Clinical Data Modeling Combining Agent-Based and Epidemiological Models

Denisse Kim[1,*], Manuel Campos[1,2], Bernardo Canovas-Segura[1] and Jose M. Juarez[1]

[1]*MedAI Lab, University of Murcia, Campus Espinardo, Murcia, 30100, Spain*

[2]*Murcian Bio-Health Institute (IMIB-Arrixaca), El Palmar, Murcia, 30120, Spain*

Abstract

Hospital-acquired infections (HAIs) are a major concern nowadays, since they entail a big threat to society and an increase in healthcare costs. AI techniques show great performance in the development of effective systems to help in their control and prevention. However, many recent studies highlight the lack of available datasets for reproducing their experiments, claiming for more trustworthy medical AI models. Realistic data simulation is a valid approach for testing these models when data is publicly unavailable or when clinical data gathering is cumbersome or impossible. Main simulators often focus on implementing compartmental epidemiological models and contact networks for validating epidemiological hypotheses. However, very little attention is paid to hospital infrastructure (e.g. hospital building, policy, shifts, etc.) which plays a key role in the infection and outbreak processes. This paper proposes a novel approach for a simulation model of HAI spread, combining agent-based patient description, spatial-temporal constraints of the hospital settings, and microorganism behavior driven by epidemiological models.

Keywords

simulation model, agent-based model, hospital-acquired infection, infection control, epidemiological model

1. Introduction

Multidrug-resistant bacteria (MDR-bacteria) are bacteria that evolved and acquired resistance to antimicrobial drugs. This resistance makes the treatment more complex, increasing the risk of infection, its spread and mortality [1]. They are a growing concern, especially in the case of hospital-acquired infections (HAIs), since they entail an increase in healthcare costs and a big threat to society. Health systems must have the necessary means to be able to assess the presence of these infections in hospitals. To this end, the spatial structure of a hospital and the physical distribution of patients over time play important roles in the detection of infection outbreaks and the prevention of their spread.

In this context, machine learning (ML) and deep learning (DL) techniques present an opportunity to develop effective systems that can help in the clinical decision and planning process.

*Corresponding author.

✉ denissemisoo.kim@um.es (D. Kim); manuelcampos@um.es (M. Campos); bernardocs@um.es
(B. Canovas-Segura); jmjuarez@um.es (J. M. Juarez)

🆔 0000-0002-1790-372X (D. Kim); 0000-0002-5233-3769 (M. Campos); 0000-0002-0777-0441 (B. Canovas-Segura);
0000-0003-1776-1992 (J. M. Juarez)

The development of these systems requires access to big volumes of high-quality data, both for training and validating. However, several concerns might compromise the use of health data in AI research, such as the quality of accessible data and the risk of bias, the protection of individual privacy, or the individual loss of autonomy, among others [2].

Some approaches to preserving privacy are the anonymization or pseudo-anonymization of data. This could encourage the data usage, though not without challenges (e.g. data triangulation) [3]. Part of the solution to this problem is the simulation of realistic data, since it has two main benefits: from a public-health perspective, they allow predictive analysis and early evaluation of hospital policies in different scenarios; from a medical-AI perspective, they are useful for implementing and evaluating future ML and DL techniques in a more fair and reliable way.

This paper presents a simulation model to study the propagation of infections within a hospital. The core of the simulation is to study the movement of a population of patients inside a hospital, and to allow the analysis of the spread and outbreak of an infective disease in this population. Therefore, it consists in the combination of an agent-based model (micro-scale model), the dynamics of an epidemiological compartmental model (macro-scale model), and the policies and the physical structure of a healthcare environment.

2. Simulation model and structure

Our proposal is based on a discrete event time model in order to simulate a bacterial infection spread among patients in a hospital setting. In each step of time, hospitalized patients are going to move around the hospital, and they are going to be involved in the transmission of a bacterial infection as well as in its potential outbreaks. To achieve this, the simulation model consists in 3 main components: the input parameters, the core of the simulator (i.e. micro-scale modeling via an agent-based model, macro-scale modeling with the use of a compartmental model, and spatial-temporal constraints of the hospital settings), and the simulation outputs. The framework of the simulation model is presented in Figure 1 and the components are explained below.

2.1. Macro-scale model: epidemic model

Compartmental models are modeling techniques often applied in the simulation of infectious diseases. The population is assigned to labeled compartments and they can progress from one to another. In these models, the population dynamics are well known [4], which makes them suitable for performing predictions and estimations of epidemiological parameters. Among the many variations of compartmental models (e.g. SIR, SIS, SEIR, etc.), the SEIRD model is the most suitable, since it adjusts well to the disease progression we want to study. Therefore, to represent the evolution of the disease, we have adjusted and applied this model by assigning each patient a state of health at a time: susceptible (S), exposed (E), infected (I), recovered (R), deceased (D), or non-susceptible (NS), which represents those that have immunity.

Population are admitted at the hospital in state S, I, or NS. They can get infected while in state S. If this happens, they go to state E, which means that they are incubating the disease, but are not contagious. They are going to remain in state E a period of time that depends on the incubation duration of the infection. Once this period is over, they go to state I, thus they
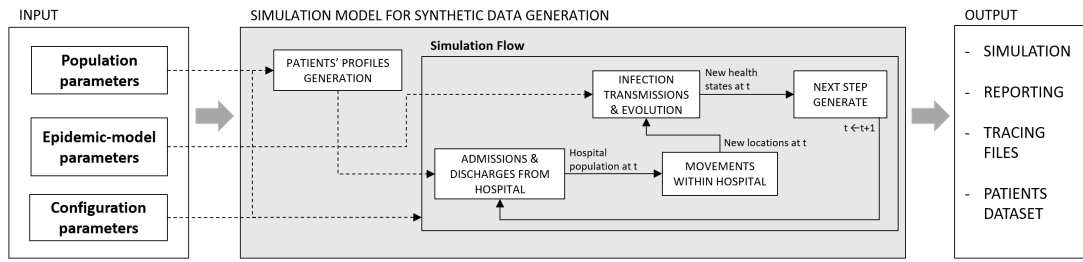
Figure 1: High scale representation of the simulation model.

can infect others and contaminate the environment. If they survive the disease, they go to state R and, if not, they go to state D.

2.2. Micro-scale model: agent-based model

Agent-based approaches ease micro-scale simulations, describing individual and their interactions. Unlike other approaches, like [5], patients are the only agents in the model. This decision helps us to focus on the evolution of the infection process based on solid epidemiological models, avoiding unverified factors of contagion vectors like healthcare workers or visitors.

With the patient simulation proposed, we track their stays from the arrival to the hospital to their discharge. Each patient has a unique ID and a set of attributes that include the localization where they are, age, gender, length of stay (LOS) in the hospital, health state, incubation period, infection duration, and applied treatment. We have considered only adult patients and we track their movements around the hospital in each time step through the different areas and services.

Patients can interact with other patients and with the environment. Interactions between patients take place when they share a room. During these interactions, a sick patient can infect another with probability p_{pp}. Interactions with the environment happen when a patient has spent an amount of time in the same place, and they can infect it with probability p_{pl}. This probability is going to be higher if the infected patient has not started the treatment yet, and lower if they have been on it for more than 3 days. Finally, a contaminated place can infect a susceptible patient with probability p_{lp}.

Regarding the recovery, infected patients may have a quick recovery without treatment or a longer one with the need of treatment. Both have a probability of success and a duration. In case of non-recovery, a patient may die with probability p_d. All of these probabilities and time periods depend on the modeled pathogen and are part of the parameters explained in Section 3.

2.3. Hospital: spatial-temporal constraints and policies

The spatial distribution of the hospital plays a key role in the spread of infections. We have considered a two-story hospital, where we took into account the most likely places in which a hospitalized patient can become infected: emergency room (ER), operating rooms, radiology rooms, wards which contain several patient rooms each, and an intensive care unit (ICU). The

ER, the ICU and each room can have a user-defined number of beds. Each bed and place have a unique ID and a state indicating whether they are contaminated by the infection or not.

The places that a patient can go to are divided into two types: temporary and indefinite. Temporary places are those in which a patient is going to spend a short period of time (e.g. radiology or surgery). Indefinite places are those where a patient can stay for a long period of time (e.g. a bed, the ICU). The patients movements are spatially and temporally constrained. In order to implement these constraints as realistic as possible, we have designed a series of rules following the suggestions of medical doctors:

- Temporal constraint: in each step of the simulation, only a limited number of patients can move to each ward.
- Spatial-temporal (ST) constraint: patients must have spent a minimum number of steps without having gone to a temporary place to go back. For example, if a patient has just undergone surgery, they will not go back into the operating room right away.
- ST constraint: patients that have been in the ER or the ICU for a certain period of time can be transferred to a ward.
- Spatial constraint: when a patient goes to a temporary place, in the next step they return to the same bed where they were before.
- ST constraint: patients can change of bed in the same ward during 1 simulation step.
- ST constraint: patients in a ward can change to another ward during 1 simulation step.
- ST constraint: patients in a ward or the ER might be transferred to the ICU.

3. Parameters

The simulation model receives input parameters with user-assigned values. These parameters can be classified into three types: population parameters, epidemic-model parameters and configuration parameters.

Population parameters are those that represent the population and the characteristics of the hospital (i.e. the occupancy rate, admission rate through the ER, number of beds, rooms and wards, etc.). Among these parameters, we also find the patient's age and LOS distributions.

Epidemic-model parameters are those that configure the infection behavior and allow the change from one state of health to another (i.e. probability of contamination, probability of recovery, treatment duration, etc.). We obtained or calculated both these and the population parameters from public access data and information from the literature. Those for which there was not enough information to infer their distribution follow a triangular distribution defined by a mean value that represents its mode, a minimum and a maximum value, based on [6].

Configuration parameters comprise the settings for each run of the simulator (i.e. the cleaning frequency, the time between the movements of a patient, etc.). We defined these parameters based on the hospital size, data published by hospitals, and information obtained from an expert.

4. Simulation outputs

When working with micro and macro-scale modeling, one advantage is the derivation of results with low and high level of abstraction. Two straightforward outputs are daily statistics of the
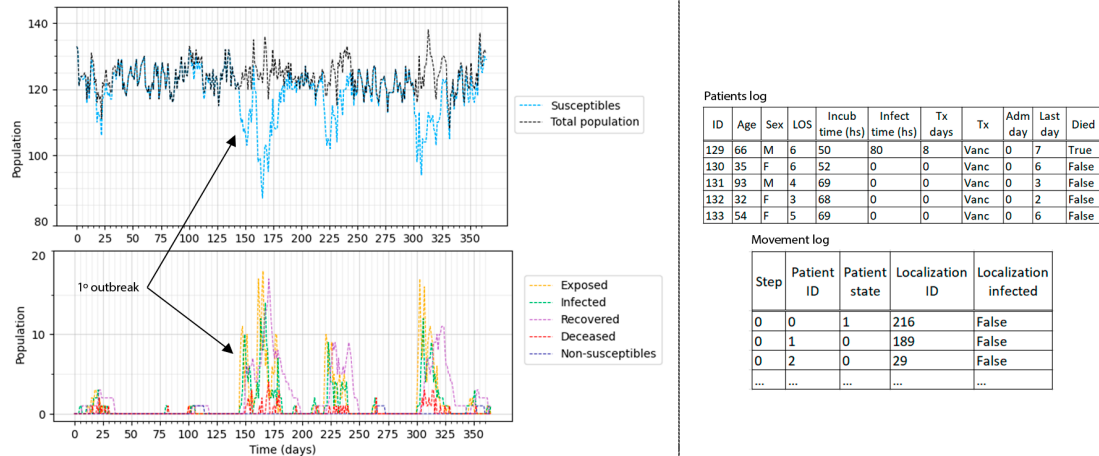
Figure 2: Simulator outputs with CDI experiments. On the left, a macro-scale model output with the infection dynamics and number of patients in each health state at high level of abstraction. On the right, a micro-scale model output with the patients log and their movements trace at low level of abstraction.

Patients log

ID	Age	Sex	LOS	Incub time (hs)	Infect time (hs)	Tx days	Tx	Adm day	Last day	Died
129	66	M	6	50	80	8	Vanc	0	7	True
130	35	F	6	52	0	0	Vanc	0	6	False
131	93	M	4	69	0	0	Vanc	0	3	False
132	32	F	3	68	0	0	Vanc	0	2	False
133	54	F	5	69	0	0	Vanc	0	6	False

Movement log

Step	Patient ID	Patient state	Localization ID	Localization infected
0	0	1	216	False
0	1	0	189	False
0	2	0	29	False
...	...	...	...	...

processes under study, and a database of the patients during the time of the simulation. This database includes information on each patient and the places from the hospital where they have been to. Another output is the extraction of aggregated information of the infection: in this case, the number of patients in each health state per day is stored. By combining this with the localization of the patients in each step, it is possible to compute any epidemiological indicator that can be calculated with these data (e.g. prevalence, incidence density, etc.) for the different areas from the hospital.

5. Experiments and results

To experiment with this simulator, we have configured a hospital with 212 beds and the input parameters according to the Santa Lucia General University Hospital from Murcia, Spain. Each step is 8 hours, since it is the duration of a standard work shift. For the infection, we have chosen the Clostridium Difficile (CD) pathogen, which is the main cause of infectious diarrhea in hospitalized patients [7]. Figure 2 presents part of the output from an execution of our simulator with CDI, this includes a representation of the number of patients in each health state and their evolution and dynamics through time, as well as an example of the output logs. The patients log includes information regarding their age, sex, LOS, duration of incubation, duration of infection, duration of treatment, treatment, admission day, last day in the hospital, and deceased check. The movement log includes, for each step, the patients' ID and health states, and their localizations' ID and contamination check.

6. Related work

Codella et al. [6] implemented an agent-based simulation model with a Markov model to study the transmission of CD in a midsized hospital. They represented different types of agents and applied the model to compare the output of several control strategies. Our model is different from this approach in that our goal is to generate a dataset with individual as well as aggregated information for their latter use in other AI implementations. Besides this, another output from our work are epidemiological indicators to help monitor the spread of an MDR-bacteria infection in a hospital setting.

Lee et al. [5] presented a software tool called the Regional Healthcare Ecosystem Analyst, that creates an agent-based model with input data of a healthcare ecosystem. It is configurable by the user, including the characteristics of the infection chosen to represent. Its aim is to serve as a virtual laboratory to help to test different policies and interventions. Practically any healthcare facility type can be represented and beds are divided into ICU units and general wards. This differs from our model in that we give more importance to space by studying an infection spread in the most common areas of a hospital setting (e.g. the ER, the ICU, etc.). Another difference is that they use subroutines to calculate the number of agents in infectious and susceptible state in each ward, and based on that, they calculate the number of new cases in that ward that day. Instead of this, we monitor all the agents present in the hospital, so that we can know when they shared a room and interacted at low level.

Haber et al. [8] combined an agent-based with a compartmental model and focused on the study of second-line drugs in a small hospital. They analyzed several interventions to reduce the use of antibiotics and the incidence of HAIs. The main difference with ours is that the infection spread is calculated with differential equations and they do not model patients movements in the hospital, nor give the same importance to space and time as we do.

7. Conclusions and future work

This paper proposes an agent-based model coupled with the infection dynamics extracted from an epidemiological model. This model can be used as a generator of synthetic clinical data on MDR-bacteria infections within hospitals. The use of an agent-based model together with the role of the hospital topology in an infection spread can enhance the detection of spatial and temporal patterns to help in the monitoring and the decision-making process. This is thanks to a more precise study of the infection process and the consequences in hospitalized patients. The capacity of tracing patients at a low level and to also obtain aggregate results from them can play a key role and be a step forward in the creation of a more explainable AI and in the generation of higher-quality synthetic data. In future we plan to carry out a thorough evaluation of the model to ensure its correct implementation, clinical meaning and utility. Once we perform this evaluation, the simulator is going to be available in an open repository.

Acknowledgments

References

[1] World Health Organization, Antimicrobial resistance, 2022. URL: https://www.who.int/news-room/fact-sheets/detail/antimicrobial-resistance.

[2] World Health Organization, Ethics and governance of artificial intelligence for health, 2023. URL: https://www.who.int/publications-detail-redirect/9789240029200.

[3] B. C. L. Paisner, At a glance: De-identification, anonymization, and pseudonymization under the gdpr, 2017. URL: https://www.bclplaw.com/en-US/events-insights-news/at-a-glance-de-identification-anonymization-and-pseudonymization-1.html.

[4] E. Hunter, J. Kelleher, A framework for validating and testing agent-based models: a case study from infectious diseases modelling., 34th. Annual European Simulation and Modelling Conference (2020). doi:https://doi.org/10.21427/2xjb-cq79.

[5] B. Y. Lee, K. F. Wong, S. M. Bartsch, S. L. Yilmaz, T. R. Avery, S. T. Brown, Y. Song, A. Singh, D. S. Kim, S. S. Huang, The Regional Healthcare Ecosystem Analyst (RHEA): a simulation modeling tool to assist infectious disease control in a health system, Journal of the American Medical Informatics Association: JAMIA 20 (2013) 139–146. doi:10.1136/amiajnl-2012-001107.

[6] J. Codella, N. Safdar, R. Heffernan, O. Alagoz, An agent-based simulation model for clostridium difficile infection control, Medical Decision Making: An International Journal of the Society for Medical Decision Making 35 (2015) 211–229. doi:10.1177/0272989X14545788.

[7] L. Meyer, R. Espinoza, R. Quera, Infección por clostridium difficile: epidemiología, diagnóstico y estrategias terapéuticas, Revista Medica Clinica Las Condes 25 (2014) 473–484. doi:10.1016/S0716-8640(14)70064-1, publisher: Elsevier.

[8] M. Haber, B. R. Levin, P. Kramarz, Antibiotic control of antibiotic resistance in hospitals: a simulation study, BMC infectious diseases 10 (2010) 254. doi:10.1186/1471-2334-10-254.