the repository and yet avoid 'vanity' publication
- What target materials should be included in (or accessible through) the remit of NatHIC

Why is a comprehensive resource necessary

Health Informatics is still an emerging discipline, even after over thirty years [10] and as such is frequently seen as

an operational adjunct and a fragmented research domain. There is evidence that students, researchers and commentators on HI developments over time had difficulty finding historic materials to:

- set the context for current work
- explain the strategic direction of HI over time
- provide details of major national and local HI initiatives that were completed or had refocused for subsequent phases

In addition, there was anecdotal evidence, strongly articulated by professionals in the field that current initiatives were, or were likely to, repeat the mistakes and not able to avoid pitfalls that had been negotiated previously [11] if they did not take into account the way such challenges were addressed in the past.

Core content

It is not logical for NatHIC to deploy scarce resources to replicate / duplicate established sources of relevant material unless they are thought to be at risk. NatHIC is in the process of exploring links and collaborative working with repositories such as US National Library of Medicine, UK National Library of Health, and Department of Health with a view to providing secure gateways to their resources from NatHIC.

In order to be readily accessible and searchable, NatHIC will require ('a smart') classification and categorisation of its content. In view of the eclectic nature and wide range of that content, which cuts across traditional library resources, further work is ongoing exploring the suitability of existing ontologies and classifications.

Currently, the inclusion of materials is organic, decisions being based predominantly on advice about the seminal nature of thematic materials; but in due course a strategic acquisition plan for materials that the community / peer group see as important will be necessary.

The materials, nature and the links to NatHIC confound current definitions so the current structured strategies to facilitate effective searching are incomplete. For example – promotional literature aimed at a lay audience that is currently used to explain a proffered guarantee relating to care records handling will no doubt be varied because of current concerns [12]. It will be necessary to retain copies of the original documents to contextualise the emotions surrounding 'opting in or opting out' of the formal shared records plans for the 'National Spine' [1].

It was suggested that NatHIC also explored the management and retention of Case Study reports describing both completed assignments and projects still in progress in addition to issues papers created by various sources over time. There are a number of active groups in the field already cataloging such information, however these are disparate and diffuse, predominantly managing only the status reports of projects they themselves funded or support (for example Department of Health Service Delivery and Organisation R&D Programme : www.sdo.lshtm.ac.uk/commissioned projects.htm or the European Commission Europa site : europa.eu). NatHIC therefore could bring together linkages to those sources to produce a full picture of research activity across the country, and by UK participants wherever the project is lead from.

Also felt to be useful for ad hoc reference and download were templates and schemas that previously were only made available with specific initiatives and not available at other times, including business case presentations, project initiation documents and audit schemas.

Format

The permissible media for submission of items for the Collection were felt to be less crucial than the content. However it was recognised that some media were made obsolete by changing technologies. It is the intention of UCLAN, over time, to explore the development of mechanisms to make best efforts to keep seminal material readily accessible. For example – degraded old documents can be made available in scanned electronic form; obsolete e-media can be re-versioned forward as required (such as VHS video formats being transferred to CD-ROM). The experiences of other locations, such as the British Library will be sought to frame this activity.

NatHIC 'material content' will consist of physical hard copy media, audio and video/CD-ROM in so far as are still operational, robust electronic documents where they are provided by a source organisation (and subsequently where necessary to preserve content from obsolescence) and also links to other repositories, sources of content.

Where possible, donors will be encouraged to provide electronic media and direct access to back catalogues.

It was identified that users of NatHIC may only have a requirement for selected materials from within a journal / report and that partial download facilities should be available wherever possible. This will require a level of pre-processing which is not yet in place.

Moderation

As stated previously, NatHIC is currently taking all materials offered to it, with the knowledge that some may be extraneous, duplicate or out of scope. If the interest in NatHIC continues to rise, content may rise indeterminately and there will be a need for moderation of 'copy' in order to preserve the legitimacy of the resource. This will need to be carried out without bias by domain experts and knowledge management staff in the light of a number of factors, such as:

- content value to the domain
- scope of the topics and themes deemed appropriate
- usage rates
- copyright and license criteria

Value to the domain may change over time and will need periodic review. This could result in some material being withdrawn as no longer of significant worth or retained upon being confirmed as seminal and necessary to remain in the Collection.

Categorisation of materials will include evaluation by some, as yet un-constituted body, as relevant, in-scope and worthy of inclusion. In the immediate future, all received material will be stored. It will then be reviewed and processed for long term retention when appropriate criteria are determined.

The volume of copies submitted may require management against their usage; some key documents like national audit reports may be on frequent access by many people. A policy of retention for best quality copies being retained is already in operation, subject to estimated usage.

The mid-term requirement for moderation will necessitate drawing up rule-based policies for relevance, inclusion / exclusion and retention / withdrawal of material to ensure the integrity of NatHIC and its objective to preserve history and protect the future. This cannot be developed in isolation, but will be framed by the incoming material, potential user requirements and overall scope, and agreed by a representative editorial board, as yet un-constituted. In addition, for materials given to NatHIC, standard library protocols and existing mechanisms for clearing / establishing copyright and usage permission will be validated to determine their appropriateness; these materials include public domain materials, in-house reports etc., personal publications (for example student dissertations) or materials previously published elsewhere.

Future functionality

Facilities for self-loading of relevant publications may be useful; both by leaders in the field and for first line exposure of student theses and scientific papers. However draft terms and conditions will require definition and consultation, and processes for review will need to be operational to avoid / limit vanity publishing of dubious quality and contribution to the domain.

Grey literature, currently un-cited, ill-defined and transient may present a valuable resource to complement existing publications and complete the funnel plot described by Machan and demonstrate full coverage of knowledge across the domain. Further research will be necessary to validate this premise. In the interim, grey literature will be pursued, collected and reviewed at a later date.

Commercial bookshops around the world provide a service to new readers that indicates what the views of previous readers were on a particular publication. These views are displayed with books for sale on the shelves, but one suggestion is that a similar electronic facility could be added to NatHIC to allow users to comment on, or recognize similar materials on a topic that is in the catalog.

NatHIC aims to be inclusive and act as a portal where other established sources already have Health Informatics materials that are relevant. This will require planning and negotiation and will happen incrementally.

Critical success factors for NatHIC

Initially, success would be measured by a reduction in the number of historic documents being 'lost' to the community. Subsequently, it may be possible to monitor 'unrequited hits' for requested materials as an indicator of whether NatHIC is meeting community needs. As a proxy, an increasing number of formal arrangements with key stakeholders to routinely provide updated material will indicate increasing success.

The NatHIC has a linked informal blog [12] that will capture data on suggestions for new material and links. Monitoring unrequited searches and feedback will be reviewed periodically, so that NatHIC can target areas of interest that are currently not fully satisfied.

Results

To date, by early 2007, the Collection has over 1,500 contributions within an overall university library catalog of over 608 thousand items, including over ten thousand journals in the medicine and social sciences thematic areas. The range of media includes both contemporary hard copy and CD/DVD material and 'at risk' ephemeral media – including VHS videos and degrading paper. The 'at risk' material is being catalogued and steps will be taken to migrate it to viable media to capture the content.

Amongst the material provided are items which track the sequential phases of strategic development of NHS computing in England, specific business specifications of departmental and clinical systems, research deliverables from the sixties onwards and guidance on audit and information governance in both the NHS and European contexts.

The earliest material so far is - Berne, Eric (1968) Games people play: the psychology of human relationships [National Health Informatics Collection] Harmondsworth : Penguin, 0140027688. It is our belief that earlier material exists and is still being catalogued and prepared for NatHIC. HI in the UK is identifiable from 1961 when the first in-house hospital system was established in Manchester, although arguably, the NHS Central Registry pre-dates this as used in World War II to ensure unique patient identification. NatHIC is used by international cohorts of UCLAN HI students on Foundation degree to Masters courses; nearly 50 interested external parties have requested registration as Associates of UCLAN in order to access the NatHIC. This low number will increase as the body of content increases and as we attend events and promote the service.

Conclusions

The establishment of a national collection of Health Informatics material has been long overdue in the UK. It brings together resources that previously were (thought) lost or destroyed or considered to be of transient value. The advent of innovative electronic media makes this task feasible and sustainable in a manner not previously possible, and makes the National HI Collection accessible to a broad range of potential users. Usage over time will indicate its success. At this point in time, historic content is being added to the collection on a regular basis and arrangements with publishers for new materials to sustain

the Collection as a living archive are underway. In addition, newly published material will be incorporated as and when publishers and authors lodge reference copies with the Collection. The lessons learnt by this development will have resonance with any researcher who has previously not been able to source a required text on any media.

Acknowledgements

The authors thank colleagues working in and for the NHS, now and in the past, for their generous contributions to the content and development of the National Health Informatics Collection.

References

[1] Connecting for Health National Programme for IT, NHS in England www.connectingforhealth.nhs.uk (accessed December 2006)

[2] Health Informatics Unit outline, Lancashire School of Health and Post-Graduate Medicine, University of Central Lancashire www.uclan.ac.uk/healthinformatics (accessed March 2007)

[3] Livesey Temperance Society Collection, www.uclan.ac.uk/library/usersupport/lrs/collection/livesey/index.htm (accessed March 2007)

[4] Centre for Health Informatics Research & Development, Winchester www.chirad.org.uk (accessed March 2007)

[5] British Computer Society Health Informatics Forum www.bcshif.org (accessed March 2007)

[6] Special Quality Needs and Disability Act (SENDA 2001) www.opsi.gov.uk/acts/acts2001/20010010.htm (accessed March 2007)

[7] Mason A, Bishop P, et al. Developing a District IT Policy, Kings Fund/ Korner for Department of Health ,1983.

[8] Machan C, Ammenwerth E, Bodner T. Publication bias in Medical Informatics evaluation research : Is it an issue or not? MIE2006 Congress Proceedings, European Notes in Medical Informatics Vol II, No 2,2006.

[9] NHS Health Informatics Faculty www.informatics.nhs.uk (accessed March 2007)

[10] Roberts JM. The contribution made to the support of healthcare delivery and management by health informatics at a local and national level over a period of thirty years. Doctoral Thesis Teesside University,March 2006.

[11] BCS HI Forum, The Way Forward for NHS Health Informatics: Where should NHS Connecting for Health go from here? December 2006.

[12] UK National Health Informatics Collection www.uclan.ac.uk/nationalhealthinformaticscollection/ (accessed March 2007)

Address for correspondence

Beverley Ellis, HIU, Lancashire School of Health and Post-graduate Medicine, Harrington Building, University of Central Lancashire, PRESTON, PR1 2HE, UK bsellis@uclan.ac.uk

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

The EIPEN Project: Promoting Interprofessional Education in Health Professions

Joseph Liaskos, Antonis Frigas, Kostantinos Antypas, Dimitrios Zikos, Marianna Diomidous, John Mantas

Health Informatics Laboratory, Faculty of Nursing, University of Athens, Greece

Abstract

The Leonardo project under the name European Interprofessional Education Network (EIPEN) in health and social care, has been dealing with the challenges of Interprofessional Education (IPE). The EIPEN project tries to develop a transnational network of universities and employers in the six participating countries and at the same time to promote good practices in Interprofessional Learning and Teaching in health and social care. IPE provides opportunities for students and practitioners to learn with, from and about each other during qualifying and post-qualifying training and in their practice. IPE in health and social care includes the education and training of practitioners in human and animal medicine, dentistry, nursing, physiotherapy, occupational therapy, pharmacy and all other health professions including public and environmental health and health promotion, and social work. The outcomes of the EIPEN Project will provide means, material and guidelines for the enhancement of professional education in the multidisciplinary field of Health Informatics.

Keywords:

education, professional education, interprofessional education, multiprofessional learning

Introduction

According to the World Health Organization (WHO) Interprofessional Education (IPE) is identified [1] as an important component of primary health care. The Interpfofessional Education definition as given by CAIPE (Centre for the Advancement of Interprofessional Education) in 1997 [2] describes IPE as: “occasions when two or more professions learn from and about each other to improve collaboration and the quality of care”. The term multiprofessional education is also used to describe occasions when two or more professions learn side by side for whatever reason. The term interdisciplinary or multidisciplinary education or collaboration is referred to the combination and involvement of an assignment not necessarily working in an integrated or coordinated manner.

The interprofessional team in health has concentrated on two or at most three professions, primarily medicine, nursing and pharmacy. Interprofessional education has been invoked ever more frequently during the past thirty years to encourage collaboration in health and social care to help improve services, effect change and implement workforce strategies [3]. However, during the last decade, a renewed interest in IPE and an activity in health sciences programs appeared internationally. Inter-professionalism is very important while its importance is increasing; however there is not much literature on governance of Interprofessional Learning (IPL) [4].

Interprofessional education

Interprofessional education has been actively embraced in health professions since the 1950s, and in some cases earlier, and it is an evolutionary field of practice and research. It is seen as a way to develop collaborative practice among health and social care professions, and it plays a vital role in the patient-centred health services delivery approach. It contributes to collaborative practice and skills, aiming at patient safety. Teamwork is achieved through innovative methods of learning and practicing, improved ways of dealing with patient and technological complexity, enforcing professional relations and common understanding [3-7].

The implementation of interprofessional education is a difficult task for a various reasons: there are differences in prerequisites for admission to professional programs; the length of professional education; the extent and nature of the utilization of community and hospital resources for practice (clinical) education; students' freedom, or lack thereof, in the selection of professional courses; timetabling differences and conflicts across professional programs; faculty teaching loads; research interests of faculty; methods of administration within the various programs; and the powers vested in Deans of Faculties through statutory legislation, for example, through the power to appoint faculty members and to develop curricula [3,8].

Providing interprofessional learning experience that promotes teamwork and collaboration is a difficult task. There is a need to find academically acceptable mechanisms in order to measure the effectiveness of IPE activities. The existing attitudes of students, faculty and administration need to be changed in order to make IPE effective. The promotion of IPE as well as the measurement of its effectiveness requires that the students’ attitudes towards such work are assessed and evaluated.

Selected for best paper award.

Interprofessional education must confront particular challenges and needs seriously efforts in order to be successful. According to John H.V. Gilbert [8] these challenges include structural differences between faculty organizations; conflicting university and professional agendas; lack of adequate human resources to implement such programs, both within the university and across the community boundary; complex communication demands, within the university and with its community partners; rotation and replacement of team members; and lack of regular evaluation of interprofessional educational goals and programs.

IPE succeeds only when certain conditions are met: when the subject matter requires a team approach; when the effects of IPE can be clearly measured, for example, when critical reasoning skills are enhanced; when claims for resources to support IPE can be justified, that is, support for faculty and students is clearly necessary for success; when the skills being taught are within the competencies expected of a particular professional team; and when skills and knowledge can be explicitly taught and are clearly transferable, that is, those skills can be moved from one case to another [9].

Evaluation methods have to be developed in order to allow the assessment of the outcomes. These outcomes may include the patient, the process of interprofessional practice, individual professionals or agencies in which collaborations are carried out [10-12]. To exploit the benefits and outcomes of interprofessional education one must establish access to a wide range of resources of new knowledge and new skills. Many times, those benefits come through the shared respect and trust of the interprofessional partners who have been educated together in teams.

IPE and health informatics

Health Informatics is a multi-disciplinary field that deals with the collection, storage, retrieval, communication and optimal use of health related data, information and knowledge. The discipline utilises the methods and technologies of the information sciences for the purposes of problem solving and decision-making thus assuring quality healthcare in all basic and applied areas of biomedical sciences. Its domain covers computational and informational aspects of processes and structures in health care. Its aim is to study all the applications of informatics and computer science in Health Sciences (Medicine, Nursing, Dentistry, Biology, and Pharmacy) and health care. As a multi-disciplinary field it requires an educational approach aiming to collaborative learning. The different professionals who are graduates from specialised programs in health and medical informatics include physicians and nurses of different specialties, pharmacists, health care managers, computer scientists/ informaticians, engineers, e.tc. Moreover, practically all professionals in health care should, during their studies, be confronted with health and medical informatics education [13]. The development of educational curricula for Health Informatics follows an interprofessional approach.

In such an interprofessional educational environment (undergraduate or postgraduate), both teachers and students/ professionals from different disciplines have to be educated in a collaborative manner so as to develop the mutual understanding and respect among such multi-disciplinary groups, to enhance the existing and find new opportunities for shared learning and teaching, and advance the knowledge, skills, and attitudes of professional roles. For these aims, students will have to choose from a variety of learning activities such as team projects, which would provide opportunities for students to work together, tutorial courses focusing on a certain topic or area of health informatics theory or practice, exchange of experiences (e.g. observe team in action, discussion of real cases, placement of student or team in a real team of professionals in the clinical environment).

European Interprofessional Education Network (EIPEN) Project

The European Interprofessional Education Network in health and social care project, is supported and funded by EC Leonardo da Vinci Community action programme on vocational training (Project n^{o} 2005 UK/05/13/F/NT-162-335). It started on November 2005 and has two years duration until end of October 2007. The coordinator of the project is the Higher Education Academy Subject Centres, led by Health Sciences and Practice, based at King's College London (UK). In the network belong 16 partners from six countries. The European countries participating in the project are Finland, Greece, Hungary, Poland, Sweden, and United Kingdom, while the partners are universities, educational institutions, non governmental associations and centres related to interprofessional education [14-15]. From Greece two partners belong to the network: The Health Informatics Laboratory of the University of Athens and the, no budget holding partner, Greek Health Informatics Association (GHIA). The purpose of the project is to set up a network to develop and disseminate good practices in interprofessional education in health and social care in partner countries.

IPE provides opportunities for students and practitioners to learn with, from and about each other during qualifying and post-qualifying training and in their practice. IPE in health and social care includes the education and training of practitioners in human and animal medicine, dentistry, nursing, physiotherapy, occupational therapy, pharmacy and all other health professions including public and environmental health and health promotion, and social work. Interprofessional learning has been shown to lead to enhanced teamwork and health care. However, developing successful programmes and learning content and process is problematic. Thus partners desire to exchange teaching and management experiences as well as learning materials.

EIPEN has two interlinked aims: a) To develop a transnational, sustainable, inclusive network of people and organizations (universities and employers) in the six participating countries, and b)To share, develop, and promote effective interprofessional learning and teaching curricula,

methods and materials, good practices, for improving collaborative practice and multi agency working in health and social care. The model for network development is the successful UK Learning and Teaching Support Network, now the Higher Education Academy Subject Centres (www.heacademy.ac.uk), which uses a wide range of methods to disseminate and debate good practices in learning and teaching in Higher Education. The Subject Centres take account of the views of employers and other stakeholders, to sustain and extend their networks and to influence education policy. The network will be established over two years on a radial model. Each partner country will develop a website of resources and a programme of workshops and seminars that will link to a central interactive EIPEN web portal as well as regional and international events.

EIPEN goals

- Improved access to good practices, resources for learning and teaching, expertise, teaching materials and innovative case studies from universities, hospitals and other vocational training contexts in health and social care.
- A sustainable programme of events for presentations and debates concerning IPE.
- Dialogue with interprofessional education practitioners and peer consultation.
- New or improved sustainable national networks providing support for practitioners, teachers and students.
- Involvement of service users, students and policy makers.

EIPEN achieves these goals through the development of new networks within the health and social care sector through a programme of events at national, regional and transnational level:

- Interactive Web Portal with free membership registration, linked to national websites and resources
- Learning and teaching workshops and Seminars
- Transnational Steering Group Meetings
- International conference (2007 in Krakow)
- National learning and teaching resource databases linked to EIPEN portal
- Reports, presentations and publications
- Links with other International IPE groups

The products in two years includes the network, an interactive website with a resource database and a directory of people and organisation and discussion fora, and reports of national and international workshops and seminars on interprofessional training. The Transnational Network establishes and will be tested within 6 countries to ensure robust technology and systems and allow the design of procedures that will allow the resources of the network to be optimised. The aim of the project is to make the network available throughout the EU, but it is not possible to predict the specific longer term outcomes of EIPEN; these will emerge within the partnership. The web portal provides a facility for exploring other developments in health and social care practice and training, and the relationships being created increase opportunities for the exchange of ideas, staff and other partnerships.

Conclusion

IPE is innovative for professions, employers and education in health and social care. The innovation derives from the identification of opportunities for shared learning and teaching, the development of mutual understanding and respect among multi-professional and multi-disciplinary groups, the dissemination of interprofessional learning and teaching both in undergraduate and postgraduate education, and the advancement of knowledge, skills, and attitudes of professional roles. EIPEN develops new forms of networking within the health and social care sector for partners. EIPEN shares new products in the form of materials and methods for health and social care learning and teaching that will enhance the interprofessional education and learning in Europe.

Acknowledgments

EIPEN project is partially funded 2005-2007 by the European Commission as project No: 2005 UK/05/B/F/NT-162_335 in the Leonardo da Vinci Community action programme on vocational training. Project (network) partners are: Higher Education Academy, Health Sciences and Practice Network – King's College London (UK), Health Informatics Laboratory, Faculty of Nursing – National and Kapodistrian University of Athens (EL), Greek Health Informatics Association (EL), University and Polytechnic of Oulu (FI), Department of Mental Health Services – Oulu (FI), The Friends of the Young Association – Oulu (FI), Szechenyi Istvan University (HU), United Health and Social Institute – Gyor (HU), Department of Hygiene and Ecology – Jagiellonian University (PL), Occupational Medicine Centre of the Malopolska Region (PL), Karolinska Institute (SE), Centrum för Allmänmedicin Karolinska Intstitute (SE), The Primary Health Care Board – Stockholm County Council (SE), Centre for the Advancement of Interprofessional Education (UK), South West Peninsula Strategic Health Authority (UK), Steve Crane Consulting Ltd (UK).

References

[1] World Health Organisation (WHO). *Primary Health Care.* Report of the International Conference on Primary Health Care, Alma Ata. Geneva, 1978.

[2] Centre for the Advancement of Interprofessional Education (CAIPE). Interprofessional Education: What, How and When. CAIPE Bulletin No. 13. London, UK, 1997.

[3] Barr H. Interprofessional education: today, yesterday and tomorrow. UK: Learning and Teaching Support Network for Health Sciences & Practice, Centre for Health Sciences and Practice, 2003.

[4] Braithwaite J, Travaglia F. The ACT Health interprofessional learning and clinical education project: background discussion paper #1. The value, governance and context of inter-professional learning and practice. Canberra:, Braithwaite and Associates and the ACT Health Department, 2005.

[5] Lavin MA, Ruebling I, Banks R, Block L, Counte M, Furman G, et al. Interdisciplinary health professional education: a historical review. Advances in Health Sciences Education 2001;6(1):25-47.

[6] Atwal A, Caldwell K. Do all health and social care professionals interact equally: a study of interactions

in multidisciplinary teams in the United Kingdom. Scandinavian Journal of Caring Sciences, 2005;19: 268-273.
[7] Hall P. Interprofessional teamwork: Professional culture as barriers. Journal of Interprofessional Care. 2005; Supplement 1, 188-196.
[8] Gilbert JHV. Interprofessional education for collaborative, patient-centred practice. Nursing Leadership 2005;18(2): 32-38.
[9] Parsell G, Bligh J. The Development of a Questionnaire to Assess the Readiness of Healthcare Students for Interprofessional Learning (RIPLS). Medical Education 1999; 33: 95-100.
[10] Freeth D, Hammick M, Reeves S, Koppel I, Barr H. Effective Interprofessional Education: Development, Delivery and Evaluation. Blackwell Publishing, 2005.
[11] Freeth D, Hammick M, Koppel I, Reeves S, Barr H. Evaluating Interprofessional Education: a self help guide London: Higher Education Academy Learning and Teaching Support Network for Health Sciences and Practice, [Online] 2005 [cited 2006 Nov 25]. Available from: http://www.health.ltsn.ac.uk/publications/ occasionalpaper /occp5.pdf
[12] Freeth D, Hammick M, Koppel I, Reeves S, Barr H. A Critical Review of Evaluations of Interprofessional Education, Higher Education Academy Learning and Teaching Support Network for Health Sciences and Practice, [Online], 2003 [cited 2006 Nov 25]. Available from: URL: http://www.health.ltsn.ac.uk/publications/ occasionalpaper/ occasionalpaper02.pdf
[13] International Medical Informatics Association, Working Group 1: Health and Medical Informatics Education. Recommendations of the International Medical Informatics Association (IMIA) on education in health and medical informatics. Stud Health Technol Inform. 2004;109:226-43.
[14] European InterProfessional Education Network (EIPEN), [Online]. 2006 [cited 2007 Mar 28] Available from: URL: http://www.eipen.org.
[15] European InterProfessional Education Network (EIPEN), Centre for Health Sciences and Practice [Online]. 2006 [cited 2007 Mar 28]. Available from: URL: http:// www.health.heacademy.ac.uk/projects/eipen/

Address for correspondence

Joseph Liaskos, University of Athens, Faculty of Nursing, 123 Papadiamantopoulou str, Athens 115 27.
E-mail: iliaskos@nurs.uoa.gr

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Building ICT Capabilities for Clinical Work in a Sustainable Healthcare System: Approaches to Bridging the Higher Education Learning and Teaching Gap

Kathleen Gray, Jenny Sim

Biomedical Multimedia Unit, The University of Melbourne, Australia
Centre for Medical and Health Sciences Education, Monash University, Australia

Abstract

There is a recognised gap in information and communications technology (ICT) learning and teaching in higher education for entry-level healthcare professionals. This paper proposes a research model for understanding the dimensions of this gap. We describe methodological approaches to understanding present practices, identifying levers for change and learning by doing. We discuss issues faced in getting started and sustaining momentum on the research that is an essential prerequisite to effectively build the ICT capacity required by the clinical workforce in a sustainable healthcare system.

Keywords:

clinical informatics, competency-based education, education research, interdisciplinary communication, professional education, teaching.

Introduction

There is a recognised gap in information and communications technology (ICT) learning and teaching in higher education for entry-level healthcare professionals, and it needs to be bridged in order to build the capacity required by the clinical workforce in a sustainable healthcare system. Despite advances in educating specialist informaticians, in offering informatics electives to pre-qualification healthcare professionals, and in mapping details of ICT capabilities for healthcare professions, fundamental broadly based workforce capacity is not yet being established.

The contribution of ICT towards building a more sustainable healthcare system cannot be achieved by specialist informaticians alone, and relies upon active use of ICT capabilities in the professional practice of all clinicians. Clinical professionals' development of essential informatics capabilities is of local, national and international significance for evolving professional practice and education standards, and reform of healthcare system operations and management. Health informatics is an increasingly influential part of the working environment of "clinical staff including doctors, nurses, pathologists, pharmacists and other clinical professionals" [1]. The informatics capabilities of such staff are also a matter of interest to the public, as stakeholders in the healthcare system; increasing integration of ICT into healthcare has important implications for the quality of care, and for patient privacy and safety.

Improving the implementation of ICT in the healthcare system cannot be done only by providing these capabilities as optional extras for study by entry-level clinical practitioners; rather, it is necessary to embed informatics learning thoroughly in other aspects of professional learning and development [2, 3]. Examples of current educational interest in the learning and teaching of such capabilities in various clinical professions include allied health [4], dentistry [5], medicine [6], nursing [7] and public health [8]. But often such capabilities are addressed only in stand-alone ICT modules or elective units of study offered to students in their basic clinical training.

Nor can the healthcare system's need for technological transformation be met by learning and teaching capabilities simply profession-by-profession for clinicians. Calls to adopt a collaborative and multidisciplinary approach in promoting good health and well being [9] and to build ICT capacity across the entire health workforce [10] underscore the need for cross-professional learning and practice. ICT is recognised by a number of health educators to be an area of capability well suited to collaborative learning across professional boundaries [for example 11, 12].

Methods

A comprehensive research agenda is required to assess and address the higher education learning and teaching gap just described, so that we know what are the most promising approaches to pursue, in order to bring about the necessary educational change. This section proposes a three-dimensional model of the terrain to be bridged; it suggests a methodology for understanding the present situation, identifying levers for change and acting deliberately in response to this gap.

Present practices

One dimension of the gap is our lack of detailed knowledge about present practice in ICT education. It is important that those who educate future clinical professionals access and share research into "understanding where we are now … in order to clarify what practical steps are needed to move forward into the future" [13].

Clarifying the present situation with regard to ICT learning and teaching can provide important baseline data to allow prioritisation of research projects and evaluation of the effects of interventions. Key aspects of research to be done in this area include:

- Research into current roles / uses of ICT in educational settings for entry-level health professions: Across higher education, "There is enhanced use of educational technology … but all too often this emulates traditional didactic teaching and testing instead of promoting student curiosity and autonomy… In other words there is too often a poor alignment … between what is taught and the competencies students will need in their later lives and work settings" [14].
- Research into expectations about ICT as expressed in curriculum standards and professional accreditation processes: "Curriculum documentation can be read as giving in principle scope and support for the teaching and learning of essential clinical uses … However, documentation does not systematically address principles or processes for learning experiences that would scaffold or guide a student's transition …to those that are necessary or desirable in contemporary clinical practice" [15].
- Research into ICT applications in use in various clinical workplaces where students do their placement learning and hold their first jobs after graduation: A reality check is required to compare recommendations re informatics knowledge and skills that all clinical professionals *should* have [for example, 16] against the environments in which they actually work.

Levers for change

Another dimension of the gap is our lack of detailed knowledge about factors influencing ICT learning and teaching in higher education. We can locate obstacles to and levers for bridging the gap if we have a better understanding of the current dynamics of learning, teaching and educational provision:

- Understanding learning, especially understanding what learners need to be taught formally versus what they already know: For example "very little empirical research has actually questioned the Net Generation about their experiences with technology and worked with educational practitioners to determine the implications this has for Higher Education" [17].
- Understanding teaching, in particular the attitudes and experiences that determine orientations of key staff such as degree and year-level coordinators: For example, academics teaching core aspects of medical degree studies may find informatics "difficult to conceptualize" as a field of study and they may be equivocal about its inclusion in professional training [18]; and a nursing educator asks, "How can we expect faculty to transform nursing education for a type of practice that they have not experienced?" [19].
- Understanding what is involved in the provision of higher education, notably the pressures of globalisation, massification and privatisation on educational quality, and thus issues in competition with ICT education for attention in the operation of every clinical degree. In an example taken from physiotherapy, "The issue evoking most concern and comment is that of the ability of schools … and their professional clinical colleagues to continue to deliver appropriate clinical education within current resource constraints" [20].

Learning by doing

Another dimension of the gap is our lack of detailed knowledge about educational development prospects. Educational change is continuous and complex, and it is important that we recognise and capture the contribution of 'learning by doing' to what options there are for integrating ICT education into mainstream curriculum, and for finding interprofessional and inter-institutional ways of educating for ICT capabilities. This includes understanding what may be achieved by:

- Designing and implementing new learning and teaching resources and environments, such as plans "to integrate an entire enterprise-wide, electronic health record (EHR) system into the teaching curricula of nursing, physicians, pharmacists, and allied health schools, as well as health informatics and computer science" [21].
- Taking educational leadership roles in macro curriculum reform. Current health workforce shortages are producing many alternative scenarios for ways in which universities might partner with health services to provide frameworks for "experimentation and responsiveness in terms of preparing new types of health workers" [22].
- Strengthening research-led teaching or the teaching-research nexus, so that entry-level degree programs are appropriately informed by research into informatics – for example, there are a number of instances of research into using wireless handheld devices in clinical practice now being translated to enrich clinical teaching [23].

Such a research agenda needs to take a grounded theory approach. That is, it needs to iteratively and comparatively use data obtained from mixed methods of research - analysis of research literature and public documents, field observations and interviews, empirical studies; action research; individual and group evaluation by research participants - to provide an evidence base for the design and implementation of ICT learning and teaching for professional practice that will lead to systematic and sustained educational improvements. This approach is "uniquely suited to form the basis of research programmes that arise from theory grounded in the [health sciences] education experience, and then build toward implementation of practical educational innovations" [24]. Such an approach to research addresses recommendations for structuring clinical education [25], for developing authentic learning [26], and for supporting successful implementation of interprofessional education for collaborative practice [27]. This approach to research also accommodates the concept that educational development can overcome aspects of frag-

mentation and build a new professionalism in academics' working lives, through sponsoring a series of critical conversations [28].

Discussion

Preparing to advance upon a three-dimensional model using a toolkit of mixed methods to fill a major gap in clinical education knowledge and practice is a daunting prospect. It is no wonder the gap remains wide. In this section of the paper we review issues of when and where to start and how to sustain momentum, and we highlight some overarching considerations.

Of course, such work has been started, by many redoubtable educational researchers and practitioners, and has been in train for some time, however it continues to be hampered by issues of:

- The relatively low status of research into teaching and learning, especially in comparison with other research agendas in health and life sciences, and the consequently weak coordination and recognition of research in this field: this may be illustrated by the anomaly of strong emphasis on evidence-based practice in clinical work alongside its near absence in teaching and learning.
- The challenge of planning and resourcing complex ICT education research that is not just do-able, but meaningful in scope and scale, in order to make any significant difference to learning, teaching or curriculum: There are a myriad of political and logistical pros and cons to be weighed up in working on projects as a local initiative, versus on a state level or nationally, and / or in doing comparative studies nationally or internationally.
- The never-ending quest for sustainable long-term research partnerships with stakeholders external to the education system, such as ICT developers and vendors, healthcare agencies, health system funders: how is it possible to establish, maintain and evaluate these partnerships, and how to sustain this commitment alongside the full academic teaching and administration workload of many educational researchers?

Conclusion

Our view of the research to be done, the approach needed to bridge the knowledge and practice gap and the challenges faced in taking on this agenda rests squarely on eight key change lessons for higher education [29]:

1. "There are far more options for improvement or innovation than there is time or resources to address them."
2. "Change is not an event but is a complex and subject learning / unlearning process for all concerned."
3. "Enhancements in learning programs generate a need for improvement in the systems and infrastructure that underpin them."
4. "The most successful changes are the result of a team effort in which the most appropriate and best positioned people are involved in a process of action learning."
5. "The change process is cyclical, not linear."
6. "Change does not just happen – it must be led."
7. "Change is a mix of external forces and individual action."
8. "We must look outside as well as inside for viable change ideas and solutions."

This research needs to have a broadly-based view of the terrain, a robust methodological framework and mechanisms to share findings across boundaries and communities. Stakeholders must work together to keep this research agenda in the forefront in teaching and learning quality and research quality forums; in professional and industry forums on quality of care and workforce planning; in relevant non-governmental organisation and government forums. This work requires coordination and commitment among various stakeholders at multiple levels, and it requires substantial resources to ensure outcomes that are of significant benefit to the community, whether locally, nationally or internationally. Compounding the difficulty, Australian national classification systems for research fields do not enable ready identification of health informatics education research or give such research high visibility. However, it is only through seeking change in these conditions we will see real progress being made to bridge the higher education learning and teaching gap in building ICT capabilities for twenty-first century clinical work in a sustainable healthcare system.

The need for informatics education for non-specialists has been obvious for many years now [30, 31]. Our aim in this paper is to encourage the growth in strategic collaborations to take on all aspects of the associated research agenda over the next few years, in order to improve approaches to developing essential informatics capabilities in the entry-level education of clinical professionals. We welcome enquiries and approaches from others interested to work with us in this endeavour.

Acknowledgements

The authors wish to acknowledge influences on formulating the ideas in this paper that came from discussions during 2006 with Professor Evelyn Hovenga, Associate Professor Tangerine Holt and Dr Gregor Kennedy.

References

[1] Australian Health Information Council. Workforce capacity building national action plan. Canberra; 2003.

[2] Hersh W. The full spectrum of biomedical informatics research and education at OHSU. In: Haux R, Kulikowski C, editors. IMIA year book of medical informatics 2005: ubiquitous health care systems. Stuttgart: Schattauer; 2004. p. 167-72.

[3] Pearson A, Urquhart C. Health informatics education – working across professional boundaries. *Library Review.* 2002: 51(3/4): 200-10.

[4] Williamson D, Toth A, Grain H. Building health informatics capacity for the allied health professional workforce - case studies from the university sector. In: Proceedings of HIC 2005: Thirteenth National Health Informatics Conference;

HINZ 2005: Fourth Health Informatics Conference. Melbourne, Australia; 2005.
[5] Schleyer T. Dental informatics: A work in progress. *Advances in Dental Research.* 2003 17: 9-15.
[6] Blumenthal J, Mays B, Weinfeld M, Banks, M, Schaffer J. Defining and assessing medical informatics competencies. *Medical Reference Services Quarterly.* 2005 24(2): 95-102.
[7] Smedley A. The importance of informatics competencies in nursing: an Australian perspective. *Computers, Informatics, Nursing (CIN)*. 2005 23(2): 106-10.
[8] Karras B, O'Carroll P, Oberle M, Masuda D, Lober W, Robins L, Kim S, Schaad D, Scott C. Development and evaluation of public health informatics at University of Washington. *Journal of Public Health Management and Practice.* 2002 8(3): 37-43.
[9] Department of Education, Science and Technology (DEST). National research priorities: promoting and maintaining good health. 2005. [Monograph on the Internet] [Cited 2006 November 30]. Available from http://www.dest.gov.au/sectors/research_sector/policies_issues_reviews/key_issues/national_research_priorities/priority_goals/promoting_and_maintaining_good_health.htm
[10] Iliffe J. Building the IM & ICT capacity of the health workforce. In: Proceedings of the National Health Information Summit: transforming Australia's health system. 2005. p.121-3.
[11] Alpay L, Littleton K. Contexts for collaboration in healthcare education. *Health Informatics Journal*. 2001 7: 121-6.
[12] Carroll C, Rychlewski W, Teat M, Clawson D. Enhancing entrepreneurship and professionalism in medical informatics instruction: a collaborative training model. *Journal of the American Medical Informatics Association*. 2004 11(4): 317-9.
[13] Iliffe J. Building the IM & ICT capacity of the health workforce. In: Proceedings of the National Health Information Summit: transforming Australia's health system. 2005. p.121-3.
[14] Knapper C. Three decades of educational development. *International Journal for Academic Development*. 2003 8(1/2): 5-9.
[15] Gray K, Kennedy G, Judd T. Teaching health informatics to the net generation: a new baseline for building capability? Paper presented at MedNET 2006 Congress. Toronto, Canada. 2006.
[16] Garde S, Hovenga E. Australian health informatics educational framework. Australian College of Health Informatics. 2006.
[17] Kennedy G, Krause KL, Gray K, Judd T, Bennett S, Maton K, Dalgarno B, Bishop A. Questioning the net generation: a collaborative project in Australian higher education, In: ASCILITE 2006 Proceedings. Sydney: University of Sydney. 2006.
[18] Buckeridge D, Goel V. Medical informatics in an undergraduate curriculum: A qualitative study. *BMC Medical Informatics and Decision Making.* 2002 2: 6-16.
[19] Skiba D. Informatics competencies. *Nursing Education Perspectives.* 2004 25(6): 312.
[20] McMeeken J, Webb G, Krause KL, Grant R, Garnett R. Learning outcomes and curriculum development in physiotherapy. Australian Universities Teaching Committee / The University of Melbourne. 2005.
[21] Hovenga E, Louey-Gong D. Industry-government-academic partnership: the Australian story. In: Weaver C, Delaney C, Weber P, Carr R, editors: Nursing and informatics for the 21st century: an international look at trends, cases, and the future. Chicago. HIMSS. 2006. p. 206-9.
[22] Duckett S. Intervention to facilitate health workforce restructure. *Australian and New Zealand Health Policy.* 2005 2:14.
[23] Kho A, Henderson L, Dressler D, Kripalani, S. Use of handheld computers in medical education. *Journal of General Internal Medicine.* 2006 21(5): 531-7.
[24] Kennedy T, Lingard L. Making sense of grounded theory in medical education. *Medical Education.* 2006 40: 101-8.
[25] Henderson A, Winch S, Heel A. Partner, learn, progress: a conceptual model for continuous clinical education. *Nurse Education Today.*2006 26(2): 104-9.
[26] Barab S, Squire K, Dueber W. A co-evolutionary model for supporting the emergence of authenticity. *Educational Technology, Research and Development.* 2000 48(2): 37-62.
[27] Ginsburg L, Tregunno D. New approaches to interprofessional education and collaborative practice: lessons from the organizational change literature. *Journal of Interprofessional Care*. 2005 Supplement 1: 177-87.
[28] Rowland S. Overcoming fragmentation in professional life: the challenge for academic development. *Higher Education Quarterly.* 2002 56(1): 52-64.
[29] Scott G. Effective change management in higher education. *Educause Review.* 2003 38(6): 64-80.
[30] Staggers N, Gassert C, Skiba D. Health professionals' views of informatics education. *Journal of the American Medical Informatics Association*. 2000 7: 550-8.
[31] Coeira E. Medical informatics meets medical education: there's more to understanding information than technology. *Medical Journal of Australia.* 1998 168: 319-20.

Address for correspondence

Kathleen Gray
Biomedical Multimedia Unit
Faculty of Medicine, Dentistry and Health Sciences
The University of Melbourne VIC
Australia 3010
kgray@unimelb.edu.au

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

The Development of an Online Clinical Log for Advanced Practice Nursing Students: A Case Study

Jeffry S. Gordon [a], Ryan McNew [a], Patricia Trangenstein [a]

[a] *Vanderbilt University School of Nursing, Nashville, TN USA*

Abstract and objective

Three years ago at the Medinfo conference a prototype version of a clinical log for nursing students at Vanderbilt University was demonstrated. The purpose of the log is to document the types of clinical experiences the students are participating in as part of their academic program. We collected log data during that first year and received ongoing formative feedback from both students and faculty regarding its current feature set and desirable features for future implementations. Most of the requested new features have now been implemented. This paper describes some of the latest features of the clinical log, the advantages and disadvantages of ongoing development versus acquiring commercial products, and the procedures and results we have put in place to gather from faculty and students the features they want to see in the log. This paper also documents some of the data from early data mining.

Keywords:

education, nursing, graduate; nurse practitioners professional practice; students, nursing; computers, handheld/utilization; education, nursing, graduate/ methods; nursing education research; preceptorship

Introduction

Vanderbilt University's nursing curriculum is an Advanced Practice program designed to turn out Nurse Practitioners at the master's level in the following specialty areas: Acute Care, Adult, Family, Pediatric, Acute Care Pediatric, Midwifery, Psychiatric/Mental Health, Women's Health, and Neonatology.

Five years ago the PNP specialty faculty approached the informatics group about building an online clinical log to track their students' clinical experiences. They wanted a web-based version because it would provide more timely data to the faculty on the types of experiences their students were having. Until then, the students completed a notebook based paper log, sending it in twice a semester, and completed an op-scan questionnaire that was completed at the end of the semester. Op scan questionaires became increasingly popular in the early 1900s as an alternative to paper (Witzke et al., 1990) however, neither log type provided timely data. Notebooks could be lost in transit and it created a backlog several times a semester for faculty who had to read, make decisions on, and comment on the clinical experiences the students were having. McVeigh (1997) notes that paper logs are very inefficient. The faculty wanted the ability to check student experiences nightly to make sure there were no experience gaps. Gaps, if found, could then be dealt with in a timely manner (Alderson et al., 1999). They wanted the students to look for specific types of cases and log them as documentation that they had participated in them. The initial effort was rather primitive but it did provide the students with the ability to nightly log their clinical experiences and allowed their instructors to comment nightly on these experiences. Within two months, two other specialties (ANP, FNP) requested similar logs for them. This clearly escalated the development from "proof of concept" to the production of a professional quality web tool.

Currently the Vanderbilt clinical log supports students in 8 of the 9 specialties and will cover all 9 specialties next year. Initially we attempted to make all the specialty logs very similar. This would not only aid in coding the logs but it would make cross specialty comparisons a do-able procedure. Unfortunately, the significant differences in specialties and specialty requirements argued against keeping them very similar and as the years have progressed, even the basic features of the logs themselves have significantly diverged.

Methods

The initial logs were created by meeting with faculty in the specialty areas, demonstrating our existing log, and working from that. The datasets in the logs themselves were composed of generic data (time of patient meeting, type of insurance, gender of patient etc.), radio buttons indicating the level of participation in the encounter, checkboxes geared to the specific activities within the specialty, ICD9 code area, and a comment area. The student would treat each patient encounter as an individual record by entering the data above. The data could then later be queried by the student or the faculty member and summaries could be printed out. The faculty member could then write comments back to the student that would be included in the log. For all practical purposes, the only initial differences by specialty were the checkboxes identifying what procedures the student participated in.

After the first year, both faculty and students requested a large numbers of new features and additions they wanted

in the log. One area of immediate concern related to the checkboxes themselves. Standards and competencies from professional organizations were used to create the procedure checkboxes that was at the heart of the log. The log was created, to a great extent, to be faculty independent. It would be unacceptable to have a faculty member who was responsible for working on the log to leave, only to be replaced by another faculty member who wanted a completely different set competencies and experiences documented. The faculty examined the National Organization of Nurse Practitioner Faculties (NONPF) competencies in their specialty areas and designed their competency checkboxes from these competencies. Since NONPF was only creating some of those competencies during that time, it became apparent, particularly in the Psychiatric/Mental Health specialty that these checkboxes would have to be revisited after a year.

Results

We had three goals in the development of the clinical log.

1. The system had to be powerful enough to collect the right kinds of data. It had to meet the needs of the students and faculty and reflect the standards set by the professional organizations.
2. The system had to be easy to use. If it was difficult to navigate or cumbersome and time consuming to use the students would object. The faculty did not want the students to think of this as busy work.
3. The system had to be extremely reliable. Many of these students are not computer comfortable. They would never know whether the failure was at their end or our end but it would not matter. The frustration would be evident.

We met with each specialty of students three times formally during the year and with the faculty in each specialty at least one time formally per year. We made sure that students and faculty felt comfortable enough to approach any one of us at any time, through face to face meetings, email, telephone, or instant messenger.

The primary concern of the specialty faculty was to create a system that was easy to use by both students and faculty and where the students would not be bogged down nightly recording data as busy work. It is well documented in the literature that students are concerned about being overwhelmed with busy work while completing a log (Bardes et al, 2005) The faculty requested that a record should take no longer than 60 seconds to enter, and short of lengthy comments that goal has been met. The faculty also wanted a way in which students could flag certain records as important for their instructor to view. A checkbox was created that allows the student to indicate to the faculty member that this is a record of consequence. The students wanted to know when faculty commented on a specific encounter so an indicator was created on the student side showing which specific records had received faculty comment. The faculty members could even select the color of ink of the comment and tie that color to the flag indicator on the student side. While we assigned no importance to the color, and most faculty continued to write their comments in black, the additional color feature was a useful way for specific faculty to set priorities on comments.

The system had to have an easy to use login procedure. It was surprising that even well into the semester, students would forget their login names and passwords. In response, we created an email utility that would email them their lost password, but many students refused to use it. Since the university has a common username and password login for email, student records, and its course management system, we elected to switch the login procedures to that and this year not one student has contacted any of us about forgotten usernames and passwords.

System stability was a very critical feature. Many of these students and some faculty do not claim to be computer comfortable. Failures for any reason would cause them to balk at using the system and the students would then pressure the faculty to return to a paper log environment. Other than students forgetting their passwords, the system has never been down in three years. On several occasions faculty could not initially see a particular student but that was because the procedure we used to identify them by specialty incorrectly placed several students into the wrong specialty. Once those were identified (7 in three years) the problem was fixed manually. To summarize: The system had to be easy to use, powerful enough to provide students and faculty with the data they needed, and it had to be stable.

After about a year into development, several commercial products became available and we offered to consider switching to one of those applications... Faculty however, did not want to switch. The reasons they gave were: “we already know this log and it works.”, “why purchase something when we already have what we want”, “I like that I can make recommendations for changes and they will be implemented right away” (we quickly learned that rapid turn around on requests built strong loyalty for our efforts), “if our students have problems they know exactly where to get help”. Those were balanced against reasons for switching to a commercial product: We could get out of this business and focus our development resources elsewhere and there would be greater opportunities to share data across schools. As much as it was discussed, neither of these reasons was compelling enough to get the faculty to adopt a commercial product and the team realized that by customizing the log to their specialty requirements, we were building a product that they would use far into the future.

While great efforts were made to try to at least keep the logs of the various specialties parallel, it became apparent that this was not going to be possible. The Psychiatric/Mental Health specialty is an interesting case study on how difficult it became to keep the logs in synch. First, the PMHNP specialty uses DSM codes, not just ICD9 codes. Secondly, the manner in which they interact with their patients during their clinical hours is substantially different. Typically the other specialties’ students see any given patient one time. It is extremely rare that they would see a patient twice during their clinical experience. For these specialties, each encounter was equivalent to seeing a new patient. The PMHNP specialty’s students, however, see the

same patients over and over again week after week and they wanted the ability to "tie" the common records of a particular patient together. Retrofitting the log with this capability presented an interesting collection of challenges but they needed this for two reasons. It would make it easier to document patient progress over time if they could aggregate a specific patient's records in one place and they wanted a way to save their students' time by eliminating the need to check a host of demographic boxes on the same patient each time when the contents of those boxes essentially changed little from encounter to encounter.

Initially we provided a comment field for students to write their clinical comments on a particular encounter but the 5000 character limit was quickly discovered as inadequate for the PMHNP specialty. Furthermore the PMHNP faculty wanted their students to work off of a Microsoft Word created template that was not possible to code into the system. We then created an uploader that allowed the student to write their clinical comments in the Word template, save it under a specific record name that tied it to a particular encounter, and then upload the file to the server. The faculty side was then flagged to inform the instructor that a file attached to a specific encounter was available for viewing and comment. This year the PMHNP faculty wants the students to audio record the encounter. We have now created a methodology for recording the audio on their PDA, converting the file to mp3, and using the same uploader to attach the audio file to the appropriate encounter record.

Since certification requirements differ considerably between specialties, we had to create customized ways of tracking time and activities in an encounter. The Family PMHNP students for example, need to separate and track direct contact hours with patients under 18 versus direct contact hours with patients 18 and over, indirect hours related to time spent preparing for an encounter, and supervisory hours related to time spent conferring with a supervisor. Students were keeping this data on paper separate from the log. The instructors approached us with this issue and we immediately put it into the log. Other specialties, such and Midwifery, on the other hand, needed to track the number and types of births the student participated in.

Students approached us very early on and complained that they were in fact entering data on each patient twice: Once on paper after the patient had departed, and again online in the web interface that evening. Garrett (2006) identified the same issue. The students wanted a way to only enter data once and so we came up with a PDA solution to the problem. The student would enter the data immediately after the clinical encounter on their PDA, then they would synch the PDA nightly to avoid losing data. Once a week that synched data would be uploaded to the server and installed into the online database. We decided to force each specialty to use the web version only for one year, and then when the checkbox choices stabilized we would build them a PDA version. Constructing the PDA version would be difficult and we wanted to avoid making major changes in it each year. Currently we have PDA versions for 6 of the 9 specialties.

This year the PDA usage with the log has been as follows:

Table 1 – PDA Users

Specialty	Students	PDA users	Percent
ACNP	43	9	21%
ANP	35	6	17%
FNP	62	11	17%
PNP	34	19	56%
NMW	19	11	58%
PMHNP	28	0	0%

While 100% acceptance of the PDA solution was never expected, the relatively low percent of students using this technique was somewhat surprising. Garrett (Garrett et al., 2006) points out that the inherent restrictions on PDA can result in limited use. However, even with low numbers, the allegiance to this approach by the students practicing it was substantial. If a student started using the PDA environment to record their data, they stuck with it week after week.

Students gave a number of reasons why they chose not to use the PDA approach once they were given the option.

1. While PDA purchase was strongly encouraged and, in some cases required by each specialty, some students refused to make the purchase.
2. Some students balked at the $40 charge to purchase a PDA database program needed to make the PDA clinical log work. They opted to stay with the free web version.
3. Installing all of the software on the PDA to make it work was rather onerous. Some students claimed they couldn't get over that hump even though we created a web based Camtasia presentation that explained it in detail.
4. PDAs can be intimidating, particularly when they don't work the way the user thinks they should work and some students therefore just don't like PDAs.
5. Almost half of the students in the PMHNP program, for example, who commented on the PDA problems in their course evaluations, alluded to the lack of screen geography or difficulty in navigating the PDA clinical log application due to small screen space. This suggests that a tablet computer might overcome their reluctance here.

Last year none of the PMHNP students elected to use the PDA for data collection. When we asked them about that it was clear that the primary feature of the clinical log that they regarded as most important was the comment area and the corresponding ability to upload a Word file on the encounter to the record (Over 75% of the students were regularly uploading Word files with each log entry as

opposed to using the comment area in the log.). The PDA piece actually got in the way of their ability to do this. Comments of substantive length are cumbersome to write on a PDA and uploading a file works best when it is added to the record immediately after the record was created. Since PDA records were only uploaded once a week, the delay became an issue for these students. However, they do use the PDA to audio record their encounters. What this demonstrates is that people will use the technology they need to do their job as long as the technology helps them do their job, but they will not conform what they need to do to accommodate the technology if they perceive the technology is getting in the way. Since the PDA clinical log got in the way of the PMHNP students' ability to record the necessary data, they elected to just not use that capability.

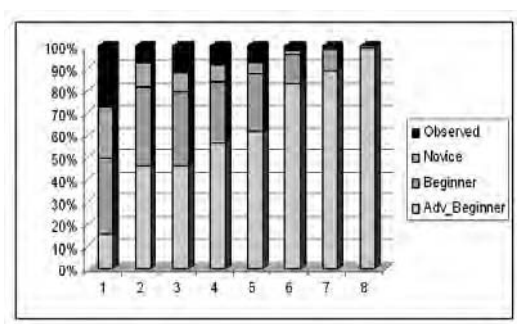
Perhaps the high percent of PDA users in the PNP and NMW groups is due to interested technologically literate and supportive faculty members who modeled behavior by using the PDAs for a variety of experiences in class. Since one of those faculty members has since left the institution, it will be interesting to see whether those numbers hold this year.

Early on in the development the specialty faculty members were asked what types of data queries they would want for both themselves and their students. It was necessary that they be very careful what they asked for because these queries are cumbersome to write. Because of limited resources, it was important that they not request queries they ultimately would not use... Rather than develop a series of queries the team created a tool to output the data into a spreadsheet table (Excel format) and taught the faculty and students how to analyze the data from there. The faculty appreciated the opportunity to learn Excel since they felt it could assist them in their research. They also said that since their students are becoming advanced practice nurses who will participate in research they too should learn Excel as a research tool. However, even with hands-on experience learning Excel in the computer lab, some students were just having a lot of trouble picking it up after a single training session. In response, the team created a web based animated training module using Camtasia and made that available on demand. Students could then go through the narrated training module whenever they were working on their analyses. In two of the specialties, the faculty showed the Camtasia presentation in class. The students commented that with the Camtasia created presentation, preparing the final analysis was very easy. The NMW specialty faculty, however; decided they wanted to create special queries outside of Excel. One of the students in the NMW program had an undergraduate degree in Computer Science and offered to program the queries. She was granted access to the DSN name, database tables, and variable names. She said she appreciated the chance to renew her skills in the area. Within several months she had a variety of queries programmed directly into the environment. This year, the NMW faculty wanted a significant number of changes to their log and we were concerned that the queries the student developed would no longer work. However, her code was well documented and it was easy to make the necessary changes to it. The other specialties have now seen the NMW queries but, much to our surprise, have not asked for them, even though we have offered building them. Apparently they still prefer the Excel approach and appreciate that we are teaching this lifelong skill to their students.

Students report that by using the output from the clinical log as a portfolio, they appeared to have an advantage in searching for the positions they wanted. One psychiatric student was able to demonstrate the types of patients he had worked with to such a degree that they offered him the position immediately. Others reported similar encounters. Several students asked if we could keep their accounts in the log active because they wanted to keep an ongoing portfolio of their more interesting cases and wanted documentation to support their activity.

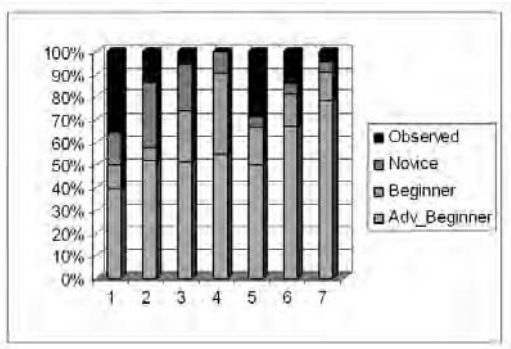
Faculty are now beginning to discover how to data mine the data tables and are seriously looking at the kinds of experiences their students are having across years. The overall data shows that the students definitely do gain clinical experience and confidence through the year. In the log the students evaluate their involvement in each encounter on a 4 point scale: (Observed, Novice, Beginner, up to Advanced Beginner with descriptor adjectives to enhance reliability of assessment). Figure 1, below, from the Family Nurse Practitioner (FNP) specialty, clearly shows that the students move, almost linearly from mostly observing to becoming more advanced in their participation over their 8 month clinical experience.

Figure 1: Type of involvement per encounter by month (FNP)



The relatively non-linear shape of the curve, below, in the Acute Care Pediatric Nurse Practitioner (PNP-AC) specialty is now serving as the basis for discussion on the types of clinical experiences this specialty is having, even though the general trend is toward more advanced experiences. Why, for example, is it relatively flat from months 2 to 5?

Figure 2: type of involvement per encounter by month (PNPAC)



We are now going to embark on a serious look across years and across specialties to guarantee that we are meeting the needs of the disciplines. We currently have approximately 490,000 individual patient encounter records in the system spanning the three years. As we add our final specialty and offer to make the log available to our partner that teaches the anesthesiology specialty, we anticipate adding over 200,000 records per year to the system.

Our next steps in the development process consist of altering the procedures we use for the PDA side, exploring alternative input devices, such as the intelligent pen and java enabled cell phones, and stabilizing the type of data collected to better allow for year to year comparisons. At the request of the program faculty, the PMHNP PDA log will be abandoned. Instead they want us to facilitate using the PDA to better audio record the patient encounters their students have and we will spend our development time with them promoting that. The rest of the specialties want the PDA piece streamlined with the data available for faculty review nightly instead of once per week. We are currently looking at two different technologies to accomplish that task. Finally, there is the ongoing conflict between standardizing the data to facilitate program to program evaluations and the customizing each specialty area to more exactly meet the needs of each particular program. This next year there will be an attempt to create more standardization where possible, but after that we are going to let the data fields stabilize for about two years to permit year to year comparisons and analysis.

Conclusion

The development of the clinical log has been an exceptional team building experience between faculty in the specialty areas and the informatics faculty and staff. Creating a professional quality "home-grown" log has given the school an excellent example of what the skills of the informatics area can accomplish. The faculty has been very pleased with the quick response they get when they want a new or improved feature and now rely on the informatics team for all of its technological needs. In return, we has learned a great deal about what works and what doesn't with students by specialty area and has developed a close professional relationship with the specialty faculty. We have been surprised at how rapidly the faculty has adopted the log. Based on prior work we anticipated relatively long lead times to change the culture (Nierenberg et al., 2007) but that turned out not to be the case. The faculty is now beginning to data mine the log to improve the clinical experiences of their students and perhaps alter their clinical content as well as modify the focus in their didactic teaching. Program alteration does not come automatically, however, and effort must be made to use the logs to improve the structure of student learning (Dolmans et al, 1999) This experience has translated into many other additional projects. We are now using the log technology for a grant supported cell phone based messaging system to encourage smoking cessation. We have created many online assessments, surveys, and questionnaires for the school and associated organizations. Finally, we are exploring new technologies, (such as intelligent pen based data collection systems and java enabled cellular phones) to accomplish many of these tasks.

References

[1] Alderson, T.J., Oswald, N.T. (1999). Clinical experience of medical students in primary care: use of an electronic log in monitoring experience and in guiding education in the Cambridge community based clinical course. Medical Education, *33*, 429-33.

[2] Bardes, C.L., Wenderoth, S., Lemise, R., Ortanez, P., Storey-Johnson, C. (2005). Specifying students-patient encounters, web-based case logs, and meeting standards of the liaison committee on medical education. Academic Medicine, *80(12)*, 1127-32.

[3] Dolmans, D., Schmidt, A., Beck, J., Beintema, M., Gerver, W.J. (1999). Does a student log provide a means to better structure clinical education? Medical Education, *33*, 89-94.

[4] Garrett, BM, Jackson, C. (2006) A mobile clinical e-portfolio for nursing and medical students, using wireless personal digital assistants (PDAs). Nurse Education Today, *26(8)*, 647-654.

[5] McVeigh, S., Williams, R., Pangaro, L. (1997). An internet home page to log students' contacts. Academic Medicine, *72(5)*, 417-18.

[6] Nierenberg, D.W., Eliassen, S.M., McAllister, S.B., Reid, B.P., Pipas, C.F., Young, W.W., Ogrinc, G.S. (2007). A web-based system for students to document their experiences within six core competency domains during all clinical clerkships. Academic Medicine, *82(1)*, 51-73.

[7] Witzke, D.B., Koff, N.A., McGeagh, A.M., Skinner, P.D. (1990). Developing a computer-based system to document and report students' clinical patient encounters. Academic Medicine, *65(7)*, 440-41.

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Personalized Case Driven Parental Education Informatics in the NICU

John Chuo[1], Pavel Sherman[2], Claire Drain[1], Casimir Kulikowski [2]

[1] *Department of Pediatrics, Robert Wood Johnson Medical School - UMDNJ, Bristol Myers Squibb Children's Hospital at Robert Wood Johnson University Hospital*
[2]*Department of Computer Science, Rutgers University, USA*

Abstract

Neonatal Intensive Care Units (NICUs) are foreign and intimidating to parents of premature infants. The high levels of anxiety and stress they can produce needs to be reduced by thoughtful advice from healthcare providers (HCPs), to educate parents about their child's condition. Unfortunately time constraints often limit HCPs to only a few minutes with each baby's parents daily – only enough to convey critical information at a high level and with limited depth. Parents searching for more detailed information themselves in the literature over the web have sometimes reported disappointing experiences. We are proposing to improve parental education by patient-centric web methods leveraging the electronic patient record with internet and cell phone technologies. This can be an important informatics resource, complementing and enhancing face-to-face communication through personalization of education and advice to the parents.

Keywords:

Neonatal Intensive Care Unit (NICU), Personalized Parent-Patient (PPP) model, personalized parental education, electronic health record (EHR), Health Care Providers (HCPs)

Introduction: NICUs and Parents

The NICU experience is an emotional and confusing time for parents; what many families expected to be a time of joy and happiness has turned into an event filled with uncertainties. Many parents entering the NICU feel overwhelmed by the intensive care unit environment, including the technology, language used, and the pace of events [1,2]. This, coupled with their own sense of loss of expectations for the birth and future of their child, can lead to miscommunication between staff and families. What may seem "normal" to the NICU staff, can feel anything but normal to the family. An important way to make the experience most constructive is to develop a trust relationship between families and their HCP through effective communication [3,4,5].

The informational asymmetry that exists between the medical staff and their patient families is a source of anxiety [5]. While every individual has their own communication and information processing styles, what is important to parents is that they have access to all of the information that they want to have, when they want it, and that they be allowed to absorb the information in their own time. Withholding information, even under the pretense of sparing the family pain, or perhaps unnecessary anxiety, does not create a trusting relationship between the staff and the family [2]. The Bayer Institute for Health Care Communication identifies education along with empathy, engagement, and enlistment as one of the key factors of successful communication. Learning and understanding is best done in steady short segments, rather than episodic large segments [9]. Benefits of an educated parent are less anxiety, less depression, shortened NICU stay, and increased satisfaction with the medical staff care [3-7].

The high acuity nature of the NICU prevents the HCP from spending the necessary time needed to provide optimal patient education. In one study, parents feel that important information was not relayed satisfactorily almost 50% of the time (i.e. either too much, too little, or not explained at all) [10]. We propose a solution for optimizing parental education in the NICU by leveraging technology to educate parents 'on the fly', engaging the HCP only for critical tasks such as content verification before sending to the parents. The system should map existing information in the EHR to existing peer reviewed knowledge resources. Our system (1) matches personal content in the EHR to existing knowledge resources (KRs), (2) presents those KRs to HCPs for approval before sending them to parents via a secure web portal, and (3) uses cell phone technology to alert HCPs and parents of new information available for sending and viewing, respectively. This strategy aims to enhance the personal meetings between HCP and parents by (a) preparing parents before the meetings with background information, (b) giving additional information that may clarify questions that parents may think of after the meetings but are unable to ask HCPs immediately, (c) expediting communication between the HCP and parents regarding important and unexpected issues that may arise during the baby's clinical course.

Table 1 Neonatal Clinical Scenario

Event		Action by HCP using EMR in the NICU			EMR Test results
Date Time	(#) Description	Add to patient's problem list	Orders	Notes	
Day 1 8 AM	(1) Infant delivered by normal spontaneous vaginal delivery with 1, 5 and 10 minute apgar scores of 9/9/9. Prenatal labs significant for GBS+,Hepatitis B unknown. Baby with respiratory distress.	Prematurity, Rule out sepsis, Hepatitis status unknown, Surfactant Deficiency Syndrome	*Test Order:* hemogram, blood culture, c-reactive protein, chest x-ray *Med Order:* Ampicillin, Gentamycin,Hepatitis Immunoglobin /vaccine	CPAP	
9 AM	(2) CXR c/w RDS,give 40% O_{2}		*Med Order:* Surfactant	Intubation*	
1 PM	(3) Hypotension, wide pulse pressures,		*Test order:* Echocardiogram		
3 PM	(4) Hypotension		*Med order:* Indomethacin		*Echo* - PDA
Day 2 8 AM	(5) Normotensive,Infant extubated and placed on CPAP			Extubation to CPAP	
1 PM	(6) Started feeds- breastmilk		*Feeding order:* start BM 1 ml every 3 hours via NGT		
Day 3-20	(7) Increased Apnea episodes Jaundice	Apnea Jaundice	*Med order:* Caffeine, Phototherapy		
Day 21 8 AM	(8) Routine hemogram significant for anemia	Anemia			Hematocrit = 25
Day 49 8 AM	(9) Ready for discharge		*Order:* Discharge		

Methods: Clinical scenario

Setting. Currently, physicians in our neonatal intensive care unit at the Bristol Myers Squibb Children's hospital enter patient information into an electronic health record system named NeoData, by Metasoft Systems, Inc. The Eclipsys 3000 TDS system is used to enter orders and view test results. Patient problem lists, orders, and test results will be obtainable by our software system. We will illustrate how our system can help educate patients' parents in our NICU. Table 1 highlights events for a typical neonate over a timeline spanning three common phases of a patient's hospital stay: Admission, Interim, Discharge.

Synopsis. TJ is a premature baby boy born to a 39 year old mother at 29 weeks gestation. The apgar scores are 9 at 1, 5 and 10 minutes after delivery. His mother's prenatal laboratory result is positive for Group B Streptococcus and her Hepatitis status is unknown. TJ's immediate medical issues are Surfactant Deficiency Syndrome (also known as Respiratory Distress Syndrome) and Patent Ductus Arteriosus, resulting in respiratory distress and hypotension. An immature immune system makes TJ vulnerable to infections;jaundice and apnea also occur. His stay in the NICU also involves extensive routine care issues that support proper growth and maturation until discharge.

Admission period. TJ's admission period (day 1-2) is one of the most clinically telling, data intensive and anxious moments. Perhaps the most important and longest HCP-parental meetings occur at this time, typically lasting 30 minutes. During this time, TJ's HCP team must not only orient his parents to the NICU environment, but also discuss prematurity and its implications, its many acute events, and be sensitive to the parent's level of confusion, stress, and medical unawareness – in TJ's case, the numerous topics are respiratory distress, sepsis, antibiotics, various diagnostic lab tests, and Hepatitis prophylaxis (event 1). The nurse at the meeting must review general nursing procedures and routine regulations as well as the operational rules of the NICU. Not surprisingly, when parents are later asked about what was discussed, many can not recall key points of the content [8]. Reasons for difficulties in knowledge retention are (1) too much, too soon, too complicated (2) emotional interference, and (3) lack of HCP follow-up.

A detailed discussion of all of TJ's acute clinical issues (prematurity, sepsis, respiratory distress syndrome) would be overwhelming. An HCP-given high level overview in lay terms of current clinical issues and forthcoming expectations is effective for the short term, but needs to be followed up by detailed information and opportunities for parents to ask follow-up questions after 'digesting' the news – all done using a family centered approach [13]. Our system enables ongoing delivery of important information, complementing what is learned during these meetings, especially the initial one, so that they can be better prepared to ask important questions in subsequent meetings.

Interim period. The interim period can be both long and unpredictable, testing the emotional fortitude and patience of parents and HCPs. In most cases, meetings with the HCPs are sporadic and dictated by clinical events rather than regularly scheduled updates. In one study, only 26% of the surveyed parents report talking to the neonatologist on a daily basis [10]. Without meaningful and well timed dialogues with HCPs, parents can easily be misinformed by the internet and non peer-reviewed literature. A study by Dhillon et al reveal that only 10% of parents feel medical information found on the internet is reliable and up to 24% found the information to be over-frightening [12]. Shortly after birth, TJ's respiratory distress worsened and prompted emergency intubation and placement on a ventilator. Soon he became hypotensive (events 3-4) and an echocardiogram revealed a patent ductus arteriosus, which

is a common life threatening condition causing hypotension. The treatment involves risky medication therapies like Indomethacin. For events 2-4, although physicians would prefer updating parents directly, having more detailed information available for the parents to review after being updated is important in the education process. Complex conditions such as apnea, anemia and jaundice could be signs of either life threatening problems such as intraventicular hemorrhages and biliary atresia, but more often are part of the normal stages of development that require patience and time. Preparing parents to expect these non-emergency, anticipated conditions ahead of time will reduce anxiety and build trust.. Educated parents tend to feel less vulnerable, more empowered and engaged [14].

The overall background of stress during this period is riddled with episodic roller-coaster moments of relief and worry that is dictated by the clinical picture. Most communicated information during this period falls into one of three categories: (1) diagnosis and prognosis (events 1,7,8), (2) policies and procedures (events 2,5), and (3) routine care (events 6,7). The timing of information delivery is important to patient treatment and parental care. Early awareness of important test results such as the presence of a hemodynamically significant Patent Ductus Arteriosus can expedite communication between HCP and parents so that time sensitive treatments can begin sooner (event 2). Our software would enable this function by sending a text message to the HCP as soon as the result appears in the EHR.

Discharge. The discharge period is one of the happiest yet intimidating times, marking the beginning of a new phase in TJ's care in the hands of his parents. Proper parental education is more important now than ever to TJ's well-being. Unfortunately, the discharge meeting has similar time and content constraints as the initial one that occurred nearly 2 months earlier. Researchers show that parents who have been educated all along about their infant's clinical issues, as our system would permit, are more prepared to take over as the baby's primary caregiver [15,16]. Following the same delivery strategy, the HCP will typically give TJ's parents an overview – this time of his special needs at home. The HCP will introduce issues that were not relevant to his hospital stay but important in the 'outside world' such as TJ's vulnerability to respiratory synctius virus infections and associated complications - therefore the need for medication prophylaxis. Proper parental education and understanding of TJ's developmental needs (i.e. sensory stimulation, engagement) at home would maximize his potential for a normal life – the discharge meeting only begins the education. Our system will supplement the issues discussed with additional information that the parents can review, so they can better learn from their pediatrician and TJ's development care team.

Proposed system design and functionalities:

Figure 1 illustrates the major components of our proposed system. A knowledge repository consists of reliable information to be delivered to parents/guardians and healthcare providers; a personalized parent-patient (PPP) model matching neonate events from the patients' electronic health records to parental roles and their information needs; an algorithm that personalizes the information delivered to parents/guardians; and a notification component for informing the appropriate parties when new information becomes available or previously delivered information is updated.

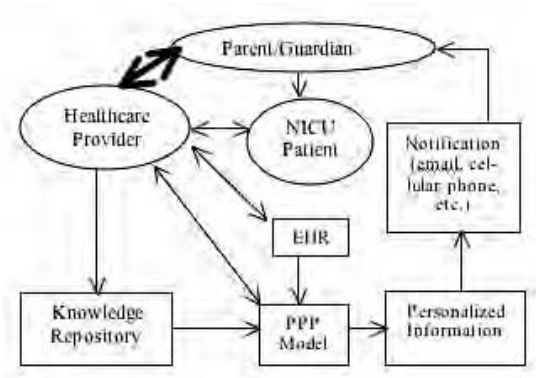


Figure 1: System Design

The knowledge repository (KR) stores all documents to be delivered and are approved by a staff member, which are reviewed and reapproved subsequently. Entering a document to the KR restarts the personalization algorithm to compute a relevance score of a new document for each PPP model. The document is marked for delivery for each PPP model for which the document's relevance score exceeds the threshold. A healthcare provider is then notified that new information is available for delivery, and on approval, the parent/guardian is notified that they can access it. The PPP model takes EHR data and maps it through its meta-data schema to categories of information delivery. Updates to an EHR will trigger the personalization algorithm to re-compute the relevance score for each document in the KR. After this is completed, the articles are recommended, approved, and delivered to the parents/guardians and/or healthcare providers associated with the appropriate child's record.

The relevance score for each document is based on weights developed by neonatologists and NICU staff based on how each document is triggered by the context of a neonate event, and answers a typical patient-parent information problem. Personalization comes from model matching the metadata of specific relevant documents to the metadata from the PPP models. If both a document and a PPP model contain the same term in their metadata, then the relevance score of the document increases by its weight (representing the informative urgency of the information for the PP context). If the document contains too many keywords not found in the PPP model, then the document's relevance score decreases by an "unlikely information" weight. Otherwise the score is left unchanged. After each document has been assigned a relevance score for each PPP model, those documents with relevance scores above a certain threshold will be queued

for approval for delivery to parents/guardians. The initial prototype is implemented with this procedure. We are currently experimenting with more sophisticated personalization algorithms to help improve the accuracy of the relevance score for a set of representative neonate scenarios (including TJ's), while reducing the burden on the HCP reviewers and approvers by summarizing displays and clusters of indices to the (as yet few, brief, and to-the-point) references. This raises many issues of informatics research we briefly address in the concluding section.

Table 2 shows how the system can improve communication between the parent and TJ's HCP. As the details of each event are recorded in the EHR, the metadata associated with each entry is compared with the metadata of the knowledge resources according to the personalization algorithm. The knowledge resources matching the EHR are queued for delivery to TJ's parents. The HCP may also choose to manually deliver certain documents from the KR. Messages are sent by cell phone to the parents immediately after the documents are approved. In some cases, the physician may not be aware of relevant updates to the EHR and the system will alert the physician of this, increasing communication efficiency. We now give an example of how the system can be used to deliver information to TJ's parents.

Table 2: Knowledge Resources for Clinical Scenario

	Knowledge Resource		**SMS Messaging**	
Event (#)	**Automatic (EHR guided)**	**Manual**	**To MD**	**To Parent**
1	NICU orientation, Prematurity, sepsis, Hepatitis prophylaxis, RDS, CPAP, CBC, CRP	Apgars, Group B Streptococcus, Blood transfusions	X	X
2	Intubation, Surfactant			X
3	Echocardiogram			X
4	PDA, Indomethacin		X	X
5	Extubation, CPAP			X
6	Store breastmilk, NGT			
7	Jaundice, Phototherapy, apnea, Caffeine	Kangaroo care		X
8	Transfusion			
9	Discharge instructions			

Discussion and conclusion:

We outline an informatics system that augments personal communications between parents and health care providers (HCPs), so that the information they "pull" from many sources can be channeled and interpreted in the context of their child's problems. The personalized technology of the "push" system will allow pertinent information to reach parents, and address their child's clinical status, in the most time sensitive manner. As we discussed, early and continuous education of parents has proven to be very effective in a NICU setting. In TJ's case, early presentation to the parents of a commonly anticipated clinical scenario during the NICU orientation, such as the likelihood of blood transfusion, will address several issues. It will allow TJ's parents time to assimilate and ask about the need for such a procedure, raising issues about TJ's safety and comfort, and the safety of the blood supply [17]. By presenting all of the information to TJ's parents as early in the process as possible, it will also allow the family time to find donors, and have them screened, for a directed blood donation, if that is something they want. Blood transfusions require parental consent. Having addressed all such relevant concerns earlier during TJ's course of treatment will allow the HCPs to care for TJ, and his family, in the most efficient way possible.

In a situation of greater emergency, such as TJ's need for intubation, the "push" system is also an effective tool to alert parents to the fact that their child's clinical status has changed. While we would not advise alerting parents of intubation via the system, it does provide the HCP with the opportunity to reach the parents and let them know that they need to call the unit to speak with their attending physician as quickly as possible. The system will also allow parents to be pushed information regarding the reasons for TJ's intubation, and the potential side effects of long term oxygen use. While we would not expect parents rushing to their child's bedside in an emergency situation to stop and read the information provided, it is reasonable to assume that they would have questions regarding the above issues. The "push" system will allow the HCP to anticipate common concerns of parents in this situation, and to address those concerns as quickly as possible. By sending the information to the parents electronically, the parents will be able to read and digest the information, and its implications, when they are ready. In this format the parents will also be able to revisit the information as often as needed, until they are comfortable with their own understanding of TJ's situation.

Currently, HCPs and administrators are the main users of EHRs. This proposal suggests leveraging the EHR as part of a personalized educational tool that uses physicians to direct the pushing of relevant educational knowledge to NICU parents. The implementation challenges involve (1) validation of knowledge resources, (2) appropriate mapping of KR to EHR content via the PPP model, (3) HIPPA and security, (4) integration into the already overtaxed physician workflow, (5) parental literacy, and (6) implement integration with EHR and other legacy systems. Our system will integrate into everyday workflow with mini-

mal perturbation of existing workflow by sending information to patients "on the fly" as they are entering orders, notes, or updating problem lists.

There are many informatics challenges in developing the system we have outlined here for the NICU. The personalization of information over multimedia databases can involve multiple steps of extracting reference literature, summarizing it, and matching it to a query based on contexts of extraction mapped to clinical databases, as investigated and carried out in the PERSIVAL project over a considerable period of time [18]. Many issues of text mining, semantic information modeling and query definition and refinement are involved in ways that present considerable challenges for a full-fledged system design. We have outlined a scenario in the NICU which, with others, are serving our group as use cases in the design of a prototype system for NICU-specific customization of information [19] in a patient-centered manner [20], but that will address issues unique to this health care context, where parents and guardians are the ones being communicated with, and where often inter-cultural and language issues must also be addressed [21]. The present paper summarizes some of the NICU specific factors that make this project such an important health informatics challenge to address.

References

[1] Young SR, Watson MA, Corff KE, Odle P, Hasse J, Bowerman JL. Parent stress and coping in NICU and PICU. J. Pediatr Nurs. 12(3): 169-77; 1997

[2] Conner JM, Nelson EC. Neonatal intensive care: satisfaction measured from a parent's perspective. Pediatrics 1999; 103(1): 336-349.

[3] Davis DL, Stein MT. Parenting Your Premature Baby. Golden, CO: Fulcrum, 2004.

[4] Mann MY, Lloyd-Puryear MA, Linzer D. Enhancing communication in the 21st century. Pediatrics 2006; 117(5): S315-S319.

[5] Bass LS. What do parents need when their infant is a patient in the NICU? Neonatal Network 1991; 10(4): 25–33.

[6] Worchel FF, Prevatt BC, Miner J, Allen M, Wagner L, Nation P. Pediatrician's communication style: relationship to parent's perceptions and behaviors. Journal of Pediatric Psychology 1995; 20(5): 633-644.

[7] Melnyk BM, Feinstein NF, Alpert-Gillis L, Fairbanks E, Crean HF, Sinkin RA, Stone PW, Small L, Tu X, Gross SJ. Reducing premature infants' length of stay and improving parents' mental health outcomes with creating opportunities for parent empowerment (cope) neonatal intensive care unit program: a randomized, controlled trial. Pediatrics 2006; 118(5): e1414-e1427.

[8] Wereszczak J, Miles MS, Holditch-Davis D. Maternal recall of the neonatal intensive care unit. Neonatal Network 1997; 16(4): 33–40.

[9] Barbara Klug Redman, PhD, RN, FAAN . The Practice of Patient Education, 10th Edition - A Case Study Approach. Elsevier Publishing, 2006

[10] Kowalsji WJ, Leef KH, Mackley A, Spear ML, Paul DA. Communicating with parents of premature infants: who is the informant? J of Perinatology (2006) 26, 44-48

[11] Koh THHG, Jarvis C. Promoting effective communication in neonatal intensive care units by audiotaping doctor-parents conversations. International Journal of Clinical Practice 1998;52:27-9.

[12] Dhillon AS, Albersheim SG, Alsaad S, Pargass NS, Zupancic JAF. Internet use and perceptions of information reliability by parents in a neonatal intensive care unit. J Perinatol 2003; 23: 420–424.

[13] Harrison H. The principles of family-centered neonatal care. Pediatrics 1993;92:643-50.

[14] Cullen R. Empowering patients through health information literacy training. Proceedings from World Library and Information Congress: 70th IFLA General Conference and Council. Aug 2004.

[15] Splingaire JM. Readiness for Hospital Discharge: Prdictors and Outcomes in NICU parents. http://stti.confex.com/stti/bcscience38/techprogram/paper_26745.htm

[16] Hospital Discharge of the High-Risk Neonate- Proposed Guidelines. Committee on Fetus and Newborn. American Academy of Pediatrics

[17] Ringer SA, Richardson DK, Sacher RA, Keszler M, Churchill WH. Variations in Transfusion Practice in Neonatal Intensive care. Pediatrics 101:194-200; 1998

[18] McKeown KR Chang SF Ciino J Fenier sk Friedman C Gravano L Hatzivassiloglou V Johnson S Jordan DA Klavans JL Kushniruk A Patel V Teufel S PERSIVAL, A System for Personalized Search and Summarization over Multimedia Healthcare Inforamtion. Proc. JCDL '01: 331-340, June 2001.

[19] DiMarco C Hirst G Wanner ? Wilkinson J HealthDoc: Customizing patient information and health education by medical condition and personal characteristics. 1st Int Wkshp on Artificial Intelligence in Patient Education, Glasgow, 1995.

[20] Gustaffson DH Hawkins R Boberg E Pingree S Serline RE Graziano F Chan CL Imapact of a patient-centered, computer-based health information/support system. Am J Prev Med 1999; 16: 1:9

[21] Piette JD Patient Education via Automated Calls 1999 J Prev Med 1999; 17:138-141.

Address for correspondence:

John Chuo, MD, MS
48 Plantation Rd. Whitehouse Station, NJ 08889, USA
chuojo@umdnj.ed

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Conceptual Model of Health Information Ethics as a Basis for Computer-based Instructions for Electronic Patient Record Systems

Mihoko Okada[a], Kazuko Yamamoto[b], Kayo Watanabe[c]

[a] *Department of Health Informatics, Kawasaki University of Medical Welfare, Japan*
[b] *Adjunct instructor,Department of Health Informatics, Kawasaki University of Medical Welfare, Japan*
[c] *Asahigawasou Medical Welfare Academy School , Japan*

Abstract

A computer-based learning system called Electronic Patient Record (EPR) Laboratory has been developed for students to acquire knowledge and practical skills of EPR systems. The Laboratory is basically for self-learning. Among the subjects dealt with in the system is health information ethics. We consider this to be of the utmost importance for personnel involved in patient information handling. The variety of material on the subject has led to a problem in dealing with it in a methodical manner. In this paper, we present a conceptual model of health information ethics developed using UML to represent the semantics and the knowledge of the domain. Based on the model, we could represent the scope of health information ethics, give structure to the learning materials, and build a control mechanism for a test, fail and review cycle. We consider that the approach is applicable to other domains.

Keywords:

learning support system, health information ethics
conceptual model, UML, electronic patient record system

Introduction

In recent years, the introduction of electronic patient record (EPR) systems into hospitals has influenced the roles and responsibilities of personnel in charge of patient records significantly. At the Department of Health Informatics, Kawasaki University of Medical Welfare, we have about 80 student enrollments each year, and about a half of them aim to work for hospitals as patient record administrators or as health care information technologists in charge of health information systems. To assist students to prepare for EPR systems, we have been developing a computer-based learning system called Electronic Patient Record (EPR) Laboratory for over five years. The details about the EPR Laboratory are discussed in our previous paper (in Japanese) [1]. The system is designed to supplement conventional lectures on health information systems given as part of our undergraduate curriculum based on Recommendations on Health Informatics Education, 2001 by IMIA WG1. Using the Laboratory, the students learn the fundamentals of EPR systems including operational aspects and the subjects connected with patient information handling. It is basically meant for self-learning. When a student failed a test, the system shows which questions were not answered correctly. Then student must review the materials and try the test again. We call this a test, fail and review cycle.

Among the subjects dealt with in the EPR Laboratory is health information ethics. We consider it to be of the utmost importance for personnel involved in patient information handling. Although recently some textbooks on the subject have become available [2-4], health information ethics is still very new, relevant guidelines are updated or published yearly, and the domain is changing and evolving rapidly. There is a variety of material on a range of topics. The materials and tests were arranged sequentially in the Laboratory; however, topics are not necessarily sequentially ordered, and a sequential order is not appropriate for such a complex subject as health information ethics. It was harder to capture and represent the domain than in conventional subjects. Further, when a student failed a test, it was not necessarily possible to indicate automatically which part or all of materials that should be reviewed.

To overcome the problems, we developed a conceptual model of health information ethics using UML. The objectives were to represent the domain of health information ethics, to give structure to learning materials including documents (descriptive texts, published guidelines), exercises and tests, and to build a control mechanism for a test, fail and review cycle into the Laboratory. The definition of health information ethics itself is beyond the scope of the present study, since there is no commonly agreed definition. Instead, we present the developed conceptual model, describe how it is applied in the EPR Laboratory, and discuss the usefulness of the approach.

Materials and methods

Electronic patient record (EPR) laboratory

The present study makes use of the EPR Laboratory as a working bench. The EPR Laboratory was developed using Cache 5.0.11 (InterSystems). It is used in our Health Data Management Practice course for the third year students and delivered in a computer room with one we b server and 140 client PCs. The system is used only for educational purposes

and is not a full-fledged EPR system. The system is designed mainly 1for the students to learn the fundamentals of EPR systems through practices, and 2for instructors to prepare learning materials. EPR Laboratory aims to support the students to learn not only operational aspects but also healthcare services delivery and underlining information technology (IT). IT topics such as information security are covered in the EPR Laboratory in order that the students understand why they should acquire IT knowledge to deal with patient information. Details of IT topics such as SQL and XML are covered by other courses.

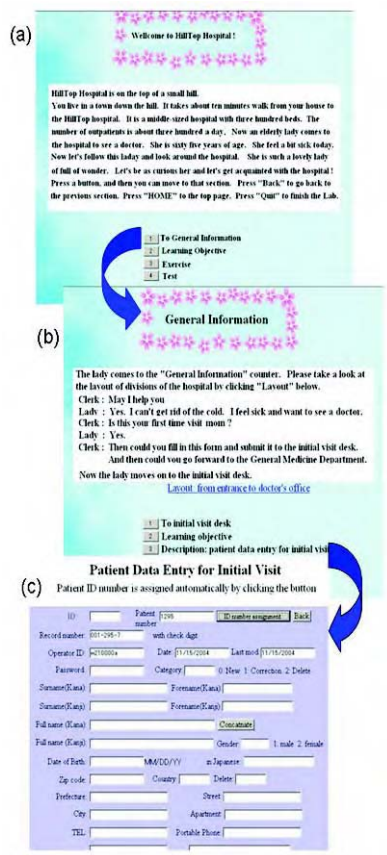


Figure 1 - EPR Labatory: a) The introductory section; b) Description about general information; c) Data entry for patient basic data

Learning materials are arranged in sections, and in each section, an objective(s), descriptive texts and exercises are given. Figure 1 shows the screens of the introductory section on the organization and functions of a hospital and patient flow. English messages in Figure 1 are prepared only for this paper, and in the actual system, all messages are shown in Japanese. The section starts with a fictitious story that an elderly lady comes to visit a hospital (Figure 1(a)). The descriptions continue as the lady moves from general information, initial visit desk, doctor's office, and so on. Figure 1(b) is the screen of general information. Figure 1(c) is the screen of the basic patient data entry, where the students practice operations. There are only two levels of access mode: the record administrator mode and the doctor mode. In the former mode, only limited operations are available including basic data entry, creation and registration of templates for data entry, semi-automatic coding of diagnoses (ICD-10), computation of hospital statistics, and so on. At certain point, there is a test, and the students may not move on to the next screen until they pass the test.

To assist instructors, facilities for preparing teaching materials and tests are provided. Questions on a test may be designed and edited using a textbox, check box, list box and so on, and the score for each question and the pass criteria of the test may be specified.

Conceptual modeling with UML

Modeling is a proven and well-accepted engineering technique in software engineering and it has been applied in a variety of health informatics studies [5-9]. In our study, a conceptual model was developed to solve the problems encountered during the development of the learning contents of health information ethics.

Results

We consider health information ethics a subject that everyone who has access to information systems with patient data should learn. Two levels of learning objectives may be distinguished: the first level for general users of health information systems, and the second level for those who are responsible for patient information handling. In the following, we will focus on the learning objectives of the second level; that is, the objectives for students who aim to be health information professionals.

A conceptual model of health information ethics

General outline of the model

Health information ethics covers both health care services (patient care) and medical and health research. The domain includes a wide range of topics and there are various types of relations among the topics. To represent the knowledge of the domain of health information ethics, we adopted the information modeling technique and developed a conceptual model using UML [10].Figure 2 shows the representation of a simplified version of the model. The model represents the scope of health information ethics covered in our current course including health ethics, information ethics, information security, EBM, and ethics for health information professionals. The conceptual model was developed to give structure to learning materials and to build a control mechanism into the test, fail and review cycle in the EPR Laboratory.

Terms and concepts

The model consists of classes and relationships among the classes. A class, rendered as a rectangle in UML, is used to model abstractions that are drawn from the domain health information ethics. In Figure 2, HealthInf and HealthEthics are examples of classes. We use a short name for a class because of space limitations. HealthInf, for example, is

short for health information. Each of the classes (abstractions) is a part of the vocabulary of the domain.

There are three kinds of relationships: dependencies, generalizations and associations. A 'dependency' is used to show one thing using another. It is rendered as a dashed line directed to the thing being depended on. In Figure 2, for example, there is a dependency from the class 'EBM' to the package 'Epidemiology', where a package (rendered as a tabbed folder) is the basic grouping with which we may organize a model. We use a package when the contents are relevant but are broad in scope and are taught outside the course. 'EBM' uses the basics of the subject 'Epidemiology' taught elsewhere. In software, a 'dependency' means a change in specification of one thing may affect another thing that uses it.

A 'generalization' is a relationship between a general thing (called the parent) and a more specific kind of that thing (the child). It is rendered as a solid line with a large open arrowhead pointing to the parent. In Figure 2, the class ClinicalInf (short for clinical information) is a child of HealthInf. A child inherits the properties of its parents, and it may also have its own properties in addition to those found in its parents. A class may have zero, one or more parents. A class that has no children is called a leaf class.

An 'association', rendered as a solid line connecting classes, is used to show structural relationships; that is, objects (to be described later) of one thing are connected to objects of another. An 'aggregation' is a special kind of association and is used to model a whole/part relationship. It is shown with a diamond at the end. For example, SecurityPolicy is a part of InfSecurity (Information Security) in Figure 2.

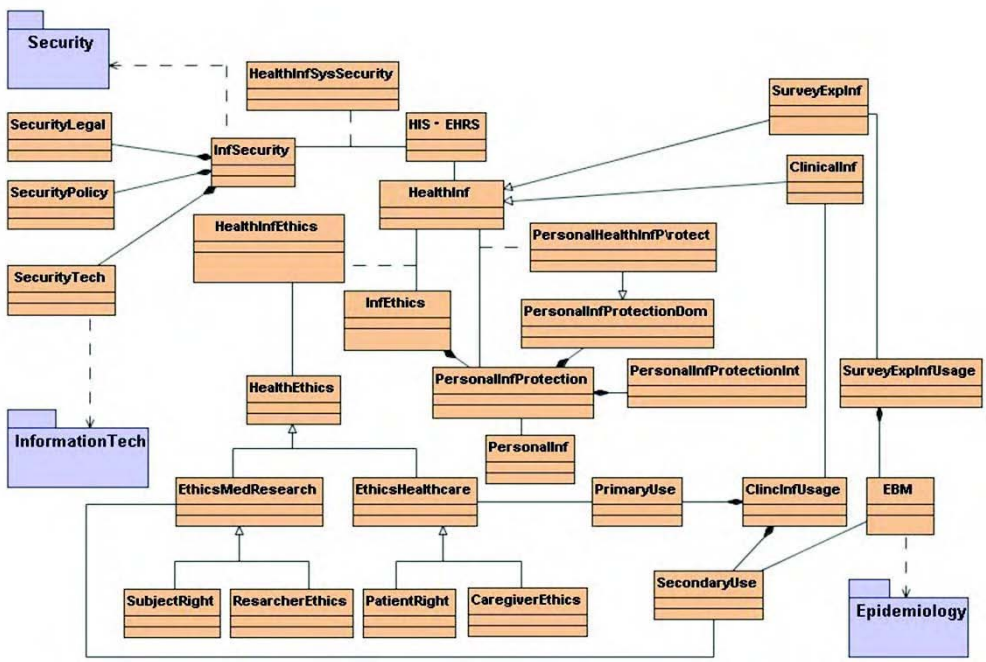


Figure 2 - A conceptual model of "Health Informatin Ethics" (simplified version)

Conceptual model in EPR laboratory

The conceptual model has been incorporated into the EPR Laboratory. Some of the implementation details are described below.

Objects in EPR laboratory

In our application, teaching materials (including documents, descriptive texts, exercises and tests) are physical things and are considered objects. Each object belongs to a class, and a class is a description of a set of objects that share the same attributes, relationships and semantics. There are also abstract classes (classes with no objects). Classes with objects may explicitly be called concrete classes. In our application, concrete classes are mostly leaf classes. In Figure 2, leaf classes are omitted because of space limitations. Below in ResearcherEthics for example, there are classes ResearcherEthicsGuideline, ResearcherEthicsText and ResearcherEthicsTest. The objects that belong to the class ResearcherEthicsGuideline are the guidelines including World Medical Association Declaration of Helsinki Ethical Principles for Medical Research Involving Human Subjects, and the important ethical guidelines on medical or epidemiological research published in Japan. There are dependency relations among the classes, e.g., from ResearcherEthicsTest to ResearcherEthicsText. This means some of the texts are used or referenced by a test, and changes in the texts may affect the test but not the reverse.

Attributes of classes in EPR laboratory

For concrete classes with learning materials as objects, the following attributes are defined, where attribute names are shown in italics. The attributes may be considered as metadata of learning materials. The materials may be documents (such as guidelines), tests and so on.

Class_name

a) title : the title of the material

b) description : the description of the material

c) responsible_entity : responsible entity for the material

d) year_month : year (and month) of publication or writing

e) depend_on : the title(s) of the material(s) that this material depends on

For a test, the following attributes are defined in addition to those defined in a) through e) above:

f) question_ids : identifications of the questions belonging to this test

g) pass_criteria : the pass criteria for this test

An example pass criteria for a test may be 'the sum of all questions is greater than 80 and Question 1 is greater than 5 and Question 7 is greater than 5'.

For a question comprising a given test, the following attributes are defined in addition to the attributes a) to e) above:

f) answer : answer(s) to this question

g) scoring_method : the method for scoring this question

Control of test-fail-review cycle

There are three levels of questions: fundamental, advanced, and applied. A fundamental question is such that an answer can be found easily in a learning material. A question is advanced if the answer may be obtained from multiple materials. A question is applied if an answer cannot be found directly from materials.

If a student fails in answering fundamental or advanced questions, they are guided to read the relevant materials that are indicated by the '*depend_on'* attribute. As an example, Figure 3 shows a fundamental question on personal data protection. The question is stored as an object of the class representing questions for the topic with the following attribute values:

a) title : question 1

b) description : Check if a student has an understanding of the outline of the OECD eight principles

c) responsible_entity : name of the composer of the question

d) year_month : 2005 Dec

e) depend_on : ResearcherEthicsGuideline. 'OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data'

f) answers : 1. e 2. b 3. a 4. g

g) scoring_method: type 3

The value of the attribute '*depend_on'* shows the material(s) that this question is dependent on; that is, the title of the guideline preceded by the class name. The guideline itself is stored as an object of the class 'ResearcherEthicsGuideline', with the following values of the attributes (shown only partly):

a) title : OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data

b) description : ……

For fundamental and advanced questions, answers are selected from a list of choices and the scores are computed automatically according to the rules specified by the attribute '*scoring_method'*.

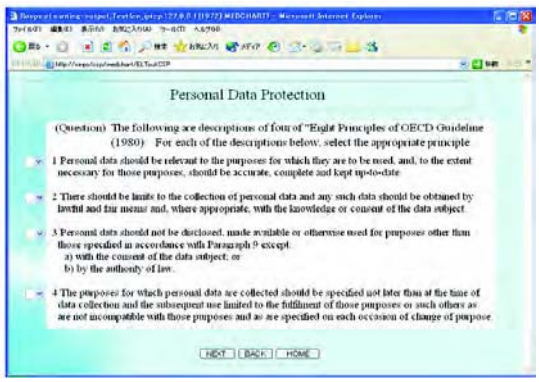


Figure 3 - A sample fundamental question on 'personal data protection'

Applied questions are given toward the end of the course. The questions are often taken from the real life problems that could be very complex. For this type of question, a student is asked to answer in writing in a textbox. The answer is sent to the instructor, and the instructor gives a score. A sample applied question is shown below.

A sample applied question:

> Assume that there was a large-scale accident, and a number of injured people, unconscious and unidentified, were carried into your hospital. Then you receive a phone call from a person saying that his relative might have been involved in the accident. He would like to know whether the named person is among the injured people brought into your hospital. Is it appropriate for you to answer his question over the phone? Discuss why you consider appropriate or not appropriate.
>
> 1. appropriate 2. not appropriate

The applied questions on personal data protection are taken from Q & A for The Guideline on Healthcare and Welfare Entities Handling of Personal Information issued by the Japanese Ministry of Health, Labour and Welfare.

Discussion and conclusion

The EPR Laboratory has been developed for over five years to assist students to prepare for Electronic Patient Record (EPR) systems. The system is meant to supplement conventional lectures. Among the subjects covered in the EPR Laboratory, is health information ethics. For a newly emerging concept such as health information ethics, it was necessary to have a method to represent the knowledge of the domain and organize learning materials in a systematic way. Further, it was also necessary to build a mechanism to tell the students about their weak points when they failed a test so they could review materials more appropriately.

A conceptual model of health information ethics was developed to organize the contents of health information ethics in a way that expresses the semantics of the domain more visibly. Modeling is a proven and well-accepted engineering technique in software engineering, and many aspects of modeling techniques for software apply to our application as well.

The contents of health information ethics in the EPR Laboratory are intended to educate students who aim to be health information professionals. We consider that ability for making decision on real problems of health information ethics may only be acquired through practice in health care settings but, at the same time, we consider students should acquire the fundamental principles before they are involved in the real-life decision-making. The Laboratory may also be used to educate students aiming for to be health care professionals, or for training of health care professionals.

Incorporation of the model into the EPR Laboratory provides a template for storing learning materials in an orderly way, and builds a control mechanism for a test, fail and review cycle. It also provides a means to thoroughly examine the necessity and suitability of the contents. We consider that the approach is applicable to other domains.

References

[1] Watanabe K., Yamamoto K., Okada M., and Takaue R. A computer aided instruction for electronic patient records in a hospital. Japan Journal of Medical Informatics 2006: 25(4): 249-56. (published in March 2006 in Japanese)

[2] Chapman AR. (Ed). Health Care and Information Ethics: Protecting Fundamental Human Rights. Kansas City: Sheed & Ward, 1997.

[3] Anderson J.G. and Goodman KW. Ethics and Information Technology: A Case-Based Approach to a Health Care System in Transition. New York: Springer, 2002.

[4] Harman LB. Ethical Challenges in the Management of Health Information. Maryland: Aspen, 2001.

[5] Okada M., Hashimoto H., and Ohida T. Domain analysis and modeling to improve comparability of health statistics. In: Proc. MEDINFO2001, Patel V. et al., eds. Amsterdam: IOS Press, 2001; pp.1374-78.

[6] Burek P., Hoehndorf R., Loebe F., Visagie J., Herre H., and Kelso J. A top-level ontology of functions and its application in the Open Biomedical Ontologies. Bioinformatics 2006: 22(14): e66-73.

[7] Garde S., Knaup P., and Qumquad H.R. A UML-based approach for remodeling of legacy systems in health care. Int J Med Inform 2003: 70(2-3): 183-94.

[8] Ruggieri A.P., Elkin P.L., Solbrig H., and Chute C.G. Expression of a domain ontology model in unified modeling language for the World Health Organization International classification of impairment, disability, and handicap, version 2. Proc AMIA Symp 2001: 573-7.

[9] Bornberg-Bauer E. and Paton N.W. Conceptual data modelling for bioinformatics. Brief Bioinform 2002: 3(2):166-80.

[10] Booch G., Rumbaugh J., and Jacobson I. The Unified Modeling Language User Guide, 2nd Edition. Reading: Addison-Wesley, 2005.

Address for correspondence

288 Matsushima,
Kurashiki City,
Okayama 701-0193,
Japan

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

PDA-based Informatics Strategies for Tobacco Use Screening and Smoking Cessation Management: A Case Study

Suzanne Bakken[a,b], W. Dan Roberts[c], Elizabeth Chen[b], Joann Dilone[a], Nam-Ju Lee[a], Eneida Mendonca[b], Marianthi Markatou[d]

[a] *Columbia University School of Nursing, New York, New York*
[b] *Department of Biomedical Informatics, Columbia University, New York, New York*
[c] *Adelphi University, Garden City, New York*[d] *Department of Biostatistics, Mailman School of Public Health, Columbia University, New York, New York*

Abstract

The purpose of this case study is to describe three incremental personal digital PDA-based informatics strategies aimed at improving screening for tobacco use and guideline-based tobacco cessation management: 1) PDA clinical log with tobacco cessation diagnoses and plan of care options, 2) PDA decision support system, and 3) PDA decision support system with infobuttons - context-specific links to the National Cancer Institute's Cancer Information Services tobacco cessation information. These strategies were implemented within the context of an evidence-based advanced practice nurse curriculum at the Columbia University School of Nursing.

Keywords:

tobacco cessation, decision support, nursing informatics

Introduction

Tobacco use is a significant global issue. Cigarette smoking is the leading cause of lung cancer and mortality related to lung cancer and has been implicated in other cancers such as larynx, oral cavity and pharynx, esophagus, bladder, stomach, pancreatic, liver, renal cell and renal pelvis, cervix, and acute myeloid leukemia [1, 2]. Tobacco use contributes to 430,000 deaths per year in the U.S., which according to Healthy People 2010 represents 5 million years of potential life lost [3]. Healthy People 2010 further reports that fifty billion dollars per year of direct medical costs are a result of cigarette smoking. In the U.S. as in other countries, the goal of reducing tobacco use remains an urgent necessity for improving both quality and length of life. Meeting Healthy People 2010 goals related to tobacco cessation is estimated to prevent an addition 7.1 million premature deaths [2].

The purpose of this case study is to describe three incremental personal digital PDA-based informatics strategies aimed at improving screening for tobacco use and guideline-based tobacco cessation management: 1) PDA clinical log with tobacco cessation diagnoses and plan of care options, 2) PDA decision support system, and 3) PDA decision support system with infobuttons - context-specific links to the National Cancer Institute's Cancer Information Services tobacco cessation information. We also provide descriptive data regarding use of the first and second strategies.

These strategies were implemented within the context of an evidence-based advanced practice nurse (APN) curriculum at the Columbia University School of Nursing. The first strategy was developed to serve as a control for the second strategy in a randomized controlled trial that is currently in progress. The third strategy is in the design phase for a second randomized controlled trial.

Tobacco cessation and nursing practice

Although tobacco use prevention and smoking cessation counseling are within the purview of nursing practice, actual or perceived barriers frequently deter nurses from undertaking such interventions. McCarty et al. reported reasons given by nurses for not advising patients about smoking cessation: lack of a nursing time, lack of knowledge related counseling techniques, and reluctance to ask a patient to give up a coping mechanism during a hospitalization were given as reasons for nurses not to advise patients about smoking cessation [4]. They also found that nurses often felt counseling was ineffective. Current smokers in this study group were less likely to give smoking cessation advice. In another study, McCarty et al. found that only 50% of patients identified as smokers received smoking cessation advice [5]. Similarly, Gomm et al. reported nurses' attitudinal barriers of 'not my place', 'they'll quit if they want to', and not wanting to add to the patient's 'stress' [6]. In a study of pregnant women, midwifery assessment of smoking status and providing advice to pregnant women was accepted as part of the routine care [7]. Further, the women in the study didn't believe that assessing smoking status or providing advice altered the relationship between them and their midwife. Therefore, the nurse's perception may not be aligned with the perception of the patient about smoking cessation advice.

According to the U.S. Surgeon General's report [8], approximately 36 million American smokers will have at

least one outpatient visit each year but only about half of these smokers will receive smoking cessation advice at their visit. This is consistent with data from the Robert Wood Johnson Foundation that suggests that almost all obstetricians report asking about tobacco use in pregnancy, but fewer than half go on to discuss cessation strategies and offer self-help materials [9]. Smoking cessation advice and treatment is within the purview of the APN. Consequently it is important to design strategies that support guideline-based care and to integrate these into the curriculum so that APN students can address this important health concern during and after their clinical training.

Decision support and guideline-based care

For more than two decades computers have assisted in the provision of reminders to clinicians regarding standardized protocols or guidelines [10-12]. A number of RCTs have demonstrated that computer-based reminders decrease errors of omission [10, 13, 14] and increase compliance with preventive care guidelines [15, 16]. Several recent systematic reviews also suggest that such systems impact clinician adherence to guidelines [17-19].

Additional studies support the potential for decision support to affect clinician behavior related to smoking cessation. A study of physicians in Vermont revealed that even though they were familiar with the U.S. Public Health Service guideline [8] on tobacco use and tobacco dependent treatment and had positive attitudes toward it, they were unfamiliar with smoking cessation resources [20]. The authors concluded that interventions to improve clinician adherence to the guideline should address the inaccurate perception that smokers are unreceptive to counseling and increase knowledge of smoking cessation resources and that a decision support system could meet these needs. An randomized controlled trial of a multifaceted intervention that included computer-generated prompts for care provision by clinic staff found that patients in the experimental group were more likely to report receiving smoking cessation advice by nursing and anesthesia staff and to receive nicotine replacement if nicotine dependent [21].

PDA clinical log at the Columbia Univ. School of Nursing

The PDA clinical log has been in use for more than four years and was designed to serve multiple purposes. These include: student documentation of clinical encounters using standardized nursing terminology and other health-care-related coding systems; use of benchmarking reports to learn to critically examine one's practice over time; and faculty monitoring of student performance. Key steps that were required to create a system that supported these multiple purposes were: design of the system architecture, selection of data elements and standardized terminologies for the data elements, design and implementation of the user interface, design and implementation of the database and knowledge base, and design and implementation of reports. These details have been described in detail elsewhere [22, 23].

APN students enter data using Palm OS devices and synchronize with a central database through Ethernet cradles, WiFi, Internet, or cellular network depending upon their location and type of device (e.g., Palm TX, Treo 700). Faculty and student reports are generated every two weeks.

Use of the PDA Clinical Log is a required part of the APN clinical curriculum. APN student participation in the randomized controlled trial that is testing the affect of a PDA support system on clinical care processes is optional.

Infobuttons and infobutton manager

Infobuttons are context-specific links between a clinical information system and web-based information resources that attempt to anticipate users' information needs and to automatically satisfy those needs [24-27]. Infobuttons are built upon three premises: high-quality, web-based resources exist and are continuously updated by their developers; clinical information systems should access existing web-based information resources, rather than develop and maintain content; and context-specific links within clinical information systems can provide more efficient access to web-based information resources. Infobuttons are currently deployed within two clinical information systems at Columbia University Medical Center: WebCIS and Eclipsys, as well as at two other institutions. In addition, they are accessible via the PDA in an experimental system called PalmCIS [28]. Cimino has recently described the use and positive influence of infobuttons on clinical care [29].

Materials and methods

PDA clinical log with tobacco cessation diagnoses and plan of care options

To create an appropriate application to serve as the control for the PDA decision support system, we needed to make sure it was possible to document all aspects of guideline-based care, i.e., the diagnosis of tobacco dependence and the associated plan of care items for Diagnostics, Procedures, Prescriptions, Patient Teaching and Counselling, and Referrals. First, the U.S. Public Health Service guideline was decomposed into its constituent elements by an expert in tobacco cessation and then reviewed by an additional expert in tobacco cessation and an expert in guideline representation [30]. Then, we represented the diagnostic and intervention concepts with codes from existing terminologies or created new codes in instances where existing terminologies were insufficient to represent the concepts. Next, the guideline-related terms were then organized into existing lists of terms in the menus for diagnoses and the five categories in the plan of care.

PDA decision support system

The PDA decision support system for tobacco screening and smoking cessation management provides decision support for a short clinical intervention based upon the Ask, Advise, Asses, Assess, and Arrange smoking cessation model [8, 30, 31]. As shown in Figure 1, there are two

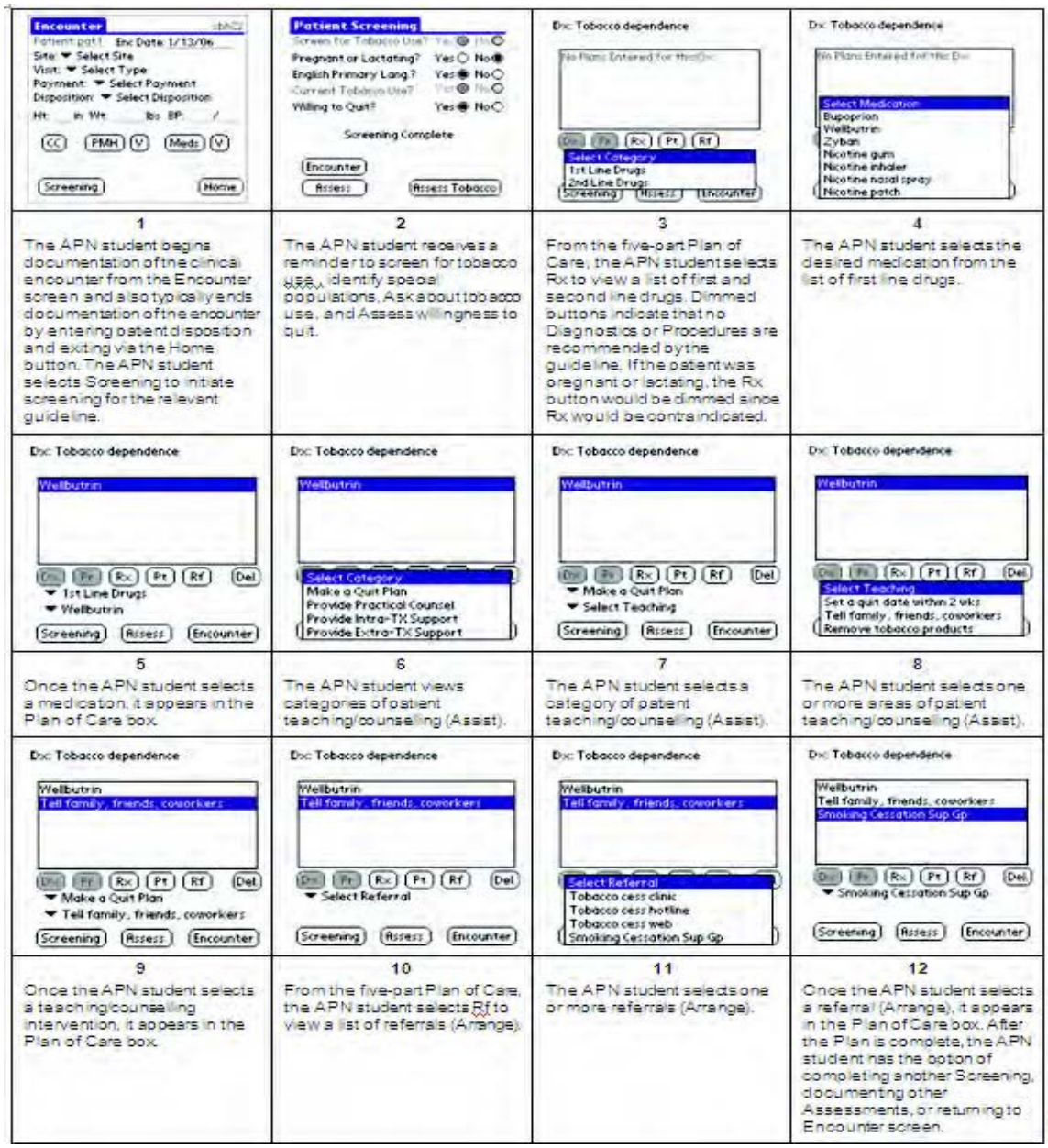


Figure 1 - PDA decision support system for tobacco cessation

aspects to the decision support: the student is prompted to screen and is presented with a guideline-based documentaion template for a plan of care that is tailored according to the smoker's willingness to quit and whether or not they are pregnant or lactating.

PDA decision support system with Infobuttons

Infobuttons are enabled though an Infobutton Manager, which provides a standardized interface for matching user contexts to information resources. This is done through six different methods for accessing the information resources: simple link, concept-based link, simple search, concept-based search, intelligent agent, and calculator [24].

For the PDA decision support system with infobuttons, our design is based upon the use of simple links and concept-based links to access the National Cancer Institute's Cancer Information Service tobacco cessation information within the Patient Teaching and Counselling and Referrals sections of the guideline-based document template. For example, if an APN student wishes to access patient edu-

cation materials in Spanish about making a Quit Plan, she will select the infobutton associated with that intervention and through cellular telephone technology access the Cancer Information Service resource of relevance.

Because the Cancer Information Service resources are not designed to optimize viewing on small devices such as PDAs and cellular telephones, we must also develop and implement an approach for displaying this information in a manner suitable for these small devices.

Results

Descriptive data are reported for Fall semester 2005 (PDA clinical log with tobacco cessation diagnoses and plan of care options) and for Fall 2006 (PDA decision support system) so that they reflect similar points in the APN curriculum. The third strategy is still in the design phase.

PDA clinical log with tobacco cessation diagnoses and plan of care options

In Fall of 2005 when all APN students received the PDA clinical log that included tobacco cessation diagnoses and plan of care options, 1132 encounters were documented using the PDA clinical log for persons more than 9 years old. Only one encounter included a tobacco dependence diagnosis. The associated plan of care included teaching and counselling related to “Make a Quit Plan”.

PDA decision support system

In Fall of 2006, 13 students randomized to received decision support for tobacco cessation entered 150 encounters in which they received a reminder to screen. Patients were screened in 64% (n=96) of the encounters with reminders. Fifteen of the 96 screened were identified as current smokers and 8 indicated that they were willing to quit. However, only one tobacco-related plan of care was documented in the guideline-based documentation template.

The majority of those screened were female (75%) and either Hispanic (42%) or Black (27%). Forty-four percent were between 19-35 years of age. Of those screened, 30% were pregnant or lactating and 42% did not speak English as their primary language. Among those identified as currently smoking, 33% were pregnant or lactating and 73% did not speak English as their primary language.

Discussion

The screening rate for those APN students who received the PDA decision support reminder to screen could be viewed as encouraging, but this only resulted in an associated plan of care in one of eight patients identified as willing to quit smoking. Further study is needed to determine if this is an artifact of the documentation process or an actual quality of care issue. Infobuttons have the potential to provide additional decision support through provision of access to Cancer Information Service resources. The effect of the PDA decision support system on screening rates and adherence to guideline recommendations is currently being tested in a randomized controlled trial in which the control groups receive PDA decision support for either depression or obesity management. A second randomized controlled trial will compare the PDA decision support system and the PDA decision support system with infobuttons on use of Cancer Information Service resources by APN students and the patients they refer to the resources.

Conclusion

There has been little work focused on decision support for APN practice or specifically for tobacco screening and smoking cessation management. Our case study describes our incremental approach to designing PDA-based informatics strategies for screening and management of tobacco dependence within the context of an evidence-based APN curriculum. It is vital that informatics methods and information technologies be conscientiously applied to address this significant health issue from both the clinician and consumer perspectives. Our hope is that by integrating these approaches into the curriculum, the benefits will affect not only the patients the nurses care for during their APN education but also the patients they will care for after graduation through application of such tools into their APN practice.

Acknowledgements

This research is funded by the National Institute of Nursing Research (R01 NR008903) and the National Cancer Institute (R21 126325). The authors thank the other members of the research teams for their contributions.

References

[1] Sasco AJ, Secretan MB, Straif K. Tobacco smoking and cancer: a brief review of recent epidemiological evidence. Lung Cancer 2004;45 Suppl 2:S3-9.

[2] The Health Consequences of Smoking: A Report of the Surgeon General. Washington, DC: Department of Health and Human Services May 27, 2004.

[3] Office of Disease Prevention and Health Promotion. Healthy People 2010: Understanding and Improving Health. Washington, DC: Office of Disease Prevention and Health Promotion (ODPHP), United States Department of Health and Human Services; 2000.

[4] MCarty M, Zander KM, Hennrikus D, Lando H. Barriers among nurses to providing smoking cessation advice to hospitalized smokers. Am J Health Promot 2001;16(2):85-87.

[5] McCarty M, Hennrikus D, Lando H, Vessey J. Nurses' attitudes concerning the delivery of brief cessation advice to hospitalized smokers. Prev Med 2001;33(6):674-681.

[6] Gomm M, Lincoln P, Egeland P, Rosenberg M. Helping hospitalised clients quit smoking: a study of rural nursing practice and barriers. Aust J Rural Health 2002;10(1):26-32.

[7] McCurry N, Thompson K, Parahoo K, O'Doherty E, O'Doherty A. Pregnant women's perceptions of the implementation of smoking cessation advice. Health Educ J 2002;61(1):20-31.

[8] Centers for Disease Control and Prevention. Surgeon General's Report: Reducing Tobacco Use. Atlanta, GA: Centers for Disease Control and Prevention; 2000.

[9] The Robert Wood Johnson Foundation. Substance Abuse: The Nation's Number One Health Problem. Princeton, NJ: The Robert Wood Johnson Foundation; 2001.

[10] McDonald CJ. Protocol-based reminders, the quality of care, and the non-perfectability of man. New Engl J Med 1976;295:1351-1355.
[11] McDonald CJ, Hui SL, Smith DM, Tierney WM, Cohen SJ, Weingerger M, et al. Reminders to physicians from an introspective computer medical record: A two-year randomized trial. Ann Intern Med 1984;100:130-138.
[12] Barnett GO, Winickoff RN, Morgan MM, Zielstorff RD. A computer-based monitoring system for follow-up of elevated blood pressure. Med Care 1983;21(4):400-409.
[13] Burack RC, Gimotty PA, George J, Stengle W, Warbasse L, Moncrease A. Promoting screening mammography in inner-city settings: A randomized controlled trial of computerized reminders as a component of a program to facilitate mammography. Med Care 1994;32(6):609-624.
[14] Overhage JM, Tierney WM, Zhou X, McDonald CJ. A randomized trial of "corollary orders" to prevent errors of omission. J Am Med Inform Assoc 1997;4(5):364-375.
[15] McPhee SJ, Bird JA, Fordham D, Rodnick JE, Osborn EH. Promoting cancer prevention activities by primary care physicians: Results of a randomized, controlled trial. JAMA 1991;266(4):538-544.
[16] McDonald C, Hui S, Tierney WM. Effects of computer reminders for influenza vaccination on morbidity during influenza epidemics. MD Comput 1992;9:304-312.
[17] Johnston ME, Langston KB, Haynes RB, Mathieu A. Effects of computer-based clinical decision support systems on clinician performance and patient outcome: A critical appraisal of research. Ann Intern Med 1994;120(2):135-142.
[18] Shea S, DuMouchel W, Bahamonde L. A meta-analysis of 16 randomized controlled trials to evaluate computer-based clinical reminder systems for preventive care in the ambulatory setting. J Am Med Inform Assoc 1996;3(6):399-409.
[19] Zielstorff RD. On line practice guidelines: State of our knowledge and a look to the future. J Am Med Inform Assoc 1998;5(3):227-36.
[20] Marcy TW, Skelly J, Shiffman RN, Flynn BS. Facilitating adherence to the tobacco use treatment guideline with computer-mediated decision support systems: physician and clinic office manager perspectives. Prev Med 2005;41(2):479-87.
[21] Wolfenden L, Wiggers J, Knight J, Campbell E, Spigelman A, Kerridge R, et al. Increasing smoking cessation care in a preoperative clinic: a randomized controlled trial. Prev Med 2005;41(1):284-90.
[22] Bakken S, Cook S, Curtis L, Desjardins K, Hyun S, Jenkins M, et al. Promoting patient safety through informatics-based nursing education. Int J Med Inf 2004;73(7-8):581-589.
[23] Klein WT, Bakken S. Design and Implementation of a Student Clinical Log Database and Knowledge Base. In: NI 2003; Rio de Janeiro, Brazil; 2003.
[24] Cimino J, Li J, Allen M, Currie L, Graham, M, Janetzki V, et al. Practical considerations for exploiting the world wide web to create infobuttons. Medinfo. 2004 2004(277-281).
[25] Allen M, Currie LM, Graham M, Bakken S, Patel VL, Cimino JJ. The classification of clinicians' information needs while using a clinical information system. AMIA Annu Symp Proc 2003:26-30.
[26] Currie LM, Graham M, Allen M, Bakken S, Patel V, Cimino JJ. Clinical information needs in context: an observational study of clinicians while using a clinical information system. AMIA Annu Symp Proc 2003:190-4.
[27] Lei J, Chen ES, Stetson PD, McKnight LK, Mendonca EA, Cimino JJ. Development of infobuttons in a wireless environment. AMIA Annu Symp Proc 2003:906.
[28] Chen ES, Mendonca EA, McKnight LK, Stetson PD, Lei J, Cimino JJ. PalmCIS: A wireless handheld application for satisfying clinician information needs. J Am Med Inform Assoc 2004;11(1):19-28.
[29] Cimino JJ. Use, usability, usefulness, and impact of an Infobutton Manager. Proc American Medical Informatics Association Symp 2006:151-155.
[30] Roberts WD, Patel VL, Stone PW, Bakken S. Knowledge content of advanced practice nurse and physician experts: A cognitive evaluation of clinical practice guideline comprehension. In: 9th International Congress on Nursing Informatics; 2006; Seoul, Korea; 2006.
[31] Bakken S, Jenkins M, Choi J, Hyun S, John R, Joyce M, et al. Usefulness of a Personal Digital Assistant-based advanced practice nursing student clinical log: Faculty stakeholder exemplars. In: 9th International Congress on Nursing Informatics; 2006; Seoul, Korea; 2006.

Address for correspondence

Suzanne Bakken, RN, DNSc
Columbia University
Mailbox 6
New York, NY 10032
e-mail: suzanne.bakken@dbmi.columbia.edu

This page intentionally left blank

Chapter 12.

Poster Contributions

Selected for
Best Poster Awards 2007

This page intentionally left blank

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Needs Assessment for the Computer-interpretable Hypertension Guideline at Public Health Centers in Korea

EunJung Lee[a], SoYoung Kim[b], InSook Cho[c], JiHyun Kim[a], JaeHo Lee[d], Yoon Kim[b]

[a] *Center for Interoperable EHR, Korea*
[b] *Department of Health Policy and Management, Seoul National University College of Medicine, Korea*
[c] *Department of Nursing, Inha University College of Medicine, Korea*
[d] *Department of Emergency Medicine, University of Ulsan College of Medicine, Korea*

Abstract

Computer-interpretable hypertension guidelines can make for the improvement of blood pressure control rate by assisting clinicians at the point-of-service to behave according to the evidence-based guidelines. We surveyed 117 public health centers in Korea to evaluate current hypertension management status with clinicians' performance, and to analyze the needs for the computer-interpretable hypertension guidelines. Hypertension control rate was 57%, while clinicians overestimated it as 79.5%. 60.4% of patients were treated with 2 or more antihypertensive medications, and most frequently used drug class was calcium channel blocker. Inappropriate prescription rate of contraindicated patients was 2.6%. Two-thirds of clinicians agreed to implement the computer-interpretable hypertension guideline. Implementation of computer-interpretable hypertension guideline is considered as a way to improve the hypertension control rate and to reduce the inappropriateness of the therapeutic choice.

Keywords:

needs assessment, hypertension, guideline, practice guideline

Introduction

For the prevention of cardiovascular diseases, hypertension management takes important portion. Yet world-wide hypertension control rate is known to be less than one-third, especially in Korea, it is about 17%. Computer-interpretable guideline is suggested as an intervention tool for the quality improvement of patient care as modifying clinicians' behavior. This is a preliminary study for developing computer-interpretable hypertension guideline, which is designed to improve hypertension control rate in primary health care. We intended to analyze current hypertension management status, clinicians' performance and needs for the CIGs.

Materials and methods

We surveyed 117 public health centers and 154 clinicians in Gyeonggi province. We extracted medical records of hypertension patients from the centers' clinical database during a 10-month period. Data included blood pressure, medication, comorbid disease, and others. Questionnaire was answered by the clinicians to assess their adherence to hypertension guidelines and needs for the CIG program.

Results

38,474 patients' data and 41 clinicians' survey was analyzed. 61% of clinicians agreed to implement the computer-interpretable hypertension guideline. Hypertension control rate was 57%. Clinicians overestimated as 79.5% of patients had controlled blood pressure. Over 60% of patients were treated with 2 or more antihypertensive medications, and calcium channel blocker was the most frequently prescribed. Based on JNC7 report, inappropriate prescriptions, such as monotherapy for the stage hypertensive patient, beta blocker only for the diabetes patient, combination of beta blocker and non-dihydropyridine calcium channel blocker, thiazide diuretics for the gout patient, and beta blocker for the asthma patient, were 2.6% (N=1001).

Conclusion

Clinicians overestimated the proportion of their patients with controlled blood pressure. Implementation of computer-interpretable hypertension guideline is considered as a way to reduce the inappropriateness of the therapeutic

Address for correspondence

EunJung Lee, M. D. toro0117@snu.ac.kr

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Development of a Personal Medical Recorder on a Cell Phone

Akihiro Takeuchi[a], Katsura Kobayashi[b], Noritaka Mamorita[b], Noriaki Ikeda[a]

[a] *Departments of Medical Informatics, School of Allied Health Sciences, Kitasato University, Japan*
[b] *Graduate School of Medical Sciences, Kitasato University, Japan*

Abstract

Paper medical records have effectively been used in chronic diseases without information technology. To facilitate self-control in hemodialysis and observe a patient's condition continuously, we developed a mobile phone-based personal medical recorder for patients suffering chronic renal failure. The application is based on Java2 Micro Edition and operates like a scheduler. The application stores laboratory data, such as BP, BUN, creatinine, HbA1c, etc., and other pertinent clinical comments into memory on a cell phone. The application can also customize, add or delete items (laboratory data, medications, questions, etc.). Detailed graphic displays of the data are shown. The data can also be sent to a PC with infrared communications. In a usage trial, patients were favorably receptive about this application and indicated that they wanted to continue using it.

Keywords:

cell phone, medical records, infrared transmitter

Introduction

Using a diary, a chronically ill person can gain an overview of his medical situation [1]. Although patients with chronic renal failure can maintain an adequate daily nutritional intake, they might also be able to recognize their own pathophysiological state and be aware of the settings for their own hemodialysis. Although data collection was more reliable with a palmtop computer than with paper diaries [2], there have been no reports presenting a PDA/cell phone-based clinical recorder for chronic hemodialysis other than for dietary monitoring [3]. Therefore, we developed a cell phone-based medical recorder for the self-management of chronic diseases.

Materials and methods

An application was designed as a self-controllable tool for chronic diseases. It consists of a calendar canvas, dataentry canvas, plot canvas, memory control module and an infrared module (Figure 1, left). The plot canvas shows graphs of clinical and laboratory data, such as BP, BUN, creatinine, HbA1c, etc. These data are stored in text form in the memory of a cell phone (maximum 200 kB). The data size of this schedule is estimated to be about 100 bytes. The monthly or daily data are sent to a PC by an infrared module.

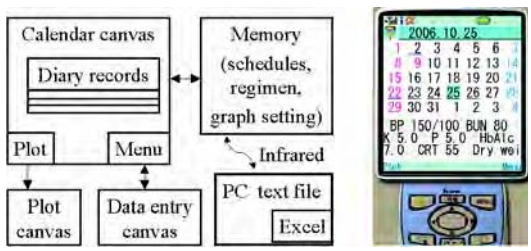


Figure 1 – Overview of the application and a calendar canvas

Results

The application was based on Java2 Micro Edition and i-Appli Tool (NTT DoCoMo, Inc.). The application (medData.jar 100 kB) operates like a scheduler. (Figure 1, right). All of the data, BP, laboratory data, medications, etc. on each day are listed on the data entry canvas. The patient interactively and personally types in numeric data and/or comments. Detailed graphic displays of data are shown sequentially on a plot canvas in years, months, weeks, or days. Color and scale customization of the plot canvas is possible for each set of laboratory data. Customizing items (laboratory data, medications, questions, etc.) is done on the data entry canvas. The application was adapted for several patients to use at the same time in one trial. The patients were favorably receptive about this application and said that they would like to continue using it. This personal medical recorder based on a cell phone could be useful as a self-management tool for patients with chronic diseases.

References

[1] Baert H, Gielen P, and Smet M. A health diary for the chronically ill. WHO Reg Publ Eur Ser 1992: 44: 328-31.
[2] Hyland ME, Kenyon CA, Allen R, et al. Diary keeping in asthma: comparison of written and electronic methods. *BMJ* 1993: 306: 487-9.
[3] Sevick MA, Piraino B, Sereika S, et al. A preliminary study of PDA-based dietary self-monitoring in hemodialysis patients. J Ren Nutr 2005: 15: 304-11.

Address for correspondence

take@kitasato-u.ac.jp, Kitasato, Sagamihara, Japan

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Implementing and Evaluating a Laboratory Information System to Optimize the Treatment of Tuberculosis Patients in Peru

Joaquin A. Blaya[a,b], Sonya S. Shin[b,c], Martin JA Yagui[d], Luis Asencios[d], Javier Vargas[d], Carmen Suares[e], Gloria Yale[f], Hamish SF Fraser[b,c]

[a] *Harvard-MIT Division of Health Sciences & Technology, USA,* [b] *Partners In Health, USA,* [c] *Division of Social Medicine &Health Inequalities, Brigham & Women's Hospital, USA,* [d]*Instituto Nacional de Salud, Perú,* [e]*Dirección de Salud IV Lima Este, Perú,* [f]*Dirección de Salud V Lima Ciudad, Perú*

Abstract and objective

Multi-drug resistant tuberculosis patients in resource-poor settings experience large delays in starting appropriate treatment and may not be monitored appropriately due to an overburdened health care system, communication delays, and missing or error-prone data. A web-based laboratory information system "e-Chasqui" has been implemented in Peru to alleviate these problems by improving the timeliness and quality of laboratory data. It has been deployed in the national TB laboratory, two regional laboratories and twelve health centers. High user satisfaction and heavy use has led to e-Chasqui is being expanded to more institutions. A study is being performed to measure its impact and generalizability.

Keywords:

clinical laboratory information system, computerized; evaluation studies; developing countries

Introduction

Treatment for MDR-TB in Peru is often delayed by over three months after initial presentation.(1) These potentially dangerous delays occur because of the long test processing time, cumbersome collection and communication procedures, and loss of specimens and test results. Similar problems are prevalent in other settings including South Africa.(2). An electronic information system can improve the handling and communication of these data between institutions. Decreasing treatment delays and ensuring an appropriate drug regimen should improve outcomes and reduce transmission.

Methods

Partners In Health has developed a web-based medical record system (PIH-EMR)(3) to support the treatment of TB, with data on 15523 patients. We created a web-based laboratory information system, "e-Chasqui" to connect laboratories to health centers to reduce delays and facilitate communication and analysis. e-Chasqui includes tools to improve data quality, notify health centers of new results, alert physicians of high-risk patients and create laboratory reports. Here we report on the system's implementation, use, and preliminary results from its evaluation.

Results

e-Chasqui has been implemented in the national reference laboratory, two of five regional laboratories in Lima and twelve health centers. Since its implementation in November, 2005, 19900 smear, 21196 culture and 3076 drug sensitivity test results have been entered. In 2006 all health centers have viewed 100% of their results online. Due to user satisfaction and heavy use, public officials have asked to expand the system to 3 other laboratories and over 10 other health centers.

Conclusions

This experience demonstrates the possible benefits of implementing a web-based laboratory information system in a low resource setting. A prospective randomized evaluation is being performed to measure its impact on delays, errors, and quality of care, including time to prescribe an effective drug regimen.

References

1. Yagui M, Perales MT, Asencios L, et al. Timely diagnosis of MDR-TB under program conditions: is rapid drug susceptibility testing sufficient? Int J Tuberc Lung Dis. 2006 Aug;10(8):838-43.
2. Hurtado R. Personal Communication. 2006.
3. Fraser H, Jazayeri D, Mitnick C, Mukherjee J, Bayona J. Informatics Tools To Monitor Progress And Outcomes Of Patients With Drug Resistant Tuberculosis In Peru. Proc AMIA Symp. 2002:270-4.

Address for correspondence

HSF Fraser Hamish_Fraser@hms.harvard.edu

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Risk analysis – A Tool for IT Development and Patient Safety
A Comparative Study of Weaknesses Before and After Implementation of a Health Care System in the County Council of Ostergotland, Sweden

Annica Öhrn[a], Gunilla Eriksson[b]

[a] *Patient safety office of county council of Ostergotland, Sweden*
[b] *Laboratory medicine unit of county council of Ostergotland, Sweden*

Abstract

There is a lack of tools to secure patient safety considering information security and health care systems. Risk analysis can be used as a systematic method for identifying risks before a medical error occurs. In healthcare is this type of analysis uncommon use. An indicator for this need of systematic proactive work is the amount of adverse advents that are reported in the deviation system .In this comparative study of a health care system during 2004 and 2006, one of the results is an reduction in the numbers of risks due to technique & equipment and in the area of training & competence. This study gives an indication of that risk analysis helps the health care organisation to develop and manage routines for reducing possible risks.

Keyword:

information system, safety management, risk management/methods, medical errors, systems analysis, quality of health care

Introduction

In the county council in Ostergotland, Sweden has a supporting system for requests and laboratory reports, called LR, and been implemented during the last year and a half. A risk analysis has been performed before and after implementation and is planned to continue regulary for the coming years. The aim of this study was to perform a comparative risk analysis for a recently implemented health care system.

Methods

Risk analysis has been performed before and after the implementation of the system. Risk analysis is a systematic tool, including: process mapping, risk identification, determination of the severity and probability of each risks and action plans. The analysis has a patient safety perspective and has been performed with a trained risk analysis team including members from different roles/specialities. The risks were categorised in areas of six potential sources; technique & equipment, rules/policies/procedures, environment, training & competence, barriers and communications & information.

Results

The risk analysis shows that the total risk points are lower during 2006 for the areas technical solutions and procedures compared to the points of the same areas during 2004.

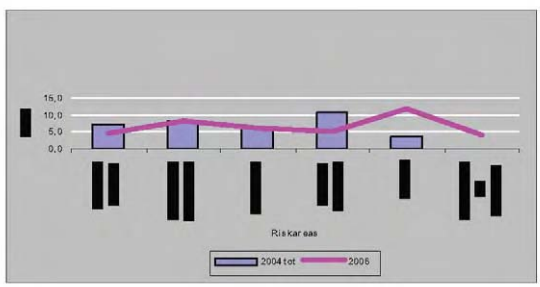


Figure 1 – Risk profile, comparison 2004-2006 LR

Discussions

The risk analysis performed demonstrated differences between the results obtained prior and after the implementation. This was especially true within the sources; technique & equipment, training & competence as well as for barriers. The risk profile within technique & equipment demonstrated that most of the risks prior to the implementation have been dealt with. Training of the personal drastically decreased the risk profile within training & competence. However, increased risks were found for the source barriers as a consequence of the introduction of new barriers, which reduced the accessibility of laboratory results. This was a consequence of the Swedish legislation for patient related information in medical records, which prevents full access of a patient's laboratory test data without the patient consent.

The risk profile for the source; rules/policies/procedures is unchanged. The numbers of risks have decreased but the risks with a high score, which were identified prior to the implementation, were never dealt with.

Conclusion

Risk analysis is a useful method for identifying and managing possible patient safety risks and also a method for proactive work in health care organisations. Systematic and regular risk analysis is preventive of managing new routines. This may be a constructive tool for the patient safety work

References

[1] Stalhandske E, DeRosier J, Patail B, Gosbee J. How to make the most of failure mode and effect analysis, Biomed Instrum Technol. 2003 Mar-Apr;37(2):96-102

Address for correspondence Patient safety office in county council of Ostergotland
University hospital, S-581 85 Linkoping, Sweden
Annica Öhrn e-post: annica.ohrn@lio.se

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Multi-label Text Classification of German Language Medical Documents

Stephan Spat[a], Bruno Cadonna[a], Ivo Rakovac[a], Christian Gütl[b], Hubert Leitner[c], Günther Stark[c], Peter Beck[a]

[a] *Institute of Medical Technologies and Health Management, JOANNEUM RESEARCH Forschungsgesellschaft mbH, Graz, Austria*
[b] *Institute for Information Systems and Computer Media, Graz University of Technology, Graz, Austria*
[c] *Steiermärkische Krankenanstaltenges. m.b.H., Graz, Austria*

Abstract and objective

At nearly every patient visit, medical documents are produced and stored in a medical record, often in an unstructured form as free text. The growing amount of stored documents increases the need for effective and timely retrieval of information. We developed a multi-label text classification system to categorize free text medical documents (e.g. discharge letters, clinical findings, reports) written in German into predefined classes. A random sample of 1,500 free text medical documents was retrieved from a general hospital information system and was manually assigned to 1 to 8 categories by a domain expert. This sample was used to train and evaluate the performance of 4 classification schemes: Naïve Bayes, k-NN, SVM, and J48. Additional tests of the effect of text preprocessing were done. In our study, preprocessing improved the performance, and best results were obtained by J48 classification.

Keywords:

machine learning, classification, medical records, multi-label

Introduction

At nearly every patient contact with healthcare-providers, medical documentation is generated and stored in medical or nursing records, often as free text. With the increasing amount of stored, unstructured free text information, the need for effective and timely retrieval of relevant information is growing. In this work, we describe the development and the evaluation of an information system for multi-label classification of medical documents into predefined classes.

Methods

A random sample of 1,500 unstructured, free text documents written in German was extracted from an electronic medical record (EMR) of a general hospital in Austria. A domain expert (physician) manually classified the retrieved documents into one or more of the following classes: surgery, vascular surgery, casualty surgery, internal medicine, neurology, anesthesia and intensive care, radiology and physiotherapy. In average, 1.47 labels were assigned to a document. We built an automated multi-label text classification system in Java based on Weka [1], an open-source machine-learning framework. Four different kinds of classification schemes were compared: Naïve Bayes, k-NN, SVM and J48. 10-fold cross validation was used for evaluation. Moreover, the influence of text preprocessing (e.g. stop-word-removal, stemming, lowercasing) was studied.

Results and conclusion

Results for the F-measure [2] with and without preprocessing are shown in figure 1. J48 performed best, followed by 1-NN, SVM, and Naïve Bayes. The results were improved by text preprocessing. The best classification scheme (J48) with text preprocessing achieved an F-measure of 0.886.

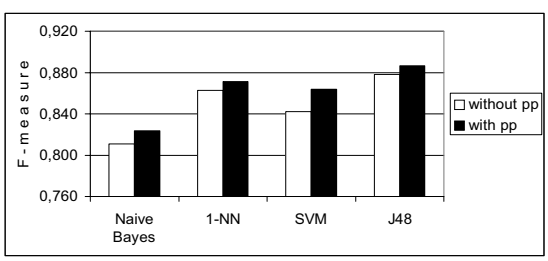


Figure 1 - F-measure with and without text preprocessing (pp)

Results show that it is possible to classify medical documents written in German originating from a general hospital with automated machine-learning classification schemes with promising results, comparable with [3]. This classification system is used in a prototype of an information retrieval system for score-calculation, thus influencing the display order of search results. Further studies are needed to evaluate the accuracy of the developed system in other hospitals as well as the user-perceived benefits of this prototype.

References

[1] Witten IH, Frank E: Data Mining: Practical machine-learning tools and techniques, 2nd Edition, Morgan Kaufmann, San Francisco, 2005.

[2] Hripcsak G, Rothschild AS: Agreement, the F-Measure, and Reliability in Information Retrieval. J Am Med Inform Assoc. 2005; 12(3): 296-298.

[3] Wilcox A, Hripcsak G: Classification algorithms applied to narrative reports. Proc AMIA Symp. 1999; 455-9.

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

What Health Influences are Caused by EMR Working?- In Case of Japanese Nursing Situation

Yukie Majima [a], Yasuko Maekawa[a]

[a] *School of Nursing, Osaka Prefecture University, Osaka, Japan*

Abstract

The purpose of this study was to determine how nurses used visual display terminals (VDT) and whether the use of VDTs had any influence on their health. In our study, the use of the EMR and the health influences of introduction of the EMR were surveyed using a questionnaire in a general hospital. The results indicated that VDT works were conducted in nurse stations rather than at the bedside. Only about 10% of nurses did exercises to prevent neck, shoulder and back pain before starting work. Since more than half of the nurses remained standing while inputting, extra guidelines should be developed for this aspect as the MHLW guidelines currently only deal with seated operations. Health influences after introduction of the EMR, more than half the subjects complained of worsened symptoms including eye strain, stiff shoulders and neck, and general fatigue, all of which are signs of VDT syndrome.

Keyword:

nurses, VDT works, electronic medical record system, health influences

Introduction

Currently in Japan, the use of EMR and digital ordering systems in hospitals is increasing. Since medical service is one of the seven main fields of the e-Japan priority policy program, it is assumed that digitalization will be promoted in medical facilities. Accordingly, appropriate workplace health measures should be taken in response to changes in the working environment of nursing personnel.

Methods

1) Study period: March 8 to 29, 2005. 2) Subjects: Nursing care personnel (378 individuals) working at a private general hospital (477 beds) in Japan which had introduced an EMR around a year previously. 3) Method and contents of study: Self-reported questionnaire study. A questionnaire was delivered to the subjects and was recovered on completion under ethical cares. It consisted of questions regarding the VDT working situation of EMR((1)posture, (2)place, (3)adjust materials around PC, (4)input device, (5) time per one use of EMR, (6)brightness around PC on each shift, (7)setting-up exercise before working) and the health influences (eye strain, low vision, dried eyes, stiff neck and shoulders , numbness of fingers, back pain, fatigue, headache, stress, gloomy) after using the EMR.

Results

323 questionnaires delivered were recovered (85.4% recovery). 193 subjects (59.8%) were in their 20s, 80 (24.8%) were in their 30s, 27 (8.4%) were in their 40s and 19 (5.9%) were in their 50s. Four subjects (1.2%) did not reveal their age. The average number of years working at the hospital was 5.67.

1) The VDT working situation of EMR

VDT works were conducted in nurse stations rather than at the bedside. Since more than half of the nurses remained standing while inputting records.

2) The health influences after introduction of the EMR

Only about 10% of nurses did exercises to prevent neck, shoulder and back pain before starting work. More than half the subjects complained of worsened symptoms including eye strain, stiff shoulders and neck, and fatigue, all of which are signs of VDT syndrome (Figure 1).

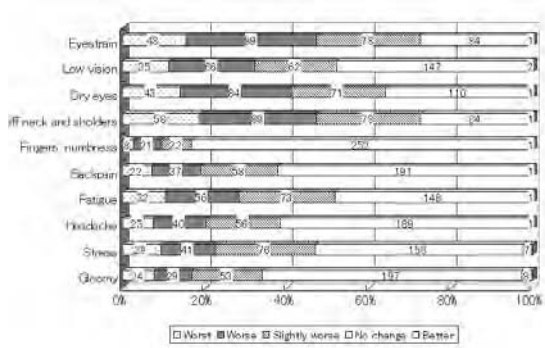


Figure 1 - Subjective signs of health influences after using EMR

Conclusion

Nurses' VDT working situations are different form those of office workers etc. Workplace health measures for VDT operations by nurses should be urgently revised based on the results of this study.

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Using PDA to Transform the Long MDS-HC Evaluation Form into a Favored System

Chiao-Ling Hsu, RN, MS, Yu-Yin Kuo, RN, MS, Polun Chang, PhD

Institute of Biomedical Informatics (formerly Health Informatics and Decision Making), National Yang-Ming University, Taiwan

Abstract

The MDS-HC has been an effective home care evaluation form. However it was not yet accepted in Taiwan because it is too long for our over-burdened home care nurses. We used a self-developed PDA information representation model to design the PDA MDS-HC support system and used the Technology Acceptance Model to examine its potential acceptability. The results showed a well-designed PDA could greatly improve the usability of a originally un-favored paper system.

Keywords:

MDS, home care, PDA, information representation

Introduction

The Resident Assessment Instrument, consisting of a Minimum Data Set (MDS) evaluation form, Client Assessment Protocols (CAPs) and Triggers, has been well internationally known, applied, and even used for the insurance claim, in the long term care [1]. The MDS-HC is vary complete and covers 17 main categories and as many as 64 evaluation items. However, currently its professional completeness doesn't make it a formal evaluation form for the home care in Taiwan because it is too long for our over-burdened home care nurses.

PDA has been a potentially useful tool for mobile nursing, especially for those in home care settings. But its interface features, such as difficult data entry in writing and small display screen, still confine its use. We developed a PDA information representation approach which had been practically proven to be easy to use by nurses [2][3]. A modified version was latter developed for the long forms.

The purpose of this study was to use the modified PDA information representation approach to design the PDA-based MDS home care evaluation support system and to examine the home care nurses' acceptance of the PDA-based MDS-HC system.

Methods

The modified PDA information representation approach is mainly composed of 4 principles:

1. The screen display area is separated into three sections for the users to clearly see and to easily navigate through the form.

2. User should be able to switch to any item in less than 3 taps.

3. The hand-writing should be replaced, if possible, by tapping for data entry.

4. User should be well reminded what tasks have been done.

A team of nursing user, nursing programmer and experienced PDA programmers was organized and the prototyping approach was used to develop the system. A convenient sample of 24 subjects, who were trained of the MDS, was used for testing. 3 representative home care scenarios were written for users to test the PDA system. Davis' Technology Acceptance Model was used to evaluate users' acceptance of the system in terms of their perceived ease of use and usefulness.

Results

The representative PDA screen was shown in Figure 1, in which uses could navigate the system through tapping the upper main category and subcategory buttons, and enter data in the lower half data entry area. The highlighted buttons mean answered questions.

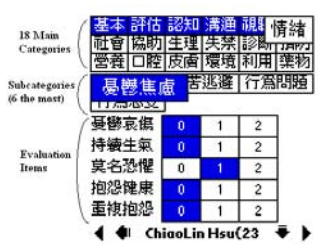


Figure 1. The representative screen shot.

Figure 1 - The representative screen shot

58% of subjects never used the MDS-HC form. Most of the subjects have BS degree and aged from 20-50. 100% agreed the system was easy to use and 92% agreed the system was useful.

Conclusions

Our study showed that a well-usability-engineered PDA system can well improve the usability of both the professional forms, which was too long to be practically accepted, and PDA, which small-display interface usually hinders its popularization.

References

[1] Morris JN, Fries BE, Bernabei R, Steel RK, Ikegami N, Carpenter GI. RAI-Home Care (RAI-HC) assessment manual for version 2.0. Boston: Inter RAI; 1999.

[2] Chang P, Hsu YS, Tzeng YM, et al. Development and Pilot Evaluation of User Acceptance of Advanced Mass-Gathering Emergency Medical Services PDA Support Systems. MedInfo 2004 Symposium, p:1421-1425.

[3] Chang P. Tzeng YM, Wu SC, et al. Development and Comparison of User Acceptance of Advanced Comprehensive Triage PDA Support System with a Traditional Terminal Alternative System. American Medical Informatics Association 2003 Symposium, p:140-144.

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

A Sustainable, Multi-Organizational Model for Decision Support During Public Health Emergencies

Julie J. McGowan[a,b], Margaret W. Richwine[a], J. Marc Overhage[a,b,c]

[a] *School of Medicine, Indiana University, Indiana*
[b] *Regenstrief Institute, Inc., Indiana*
[c] *Indiana Health Information Exchange*

Abstract

In an effort to provide decision support during times of public health emergencies, Regenstrief Institute, Inc. and the Indiana University School of Medicine, in partnership with the county and state health departments, created a sustainable model to deliver information to health care providers and public health workers. The model leverages extant systems and processes, including active surveillance through electronic laboratory reporting, delivery of health information as part of the Indiana Health Information Exchange, and evidence based utilities and Blog technology to create a public health utility with disease specific information and epidemiologic reporting requirements.

Keywords:

disease outbreaks; bioterrorism; information dissemination; decision support systems, clinical

Introduction

Knowledge at the point of decision making is critical to improving health care and this is especially true in public health when medical decisions in response to a bioterrorist event or an emerging epidemic could have a much greater impact. However, the ability to access such information has been extremely problematic because of the different levels of connectivity and training available to public health workers and health care providers across large areas.

To address this issue a partnership was formed among the Indiana State Department of Health and the Marion County Health Department, the Indiana University School of Medicine Libraries (IUSML) and the Indiana Health Information Exchange to establish a cost-effective model for both alerting health care providers and public health workers to developing threats and providing them the needed information for both treatment and reporting.

Methods

The model, predicated on an earlier construct [1], was built on three existing programs. The Regenstrief Institute, Inc. in conjunction with the Indiana Network for Patient Care and the Indiana State Department of Health developed an active surveillance of reportable conditions program with electronic laboratory reporting. [2] I3, the Indiana IAIMS (Integrated Advanced Information Management Systems) Initiative, through the Indiana Health Information Exchange and its DOCS4DOCS utility, provided a framework to notify health care providers about real or suspected public health emergencies. [3] The IUSML offered evidence based decision support using Web technologies.

The goals of the project were to insure rapid notification of public health problems to a large yet appropriate group of health care providers and to provide access to quality filtered knowledge supporting treatment guidelines and reporting requirements. Because the eventual objective was to operationalize the model for long term adoption, it was essential to develop sustainable methods.

Electronic laboratory reporting uses LOINC coding and HL7 messaging to provide timely information to the health departments regarding the potential for a public health event. Notifiable conditions from laboratories around the Indianapolis metropolitan statistical area are reported to the health departments. The enhancement model facilitates provider notification and the creation of decision support information.

The health care provider indexing in the Indiana Health Information Exchange enables the rapid notification of health care providers about emerging public health threats based on location and/or condition. This insures a higher level of relevance of the warning for the provider. Simultaneously, the IUSML is contacted about the suspected problem and receives information concerning proposed optimum treatment and management guidelines and the required reporting of the specified condition.

The IUSML is responsible for searching for current evidence about management of the condition and creating a web presence containing this information in a user effectiveness format. In addition, reporting requirements and links to the health department web sites are included. To insure sustainability, a Blog utility was chosen because of its ease of document creation and maintenance, the famil-

iarity of the format for most users, and its functionality in searching for prior conditions.

In addition to the reporting and management information for health care providers, a listserv is maintained for notification of public health workers. The health departments have their own systems for notifying their affiliated sites, however the notification of access to knowledge-based information is handled by the IUSML once the Blog site has been updated and approved.

While most of the information about emerging public health events comes from the electronic surveillance functionality, some comes from the public press. While this usually involves very specific localities and has little potential to develop, decision support is also beneficial for these events and the Web site is updated as the need presents itself.

Results

As of November 2006, there were thirteen public health events that prompted the creation of knowledge support on the Web site. The mechanism for the notification of health care providers has been developed using extant resources but has as yet to gain widespread adoption, not because of the technology but because of organizational issues.

The process for the notification of public health workers was implemented with the second public health event posted and, as a result, the utility has achieved wide spread use. Anecdotal responses to the initiative have indicated that the Web site has provided critical information for both management of health care conditions and epidemiologic reporting.

Discussion

By using three extant processes, active surveillance through electronic laboratory reporting, DOCS4DOCS delivery of information to health care providers, and the evidence based medicine services of the IUSML, a sustainable model was developed to deliver decision support information to targeted health care providers and public health workers during instances of public health events.

The technology and processes were easily adapted to meet the needs of the health departments and system users. However, organizational issues regarding the health care provider notification component have precluded the full operationalization of the system. It is anticipated that these issues will be resolved within the next year and that the system will be fully implemented. However, the public health worker notification and the creation and use of the Blog utility have proven to be an effective means to providing decision support during times of emerging infectious diseases and bioterrorism events.

Conclusion

Emerging public health crises require coordination of information from a variety of sources and targeted provision of quality filtered knowledge to a wide range of health care workers. Access to evidence is critical to the management of the events for both the population and individual patients; collecting information to monitor the impact of the occurrences contributes to the knowledge base and helps mitigate future occurances. Using extant technology and promoting organizational partnerships can offer a sustainable model to enhance responsiveness to these events.

Acknowledgements

The authors performed this work at the Regenstrief Institute for Health Care and the Indiana University School of Medicine and they were supported in part by the National Library of Medicine, grant number GO8 LM008232-02.

References

[1] Kay BA, Timperi RJ, Morse SS, Forslund D, McGowan JJ, and O'Brien T. Innovative information-sharing strategies. Emerg Infect Dis 1998: 4(3): 465-6.

[2] Overhage JM, Suico J, and McDonald CJ. Electronic laboratory reporting: barriers, solutions and findings. J Public Health Manag Pract 2001: 7(6): 60-6.

[3] McGowan JJ, Overhage JM, Barnes M, and McDonald CJ. Indianapolis I3: the third generation Integrated Advanced Information Management Systems. J Med Lib Assoc 2004: 92(2): 179-87.

Address for Correspondence

Julie J. McGowan, Ph.D., FACMI
IU School of Medicine, IB-310
975 W. Walnut St.
Indianapolis, IN 46202-5121 USA

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Customized Early Warning System Based on HTN for Home Healthcare Model

Seung-Jin Jang[a], Jip-Min Jung[a], Sung-Oh Hwang[b], Young-Ro Yoon[a]

[a] *Department of Biomedical Engineering, Health Science College, Yonsei University, Wonju, South Korea*[b]*Department of Emergency Medicine, Wonju College of Medicine, Yonsei University, Wonju, South Korea*

Abstract and objective

We adopted hierarchical task network (HTN) planning in a customized early warning system for home healthcare model. It is necessary to design a customized early warning system for the patients who were in various health states, because regular or irregular report format and alarm delivery systems were diversified according to severity of health state. HTN planning is suitable for use of constraint programming so as to effectively prune the search space during the search for solution. An efficient and scalable information control is presented by use of HTN. The paper also briefly deals with a process strategy for the early warning system.

Keywords:

hierarchical task network, emergency response, home healthcare model

Introduction

As the elderly population and the demand of well-being life increasingly grow, IT-based technology allowing bio-signal measurement and assessment at home have been focused nowadays. The elderly or citizen can be efficiently controlled with respect to health status parameters which act as input features in the early warning system for health status.

Methods & results

The early warning system developed by Home Healthcare Management System Research Center (H2MSRC) in Yonsei University adopted HTN planning in order to efficiently design the planning of emergency response because of extendible, intelligible, and easily communicated properties. Events derived from 7 objective status parameters measured from 4 types of sensor devices and 3 subjective parameters results in customized emergency response according to the formalized SHOP domain model. (Ref. figure 1)

Conclusion

HTN planning was practically adopted in our early warning system for Home healthcare model and revealed the intelligent and collaborative planning ability.

Acknowledgment

This study was supported by a grant of the Korea Health 21 R&D Project, Ministry of Health & Welfare, and Republic of Korea (02-PJ3-PG6-EV01-0001).

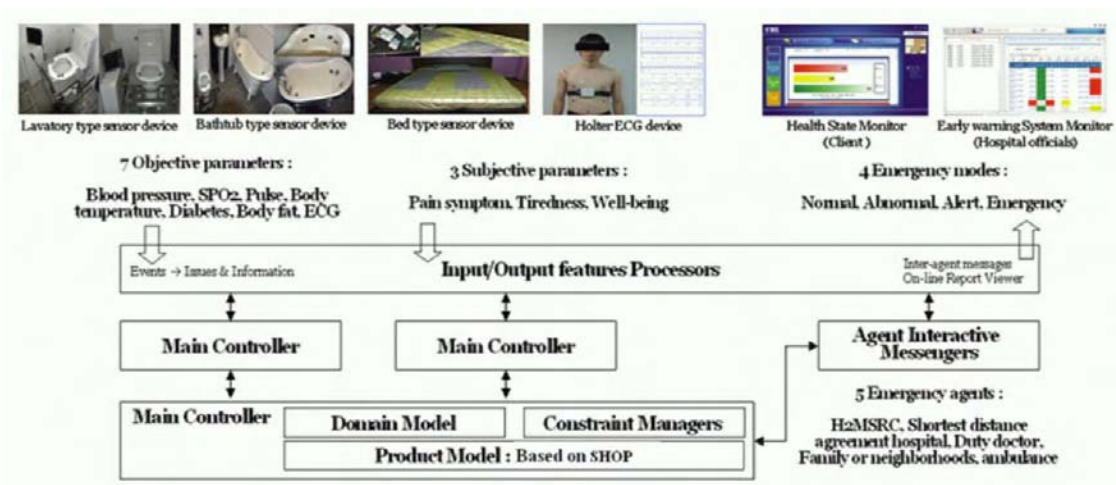


Figure 1

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Open Source Patient Data Management System for Intensive Care

Massaut J, Reper P, Hooghe L, Gottignies P

Surgical Intensive Care, Brugmann Hospital, Université Libre de Bruxelles

Abstract and objective

In Intensive Care Units, the amount of data to be processed for patients care, the turn over of the patients, the necessity for reliability and for review processes indicate the use of Patient Data Management System (PDMS). To respond to the needs of a Surgical Intensive Care Unit, we developed a PDMS based on open source software and components.

The software was designed as a client–server architecture running on the Linux operating system and powered by the PostgreSQl data base system. The client software was developed in C. The application offers the following functions: medical notes captures, observations and treatments, nursing charts with administration of medications and scoring systems functionalities. The PDMS was used to care more than two thousands patients with the expected reliability and functionalities.

Key words:

database management system, software, intensive care

Introduction

Patient Data Management Systems are mandatory in Intensive Care Unit in response to the amount of data to be processed, the turn over of the patients and the necessity for review processes. To respond to the needs of our unit we developed a PDMS based on open source software and components.

Methods

The software was designed as a client-server architecture running on the Linux operating system (SUSE Linux Enterprise Server 8.0), using the PostgreSQL relational database (v 7.2). The client software was developed in C using the GTK interface library. Remote access from remote PCs is implemented by virtual network connections (VNC), the use of VNC servers and VNC viewers.

The hardware consists in two Intel x86 servers (one master and one slave to assure the integrity of the database by replication),

14 medical grade panel PCs connected via RS232 medical bus to the patient's monitoring devices and the servers via local Ethernet network.

The software, developed on the Linux platform, offers the following functions: medical notes captures with patient's history, observations and treatments, nursing charts with functionalities for administration of medications, and scoring system possibilities for patient's classification. Interoperability between these modules is realized through access to the PostgreSQL database and not the use of local memory in the interface. The software was developed to be open source in all its components.

Results

The PDMS is in used in our unit from February 2004 and was used to care more than two thousands patients. The system is accessible at every bed through panel PCs and at desks or offices through VNC viewers on windows PCs. Its design allows an access to the database's functionalities with a high availability level (less than 5 hours of interruption over one year).

The use of open source resources was effective to customize the solution to ICU's request. The use of the C language permitted to obtain small response times but limits the portability of the system and complicated the debugging process in this critical environment.

Conclusion

PDMS based on open source software components are effective and able to respond to the needs of the ICU environment, with a high availability level.

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Why Teach Computer Security to Medical Students?

Ana M. Ferreira [ac], Ricardo Cruz-Correia [ab], Altamiro Costa-Pereira [ab]

[a] *CINTESIS - Center for Research in Health Information Systems and Technologies, Faculty of Medicine of Porto* [b] *Biostatistics and Medical Informatics Department, Faculty of Medicine, University of Porto* [c] *Department of Informatics, Faculty of Medicine, University of Porto*

Abstract

The introduction of Electronic Medical Records (EMR) within the healthcare practice can be beneficial in order to integrate and centralize heterogeneous patient information. However, there are still some problems that hinder the successful use of EMR. The concern for patient privacy is one of them. The aim of this paper is to present the results of a study that assesses attitudes of 1st year medical students towards computer security and the EMR. An anonym questionnaire was given to the students at the beginning and at the end of the academic year of 2003/2004 for them to comment on several security related scenarios. 238 questionnaires were answered at the beginning of the year whilst 222 were answered at the end of the year. The students feel, at the end of the year, that they understand better what computer security is and how to protect patient privacy information. This shows that teaching computer security to medical students, the future users of EMR, can greatly influence the success of EMR integration and therefore improve and fasten healthcare treatment.

Keywords:

computer security; education, medical, undergraduate

Introduction

The introduction of Electronic Medical Records (EMR) within the healthcare practice allows for the integration of heterogeneous information that are usually scattered over different locations [1] [2]. However, there are some barriers that impede its successful integration in most healthcare practices [3] [4].

One specific barrier relates with the privacy and security of patient information [5]. The use of new information systems within healthcare stresses the need for young doctors to comprehend them from their conception so that they can be used in a beneficial way and support their future daily work. As such, all the feedback provided during their training into the medical profession is essential for the enhancement of those systems [6], moreover in terms of computer security.

The Biostatistics and Medical Informatics Department of Porto Faculty of Medicine teaches Ethics and Medical Informatics to 1st year medical students [7]. The later subject includes theoretical and practical lectures about Electronic Medical Records (EMR) and computer security.

This study aims to assess the opinions, attitudes and awareness of 1st year medical students towards computer security issues relating to EMR, and how these can affect the successful integration of EMR within the healthcare practice.

Methods

An anonym questionnaire was given to the students both at the beginning and again at the end of the academic year of 2003/2004. It was applied two times so that we could compare their attitudes before and after they had attended the Ethics and Medical Informatics' subjects.

The questionnaires introduced 3 scenarios for the students to comment. The first scenario described a breach of patient privacy to an EMR by one of their colleagues. There were two questions relating with this scenario:

- Q1.A – Is the described scenario a security breach?
- Q1.B – What would you do if you found out about this breach?

The second scenario included additional information to the first scenario. It explained that the colleague in question had shared his password with a friend and that friend was the one to access patient private information, without him knowing it. The question related with this scenario was:

- Q2 – Would you change your attitude if you found out this new piece of information?

The third scenario introduced the issue of more sensitive information (e.g. HIV, Cancer results or even VIP related) and how this information must be protected. The question presented within the questionnaire was the following:

- Q3 – Do you think this kind of information requires stronger security measures than other types of information?

The answers to these questions were inserted into SPSS® and analysed separately.

Results

A total of 460 questionnaires were filled by the students. 52% (238) were answered before the lectures started whilst 48% (222) after the lectures finished. Table 1 shows the results obtained from the applied questionnaires.

Table 1 – Results obtained from the questionnaires

		Before the lectures % (n)	After the lectures % (n)
Answered questionnaires		*(238)*	*(222)*
Q1.A	*Valid answers*	*98 (232)*	*98 (217)*
	Yes	100 (232)	100 (217)
Q1.B	*Valid answers*	*61 (144)*	*60 (132)*
	Reason with	54 (77)	44 (58)
	Inform	40 (58)	50 (66)
	Others	6 (9)	6 (8)
Q2	*Valid answers*	*62 (148)*	*62 (138)*
	No	74 (109)	83 (115)
	Yes	26 (39)	17 (23)
Q3	*Valid answers*	*43 (103)*	*91 (204)*
	No	44 (91)	38 (77)
	Yes	55 (112)	62 (127)

For Q3, the main reason given by the students that felt no extra security measures were needed to access more sensitive information is that all security measures must be effective for all cases, independently of the patient or healthcare performed. The majority of the students that thought extra security measures were necessary agreed that this would provide for the protection of certain social groups from discrimination.

Conclusion

According to this study's results, after Medical Informatics and Ethics' lectures, students feel more conscientious to report privacy breaches to responsible parties (Q1.B). They understand better what computer security is and how to behave in order to protect confidentiality of electronic information. They consider indirect disclosure of sensitive information, such as with another person's password, a serious fault (Q2). Further, at the end of the year, students become more aware for the need of different protection levels of security depending on how sensitive information can be (Q3).

We believe that the introduction of Medical Informatics and Ethics early in the degree of the Medical course has an influence in the awareness and attitudes of first year medical students towards computer security and EMR. This can greatly influence the success of EMR integration whilst improving and fastening healthcare treatment.

References

[1] Waegemann C. EHR vs. CPR vs. EMR. Healthcare Informatics online. 2003.

[2] Cruz-Correia R, Vieira-Marques P, Costa P, Ferreira A, Oliveira-Palhares E, Araújo F, et al. Integration of Hospital data using Agent Technologies – a case study. AICommunications special issue of ECAI. 2005;18(3):191-200.

[3] Sprague L. Electronic health records: How close? How far to go? NHPF Issue Brief. 2004 Sep 29(800):1-17.

[4] Miller RH, Sim I. Physicians' use of electronic medical records: barriers and solutions. Health Aff (Millwood). 2004 Mar-Apr;23(2):116-26.

[5] Knitz M. HIPPA compliance and electronic medical records: are both possible? Graduate research report: Bowie State University. Maryland in Europe; 2005.

[6] Davis L, Domm J, Konikoff M, Miller R. Attitudes of First-year Medical Students Toward the Confidentiality of Computerized Patient Records. JAMIA. 1999; 6: 53-60.

[7] Freitas J, Cruz-Correia R, Almeida F, Costa-Pereira A. Evolution of Medical Informatics teaching in a medical undergraduate course. MedNet 2003, Eighth World Congress on the Internet in Medicine; 2003; Geneva - Switzerland; 2003. p. 312-4.

Author Correspondence

Ana M. Ferreira - First author's contact: CINTESIS – Center for Research in Health Information Systems and Technologies, Faculty of Medicine of Porto, Al. Prof. Hernani Monteiro, 4200-319 Porto, Portugal
(Phone: +351 22 551 3613; Fax: +351 22 551 3613; email: amlaf@med.up.pt)

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Application of Wireless and Mobile Computing Technologies to Improve the Efficiency of Patient Care and Education: The Role of Medical Engineering and Information Technology

Lin Guo

Centre for Adult and Paediatric Gastroenterology, Institute of Cell and Molecular Science, Queen Mary's School of Medicine and Dentistry Barts and The London, , Queen Mary, University of London, UK.

Abstract

This study explored the potential of the application of wireless and mobile computing technologies to be used in improving the efficiency of patient care and education and future developments in information and communication technologies to support healthcare professionals and medical students in healthcare research, medical education and training. The design used for this study was a systematic review of published materials obtained from EMBASE and MEDLINE online databases, and the Cochrane Library database, including personal observations. Today, more than 50% of healthcare professionals and medical students are using Personal Digital Assistant with expected growth of more than 75% by year-end 2007. In addition, wireless and mobile computing technologies allows Personal Digital Assistant to connect directly to networks or the Internet. Studies relating to processes of patient care and should evaluate mobile computing technologies as a potential timesaving tool. Wireless and mobile computing technologies is only beginning to take its first step in improving patient care and education. They have shown a positive impact on patient safety, health care efficiency, and ultimately patient satisfaction.

Keywords:

wireless and mobile computing technologies, patient care and education, efficiency, personal digital assistant

Introduction

Wireless is a term used to describe telecommunications in which electromagnetic waves (rather than some form of wire) carry the signal over part or all of the communication path.[1] Mobile Computing is a generic term describing your ability to use technology 'untethered', that is not physically connected, or in remote or mobile (non static) environments.[2] The term of wireless and mobile computing technologies is evolved in modern usage such that it requires that the mobile computing activity be connected wirelessly to/through the application of internet or to/through a private network. This connection ties the mobile device to centrally located information and/or application software through the use of battery powered, portable, and wireless computing and communication devices.[1, 2] This includes devices like laptops with wireless LAN or wireless WAN technology, smart mobile phones, wearable computers and Personal Digital Assistants (PDAs) with Bluetooth or IRDA interfaces.

This study explored the potential of the application of wireless and mobile computing technologies to be used in improving the efficiency of patient care and education and future developments in information and communication technologies to support healthcare professionals and medical students in healthcare research, medical education and training.

Design and methods

The design used for this study was a systematic review of published materials obtained from EMBASE and MEDLINE online databases, and the *Cochrane Library* database, including personal observations. Materials that match the set criteria on the application of wireless and mobile computing technologies in medical engineering and information technology research for healthcare professionals and medical students to improve the efficiency of patient care and education were selected and analysed following the United Kingdom National Health Service Centre for Reviews and Dissemination Guidelines. A variety of data collection approaches were developed to ensure data were collected on the various aspects.

Results

Wireless and mobile computing technologies is seen to be convenient to get in touch, compact, fast and portable, but problems are attached to the levels of security, confidentiality and scalability of the hardware. It is also apparent that most commonly used wireless and mobile computing technologies within the health contexts is the Personal Digital Assistant. Today, more than 50% of healthcare professionals and medical students are using Personal Digital Assistant with expected growth of more than 75% by year-end 2007.[3] Not only can the healthcare professionals and medical students use the tool on the Personal Digital Assistant as a quick look-up resource at the bedside, but can also flag a topic for further research when back in the office. Changes in treatment guidelines, concerns about patient safety, efforts to contain costs, time limitations, and better informed patients make it critical to have clinical reference information at the point of patient care and edu-

cation.[4] In addition, wireless and mobile computing technologies allows Personal Digital Assistant to connect directly to networks or the Internet.

Discussion

According to the results of this systematic review of this study, wireless and mobile computing technologies has become a valuable resource for both healthcare professionals and medical students over the past decade. This has enabled new perspectives to be developed on the interface between hardware, software and education and training processes for those involved in delivering healthcare research. Studies relating to processes of patient care and should evaluate wireless and mobile computing technologies as a potential timesaving tool, as they can be synchronised with hospital information systems to facilitate retrieval of patient information. The findings also indicate that the healthcare professionals and medical students would benefit from some technologies that they can easily use to search for such sources as database, E-journals and the Internet. At a theoretical level, wireless and mobile computing technologies such as Personal Digital Assistant is ideal for meeting these needs. Further study on processes of patient care and education should also explore wireless and mobile computing technologies as vehicles for disseminating evidence-based guideline recommendations.

Conclusion

Wireless and mobile computing technologies is only beginning to take its first step in improving patient care and education. They have shown a positive impact on patient safety, health care efficiency, and ultimately patient satisfaction. The integration of the Internet and wireless and mobile solutions will transform the use of information and communication technologies in patient care and education and take it in the role of medical engineering and information technology. The potential of wireless and mobile computing technologies is vast and its principles, application and practices are seen as a choice for healthcare professionals and medical students to be in the right place at the right time. In the future we will see more valuable resources in improving patient care and by developing wireless and mobile computing technologies.

References

[1] https://SearchMobileComputing.com, accessed on 4th November 2006

[2] http://en.wikipedia.org/wiki/Mobile_computing, accessed on 5th November 2006

[3] Osonnaya C and Osonnaya K. Information management in healthcare: analysis of trends and issues. International Journal of Medicine 2004; 1: 5-8

[4] Guo L and Osonnaya C. Improving the efficiency of patient care and education: the role of medical engineering and technology in gastroenterology. International Journal of Medicine 2006; 3: 149-151. (In press)

Address for correspondence

Mr Lin Guo
Centre for Adult and Paediatric Gastroenterology
Institute of Cell and Molecular Science
Queen Mary's School of Medicine and Dentistry
Barts and The London
Queen Mary, Uiversity of London
Whitechapel, London
E1 2AL
UK
E-mail: lin.guo@qmil.ac.uk or guolinlondon@yahoo.co.uk

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Comparing Messages in an Online Communication Forum for Cancer Patients with Patients' Messages to a Clinical Nurse Specialist

Annette Jeneson MA, Trine Andersen RN, Cornelia Ruland RN, Ph.D

Center for Shared Decision Making and Nursing Research, Rikshospitalet-Radiumhospitalet Medical Center, Oslo, Norway

Abstract

WebChoice is an online support system where cancer patients can exchange messages with other patients in an open communication forum as well as send personal e-mails to a clinical nurse specialist (CNS) who responds to their concerns. We compared the content of messages posted in the open forum with personal e-mail messages sent to the nurse. While patients were concerned with similar topics in both communication areas, there were differences in the types of messages sent. More patients actively used the patient-nurse email service compared to the forum, suggesting that nurses can play an important role in online systems supporting cancer patients over the Internet.

Keywords:

online systems, communication, nurse-patient relations, neoplasms

Introduction

Despite a growing interest in online support forums and patient-provider communication systems, knowledge of the relative use and benefits of these communication systems is still limited. In particular, there is little information about how patients' communications with care providers online differ from communications in discussion groups. Additionally, while a few studies have analyzed online messages between patients and physicians, online patient-nurse communication and the potential for nurses to support patients via the Internet is as of yet largely unexplored. The purpose of this study, therefore, was to compare the content of patient-nurse messages to messages posted on an online discussion forum for cancer patients. The study is part of a larger ongoing trial testing the effects of WebChoice, an internet support system for breast- and prostate cancer patients [1]. WebChoice includes tools for symptom monitoring, tailored information to support self-management, support forums for anonymous group discussion, as well as a more "private" communication area to exchange messages with a CNS specializing in cancer care.

Methods

We examined forum postings and e-mail messages by patients who logged on to WebChoice at least once from March - October 2006, resulting in a sample of 355 postings and 174 e-mails. Messages were coded according to the 'type' and 'topic' categories listed in Table 1 below. The 'type' categories are consistent with the coding schema of Klemm *et al* [2]. A message could belong to more than one category. An e-mail with questions about lymphedema and anxiety, for example, would be coded as *information seeking* for 'type', and *symptoms* and *feelings* for 'topic'. Messages were independently coded by two of the authors (AJ and TA) and 10% were coded by both to compare interrater agreement, established at 98 % for e-mail messages and 97 % for forum postings.

Results

Table 1 shows the percentages of e-mails and forum postings coded under the various 'type' and 'topic' categories.

Table 1. Percentage of messages categorized by type and topic

		E-mails (*174 total*)	**Postings** (*355 total*)
TYPE	Personal experiences/ opinions	66	75
	Information giving	2	20
	Information seeking	62	20
	Encouragement / Support	0	17
	Thanks	22	8
TOPIC	Health personnel / institutions	21	20
	Treatment / Tests / Test results	63	59
	Symptoms / Side-effects	64	66
	Energy / Fatigue / Sleep	9	10
	Feelings	14	10
	Sexuality / Partner	3	7
	Family / Colleagues / Others	5	3
	Living with cancer / Lifestyle	22	23
	Metastasis / Relapse	7	8

The average number of topic categories per e-mail or posting was 2.4 (±1.5) and 2.3 (±1.3), respectively. As seen in Table 1, patients were concerned with largely the same topics in their e-mail messages and in their forum postings, but the type of message or posting varied, with e-mails being more 'information seeking' in nature and postings being more 'supportive' and 'information giving' to others.

There were also differences in use of the e-mail component of the WebChoice system compared to use of the discussion forum in terms of the number of participants actively using these two different components. 71% of participants given access to WebChoice logged on at least once in the data collection period (43 females and 31 males). Of these active WebChoice users, more patients used the patient-nurse communication area (n=45 or 61% of WebChoice users) compared to those submitting messages to the forum (n=34 or 46% of users). This holds true for breast cancer patients as well as prostate cancer patients. 72% of female users who logged on to the system at least once between March to October 2006 sent at least one e-mail to the nurse while 58% submitted at least one posting. 45% of active male users sent an e-mail to the nurse while 29% contributed to the forum. However, while more patients wrote e-mails to the nurse compared to the number of patients who submitted a posting in the forum, those patients actively participating in the discussion forum submitted on average more postings (10.4±10.5) compared to the average number of e-mails sent to the nurse (3.9 ± 4.6).

Discussion and conclusion

This sample of messages and forum postings, although small, provide useful information for health professionals interested in online communication systems. In a number of computer-based support systems for patients the communication areas are consistently highlighted as the most popular sections of the system, yet our knowledge about patients' use and benefits of different features within these communication areas is limited. A preliminary usage analysis of WebChoice suggest the forum is so far the most visited section in this support system [3], but the fact that more patients submit messages to the nurse via e-mail compared to the forum, and that the nature of messages in these two communication areas differ, suggest that opportunities for patient-nurse communication can provide valuable support for patients beyond that of support offered by participation in online discussion groups. Patients often experience multiple symptoms during treatment and rehabilitation, yet short hospital admissions allow little time for detecting and relieving these. Moreover, side-effects of certain treatments are often worst after the patients are discharged to home.

Therefore, patients could greatly benefit from support through an Internet based system, where they can communicate with a care provider independent of scheduled hospital or doctor appointments in an environment that is readily accessible and even anonymous. In this sample almost all of the questions patients asked via e-mail could be appropriately addressed by the CNS and did not require advice from a physician or other specialist. This suggests that online communication with a nurse may potentially reduce not only needless patient suffering and worrying but also the numbers of doctor's appointments scheduled. As cost concerns and shortages of health professionals continue to rise, online peer- and professional support provided by nurses could prove a viable health care supplement that can improve delivery of high quality patient care in the future [4].

Acknowledgments

We thank Laura Slaughter and Gro Hjelmeland Grimsbø for helpful discussion. The research was funded by the Norwegian "Health and Rehabilitation" Grant # 2003/3/0429.

References

[1] Ruland CM, Andersen R (2004) Designing WebChoice - Individualized Support for Cancer Patients through the Internet. Medinfo, San Francisco, Sept 5-11, 2004, p.1840.

[2] Klemm P, Hurst M, Dearholt SL, Trone SR (1999) Cyber Solace: Gender Differences on Internet Cancer Support Groups. Comput Nurs 17(2), pp. 65-72

[3] Ruland CM, Jeneson A, Andersen T, Slaughter L, Andersen R, Schjødt-Osmo B, Moore SM. Designing Tailored Internet Support to Assist Cancer Patients in Illness Management. Submitted for publication, AMIA Nov 10-14 2007, Chicago IL.

[4] Sorrells-Jones J, Tschirch P, Liong MAS (2006) Nursing and Telehealth: Opportunities for Nurse Leaders to Shape the Future. Nurse Leader 4(5), pp.42-58.

Address for correspondence

Professor Cornelia Ruland
Center for Shared Decision Making and Nursing Research
Rikshospitalet-Radiumhospitalet HF
N-0027 Oslo
Norway
E-mail: cornelia.ruland@rr-research.no
Telephone: (+47) 23 07 54 60

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Guideline-based Visualization of Medication in Chronic Disease

Ján Stanek, Michelle Joy Davy

Advanced Computing Research Centre, University of South Australia, Australia

Abstract

Chronic disease management at the General Practice level is a challenging task requiring synthesis of information across time and possibly several practitioners. Sparsity of data, lack of structure, lack of time are limiting factors of computer use in this domain. Authors are exploring the concept of visualization of individual patient's medication using a clinical guideline to provide some structure to the problem by creating a state-transition model. This approach was shown to be promising at the practice level in previous research [1,5,6] where an overall synthesis of practice decisions was created and alerts were generated on the basis of individual transactions or states - current focus is on individual patient and the sequence of states describing his/her medication history.

Keywords:

general practice, visualization, decision support, guideline

Introduction

In managing chronic diseases, the information on the past course of the disease can be important - but many practitioners see reviewing of past data as a (necessary) evil to be done as quickly as possible [2]. With more and more general practitioners (GPs) using a computer in their everyday practice [3] it is more than tempting to use the existing data to provide support for quality and continuity of care. Typically the data is meant to be used by humans so the data is rather unstructured, with highly variable quality in terms of completeness or adherence to some coding standards. In a comprehensive review it was shown, that prescription data is the one most complete and reliable [4]. Assuming, that medication reflects significant proportion of decision-making about a case a guideline-based state-transition model was created and used for analysis at the practice level, as well as for generation of patient-related alerts [1,5,6]. While this approach shows promising results, information on patients as drawn from the model does not take into account patient-specific sequences of states. In the current work we explore the visualization of these sequences as well as generating alerts based on sequential patterns for particular patients in line with the ideas of Aigner and Miksch [13].

Materials and methods

The backbone of the study was a state-transition model focusing on treatment of hypertension in diabetes mellitus based on an American Diabetes Association guideline [7]. The data used in the study were de-identified extracts from a general practice in rural Australia holding age, gender, visit dates, problem codes, blood pressure measurements and drug prescriptions. The extracts covered 6 years and contain more than 70000 prescriptions. Two sequences of events/states were generated for each patient: prescription path and treatment path. Prescription path was shows just a sequence of drugs or they combinations as they were prescribed; while the treatment path shows a sequence of states as derived from the state-transition model

Treatment path

According to methodology described in [5] medications were clustered into 6 groups (group B has 3 subgroups):

- Group A: Angiotensine converting enzyme inhibitors (ACEi) - ATC codes C09A and C09C
- Group B1: -blockers (BB) - ATC codes C07AA, C07AB
- Group B2: Diuretics - ATC codes C03AA, C03CA or C03D
- Group B3: Non-dihydropyridine Ca-channel blockers (NCCB) - ATC code C08D
- Group C: Dihydropyridine Ca-channel blockers (DCCB) - ATC code C08CA
- Group D: -blockers, hydralazines, clonidine - ATC codes C02CA, C02DB, C02AC

In our study we deviated slightly from standard ATC coding - combination drugs were coded as if a set of separate drugs was given. This deviation was useful to simplify the model-building algorithms and did not have impact on the analysis.

Based on prescription data states were generated, using the amount and the daily dose of prescribed drug to calculate duration of treatment with a particular drug group. In this calculation we assumed, that patient starts the medication on the same day as it was prescribed, that the dosage is unchanged throughout the time covered by a particular prescription (and the patient adheres to the dosage).

Prescription path

A prescription is a sequence of prescription instances. A prescription instance is a drug, or combination of drugs as prescribed on a particular day, disregarding dosage or quantities prescribed.

All methods were implemented using Cache and Ensemble (Intersystems Inc.), graphs were generated using GraphViz software (www.graphviz.org).

Results

A prescription path and a treatment path were created for each patient. Both paths are shown simultaneously and provide different views on what was done. E.g. if a drug containing a BB and diuretic was prescribed and at the same time a potassium-sparing diuretic was prescribed the prescription instance is shown as B1+B2+B2; while the treatment path will show B1+B2 combination. A particular prescription path can have several treatment path counterparts (Figure 1).

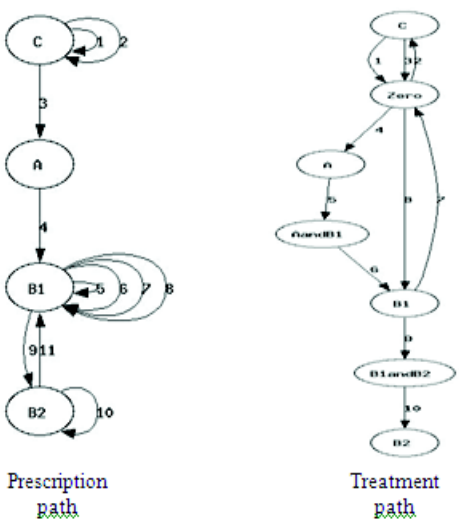


Figure 1 Prescription path vs. Treatment path

Alerts

Visualisation of the paths should be in most cases enough for an experienced GP to see any unusual patterns, however the results can be analysed in background and selecting only patients with an unusual pattern. Analysing treatment paths allows to create an additional class of alerts, taking into account more than just one state or one transition. More than just the preceding state (eventually the whole path) can be taken into account in launching an alert - e.g. transition from A to B1 as well as transition from B1 to A might be OK per se, but if this is repeated more than **n** (usually 2-4) times, it may be considered an unusual pattern and an alert should be launched.

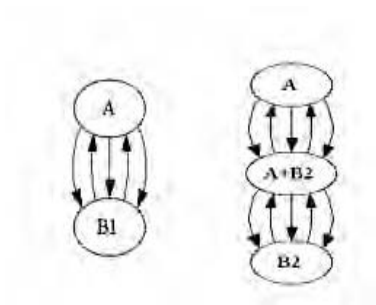


Figure 2 Cyclic patterns in treatment paths

Discussion

Guideline can be utilized at two levels in data analysis. First approach is to use the guideline to structure the problem space and then fit the data into the resulting of structure. Svátek and íha [8,9] used a guideline to lead a data-mining exercise in their analysis of guideline compliance in hospital environment. This idea is similar to what we use, however it was applied in a much better controlled environment of a large hospital. Others [10, 11, 12] focussed more on intentions and worked in a hospital, data rich setting. Major advantage of this approach is, that results are presented to the physician in the context of the guideline making it easier to recognise and appreciate the rationale behind the graphs and alerts.

Alternate approach is to use data mining techniques to raw data and then compare the results to a guideline. Possible method of creating similar results to ours is path or workflow mining (e.g. [14]). This approach can be exploited either to improve the guideline or to detect exceptions in clinical workflow.

Conclusion

Authors extended the scope of previous research [1] by adding visualizations for individual patients as well as new types of alerts taking into account more than just one transition or just one therapeutic state. This approach is to be validated by a clinical study in next future.

Acknowledgments

The study protocol was approved by the Human Research Ethics Committee of the University of South Australia (protocol P005-04) and undertaken with a Memorandum of Understanding with Lubims Pty. Ltd. Special thanks to Intersystems Inc. for providing licenses for Cache and Ensemble.

References

[1] Gadzhanova, S., Iankov, I.I., Warren, J.R., Stanek, J., Misan, G.M., Baig Z., Ponte, L.: Developing High-Specificity Antihypertensive Alerts by Therapeutic State Analysis of Electronic Prescribing Records. JAMIA, 2006 Oct 26 [epub. ahead of print]

[2] Sullivan, F., Wyatt, J.C. How computers help make efficient use of consultations. BMJ 2005;331:1010-1012.
[3] Western M, Dwan K, Makkai T, Del Mar C, Computerisation in Australian general practice. Aust Fam Phys 2003; 32(3): 1-6.
[4] Thiru K, Hassey A, Sullivan F, Systematic review of scope and quality of electronic patient record data in primary care. BMJ 2003; 326: 1070-80.
[5] Warren JR, Gadzhanova S, Stanek J, Iankov II, Understanding caseload and practice through analysis of therapeutic state transitions. Proc AMIA Annu Fall Symp 2005.
[6] Warren JR, Stanek J, Gadzhanova S, Iankov I, Misan G, Inferring 'therapeutic states' of patients from community electronic prescribing data. electronic Journal of Health Informatics 2006; 1(1): e5.
[7] American Diabetes Association, Hypertension management in adults with diabetes. Diabetes Care 2004; 27:S65-S67.
[8] Ríha A., Svátek V., Nmec P., Zvárová J.: Medical guideline as prior knowledge in electronic healthcare record mining. In: 3rd International Conference on Data Mining Methods and Databases for Engineering, Finance and Other Fields, 25-27 September 2002, Bologna, Italy.
[9] Svátek, V., Ríha, A., Peleška, A., Rauch, J.: Analysis of guideline compliance – a data mining approach. Praha 13.04.2004 – 16.04.2004. In: Kaiser, K, Miksch, S., Tu, S. W. (ed.). *Computer-Based Support for Clinical Guidelines and Protocols.* Amsterdam, IOS Press, 2004, s. 157–161. ISBN 1-58603-412-X.
[10] Advani, A.,Shahar, Y., Musen, M.A.: Medical Quality Assessment by Scoring Adherence to Guideline Intentions. 2001 Annual AMIA Conference, Washington, DC, Hanley and Belfus 2001.
[11] Chan, A.S., Coleman, R.W., Martins, S.B., Advani,A , Musen,M.A., Bosworth, H.B., Oddone, E.Z., Shlipak, M.G., Hoffman,B.B., Goldstein, M.K.: Evaluating Provider Adherence in a Trial of a Guideline-Based Decision Support System for Hypertension.MedInfo 2004, San Francisco, California, Sept 7 - 11. 2004.
[12] Marcos, M., Berger, G., van Harmelen, F., ten Teije, A., Roomans, H., Miksch, S: Using Critiquing for Improving Medical Protocols: Harder than It Seems. In: Quaglini, S., Barahona, P., and Andreassen, S. (Eds.): AIME 2001, LNAI 2101, Springer-Verlag Berlin Heidelberg 2001, pp.431–441
[13] Aigner, W, Miksch, S.: Supporting Protocol-Based Care in Medicine via Multiple Coordinated Views. Proceedings of the Second International Conference on Coordinated & Multiple Views in Exploratory Visualization (CMV'04), 2004.
[14] van der Aalst, W., Weijters, T., Maruster, L.: Workflow mining: discovering process models from event logs. IEEE Transactions on Knowledge and Data Engineering, 16 (9), September 2004.

Address for correspondence

Dr. Jan Stanek
University of South Australia
Advanced Computing Research Center
Mawson Lakes Blvd.
Mawson Lakes, SA 5095
Australia.
E-mail: jan.stanek@unisa.edu.au

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Efficiency and Safety of New Radiofrequency Identification System in Japanese Hospital

Yuichiro Saito[a], Takashi Hasegawa[b], Tetsuo Sakamaki[c]

[a]*Division of Cardiovascular Medicine, Gunma University Hospital, Maebashi, Japan*
[b]*Informatics Education Center , International University of Health and Welfare, Tokyo, Japan*
[c] *Medical Informatics and Decision Sciences,Gunma University Hospital, Maebashi, Japan*

Abstract and objective

Radiofrequency identification (RFID) uses radio-frequency tags attached to people or objects to provide identification, tracking, and security under the general heading of automatic identification. New RFID system (UHF band, 953 MHz) has been available since April, 2005 in Japan. We tested efficiency and safety of new RFID system in our hospital. Electric fields produced by new RFID had no significant effects on cardiac pacemakers, implantable cardioverter defibrillators (ICD), and other medical devices□@(ex. Electrocardiogram recorder, cardiac monitor, intra-aortic balloon pumping, infusion pump, and respirator) in our hospital. New radiofrequency tags seemed to provide extensive patient identification and to track capital equipments within our hospitals. Healthcare systems today are increasingly complex and interrelated processes, while new RFID technologies will provide opportunities for enhanced patient care and safety in Japanese hospital.

Keywords:

cardiac pacemaker, implantable cardioverter defibrillators, healthcare, patient safety

Introduction

Radiofrequency identification (RFID) has recently begun to receive increased interest in supply chain, in order to increase the efficiency and visibility of material and information flows. RFID may address to improve safety and increase in productivity. New RFID system (UHF band, 953 MHz) has been available since April, 2005 in Japan. However, there has been no attempt to determine whether electric fields produced by new RFID can influence cardiac pacemakers, implantable cardioverter defibrillators (ICD), and other medical devices in our hospital. We further examined to track new IC tags attached to people and medical equipments in hospitals.

Our data may provide the potential benefits, the area of applications, and the corresponding strategies of RFID in hospital environments.

Methods

We examined effects of RFID system on 7cardiac pacemakers and 5 ICDs in 0.18 % salt solution, similar electric condition to human body. The experimental conditions for interference induced by RFID were between a homogeneous electric field perpendicular to the area formed by the antenna and radio-frequency tags.

We next tested whether Electrocardiogram (ECG) recorder, ECG monitor, intra-aortic balloon pumping (IABP), infusion pump, and respirator in our hospital could work normally under new RFID system.

Furthermore, we evaluated to track new IC tags attached to people and medical equipments in our hospital.

Results

Electric fields produced by RFID had no significant effects on cardiac pacemakers. There was no differentiation between a unipolar and a bipolar system. The RFID systems did not interfere with ICDs.

ECG recorder, ECG monitor, IABP, infusion pump, and respirator were worked normally under operation of RFID systems.

Radiofrequency tags attached to people and capital equipments within our hospitals provide more extensive identification than traditional bar coding can.

Conclusion

RFID may be ultimately used for many of the functions currently carried out using bar coding if the cost of RFID comes down. Healthcare systems today are increasingly complex. RFID is a technology that will have a profound impact on effective and safe patient care in Japanese hospitals in near future.

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Development of Hypertension Management Ontology for Guideline-based Clinical Decision Support System

JiHyun Kim[a], InSook Cho[b], EunJung Lee[a], JaeHo Lee[c], Yoon Kim[d]

[a] *Center for Interoperable EHR, Korea*
[b] *Department of Nursing, Inha University College of Medicine, Korea*
[c] *Department of Emergency Medicine, University of Ulsan College of Medicine, Korea*
[d] *Department of Health Policy and Management, Seoul National University College of Medicine, Korea*

Abstract

For knowledge representation of guideline based hypertension management, hypertension management ontology was developed. Ontology was adopted as a method for enabling to share and to reuse knowledge of a domain with consistency in this study. This work is an essential course in representing knowledge in a computerized method and enables to reuse knowledge in interested domain areas. On the basis of published guidelines including JNC 7, concepts related to hypertension management knowledge and relations between concepts were made and defined. Entities of concept were defined. This ontology includes high level 11 classes and about 300 concepts and it defines instances of each concept. This work will be use useful to build a tool in order to acquire specific domain knowledge and contribute to spreading knowledge sharing and standardized clinical guidelines.

Keywords:

ontology, hypertension management, computerizing clinical guideline, knowledge representation

Introduction

In Korea, the Government leads Electronic Medical Record (EMR) development for the 152 primary health centers, which is planned to be expanded to Electronic Health Record (EHR) system. As a part of this work, development of clinical decision support system (CDSS) is in the process of advancing. This study describes a part of constructing the guideline based decision support system for hypertension management in primary care setting and encoding hypertension management knowledge. This is essential for representing knowledge in a computerized method. It is for reusing knowledge in interested domain areas.

For knowledge sharing nationally, knowledge management method, such as ontology is main issue on the aspect of clinical decision support system and others. Ontology is a methodology to formalize a shared understanding of a domain and enables to share and reuse of knowledge consistently between software applications and humans.

Materials and methods

This ontology is primarily based on 12 published hypertension guidelines including JNC 7. Concepts related to hypertension management were defined at the modeling stage and defined concepts were classified according to semantic category. Relations and attributes of concepts were defined. We authored knowledge in Protégé 3.1.1 environment.

Results

About 300 concepts related to hypertension management were defined. A medication prescription, laboratory test, physical examination, and other related conditions were included in related concepts. Based on the concepts defined, high level classes which contained medication, clinical finding, test, patient education, problem, patient case, event, rule, therapy adjustment, temporal predicate, and eligibility criteria were made and defined. Each class included subordinate concepts of is-a relation. Then, attributes of concepts were defined. Completed ontology had 11 classes and about 300 concepts.

Conclusion

Through this approach, Creation of a tool to acquire particular domain knowledge is expected to be easier. Furthermore, the ontology of hypertension management for guideline based clinical decision support system is considered to contribute for sharing and disseminating knowledge and to spreading standardized clinical guidelines through reusing knowledge.

Address for correspondence

Center for Interoperable EHR, Annex Building, Seoul National University College of Medicine, 199-1 Dongsoong-Dong, Jongno-Gu, Seoul, 110-810, Korea, Kijii90@snu.ac.kr

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Secure Remote Access for Web Based Clinical Information System Using Policy Control of PCs and Healthcare PKI Authentication

Katsuya Tanaka[a], Mayumi Yoshida[b], Ryuichi Yamamoto[b]

[a] *Department of Planning, Information and Management, University of Tokyo Hospital*
[b] *Interfaculty Initiative in Information Studies, Graduate School of University of Tokyo*

Abstract and objective

This paper describes a robust method of secure remote access for Web based clinical information system using Health-care PKI authentication. It enables medical staffs to refer the medical data using his own PCs from their home or their mo-bile phones over the Internet. This system contributes to re-duce strain of medical staffs especially in such institutes where intensive care is carried 24h/365d. For this purpose, it must permit the remote access to the hospital information system from PCs in various conditions, it is necessary to establish a secure connection using VPN or SSL, and control the policy of client PCs for the prevention of computer virus effect etc., and at the last, confirm user authentication with strict identification using Healthcare PKI (HPKI).

Keywords:

healthcare PKI, clinical information system, secure remote access control, VPN

Introduction

In Japan, healthcare ICT is promoted by Japanese government and in this year, the government made "New IT Reform Strategy" and "The Action Plan 2006". It emphasizes the reengineering in the healthcare field using ICT. It includes some actual plan such as constructing secure and reliable network for regional and inter-regional cooperation, introduc-ing smart cards for healthcare, and development of Health-care PKI (HPKI) which is based on ISO 17090. Actually we started to use HPKI in 2004, as a demonstration experimental system in the University of Tokyo Hospital for secure remote access to the medical data. HPKI was used in this system for authentication purpose, with verification the attributes of the national licenses of healthcare professionals, such as medical doctors, registered nurse, etc. By using this system, a physician can access the medical data of his patients with his own PCs even in his home or his mobile phones.

Methods

In our hospital, medical staffs are able to access to medical records using web based system in addition to special client PCs. Mainly based on the web interface, the following parts were also developed.

1. Certification Authority (CA)

For the convenience of user registration operation, All CA function was implemented in a note PC. It is also used for registering HPKI certificate to USB token.

2. VPN gateway

The connection between the hospital network and a client PC is established using IPSec. We picked a VPN gateway, Cisco Systems/VPN 3005.

3. Policy Control System

At the establishment of VPN connection, the server side policy is downloaded to the client PC, and during the connection, client PC is under the predefined policy. We picked Zone lab/Integrity.

4. USB Token for HPKI

At the insertion of this USB token, HPKI certificate is copied to the certain repository in Windows, and on the removable of token, the repository is cleared.

5. Reverse Proxy Server

For the verification of the HPKI certificate, especially hcRole attribute, user access is only allowed via this SSL reverse proxy server. At the connection to an internal web server in hospital, the user certificate in client PC is pushed and verified. For the mobile access over the Internet not using HPKI, we developed CHTML converter gateway. The identification number of each mobile phone is stored at the registration in server side, and the access of unknown device is prohibited.

Results

Over 30 doctors participated in this experiment and the developed system is now in use. This system is effective to check the results of emergency laboratory tests, or radiological images on PACS system from remote place. This system requires various client software, such as IPSec client software, USB token driver and policy control client, so it is difficult to manage for the user to install or operation. We are now discussing for the improvement of this system, such as using smart card for the HPKI token, SSL-VPN technology for an easy management of client PCs.

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Infobuttons: A Study of Usability

Lily Gutnik[a], Sarah Collins RN, BSN[b], Leanne M. Currie RN, DNSc[a,b], James J. Cimino MD[a], and Vimla L. Patel, PhD.DSc[a]

[a]*Department of Biomedical Informatics, Columbia University, USA*
[b]*School of Nursing, Columbia University, USA*

Abstract and objective

Studies of clinician's information needs while treating patients have shown that the resolution of these needs is often deferred or fails, which may lead to medical errors. The Infobutton Manager was developed to help improve the resolution of information needs by providing users with links to on-line health information resources. The aim of this study was to determine the usability of the Infobutton interface to resolve clinicians' information needs. We provided clinicians with typical case scenarios using a computerized order entry system (CPOE) and Infobuttons and asked them to verbalize their thought processes as they were using the CPOE. We video-recorded the computer screens as the users worked, conducted brief exit interviews, and analyzed these data. Results indicated that the participants found the resources provided by Infobutton helpful and easy to use.

Keywords:

usability testing, clinical information needs, information retrieval decision support

Introduction

Though technology for clinical care is created to facilitate the clinician's workload and enhance patient care, tools that are inadequately designed can actually have adverse effects [1]. Inadequate design can result from a disconnect between the design process and the needs of the end-users [2]. Infobutton is a tool designed to provide clinicians with access to on-line health information resources to quickly resolve their information needs [3]. This study explored the usability of a new user Infobutton interface which was designed to be easier to read, more concise, and have more consistent navigation.

Methods

We conducted the present laboratory study to seek in-depth feedback from clinicians about the design and usability of Infobuttons within a CPOE system. Each participant was given three typical scenarios to be interpreted using the CPOE system and Infobutton. The computer screen was video recorded using Morae™ software. Participants were asked to "think aloud," that is to verbalize their thoughts. At the end of the session, the researcher conducted a brief exit interview. A previously established coding framework was applied to the data to characterize information needs.

Results

Two nurses, one physician, and one physician's assistant took part in the study, yielding a total of 79 information needs. Twenty-three of these needs (29%) were related to drug information, 28 (35%) concerned institutional procedures or policies, and the remainder were related to patient care and the treatment plan. Forty-eight information needs (60%) were successfully resolved using Infobuttons, 15 were deferred, and 14 needs failed to be resolved. Fifty-one needs (65%) were from an external source (e.g., Micromedex) and 26 (33%) were from an internal source (e.g., the local intranet). Exit interviews revealed an overall satisfaction with the resources provided by and the usability of Infobuttons. Excess information making navigation difficult was one of the problems identified.

Conclusion

Although clinicians show a high occurrence of information needs as they treat their patients, many of these needs can be met with Infobuttons All participants agreed that the information resources provided by Infobutton were valuable and easy to understand. This study imparted us with insight on the questions clinicians need to have answered for effective decision-making, and accentuated the need to be mindful of information overload.

References

[1] Allen M, Currie LM, Bakken S, Patel VL, Cimino JJ. Heuristic evaluation of paper-based Web pages: a simplified inspection usability methodology. J Biomed Inform. 2006; 39(4):412-23.

[2] Kushniruk, AW, Patel, VL. Cognitive and usability engineering methods for the evaluation of clinical information systems. J Biomed Inform. 2004; 37:56-76.

[3] Cimino JJ. Li J. Bakken S. Patel VL. Theoretical, empirical and practical approaches to resolving the unmet information needs of clinical information system users. Proc Ann AMIA Symp:170-4, 2002

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

The Application of A Clinical Data Warehouse to the Assessment of Drug-Warfarin Interactions

Qiyan Zhang[a*] PhD., Yasushi Matusmura[a] MD. PhD., Hiroshi Takeda[a] MD. PhD.

[a] *Department of Medical Informatics, Graduated School of Medicine, Osaka University*

Abstract and objective

Studies on drug-warfarin interactions in human subjects were typically based on single case reports or extrapolated in healthy volunteers. In this study, we proposed a method to apply an institutional clinical data warehouse (CDW) to address this issue in a real-world setting. A case-control study was conducted by using the CDW in Osaka University Hospital from 2000/01/01/-2005/12/31. We randomly selected the steady-state outpatients who were under warfarin mono-therapy as Group A and those who were on the pre-existing warfarin therapy but with only the added medication X as Group B. The difference between the PT-INR values on one-month interval in Group A and those before and after taking medication X for one month in Group B was compared. The warfarin-Allopurinol interaction was illustrated as one example. We identified 25 cases in Group A and 15 cases in Group B respectively. The difference of the PT-INR values on one-month interval between the two groups was not significant (P=0.394,Mann-Whitney test), indicating that no warfarin-Allopurinol interaction was present. This method can be used as an alternative approach to assess drug-warfarin interactions.

Keywords:

clinical data warehouse, warfarin, drug interaction, PT-INR

Introduction

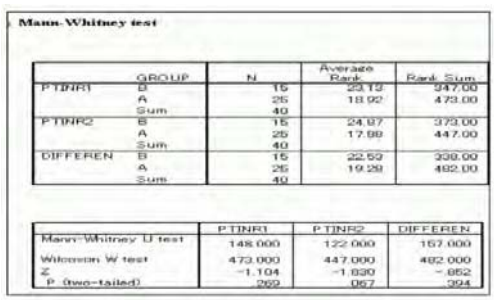
Warfarin is an effective and commonly used oral anticoagulant agent for the treatment and prevention of thromboembolism in a variety of conditions. The risk of major complication- hemorrhage may be increased when concomitant drug therapy is required. Previous studies on drug-warfarin interactions in human subjects were based on single case reports or extrapolated in healthy volunteers. Clinicians need to balance the therapeutic benefits with the bleeding risk through monitoring patient's PT-INR values during warfarin therapy. In this study, we proposed a method to apply an institutional clinical data warehouse (CDW) in a real-world setting to reduce certain practical and ethical problems faced by studies drawn from healthy volunteers and aimed to provide an alternative approach to assess drug-warfarin interactions.

Methods

The CDW in Osaka University Hospital was built in 1995 as a subject-oriented database. The outpatient's prescription data and PT-INR results were picked up from 2000/1/1/-2005/12/31 through Business Objects 6.5.1. Data were processed via Microsoft Access 2003 and statistic analyses were performed via SPSS 11.0 Japanese Version for Windows. A case-control study was conducted. The steady-state outpatients who took warfarin consecutively for at least one month were considered as the eligible study subjects. We randomly selected those who were under warfarin mono-therapy as Group A and those who were on the pre-existing warfarin therapy but with only the added medication X as Group B. Then we compared the difference between the PT-INR values with one-month interval in Group A and those before and after taking medication X for one month in Group B. The warfarin-Allopurinol interaction was demonstrated as one example.

Results

For the chosen example, we identified 25 cases in Group A and 15 cases in Group B respectively. The difference of the PT-INR values on one-month interval between the two groups was not statistically significant (P=0.394,Mann-Whitney test), indicating that no warfarin-Allopurinol interaction was present. **Figure1**



Mann-Whitney test

	GROUP	N	Average Rank	Rank Sum
PTINR1	B	15	23.13	347.00
	A	25	18.92	473.00
	Sum	40		
PTINR2	B	15	24.97	373.00
	A	25	17.88	447.00
	Sum	40		
DIFFEREN	B	15	22.53	338.00
	A	25	19.29	482.00
	Sum	40		

	PTINR1	PTINR2	DIFFEREN
Mann-Whitney U test	148.000	122.000	157.000
Wilcoxon W test	473.000	447.000	482.000
Z	-1.104	-1.830	-.852
P (two-tailed)	.269	.067	.394

Figure 1 - Mann-Whitney Test

Conclusion

This method can be used as an alternative approach to assess drug-warfarin interactions.

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Implementation of An Integrated Network for Health Research in Quebec.

Allen Huang[a], Jacques Lemieux[b], Jean-Claude Bouchard[b], Michel Bourque[b], Robyn Tamblyn[b]

[a] *McGill University Health Centre, Montreal, Canada*
[b] *Clinical and Health Informatics Research Group, McGill University, Montreal, Canada*

Abstract

Health data warehouses represent a valuable resource for health research. We have developed an infrastructure capable of providing health researchers in the province of Quebec with a toolkit to access the clinical data warehouses contained in the major academic health centres and the provincial health administrative systems. This demonstration will highlight the components that allowed the successful implementation of an integrated network to accomplish this task. Acceleration of the pace and increases in the volume and quality of health research within the province, in other jurisdictions and possibly world-wide is now an attainable goal.

Keywords:

health databases, information management, patient data privacy, database management systems, computing methodologies, health services research

Introduction

Health research done in the conventional, paper-record environment is tedious, expensive and reliant on data with varying quality. Large repositories of health data are invaluable resources to researchers and planners in health care. In the province of Québec, Canada, these repositories are housed in clinical data warehouses within the large teaching hospitals and the administrative data warehouse at the Régie de l'assurance maladie du Québec (RAMQ) – the provincial health services payer. The Infostructure de Recherche Intégrée en Santé (IRIS) - Québec project is a Canada Foundation for Innovation funded initiative to construct an integrated network for health research in the province. Its goal was to create secure access to these data warehouses, enable the linkage of patient records through the use of a provincial Master Person Identifier (MPI) and ensure that resultant datasets returned to researchers conform to privacy standards.

Methods

The IRIS-Québec architecture is a distributed, federated data warehouse model. The RAMQ already manages the MPI. The research warehouses of clinical data were constructed to ensure the highest standards of data quality. People: Researchers wishing to access these data warehouses no longer have to queue for specialized data analysts and programmers. The researchers' toolkit is a web-based, user-friendly interface that drives a powerful system with the following functions: selection of variables from the extensive data dictionary across the warehouses, building of complex queries using logical operators, a temporal relation tool to define time-dependencies of variables, a crosstabs manager, and a data extraction manager. Previously onerous and lengthy authorization steps have been streamlined into an electronic approval process which is reliable, timely and track able. Privacy: A novel "inference controller" was developed to ensure individual data privacy. This software computes the probability of the presence of an unique, potentially re-identifying profile when multiple databases are linked. The researcher's toolkit allows the dynamic modification of data query parameters in order to achieve the desired data precision without violating privacy rules. Processes: Data sharing agreements were established with each partner institution. Software agents were installed at each data source to manage queries, connectivity, linkages, and data flows. Each agent contains a copy of the inference controller engine. Performance: is maintained by de-coupling the phases of a research project from the raw data extractions. The creation of the complete cohort data cube is triggered only on submission of the finalized project profile. We will demonstrate the operation of the Toolkit through sample queries.

Conclusion

The architecture of our system can potentially be extended to other biomedical information sources such as genomics, proteomics and geneology databases and create new capacity for health and bio-medical research. This architecture can also be potentially replicated in other jurisdictions. Scaling such a system to the international level can result in enabling health and bio-medical research in the global community.

Address for correspondence

Allen Huang, MDCM
McGill University Health Centre
687 Pine Avenue W, Room M8.12
Montreal, Quebec
Canada H3A 1A1

MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

OpenECG: Promoting Interoperability Through the Consistent Implementation of the SCP-ECG Standard in Electrocardiography

Catherine Chronaki[a], Franco Chiarugi[a], Ronald Fischer[b]

[a]*Institute of Computer Science, FORTH, Heraklion, Crete*
[b]*Department of Biometrics, MHH, Hanover, Germany*

Abstract and objective

The OpenECG Network (www.openecg.net) has been created to promote interoperability in electrocardiography with tutorials, specifications, open source tools, data sets, converters, and interoperability testing. ECG vendors, members of professional organizations, researchers, and other stakeholders participate in the OpenECG network to exchange views and receive assistance in implementation. In 2006, members are more than 700 individuals from 58 countries. A specific focus area for OpenECG that concerns diagnostic quality resting electrocardiograms (ECGs) is SCP-ECG, the European standard (EN1064:2005). An online interoperability testing service assists members in consistently implementing SCP-ECG and effortlessly integrating electrocardiographs with eHealth systems. OpenECG is a case of best practice in interoperability that should be followed by medical devices and sensors for effective personalized health monitoring.

Keywords:
interoperability, standards, telemedicine, eHealth services, medical devices, electrocardiography

Methodology

Electrocardiography is the most frequently applied non-invasive examination for early detection of heart disease, a leading cause of morbidity and mortality in western countries. It is estimated that more than 100 million ECGs are recorded annually in Western Europe. The ECG allows early detection and follow-up of heart disease, but today the operation of most ECG devices is still based on proprietary protocols and file formats.

SCP-ECG is a standard communication protocol that specifies the interchange format and a messaging procedure for ECG equipment-to-computer communication and for retrieval of ECG records from the computer to the ECG equipment (if needed). Since March 2005, SCP-ECG (EN1064:2005) is the European standard for high quality diagnostic ECG exchange and if consistently implemented ensures interoperability. Nevertheless, it is rather difficult for integrators to implement SCP-ECG correctly and there are variations in implementations, which can be a barrier to interoperability. Although a number of ECG device manufacturers and integrators have implemented the SCP-ECG standard, most implementations are not fully accurate. This is due partly to misconceptions and partly to the lack of widely publicized conformance levels and IHE-like integration statements that exist for modalities in radiology.

In 2003, OpenECG established an online conformance testing service to support the OpenECG community at large in implementing interoperable eHealth systems with SCP-ECG support. A member may submit an ECG file in an alleged SCP-ECG format and receive a list of errors and warnings. If no errors are detected, the submitter may request a certificate that is granted after thorough manual review of the file.

Results

The SCP-ECG conformance testing service has been used extensively by members and in many occasions an interoperable solution was achieved with support from the help desk. In the period 2003-2006, more than 1700 ECGs were submitted for conformance testing by members in more than 20 counties worldwide. Leading is Italy with 11 members, who have submitted 37.3% of the tests. After Italy, most ECGs have been submitted by Greece (17.74%) and Hungary (11.89%). ECG devices and eHealth services have been tested, improved, and validated using online tools and support from the OpenECG helpdesk. In 2005, a web service variant of the conformance testing service was integrated to the ECG viewer that won the first prize in the OpenECG programming contest. After certain limitations of the software were identified and amended, a new version of the ECG viewer was released and is currently available at the OpenECG open source repository.

Conclusions

The OpenECG network promotes best practice in interoperability for ECGs. Innovative eHealth services capable of managing personal wellness profiles call for plug-interoperability of medical devices, which is an issue of patient safety, key to advancing quality and cultivating consumer trust in the next generation of ambient intelligent working and living environments.

Subject Index

This page intentionally left blank

Author Index

This page intentionally left blank

This page intentionally left blank

This page intentionally left blank