![Mathematica — Progress Together]

Final Concept Report

Exploring the Feasibility of Using Audit Log Data to Quantitate Burden as Providers Use Electronic Health Records

September 5, 2019

Genna Cohen, Llew Brown, Megan Fitzgerald, and Anita Somplasky

This page has been left blank for double-sided copying.

CONTENTS

TABLES

FIGURES

ACKNOWLEDGMENTS

This page has been left blank for double-sided copying.

EXECUTIVE SUMMARY

Electronic health records (EHRs) have been widely adopted in the United States, but there is growing concern that they have unintentionally burdened clinicians, inhibiting their ability to deliver health care efficiently and effectively. In response to these concerns, the Office of the Assistant Secretary for Planning and Evaluation (ASPE) contracted with Mathematica to examine the feasibility of leveraging EHR audit log data to measure clinician burden associated with administrative, clinical documentation, and clinical review tasks.

Mathematica examined peer-reviewed and gray literature, spoke with subject matter experts (SMEs), and convened a Technical Expert Panel (TEP) to answer the following three specific research questions:

1. Does evidence exist that EHR audit log data can be compared across disparate systems when seeking to measure clinician burden?

2. Are there common features across audit log data that might be leveraged for national reporting of physician/clinician burden from the currently installed EHR base?

    - Are there comparable data in audit logs across EHRs and across different implementations of EHRs?

    - Are there comparable data over time?

    - Are there audit log data that are not relevant and which do not easily support measurement of clinical burden?

3. Can these measures be implemented and collected in a manner that allows the U.S. Department of Health and Human Services (HHS) to measure changes in burden over time?

    - What is the quality and variability of the data?

    - What can be done to improve data quality?

This is the last in a series of four iterative reports, adding supplemental literature and interviews with two SMEs to prior analyses. This final report also incorporates feedback from the project's second TEP meeting.

Summary of key findings

- There are no published studies comparing EHR audit log data across disparate systems specifically to measure burden, and very few instances of audit log data being compared across systems for other purposes.
- There is growing evidence suggesting EHR audit logs can be a reliable and accurate source of information regarding clinicians' time allocation within the EHR.
- However, audit log data are not a standalone source of information and will likely need to be aggregated and/or linked with administrative data to create a meaningful model of user behavior.
- Standards and regulations governing audit log data provide some indication of potential similarities in such data across EHR vendors, yet there is wide variability in how different vendors construct audit logs.
- More foundational work is required to develop measures of burden that could scale effectively for use in a national measurement program (for example, developing a standard approach for categorizing tasks and identifying ways to contextualize audit log data).
- Government agencies are uniquely positioned to advance the comparability of audit log data through changes to the EHR certification process and leveraging their authority to convene a multi-stakeholder group to develop best practices.

We found growing evidence to suggest that EHR audit logs can be a reliable and accurate source of information regarding clinicians' time allocation within the EHR, based on studies that compared audit log data to other observational methods, as well as reports from SMEs. However, those with experience constructing measures from raw audit log data cautioned that the initial effort required to make sense of it might not scale effectively to a national measurement program. Moreover, although the standards and regulations governing audit log data indicate that there are potential similarities in such data across EHR vendors, SMEs suggest that the standards are not detailed enough to ensure comparability between vendors and products. This could make it difficult to construct measures from multiple vendors.

Respondents universally agreed that clinicians spend too much time completing tasks in the EHR, but they were split in their attribution of this burden to the EHR directly or to outside forces (such as policies affecting health care delivery and its subsequent documentation, like federal and state reporting requirements). Three key aspects of EHR use were considered to be particularly burdensome and may be amenable to quantification using EHR audit log data: (1) managing "in-basket" messages, (2) volume of time spent using the EHR (particularly after work hours), and (3) usability of the EHR interface (see Table ES.1). Respondents noted that not all EHR use was burdensome, and that it would be difficult to identify when valuable work ended and non-value-added work began. Respondents believed measures derived from audit log data might be more meaningful for understanding EHR use and identifying outliers (for example, clinicians who spent more time using the EHR than their peers) rather than definitively determining if that use was burdensome. Respondents further cautioned that finding the right comparison group from which to identify outliers might prove challenging in light of how

specialty, setting, time in practice, office staffing models, and personal preferences influence EHR use and thus the interpretation of audit log data. Alternatively, measures derived from audit log data could hold promise for longitudinal evaluations of change over time (for example, after a change in policy designed to reduce burden). Ultimately, respondents were skeptical that such data alone would provide enough context to understand burden without also incorporating administrative, survey, or other data.

Our work was informed by a supplemental literature review that focused on the three sources of burden associated with EHR use we identified as potentially valuable for measurement (in-basket management; volume of time spent using the EHR, especially after work hours; and usability of the EHR interface) as well as additional interviews with vendor experts with a focus on audit log data, the relevant standards that might influence their granularity, and potential strategies for data access.

Table ES.1. Summary of potential sources of burden and related measure concepts

	In-basket management	Volume of time spent using the EHR	Usability of the EHR interface
Broad measure concepts	• Time spent interacting with in-basket (send or read/receive) • Volume of messages (sent or received)	• Time spent actively using the EHR • Time spent actively using the EHR before or after clinical visits	• Number of screens/tabs per visit
Narrow measure concepts	• Time spent authoring messages or proportion of authored messages that are not opened, or are opened for less than 1 second • Time spent authoring messages or proportion of authored messages that do not result in subsequent activity • Proportion or volume of messages auto-generated by the EHR • Number of messages requesting co-signature for an activity performed by another licensed clinician	• Proportion of clinicians whose time spent on documentation for an initial patient encounter is "X" % greater or lower than the mean time for visit documentation, among their same-specialty peers, adjusted for patient characteristics	• Number of screens to complete an important task, such as documenting physical exam findings or renewing a prescription
Considerations	• Doesn't account for messaging outside the EHR (fax, paper referrals) • Also driven by user preferences, workflow and staffing (including team-based care), clinical specialty, and organizational protocols	• Difficult to account for different practice styles and preferences (for example, choosing to document in the evening) • Distinguishing active and inactive time similarly complex • May require linking to other data sources, including administrative • Also driven by regulatory requirements for documentation	• Difficult to capture and compare different data entry methods (for example, manual entry, voice-to-text, use of templates, or pulling forward prior notes) • Data may be too granular to easily analyze • Also driven by user documentation and workflow preferences, and patient complexity

I.   INTRODUCTION

The implementation of health information technology (HIT) and adoption of electronic health records (EHRs) have increased substantially in the past 10 years. These changes reflect both the natural diffusion of innovation and a potential boost from the passage of the HITECH Act, implemented as part of the American Recovery and Reinvestment Act of 2009 (Mennemeyer et al. 2016; Adler-Milstein and Jha 2017). The HITECH Act aimed to stimulate the economy and encourage the adoption of tools that can improve the quality and efficiency of health care delivery. The Act included incentive payments for providers who demonstrated the adoption and meaningful use (MU) of EHR systems. To date, the Medicare and Medicaid programs have paid over $38 billion in incentive payments to eligible professionals and hospitals for their use of EHR systems (Centers for Medicare & Medicaid Services [CMS] 2018). As of 2017, 80 percent of office-based physicians and 96 percent of non-federal acute care hospitals have adopted a MU certified EHR (Office of the National Coordinator for Health Information Technology [ONC] Dashboard Quick Stats 2017).

Even with widespread adoption and implementation, EHRs have had mixed results in improving the value of clinical care. Some have realized the potential benefits of EHR adoption, such as lower costs, increased efficiency, and improved quality and access to care, but the results are not universal (Campanella et al. 2015; Jones et al. 2014). Furthermore, the HITECH Act and subsequent regulatory mandates require that clinicians document patient care to an extent that many consider burdensome (Friedberg et al. 2013; Jha and Iliff 2019; Tutty et al. 2019). As noted in a recent *New Yorker* article, the aggregate effect of well-intentioned EHR requirements, (such as mandating completion of a particular field) have the unintended consequence of overwhelming clinicians and impeding their ability to deliver care to patients (Gawande 2018). These observations are supported by peer-reviewed literature as well. At least 70 percent of clinicians using EHRs blame the system for their administrative burdens (Jamoom 2017). ONC recently noted that clinicians blame EHRs for increasing regulatory and administrative burden that results from "an ever-increasing, wide ranging, and often poorly coordinated body of requirements to deliver, and receive payment for, patient care" (Burden Report ONC 2018). Clinicians have expressed concerns that EHRs result in less time for patient-clinician interaction, increased data entry tasks, and longer clinician workdays (Payne et al. 2015).

Research shows that EHRs are one of many interrelated factors contributing to increasing rates of clinician burnout, including hectic work environments, misalignment between clinicians' values and their organizations' administrators, time and efficiency pressures, and a lack of autonomy (Dyrbye et al. 2017; Gardner et al. 2018; Lee et al. 2019). EHR-related stress stemming from increased clerical burden, cognitive load on clinicians, and frequent interruptions and distractions has been cited as an important component of clinician burnout. A 2014 study of 6,375 U.S. physicians in active practice across all specialties showed that physicians who used EHRs or computerized provider order entry (CPOE) were less likely to be satisfied with the time spent on clerical tasks. In addition, they were at higher risk for professional burnout after adjusting for age, sex, specialty, practice setting, and hours worked per week. Clinicians perceived that over 60 percent of the time spent on documentation tasks in the EHR is not related

to patient care, and clinicians overall reported dissatisfaction with the burden associated with their EHRs (Shanafelt et al. 2016). In a survey of Rhode Island physicians, 70 percent reported HIT-related stress, with the highest frequency in primary-care-oriented specialties, and 26 percent reported burnout (Gardner et al. 2018).

To date, health services researchers have used a variety of methods to measure physicians' workload, satisfaction, and burden associated with EHRs, including surveys, self-reports and diaries, observational studies, and focus groups. Several studies focusing specifically on clinicians' use of EHRs rely on surveys and self-reports (DesRoches et al. 2008; Sockolow et al. 2011; Friedberg 2013; Shanafelt et al. 2016). Some studies also use direct observations (including time-and-motion observations) to measure clinicians' interactions with the EHR (Ballermann et al. 2011; Sinsky et al. 2016). In addition, the National Center for Health Statistics (NCHS), with support from ONC, is developing a survey to better understand EHR burden (NCHS 2017). One frequently raised concern about the survey and observational studies relates to the Hawthorne effect: that is, a participant might change their behavior as a result of being studied (Parsons 1974). This in turn could limit the reliability and validity of the corresponding findings. In addition, observational and self-report designs rely on time-consuming data collection, can be expensive to develop and implement, and risk adding to the burden of participants.

To address these methodological concerns, researchers have begun to explore *computational ethnography—*"a family of computational methods that leverages computer or sensor-based technologies to unobtrusively or nearly unobtrusively record end users' routine, in situ activities in health or healthcare related domains" (Zheng et al. 2015). Computational ethnography includes the possibility of leveraging EHR event logs, also known as audit logs—an automated tracking feature that monitors access to and activity within EHR systems for administrative and security purposes—to measure the time and use patterns of clinicians interacting with EHRs (Adler-Milstein and Huckman 2013; Ancker et al. 2014, 2017; Arndt et al. 2017; Kannampallil et al. 2017; Tai-Seale et al. 2017). Audit log data may contain information used for security or software operations. Collectively, this information may be leveraged to measure general aspects of use or burdensome use specifically.

Given both the emerging research and the ubiquity of the audit log data source, the Office of the Assistant Secretary for Planning and Evaluation (ASPE) anticipates that EHR audit logs may be a uniquely rich data source that could be efficiently leveraged to measure EHR burden on a national scale. ASPE contracted with Mathematica to explore aspects of EHR burden that could be quantitated, as well as the capabilities and commonalities of commercial audit log data to assess the feasibility of developing reliable and valid measures of clinicians' burden across different health systems and EHR vendors.

A.   Research questions

Although there is evidence that EHR audit logs can help capture user behavior, more research is needed to evaluate the feasibility of building a national measurement strategy around this data

source, as well as how to prioritize different dimensions of burden. To achieve these goals, this project addresses three research questions:

1. Does evidence exist that EHR audit log data can be compared across disparate systems when seeking to measure clinician burden?

2. Are there common features across audit log data that might be leveraged for national reporting of physician/ clinician burden from the currently installed EHR base?

    - Are there comparable data in audit logs across EHRs and across different implementations of EHRs?

    - Are there comparable data over time?

    - Are there audit log data that are not relevant and which do not easily support measurement of clinical burden?

3. Can these measures be implemented and collected in a manner that allows the U.S. Department of Health and Human Services (HHS) to measure changes in burden over time?

    - What is the quality and variability of the data?

    - What can be done to improve data quality?

To guide our exploration of these questions, we focus on three use cases for clinician interaction with EHRs:

1. Administrative tasks, such as accessing patient demographic information and responding to messages

2. Clinical documentation tasks, such as entering encounter notes and completing orders

3. Clinical review tasks, such as reviewing prior encounter notes or medications

These use cases align with different EHR tasks outlined in a recent article (Arndt et al. 2017) analyzing audit log data from a commercial vendor in an academic medical center and reflect activities that are both common and central to patient care.

We will describe our findings in four iterative reports that synthesize evidence from peer-reviewed and gray literature, as well as findings from telephone discussions with subject matter experts (SMEs) and meeting with a Technical Expert Panel (TEP), to provide an initial assessment of the feasibility of developing measures of clinician burden from audit log data (Figure 1). This is the final, fourth-round concept report.

Figure I.1. Sequence of concept papers and reports

Round 1 Concept paper	Round 2 Revised concept paper	Round 3 Final draft concept report	Round 4 Final concept report
•Scan peer-reviewed and gray literature •Interview subject matter expertes by telephone	•Summarize TEP input •Extend and contextualize initial findings	•Propose prototype measures of clinician burden from audit log data and address feasibility •Provide recommendations on the feasibility of developing prototype measures	•Synthesize input from second TEP meeting and information from other sources •Incorporate comments from the contracting officer's representative

B.   Methods

We utilized a literature scan, a TEP, and semi-structured interviews to explore the feasibility of using audit log data to understand EHR burden in the context of use cases that relate to direct patient care of clinicians working in ambulatory care settings.

Specifically, we evaluated peer-reviewed and gray literature that described or relied on audit logs and conducted a brief review of standards that govern audit logs (though not the focus of our work). We also reviewed recent efforts at measuring burden. This literature served as a foundation to identify SMEs with expertise in EHR burden or audit log data. In addition, we identified gaps in the literature that warranted further exploration with a TEP and through interviews.

We consulted ASPE, ONC, and CMS to identify five SMEs with the broadest experience and complementary perspectives who could address our research questions, including physician informaticists and researchers, EHR vendor executives, and national thought leaders. We invited these five SMEs to serve as members of the TEP and provide ongoing feedback on data collection, interpretation, concept paper and report content, and next steps. These experts are:

- Julia Adler-Milstein (Ph.D.), associate professor, University of California, San Francisco

- Farzad Mostashari (M.D.), former national coordinator for health information technology at HHS and co-founding chief executive officer of Aledade

- J. Marc Overhage (M.D., Ph.D.), chief medical informatics officer and vice president of intelligence strategy at Cerner

- Vimla L. Patel (Ph.D.), senior research scientist and director at the New York Academy of Medicine Center for Cognitive Studies in Medicine and Public Health

- Christine Sinsky (M.D.), vice president of professional satisfaction at the American Medical Association

We asked these panelists to review our article list and suggest other sources that would enhance our understanding of EHR burden. We also asked them to suggest other experts who would be valuable to speak with.

In consultation with ASPE, ONC, and CMS, we finalized a list of additional SMEs to participate in a 45-minute telephone interview about the potential feasibility of measuring EHR burden with audit log data. These nationally recognized experts shared their perspective based on experience at clinical, vendor, policy, and research organizations. The SMEs include a former national coordinator for HIT, EHR vendor representatives, and chief health information officers of health systems. Between the Quantitating the Burden of Electronic Health Records contract and the Policy Analysis and Decision-Making Capacity contract, we conducted 18 individual and group interviews for this report.

To conduct the interviews, we drafted a semi-structured discussion guide with questions designed to elicit information from respondents on key tasks that clinicians complete in the EHR related to patient care, manifestations of EHR burden, and the capabilities and limitations of EHR audit log data, as well as a national measurement strategy to better understand EHR burden using audit log data (see Appendix A). We incorporated input from ASPE, ONC, CMS, and the TEP to ensure the questions were balanced and covered important topics identified in the literature. Because of the diverse and varied background of participating SMEs, we tailored the protocol to each respondent prior to the discussion in order to best leverage their unique perspectives. For example, we prioritized questions about dimensions of burden with clinical workflow experts and prioritized questions about the strengths and limitations of audit log data, including comparability across systems and over time, with EHR vendor representatives.

A two-person team conducted each interview, and, after receiving consent, recorded the discussion. The lead researcher focused on developing rapport with the respondent and making sure that responses to questions were comprehensive. The second researcher took notes during the conversation and consulted the recording as needed to ensure the relevant details were captured.

The team developed codes that the second researcher applied to each segment of text to capture the topic discussed (for example, "Key EHR Tasks" or "Dimension of Burden"). The team then used Atlas.ti analytic software to consolidate information that pertained to the same codes using code reports (for example, one code report capturing all text segments related to "Key EHR Tasks" from all SMEs). This information was later summarized in analytic matrices to organize data and facilitate identification of themes. Specifically, individual responses made up the rows of the matrices; topics related to burden and the potential for creating measures of burden from audit log data made up the columns. Analysis of these matrices supported identification of areas of agreement and disagreement by comparing information on the same topic across respondents.

We summarized our findings in a draft concept paper, which we shared with experts at ASPE, ONC, and CMS before refining and sharing it with the TEP. We convened a two-hour virtual meeting on Monday, March 18, 2019, to discuss our findings, areas for future work, and planned follow-up activities. After the meeting, we reviewed additional literature related to the standards

that govern audit log data and the sources of burden associated with EHR use we identified. We also conducted two supplemental interviews to gain deeper insights into the technical issues of accessing and comparing audit log data across settings and vendors. Subsequently, we convened a second two-hour virtual TEP meeting on Wednesday, July 24, 2019, to refine measure concepts and recommendations for next steps.

II.   LITERATURE

A.   Audit log data background

In general, when a clinician or practice staff person enters information or moves across the modules within an EHR, the system creates a record that captures the information associated with the event. Growing evidence suggests that such logs can be a reliable and accurate source of information regarding clinicians' time allocation within the EHR (Arndt et al. 2017; Kannampallil et al. 2017; Tai-Seale et al. 2017; Wu et al. 2017; Zheng et al. 2015). EHR log data can be used to identify patterns in clinicians' interactions with the EHR when providing both direct patient care (patients seen in office visits) and virtual or remote activities (electronic communication with patients, billing and coding, documentation) over a certain period in diverse clinical settings.

Most studies to date using EHR data rely on a single vendor, often within the same health care setting (Arndt et al. 2017; Hribar et al. 2017). To our knowledge, no large-scale studies comparing EHR logs obtained from different vendor systems and diverse health care settings have been published. Although one study compared actions taken in practices using two different commercial vendors (Epic and GE), the study focused narrowly on alerts at three physician practice sites in Texas (Murphy et al. 2016) and did not provide a broader framework for delineating different EHR tasks. Likewise, a recent study compared EHR usability and safety in four different emergency rooms using two different commercial vendors (Epic and Cerner). However, the study only included discrete diagnostic imaging orders, laboratory orders, and electronic prescriptions. Nonetheless, this study identified more variation between sites using the same vendor than across sites using different vendors in the number of clicks as well as time spent (Ratwani et al. 2018).

The language used to describe EHR tasks across studies highlights potential variability in the granularity of the data captured in the logs obtained from different vendors. For example, four published studies using log data measured tasks related to a patient's problem list using different frameworks: Ancker et al. (2014) categorized log data related to problem lists by examining three measures, one each for adding, replacing, or dropping problems, and Adler-Milstein and Huckman (2013) combined these measures into one measure for "revising" the problem list and included a separate measure for reviewing the problem list. In contrast, Arndt et al. (2017) and Cohen et al. (2019) constructed a single measure related to reviewing or editing the active problem list. Moreover, a fifth study (Chen et al. 2016) that categorized EHR use as a chart review, orders, documentation, or "other" could conceivably split tasks related to a patient's problem list across these different categories. It is unclear whether the data used to create aggregate measures of interacting with the problem list reflect (1) the authors' choices to roll up more granular data elements for analytic purposes, or (2) differences in available data. In addition, it is unclear whether the authors were relying on audit log data specifically or other metadata that EHR vendors store in the course of use.

B.   Standards

Several standards and regulations govern audit log data and provide some indication of potential similarities in this data across EHR vendors. However, EHRs produce multiple system-generated logs and transaction-level records, not all of which are audits of user activity. Table II.1 includes a brief list of prevailing standards.

Table II.1. Key audit log standards

Standard	Overview
ASTM E2147	The American Society for Testing and Materials (ASTM) International established ASTM E2147, which "specifies how to design an access audit log to record all access to protected health information (PHI) maintained in EHRs and includes principles for how to document and disclose PHI to external users" (ASTM International 2009). ASTM E2147 defines a minimum set of data elements that audit logs must contain, which includes type of action (addition, deletion, change, queries, print, copy), date and time of event, patient identification, user identification, and identification of the patient data that was accessed (CMS 2014).
Audit Trail and Node Authentication	Integrating the Healthcare Enterprise (IHE) developed the Audit Trail and Node Authentication standard, which provides an audit mechanism reflecting security guidelines for Health Information Exchange (IHE 2019).
ISO 27789	The International Organization for Standardization (ISO) established ISO 27789 in support of a "common framework for audit trails for [EHRs], in terms of audit trigger events and audit data, to keep the complete set of personal health information auditable across information systems and domains" (ISO 2019).

The Health Insurance Portability and Accountability Act (HIPAA) Security Rule and Meaningful Use (MU) regulations incorporate common standards for EHR audit logs. Specifically, the HIPAA Security Rule requires that health care organizations implement "audit controls" for EHRs and other systems that interact with electronic protected health information (Cothran and Reilly 2017). In turn, Stage 1 of the voluntary MU certification program specified compliance with the HIPAA Security Rule, and Stage 2 specified that EHR vendors' audit reports include "entries in the audit log according to each of the data specified in the [ASTM E2147 standards]" (Sittig 2017). Moreover, ONC's final rule for the 2015 edition of the EHR certification criteria requires audit log data to include information specified in the standards developed by the American Society for Testing and Materials (ASTM) International (ONC 2015). Thus, in order to comply with MU, vendors' audit log data must align with the ASTM E2147 standard for content and structure.

At the time of this study, the latest proposed rule from ONC (n.d.) included plans to make the 2015 certification edition the standard requirement for all vendors. EHR developers seeking to comply with the requirements for audit log reporting can self-declare that they meet the criteria in the 2013 version of the ASTM standard cited in Table II.1, according to EHR certification test procedures published by ONC (2015). Based on the most current version available (ASTM International 2018), audit logs may contain multiple data elements and reporting features that

could be used to support the development of measures, including date and time of activity, unique identifiers for the patient and user, and the type of action performed in the record (such as creations, additions, deletions, changes, queries, accesses, or copy and paste).

Furthermore, the current ASTM standard promotes the accessibility and usability of audit log data by requiring a data dictionary that describes the data, including its "connections and dependencies," in nontechnical language for a lay audience. The standard also promotes "easy retrieval" of audit log contents for patients, advocates, and other authorized users. Finally, the standard requires audit log data to be retained as long as the medical record is maintained—for at least 10 years, or for 2 years after the legal age of majority, unless a longer period of record retention is prescribed by state, federal, or other law or regulation.

A more thorough review of technical standards and regulations was beyond the scope of this study. However, the current ASTM standard, with guidance to make specific content available and easily accessible, highlights efforts to create comparability across vendors in areas that would support use of audit log data to create national measures.

This page has been left blank for double-sided copying.

III. INCREASED FOCUS ON MEASURES

The National Quality Forum (NQF) has already taken steps toward developing measures of burden related to EHR use. Due to concerns about provider burden, NQF proposed measure concepts related to interoperability that also inform the extent of clinician burden (NQF 2017). For example, NQF suggested measures that assess the number of clicks a provider has to make to locate electronically exchanged data, and the extent to which electronically exchanged data was a direct match to the patient. In addition, the Agency for Healthcare Research and Quality (AHRQ) funded the development of an EHR usability toolkit to promote user-centered design and standardized processes for testing usability (AHRQ 2011). AHRQ also hosted a recent conference to present strategies for reducing provider burden through HIT design that considers clinical workflow and cognitive workload (National Web Conference on Reducing Provider Burden Through Better Health IT Design 2018).

Even more recently, ONC's Strategy on Reducing Regulatory and Administrative Burden Relating to the Use of Health IT and EHRs (2018)—hereafter referred to as the Burden Report— outlined priority sources of burden associated with EHR use. Several of which we note may be amenable to measurement using audit log data. For example, the Burden Report highlighted clinicians' concerns that complying with documentation requirements obliges them to complete documentation outside of work hours; audit log data can likely be leveraged to understand the times of day when clinicians complete documentation. The Burden Report also underscored the potential for copying and pasting to "[contribute] to meaningless data accumulation." Although audit log data may not be an ideal source for assessing the value of a particular point of data, and although there are times when copying and pasting prior notes may be appropriate, we note audit log data may be able to distinguish between information that is entered *de novo* and information that is copied and pasted from prior notes.

Furthermore, the Burden Report identified clinician "inundation" with pop-up alerts, the volume of which we note may be measured using audit log data, as could the required clicks to complete necessary actions (the reduction of which the Burden Report identifies as a priority for workflow optimization). Finally, as noted in the Burden Report, CMS is implementing several documentation policies in the CY 2019 Medicare Physician Fee Schedule (PFS) final rule aimed at clarifying documentation requirements so that data "already present in the medical record need not be re-documented but rather can be reviewed, updated, and signed off on by the billing practitioner." We believe the impacts of this policy can be measured using audit log data to determine whether resultant documentation changes are indicative of reduced burden. Indeed, Leigh Burchell, vice president of policy and government affairs of Allscripts, recently commented on the Burden Report in support of "EHR [audit] logging functionality…to gather end user data and gauge the cognitive tax score the EHR is putting on the end user" (Burchell 2019).

This page has been left blank for double-sided copying.

IV. FINDINGS

In this section, we describe key findings from SMEs and the first TEP meeting on scenarios for EHR use; making sense of audit log data; understanding burden; and potential measures of burden. These potential measures of burden focus on in-basket management, volume of time spent using the EHR (particularly after work hours), and usability of the EHR interface.

A. Scenarios for EHR use

To better understand the different ways clinicians use EHRs to document, retrieve, review, and share information, we asked respondents about the key tasks that clinicians complete in the EHR related to patient care. We offered an opportunity for open-ended response and comments on how clinicians use the system to perform administrative, clinical documentation, and clinical review tasks.

Few respondents had a formal way of describing the key patient care tasks outpatient clinicians performed in the EHR. Among respondents that did, there was agreement on the importance of clinician time spent reviewing information, submitting orders, and updating patient information. Additionally, one respondent, a practicing physician and health system chief information officer, suggested that it would be meaningful to categorize tasks according to the type of information being reviewed, with a particular focus on (1) problem lists, (2) medications, and (3) allergies and adverse reactions. Although these tasks do not encompass all areas of EHR activity, measures of burden should reflect these priorities.

B. Making sense of audit log data

A vendor respondent explained that EHRs captured different elements of data in the course of clinical use, differentiating between: the highest aggregate level of log data that captures access to records (for example, who logs in and when); the middle level that captures transactions (for example, adding information to a patient note); and the most granular level that captures mouse movement, scrolling, and the module or screen with which people interact. This description was consistent with reports from researchers and clinicians who worked with log data from a single vendor or setting.

Respondents strongly cautioned that audit log data were difficult to interpret, with one EHR vendor representative likening the "truly raw data" to "drinking from a fire hose." Many respondents utilized additional data sources to initially make sense of audit log data due to the difficulty in associating granular data with actions taken in the EHR and to the context surrounding EHR tasks that audit log data could not capture. For example, clinician researchers who compared audit log data with user observations noted that the comparison revealed divergent volumes of activities, with audit log data recording "a trail of four things" that failed to correspond to the user's mental model of completing a single task. A vendor representative shared a related concern, noting that a "black and white row" of audit log data might not be easy to understand without adding in the user's perspective, especially across products. For example, researchers must take an audit log entry related to patient diagnoses in the EHR and decide

whether it is appropriate to categorize the activity as an update to a patient's past medical history, a focus of the current visit, or some other action involving a diagnosis. While TEP members were sympathetic with the described effort to map audit log data to observable tasks, one member believed this mapping was a "one-time" exercise per vendor and did not think it would pose an insurmountable barrier.

Respondents described a range of factors that must be considered to put audit log data in context. They cautioned that specialty, setting, time in practice, EHR design, and office-staffing models influence EHR use and thus the interpretation of audit log data. As one respondent explained, a new clinician with a small patient panel may spend more time documenting per patient than a clinician with a larger panel. This variation would reflect the available time in the new clinician's schedule and not an inherent characteristic of that user's documentation style or the EHR's affordances. Indeed, a forthcoming study that used audit log data to map use of a single vendor's EHR across the country found "more variability within specialties than across specialties. Since these providers are all using the same EHR this variability must arise from other factors such as configuration differences, implementation specifics, practice configuration (e.g. how tasks are divided among the care team), individual provider choices" and other related factors (Overhage and McCallie, under review). This interpretation is bolstered by a mixed methods study that examined variation in EHR use within primary care practices, finding "substantial variation in documentation for 5 categories of clinical information which was perceived to result from optionality in the EHR design and varied implementation practices" (Cohen et al. 2019).

EHR vendor representatives and clinician researchers who used EHRs and worked with EHR data shared observations about system design issues and limitations in audit log requirements that add to the challenges of interpreting audit log data. Upgrades to vendor products could affect audit log data and challenge interpretation. In particular, upgrades could add or remove steps the end user needs to take in the EHR to complete a particular task; these steps would be reflected in the audit log data but would need to be mapped back to existing analytic frameworks to reflect, for example, that entering information in a newly available field was in fact updating the medication list and thus should be considered clinical documentation and not administrative. One researcher thought that data would be roughly comparable over time after the analytic frameworks had been updated, but was not certain at what level of fidelity. Leaders of software companies that use EHR data to enhance user experiences said that it's feasible to compare encounter-level data (such as patient demographics, lab results, and vital signs) across EHR products. But they noted wide variation in the content of audit logs, such as the approach to tracking time zones and unique patient identifiers, as well as a lack of technical guidance on decoding and interpreting the content of audit logs. Furthermore, an EHR representative cited limitations in audit log requirements, stating that the auditing functions required through ONC certification are primarily focused on what would potentially be included under disclosures. Vendors can choose what to audit above and beyond certification. For example, creating a note and signing a note are auditable events, but checking boxes or other activities performed within the note are not.

In addition, a given vendor or institution may not realize the nuance specific to their EHR system or setting and may be unaware that sharing it could aid in audit log data interpretation. For

example, one vendor representative said that their system did not allow users to edit notes after they were signed, so users often left notes unsigned for extended periods of time. The representative explained that audit log data indicating "the date the note was created and the date the note is closed will have no reflection on patient care or even how much time the provider really spent on the note." However, a TEP member added that vendor consolidation, particularly for inpatient systems, may minimize the practical significance of this issue, as it may be possible to map audit logs for fewer than 10 vendors and glean insight into most EHR users in the country. Additionally, as one TEP member noted, a measure that assesses variation of EHR use across vendors may be more meaningful than a measure that produces a central value (for example, finding more value in noting the range of time users spend documenting during a patient visit rather than focusing on the computed average).

More information is necessary to understand the effort to maintain audit log definitions over time as well as the best way to translate records of use into usage behavior that provides adequate context for understanding burden. Nonetheless, these efforts to understand the activities recorded in audit logs may yield dividends. One respondent who worked with data from different vendors believed a sufficiently flexible measure should be able to preserve meaning across the systems.

C.   Understanding EHR use and burden

We defined *burden* for respondents as "work that does not add value." This definition resonated with many respondents. One EHR medical director and practicing clinician offered a related definition that burden referred to extra steps taken in the EHR beyond what was needed to document for good medicine. However, several people from both clinical organizations and vendors noted that burden was not a binary state, but rather a continuum. One EHR medical director added that burden was contextual and would change over time as users' perspectives on value changed, giving the example that if current consent procedures changed, then pursuing informed consent the way it is commonly done now could be perceived as burdensome.

Respondents likewise cautioned that *burden* is a laden term. A clinician thought-leader elaborated: "Mixed in with burden is obligation, professional responsibility to patient and institutions, the requirement to be transparent, the responsibility to share, and to make available data that allow quality assurance and quality measurement and accountability." This same respondent added that because value does not always accrue to the individual completing documentation, but rather to society more generally, it would be important to balance the quantification of burden with the quantification of benefit from information sharing, effective clinical decision support, and other advantages of using electronic rather than paper records. A vendor representative likewise noted, "There are different value levels depending on who the stakeholder is within the system."

Moreover, respondents universally agreed that the burden popularly attributed to EHRs was often a reflection of increased regulatory and administrative requirements and not an inherent component of EHR use. According to several respondents, the changing expectations of health care delivery resulted in having to deliver more preventive care and document more information about overall health (in contrast to more narrowly addressing a presenting problem that

motivated an acute visit) and coincided with a push toward EHRs. Indeed, public comments on the Burden Report were mostly related to HIT usability and documentation, with EHR reporting the third most common area of comment. The report had one reference about the potential to leverage audit log data, but we saw few comments related to this topic. As several respondents articulated, EHRs' hard stops and other affordances created a "forcing function" to ensure that EHR users met these new expectations as well as attendant regulatory and administrative requirements. As a result, EHRs became the target of popular ire. One practicing clinician and EHR medical director explained, "If we're being asked to do more than we're humanly able to do, then it's going to be difficult to feel good about it at any point." However, several respondents noted that EHRs may have contributed to the creation of regulatory burden by fostering the perception that discrete fields and alerts about potential safety events could be smoothly built into the EHR, but few people understood how burdensome these additional requirements could be, especially in the aggregate. Nonetheless, several sources of burden associated with EHR use arose that may be valuable to measure using audit log data.

D.   Potential sources of burden associated with EHR use for measurement

1.   In-basket management

One of the most commonly identified sources of burden associated with EHR use related to the administrative use case is the volume of messages sent within the EHR, often referred to as the in-basket. Recent survey and analyses of EHR work performed by physicians in a multispecialty practice found that above-average system-generated in-basket messages per week was one of the factors most strongly associated with physicians' having burnout symptoms, intending to reduce work hours, and having poor life satisfaction (Tai-Seale et al. 2019). Indeed, many clinical respondents identified in-baskets as a source of burden. They highlighted several reasons they were "drowning" in messages that did not seem valuable and also risked crowding out messages with important safety implications.

First, clinical respondents observed that the in-basket fragmented important information and they often received results in a piecemeal fashion rather than as a batch. In one example, this problem was compounded by a system that shared preliminary test results as well as subsequent updates—one order would generate multiple in-basket items. As the practicing EMR medical director of a large health system noted, it would be better to "set workflows so the clinician receives a message with all the necessary data and can make a decision on it and reply to it in one step, instead of having to bounce it around." Second, respondents observed that workflows requiring clinicians to co-sign tasks taken by others felt burdensome. Finally, respondents observed that in-basket messages that first went to physicians should instead have been routed to other staff, which added to overall volume.

TEP members validated the significance of the burden associated with in-basket management, though they emphasized that some pain points have value and thus, not all in-basket use is burdensome. Furthermore, in-basket use is also driven by user preferences, workflow, clinical specialty, and organizational protocols, all of which affect the amount of time that physicians spend managing their in-basket. For example, one respondent reduced in-basket traffic by explicitly asking patients to refrain from sending thank-you messages to providers.

In-basket management is more likely to materialize when completing administrative tasks, such as messaging or tasks performed in response to actions of others involved in the care of the same patient. Although not explicitly addressed by respondents, burden associated with in-basket management is likely more common among the following clinicians: those who place orders; those who work on large teams (and thus have the potential to be included on messages from a larger number of teammates); those who supervise trainees (and thus have more orders and notes to cosign); and those with large panels or a large number of complex patients. In one study that quantified the alerts received by three physician practices in Texas, primary care practitioners received more than twice the amount of notifications than specialists (Murphy et al. 2016). In addition, this volume is likely to grow with the increasing adoption of patient portals and electronic information exchange (Friedberg et al. 2013).

Audit log data may be useful for measuring in-basket burden by capturing time spent interacting with the in-basket as well as the number of messages a clinician sends or receives. Several respondents described that measures of such activity were available either through their EHR vendor or their health system. However, it was not clear the extent to which the measures captured the time spent directly interacting with in-basket management and the related activities users complete to support their in-basket management (for example, navigating to the medication list or lab results before returning to the in-basket to complete the message). Although this may be possible to capture algorithmically, it nonetheless adds assumptions to the interpretation of the measure. Furthermore, respondents did not offer guidelines for determining an appropriate or burdensome amount of time spent interacting with the in-basket or an appropriate or burdensome number of messages. Instead, they used such measures benchmarked within their organization to identify individual users who were outliers in the time they spent or volume of messages they sent/received and thus might benefit from assistance.

TEP members provided several suggestions for moving forward with audit-log-derived measures of burden related to in-basket use. As a first step, one TEP member suggested it would be valuable to develop a national strategy for broadly measuring in-basket time and volume consistently across settings, noting that efforts to parse value from non-value-added time could follow as a next step. Echoing the need for subsequent work to aid interpretation of this measure, one TEP member added that there are factors external to the EHR—such as faxes and paper referrals—that contribute to burden. A practice may look very efficient due to low in-basket volume, even though it is "drowning in inefficient paper processing." This would be especially likely if, for example, the practice does not use its EHR for outbound referrals, and therefore may look artificially efficient due to low in-basket management.

Several other TEP members instead suggested it would be valuable to start with narrow cases where the rationale for classifying work as burdensome was more obvious, such as measures related to the composition of messages that are not opened, are opened for less than one second, or are opened but not acted on. Measures could describe the number or proportion of these low-value messages, or could describe time spent composing such messages. However, it may be challenging to determine the window of time for those subsequent actions to take place. For example, a message that does not immediately trigger a subsequent action when opened may become valuable in the future if a patient returns for a related reason. TEP members also

suggested measures of the number and types of interruptions that affect task completion, noting that in-basket burden might be disrupting work.

Table IV.1. Potential measure concepts regarding In-basket management

	In-basket management
Representative Quote	"If you have too many messages, or the messages go too broadly, you end up with people drowning in in-Baskets, and they literally can't keep up. You take something that was intended to help people communicate, and you make it impossible." – Health System CMIO
Broad Measure Concepts	• Time spent interacting with in-basket (send or read/receive) • Volume of messages (sent or received)
Narrow Measure Concepts	• Time spent authoring messages or proportion of authored messages that are not opened, or are opened for less than one second • Time spent authoring messages or proportion of authored messages that do not result in subsequent activity • Proportion or volume of messages auto-generated by the EHR • Number of messages requesting co-signature for an activity performed by another licensed clinician
Considerations	• Doesn't account for messaging outside the EHR (fax, paper referrals) • Also driven by user preferences, workflow and staffing (including team-based care), clinical specialty, and organizational protocols

2. Volume of time spent using the EHR

Respondents believed overall time spent using the EHR would approximate burden, noting the intuitive appeal of the time concept, the relative ease of measuring documentation time through audit log data, and the published research demonstrating that more time spent documenting was associated with clinician burnout. Furthermore, two respondents described EHRs as the source of this burden by enabling users to quickly copy and paste (or otherwise pull forward) past information, contributing to excessive note length. One respondent described this behavior as "irresponsible transferring of work to somebody else," which was not as easy prior to the diffusion of EHRs. However, as noted earlier, the CY 2019 Medicare PFS final rule encourages pulling information forward to reduce burden from re-documenting information already present in the medical record.

A forthcoming study of 155,000 ambulatory adult medical specialty providers found that, of the 13 minutes and 20 seconds users spent using the EHR at every patient visit, users on average spent 3 minutes and 36 seconds (27 percent) on documentation—this was the second most common activity behind chart review, which averages 4 minutes and 27 seconds (33 percent) per visit (Overhage and McCallie, under review). Respondents observed that specialists, particularly those who primarily conduct procedures or have relatively homogenous patient panels, likely didn't experience this dimension of EHR burden as acutely as primary care physicians (including pediatricians and obstetrician-gynecologists) whose diverse patient panels require longer, complex notes that are less amenable to templates and order sets. Respondents did not explicitly identify tasks that would be more likely to induce burden from spending too much time in the EHR.

Respondents suggested that time spent documenting outside the clinic could be a particularly meaningful measure of burden. They sometimes referred to this concept as "pajama time" because it often occurs in the evening or as "work outside of work" because it occurs outside of clinic hours. While this concept also had intuitive appeal and seemed feasible to measure in audit log data, respondents raised operational and theoretical concerns. Operationally, one vendor respondent described several competing definitions of after-hours documentation, including time spent documenting on days without appointments, after 6 or 7 p.m., or more than 30 minutes before or after a scheduled visit. Theoretically, one respondent cautioned that while documentation after hours may be burdensome, it was not unique to the EHR, as clinicians commonly took paper charts home to complete documentation in the evenings and weekends before EHRs emerged. Moreover, time spent documenting or otherwise using the EHR may be due to variation in personal approaches to the work, such as executing the briefest note that captures the encounter versus the most comprehensive note that summarizes relevant content located elsewhere within the medical record. Likewise, a vendor representative noted "some people like to work after hours. It is a measurable event, but it would be difficult to make any conclusions without assumptions."

Respondents from both clinical and vendor backgrounds noted a limitation of relying on audit log data for measuring time spent documenting or documenting after hours, which assumes that users are engaged the entirety of the time they are logged in. Related specifically to pajama time documentation, one chief medical information officer highlighted the complexity of measuring time spent actively documenting in the evening compared to time spent logged in to the EHR in the evening—because clinicians often completed documentation while watching television, they might feel as if they had spent a long time documenting, but in reality, between commercials there were 10- to 15-minute lulls that should not be considered time spent documenting. A TEP member described this as a "wall time" phenomenon, meaning that the times displayed on a wall clock at the beginning and end of a session might be too generous for calculating time worked. Conversely, periods of seeming idleness could, in some instances, reflect time spent processing information and would not necessarily reflect time spent engaged in tasks outside the EHR. Researchers suggested that measurement could mark inactive periods using a cap of 30 to 90 seconds of idle time. A health system medical director noted that too lax a cap could make tasks look artificially long, whereas too strict a cap could under-measure burden from complex tasks that don't involve a lot of activity that would register in audit log data. To help balance these competing concerns, one vendor developed a "two-tiered categorization" system that considers users active if they are logged in to the EHR and their activity records are shorter than 45 seconds apart, or if they spend longer than 45 seconds on an activity but complete three or more mouse clicks per minute, mouse movement of 1,700 pixels or more per minute, or 15 or greater keystrokes per minute (Overhage and McCallie, under review).

Furthermore, as with measures of in-basket messaging, it may be difficult to determine an appropriate amount of time to spend in the EHR as compared to a burdensome amount. For example, one health system EHR medical director described an "efficiency score" that their vendor provided to describe time spent in the EHR to complete certain tasks. This respondent noted that the score portrayed him as inefficient because he spent time in the EHR troubleshooting problems that didn't result in completed patient care tasks. In contrast, he noted

that a more senior clinician in their practice who cosigned multiple orders at a time looked efficient, when that clinician was actually being credited for others' work. A measure of burden that relies on time spent in the EHR might be limited by similar difficulties. Although it may be meaningful to measure time spent using the EHR to determine if efforts to reduce regulatory burden are successful over time, this kind of overall measure could mask changes in *how* clinicians spend time using the EHR and limit its utility.

TEP members described the interpretation of a broad measure of time spent in the EHR as a critical issue, noting that spending the least amount of time possible in the EHR wouldn't necessarily correspond to best practices. They were likewise concerned that, although it is generally considered good practice to finish documenting during a visit, some providers prefer to document later and thus after hours documentation is not necessarily problematic. They also strongly cautioned that, although it would be meaningful to calculate time spent documenting after hours, it can be difficult to get scheduling data from practices. This could make it challenging to determine how much work occurred outside scheduled hours. However, broad measures may gain meaning longitudinally as a way to capture the effect of changing federal or organizational policies.

Narrower measures of time spent using the EHR may approximate burden by comparing the amount of time that users in the same specialty and practice are logged in to the EHR (normalized for the number of patients) to identify outliers. This could create a binary indicator of burden by flagging those whose experiences are outside the norm. However, TEP members were not enthusiastic about this concept, believing it required too many assumptions regarding documentation preferences. For example, a physician who preferred to invest more time than their peers when documenting an initial patient visit might look burdened according to this measure.

Table IV.2. Potential measure concepts regarding volume of time spent using the EHR

Volume of time spent using the EHR	
Representative Quote	"There's too much information – it's a needle in the haystack problem... If you don't know how to use the tools well to find things, you'll spend a lot of time manually scrolling through a chart trying to find a piece of information." – Health System EHR Medical Director
Broad Measure Concepts	• Time spent actively using the EHR • Time spent actively using the EHR before or after clinical visits
Narrow Measure Concepts	• Proportion of clinicians whose time spent on documentation for an initial patient encounter is "X" percent greater or lower than the mean time for visit documentation, among their same-specialty peers, adjusted for patient characteristics
Considerations	• Difficult to account for different practice styles and preferences (for example, choosing to document in the evening) • Distinguishing active and inactive time similarly difficult • May require linking to other data sources, including administrative data • Also driven by regulatory requirements for documentation

3.   Usability of the EHR interface

The final commonly suggested source of burden associated with EHR use relates to usability of the EHR interface—poor usability results in too many necessary steps to complete a task in the EHR, or external requirements for additional (often structured) documentation. Focusing on EHR workflow limitations and reporting burden in the context of medication reconciliation, an EHR medical director and practicing clinician explained that while medication reconciliation is time consuming, it is clinically important and therefore not burdensome. In contrast, he noted that having to click a button to confirm that medications were reconciled in order to record the task for reporting purposes added to the work without adding value. This respondent gave another example of burdensome EHR workflow that was created expressly for generating measures: the practice had to interrupt an automated workflow to give providers an alert so that an acknowledging click could populate a numerator for an opioid-prescribing measure.

Almost all respondents commented on limitations of EHR design that contribute to burden. They reported system rigidity in mandating certain workflows (for example, requiring formal orders for delegated tasks), as well as fragmenting information into discrete points that shift documentation from a qualitative narrative assessment (for example, as to why the patient is being seen) into structured coded data. Forcing structured coded data entry contributed to burden in several ways, including increasing click count, complicating data entry, and fragmenting information in a way that respondents found difficult to review and piece together. Although one respondent believed that younger generations of clinicians did not seem as burdened as older generations by "click anaphylaxis," several others cited it as an independent example of burden.

One chief medical information officer conversely suggested that it takes longer to update and manage structured areas of the EHR (such as problem lists and medication lists) than it does to dictate notes. He noted that areas that would purportedly have broad benefit are not "populated or maintained in a meaningful way, so you end up not having information where you need it or want it, and can't really find it."

Respondents were confident that back-end log data captured the number of clicks per visit, but TEP members cautioned that these metadata were not typically captured as audit log data and thus could not be reliably used to make comparisons across vendors. Furthermore, it was not obvious to respondents how to differentiate appropriate from inappropriate click volume. Broad measures related to usability of the EHR interface, such as the average number of screens or tabs accessed per visit, may gain meaning longitudinally as a way to capture the effect of changing federal or organizational policies. Alternatively, narrower measures of screen navigation may yield early insight into burden associated with usability by focusing on an important task, such as the number of screens required to retrieve a patient's information or to over-ride alerts when ordering a medication. One thought leader suggested leveraging methods outside of audit log data, such as a system usability testing lab, to better measure the number of clicks required to complete an important task.

A potential additional limitation of measures derived from audit log data relates to the role of voice-recognition text entry, which may be captured differently across vendor systems, as well as

keyboard shortcut alternatives to mouse clicks (Ratwani et al. 2018). As an EHR vendor representative noted, something that seems intuitive, such as how long it takes to complete a note of a certain length, is influenced by how the note was completed (for example, typing, pulled forward from a prior note, copied and pasted, or entered through voice recognition), such that it may not be appropriate to compare values across the different methods for note completion. Additionally, as a TEP member added, click data are so granular that they may not be as readily available. Because of the volume of data, for example, they may be stored at an aggregated level such as how many clicks it takes to complete a task instead of storing one observation per click. Additionally if vendors define tasks differently, the data would not be comparable across vendors.

Table IV.3. Potential measure concepts regarding usability of the EHR interface

	Usability of the EHR interface
Representative Quote	"There are a whole lot of clicks that go on for test ordering, medication ordering, for ordering of home care, hospitalizations, etc. Those things might have been done verbally in the past. There's additional time and effort required of physicians and other clinicians in that regard." – Thought Leader
Broad Measure Concepts	• Number of screens/tabs per visit
Narrow Measure Concepts	• Number of screens to complete an important task, such as documenting physical exam findings or renewing a prescription
Considerations	• Difficult to capture and compare different data entry methods (for example, manual entry, voice-to-text, or pulling forward prior notes) • Data with sufficient granularity not available and may be difficult to analyze • Also driven by user documentation and workflow preferences, and patient complexity

V. LIMITATIONS

There are inherent limitations that should be noted when interpreting these results. First, our literature search was not systematic, and the fields of both EHR burden and audit log data-based research are evolving rapidly. Therefore, we may have missed articles that could have enhanced our understanding on the topic. Although we guarded against this possibility by soliciting article recommendations from experts and mining references in articles we found valuable, it is possible that we missed analyses of burden that could highlight useful dimensions for measurement. Furthermore, it is possible that we missed examples of how audit log data were used for research that could be applied to dimensions of burden.

In addition, although we spoke with nationally recognized experts about EHR use and audit log data, we spoke with only a small number of people with expertise in any particular area. For example, many of our respondents familiar with audit log data based their knowledge on the same commercial vendor, As a result, the strengths and limitations of that particular vendor's setup may feature more prominently in our results than those of other vendors. We supplemented these discussions with representatives from middleware vendors that work with data from multiple commercial EHRs, but their work does not focus on burden reduction or measure development. In short, this expertise is no substitute for direct work with audit log data to construct measures of burden. Likewise, although many of our respondents were based in clinical organizations and continued to see patients, none had a full-time medical practice, which likely influenced their understanding of burden.

Finally, this work has not explicitly considered potential unintended consequences from focusing on clinician burden reduction and efficiency. First, it is possible that efforts that focus on clinician workflow will shift work to other staff, rather than triggering broader reevaluation of EHR usability and documentation requirements. Second, the history of HIT adoption in the United States is rife with examples of unintentional patient harm (for example, Han et al. 2005; Fry and Schulte 2019). Loosening requirements related to co-signing orders, or other changes intended to reduce burden, could similarly backfire and compromise patient safety. Given that the act of measuring a phenomenon can change user behavior by signaling certain priorities or preferences (Hawthorne 1974), patient safety issues should be considered before implementing measures that may influence EHR use.

This page has been left blank for double-sided copying.

VI. DISCUSSION

Now is an optimal time to evaluate the feasibility of leveraging audit log data to measure EHR burden, both because of widespread attention on burden in federal and private initiatives and because of the increasing prevalence of research based on audit log data. Recent studies demonstrated the feasibility and validity of using log data to measure time spent documenting, variation in documentation, and other dimensions of EHR use. Through a review of gray and peer-reviewed literature, semi-structured interviews with SMEs, and consultation with a TEP, we identified three potential sources of burden related to EHR use. These sources are important to focus on for future measure development with audit log data: in-basket management, volume of time spent using the EHR (particularly outside work hours), and usability of the EHR interface. Our findings aligned with the conclusions of the ONC Burden Report—complying with documentation and reporting requirements contributes to burden along with the facilitation of copying and pasting that increases the amount of information for clinicians to review. Respondents also echoed the Burden Report's focus on documentation after work hours, though respondents converged on in-basket management as a compelling manifestation of burden in contrast to the Burden Report's focus on pop-up alerts.

However, despite several examples of how audit log data had been leveraged to understand EHR use in specific institutions and settings, respondents were skeptical about using such data as the basis of a national burden measurement strategy. Respondents were particularly concerned about the effort required to translate granular audit log data into measures and cautioned that data were not directly comparable across vendors, at least in part because the prevailing standards lacked specificity and relied on attestation. Future consideration may be given to updating the testing requirements in ONC's EHR certification process to include submission of audit log data, demonstrating how logs produced by EHR systems line up with the capabilities described in the ASTM standard.

Respondents also noted that audit log data did not capture practice or patient context that was necessary to differentiate burden from work that adds value. Variation in patterns of use within vendors reinforces the importance of considering configuration choices at a local level. Although CMS and ONC have data on some contextual variables, such as EHR systems in use and number of billing providers in a practice, these data sources are not exhaustive and would require additional effort to link to audit log data.

Given the nascent stage of quantitating burden through measures, TEP members believed it may be useful to leverage audit log data for broad measures of EHR use as building blocks to future measures that more directly capture burden; narrow measures of EHR use derived from audit log data, where the case for burden can be made without as much interpretation; or both types of measures. Although these broad measures of EHR use and narrow measures of burden would not fully encompass the complexity of burden, they could nonetheless inform organizational workflow, federal policy, and/or vendor design changes to reduce burden, and could also provide earlier indicators of the results of such interventions.

Remaining challenges relate to defining burden and operationalizing the concept in a way that is conducive to measurement, both because the concept of burden is itself complex and subjective, and because the valuable work is often entwined with the burdensome work in such a way that it is not always apparent to clinicians which tasks are actually valuable and which are purely burdensome. Likewise, SMEs indicated that there would be high startup costs to make sense of audit log data, as well as some barriers to access, that should be more specifically examined in future research.

A.   Considerations for continued progress

This report described several potential sources of burden associated with EHR use, where it may be worthwhile to pursue measures based in audit log data, as well as their potential strengths and limitations. Because developing measures can require significant time and resources, it is important to establish a process to prioritize different potential avenues for measure development. We recommend modeling this process on a recent measurement project focused on the safe use of health IT because it is a similarly nascent area for measurement. Specifically, the NQF convened a multi-stakeholder "HIT Safety Committee" (the Committee) to identify the highest priority IT-related patient safety areas to measure, and to prioritize measures within patient safety areas to be developed through an iterative process (NQF 2016).

In order to prioritize potential health IT safety measures, the Committee focused on two criteria: importance and feasibility. Importance, in the context of considering potential measures of burden, would focus on identifying (1) the scope of the type of burden being assessed (for example, number of users in target population); (2) the strength of evidence that supports the source of burden associated with EHR use; and (3) the likelihood that measuring the issue could drive change in an organization or individual behavior ("actionability"). The Committee considered feasibility as the availability and ease of capturing data consistently across measured entities. In the context of prioritizing potential measures of burden, feasibility should likely reflect the accessibility of data in logs, and the occasional need to integrate other system-level data. An early sense of the various data sources potentially required helps gauge technical scope and level of difficulty. TEP members supported the use of importance and feasibility as criteria for prioritizing measures of burden.

The TEP discussed additional considerations for prioritizing types of burden. First, they suggested evaluating the strength of the association between the measured activity and outcomes on burden. Second, the TEP felt it would be useful to consider the reliability of the measure (that is, is the interpretation ambiguous or does it have the same meaning across sites?). Third, they suggested considering validity before rolling out any measures nationwide (that is, confidence that the measure is assessing burden and not some other construct). The TEP noted that assessing validity is most applicable when prioritizing measures where the measured activity is very likely to be attributed to burden.

In future studies, researchers should more systematically categorize important tasks that clinicians conduct in the EHR and develop a framework for understanding how these tasks relate to each other. Providers, policymakers, and other stakeholders should also develop goals

regarding how these tasks should manifest in the majority of patient care. For example, in the context of in-basket management, convening a panel to discuss the amount or proportion of time that providers should spend writing, reading, and responding to messages, would better facilitate identification of burden.

A subset of these tasks could then be used as the basis for future work with audit log data across vendors to construct more robust measures of burden. Comparing EHR audit log data across vendors and specifically attempting to construct measures could serve as a proof of concept, beyond the broad and narrow measures outlined in this report, and inform the necessary enhancements to existing standards to increase the uniformity of audit log data. It may also be valuable to survey EHR and other health IT vendors regarding audit and other log data and metadata to better understand their capabilities, especially after comparing audit log data across vendors to identify areas of similarity and areas of difference that are important to catalog. Finally, future research should explore how to categorize the different components of the EHR interface that impact burden as well as the best way to capture interface variation across vendors.

Several nongovernmental organizations, such as the EHR Association and the National Research Network for EHR Audit-Log and Meta-Data, are working internally on these issues (for example, on understanding audit log data across vendors). But it may be useful to formalize such efforts through a multi-stakeholder project on audit log measurement, convened by an entity such as NQF, ONC, or ASPE. At HHS's request, NQF gathered experts and engaged the public in projects to develop conceptual frameworks and to draft measure concepts for high-priority areas such as interoperability and patient safety. Similarly, a committee or project could be established with a specific focus on burden.

Policy changes should also support the use of audit logs for measuring burden. One TEP member suggested that ONC consider developing an implementation guide that would accompany audit log certification criteria. Such a guide would encourage a more standard approach to constructing audit logs and improve comparability of audit logs across EHR vendors. Further, ONC could modify the process for certifying an EHR's compliance with audit log criteria, which is currently achieved by attestation. As part of the testing process for audit logs, a vendor could be asked to perform a set of common activities (e.g., document blood pressure or initiate an alert for a drug-drug contraindication) and require testing laboratories to collect applicable logs for future review by ONC.

This page has been left blank for double-sided copying.

REFERENCES

Adler-Milstein, J., and R.S. Huckman. "The Impact of Electronic Health Record Use on Physician Productivity." *American Journal of Managed Care*, vol. 19, no. SP10, 2013, pp. SP345–SP352. Available at https://www.ncbi.nlm.nih.gov/pubmed/24511889. Accessed October 10, 2018.

Adler-Milstein, J., and A. Jha. "HITECH Act Drove Large Gains in Hospital Electronic Health Record Adoption". *Health Affairs*, vol. 36, no. 8, 2017, pp. 1416-1422. Available at https://www.healthaffairs.org/doi/pdf/10.1377/hlthaff.2016.1651. Accessed March 7, 2019.

Ancker, J.S., A. Edwards, S. Nosal, D. Hauser, E. Mauer, and R. Kaushal. "Effects of Workload, Work Complexity, and Repeated Alerts on Alert Fatigue in a Clinical Decision Support System." *BMC Medical Informatics and Decision Making*, vol. 17, no. 1, 2017, pp. 1–9. Available at https://doi.org/10.1186/s12911-017-0430-8. Accessed October 10, 2018.

Ancker, J.S., L.M. Kern, A. Edwards, S. Nosal, D.M. Stein, D. Hauser, and R. Kaushal. "How Is the Electronic Health Record Being Used? Use of EHR Data to Assess Physician-Level Variability in Technology Use." *Journal of the American Medical Informatics Association*, vol. 21, no. 6, 2014, pp. 1001–1008. Available at https://doi.org/10.1136/amiajnl-2013-002627. Accessed October 10, 2018.

Arndt, B.G., J.W. Beasley, M.D. Watkinson, J.L. Temte, W-J Tuan, C.A. Sinsky, and V.J. Gilchrist. "Tethered to the EHR: Primary Care Physician Workload Assessment Using EHR Event Log Data and Time-Motion Observations." *Annals of Family Medicine*, vol. 1515, no. 5, 2017, pp. 419–426. Available at https://doi.org/10.1370/afm.2121. Accessed October 10, 2018.

ASTM International. "ASTM E2147-01. Standard Specification for Audit and Disclosure Logs for Use in Health Information Systems." West Conshohocken, PA: ASTM International, 2009. Available at https://www.astm.org/DATABASE.CART/HISTORICAL/E2147-01.htm. Accessed June 26, 2019.

ASTM International. "ASTM E2147-18. Standard Specification for Audit and Disclosure Logs for Use in Health Information Systems." West Conshohocken, PA: ASTM International, 2018. Available at https://www.astm.org/Standards/E2147.htm. Accessed June 20, 2019.

Ballermann, M.A., N.T. Shaw, D.C. Mayes, R.T.N. Gibney, and J.I. Westbrook. "Validation of the Work Observation Method by Activity Timing (WOMBAT) Method of Conducting Time-Motion Observations in Critical Care Settings: An Observational Study." *BMC Medical Informatics and Decision Making*, vol. 11, 2011, p. 32. Available at https://doi.org/10.1186/1472-6947-11-32. Accessed October 10, 2018.

Burchell, L. Allscripts Public Comments on ONC Draft Strategy Report. January 24, 2019. Available at https://www.healthit.gov/sites/default/files/webform/reducing_burden_report/allscripts-reducing-regulatory-and-admin-burden-response-jan-2019.pdf. Accessed April 5, 2019

Campanella, P., E. Lovato, C. Marone, L. Fallacara, A. Mancuso, W. Ricciardi, and M. L. Specchia. "The Impact of Electronic Health Records on Healthcare Quality: A Systematic Review and Meta-analysis." *European Journal of Public Health*, vol. 26, no. 1, 2015, pp. 60–64. Available at https://doi.org/10.1093/eurpub/ckv122. Accessed February 11, 2019.

Centers for Medicare & Medicaid Services (CMS). "Eligible Professional Meaningful Use Core Measures: Measure 9 of 17." Washington, DC: CMS, 2014.

Centers for Medicare & Medicaid Services (CMS). "October 2018 Monthly Payment and Registration Summary Report." Baltimore, MD: CMS, 2018. Available at https://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/DataAndReports.html. Accessed February 4, 2019.

Chen, L., U. Guo, L. C. Illipparambil, M. D. Netherton, B. Sheshadri, E. Karu, S. J. Peterson, and P. H. Mehta. "Racing Against the Clock: Internal Medicine Residents' Time Spent on Electronic Health Records." *Journal of Graduate Medical Education*, vol. 8, no. 1, 2016, pp. 39–44. Available at https://doi.org/10.4300/JGME-D-15-00240.1. Accessed November 27, 2018.

Cohen, G. R., C. P. Friedman, A. M. Ryan, C. R. Richardson, and J. Adler-Milstein. "Variation in Physicians' Electronic Health Record Documentation and Potential Patient Harm from That Variation." *Journal of General Internal Medicine*, 2019, pp. 1–13. Available at https://doi.org/10.1007/s11606-019-05025-3. Accessed June 26, 2019.

Cothran, C., and E. Reilly. "Audit Log Demands During Litigation: Response Conundrums from a Compliance Perspective." Marietta, GA: WellStar Health System, 2017.

DesRoches, C., K. Donelan, P. Buerhaus, and L. Zhonghe. "Registered Nurses' Use of Electronic Health Records: Findings from a National Survey." *Medscape Journal of Medicine*, vol. 10, no. 7, 2008, p. 164. Available at http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=18769691. Accessed October 10, 2018.

Dyrbye, L. N., T. D. Shanafelt, C. A. Sinsky, P. F. Cipriano, J. Bhatt, A. Ommaya, C. P. West, and D. Meyers. "Burnout Among Health Care Professionals: A Call to Explore and Address This Underrecognized Threat to Safe, High-Quality Care." *NAM Perspectives*, vol. 7, no. 7, 2017, pp. 1–11. Available at https://doi.org/10.31478/201707b. Accessed February 13, 2019.

Friedberg, M. W., P. G. Chen, K. R. Van Busum, F. Aunon, C. Pham, J. P. Caloyeras, et al. "Factors Affecting Physician Professional Satisfaction and Their Implications for Patient Care, Health Systems, and Health Policy." Santa Monica, CA: RAND Corporation, 2013. Available at https://doi.org/10.1214/07-EJS057. Accessed October 10, 2018.

Gardner, R. L., E. Cooper, J. Haskell, D. A. Harris, S. Poplau, P. J. Kroth, and M. Linzer. "Physician Stress and Burnout: The Impact of Health Information Technology." *Journal of the American Medical Informatics Association*, vol. 26, no. 2, 2018, pp. 1–9. Available at https://doi.org/10.1093/jamia/ocy145. Accessed December 10, 2018.

Gawande, A. "Why Doctors Hate Their Computers." *The New Yorker*, November 12, 2018, pp. 1–24. Available at https://www.newyorker.com/magazine/2018/11/12/why-doctors-hate-their-computers. Accessed November 7, 2018.

Han, Y. Y., J. A. Carcillo, S. T. Venktaraman, R. S. Clark, R. S. Watson, T. C. Nguyen, H. Bayir, and R. A. Orr. "Unexpected Increased Mortality After Implementation of a Commercially Sold Computerized Physician Order Entry System." *Pediatrics*, 2005. Available at https://www.ncbi.nlm.nih.gov/pubmed/16322178. Accessed March 24, 2019.

Hribar, M. R., S. Read-Brown, I. H. Goldstein, L. G. Reznick, L. Lombardi, M. Parikh, et al. "Secondary Use of Electronic Health Record Data for Clinical Workflow Analysis." *Journal of the American Medical Informatics Association*, vol. 25, no. 1, 2018, pp. 40–46. Available at https://doi.org/10.1093/jamia/ocx098. Accessed November 19, 2018.

Integrating the Healthcare Enterprise (IHE). "IHE IT Infrastructure (ITI) Technical Framework—Audit Trail and Node Authentication (ATNA) Profile." *IHE*, vol.1, section 9, 2019. Available at http://www.ihe.net/uploadedFiles/Documents/ITI/IHE_ITI_TF_Vol1.pdf#nameddest=9__Audit_Trail_and_Node_Authent. Accessed August 14, 2019.

International Organization for Standardization. "ISO 27789:2013. Health Informatics—Audit Trails for Electronic Health Records." Geneva, Switzerland: International Organization for Standardization, 2019. Available at https://www.iso.org/standard/44315.html. Accessed February 14, 2019.

Jamoom, E. W., D. Heisey-Grove, N. Yan, and P. Scanlon. "Physician Opinions About EHR Use by EHR Experience and by Whether the Practice Had Optimized Its EHR Use." *Journal of Health and Medical Informatics*, vol. 4, no. 11, 2017, pp.1–19. Available at https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5084912/. Accessed October 10, 2018.

Jha, A., and A. Iliff. "A Crisis in Health Care: A Call to Action on Physician Burnout." Waltham, MA: Massachusetts Medical Society, 2019. Available at http://www.massmed.org/News-and-Publications/MMS-News-Releases/Physician-Burnout-Report-2018/. Accessed March 25, 2019.

Jones, S. S., R. S. Rudin, T. Perry, and P. G. Shekelle. "Health Information Technology: An Updated Systematic Review with a Focus on Meaningful Use." *Annals of Internal Medicine,* vol.160, no. 1, 2014, pp. 48–54.

Kannampallil, T., C. Denton, J. Shapiro, and V. Patel. "Efficiency of Emergency Physicians: Insights from an Observational Study Using EHR Log Files." *Applied Clinical Informatics,* vol. 9, no. 1, 2017, pp. 99–104. Available at https://doi.org/10.1055/s-0037-1621705. Accessed October 20, 2018.

Lee, P. V., D. Berwick, and C. A. Sinsky. "Building Trust Between the Government and Clinicians: Person to Person and Organization to Organization." *JAMA*, vol. 321, no. 18, pp. 1763–1764. doi:10.1001/jama.2019.4499.

Mennemeyer, S. T., N. Menachemi, S. Rahurkar, E. W. Ford. "Impact of the HITECH Act on Physicians' Adoption of Electronic Health Records." *Journal of Medical Informatics Association,* vol. 23, no. 2, 2016, pp. 375–379. Available at https://academic.oup.com/jamia/article/23/2/375/2572425. Accessed March 7, 2019.

Murphy, D. R., A. N. D. Meyer, E. Russo, D. F. Sitting, L. Wei, and H. Singh. "The Burden of Inbox Notifications in Commercial Electronic Health Records." *JAMA Internal Medicine*, vol. 176, no. 4, 2016, pp. 559–560. Available at https://doi.org/10.2144/000114329. Accessed November 19, 2018.

National Center for Health Statistics (NCHS). "2017 National Electronic Health Record Survey (NEHRS)." Available at https://www.cdc.gov/nchs/ahcd/ahcd_products.htm. Accessed July 15, 2019.

National Quality Forum (NQF). "A Measure Framework to Assess Nationwide Progress Related to Interoperable Health Information Exchange to Support the National Quality Strategy." Washington, DC: NQF, 2017.

National Quality Forum (NQF). "Identification and Prioritization of Heath IT Patient Safety Measures." Washington, DC: NQF, 2016.

Office of the National Coordinator for Health Information Technology (ONC). "21st Century Cures Act: Interoperability, Information Blocking, and the ONC Health IT Certification Program. 45 C.F.R. Parts 170 and 171." Washington, DC: ONC, n.d. Available at https://www.healthit.gov/sites/default/files/nprm/ONCCuresActNPRM.pdf. Accessed June 20, 2019.

Office of the National Coordinator for Health Information Technology (ONC). "2015 Edition §170.315(d)(3) Audit Report(s)." Updated 2017. Washington, DC: ONC, 2017. Available at https://www.healthit.gov/test-method/audit-reports#test_procedure. Accessed June 20, 2019.

Office of the National Coordinator for Health Information Technology (ONC). "2015 Edition Health Information Technology Certification Criteria. 45 C.F.R. Part 170." Washington, DC: ONC, 2015. Available at https://www.regulations.gov/document?D=HHS_FRDOC_0001-0602. Accessed June 24, 2019.

Office of the National Coordinator for Health Information Technology (ONC). "Health IT Data Summaries – EHR Adoption." Washington, DC: ONC, 2017. Available at https://dashboard.healthit.gov/apps/health-information-technology-data-summaries.php. Accessed July 15, 2019.

Office of the National Coordinator for Health Information Technology (ONC). "Strategy on Reducing Regulatory and Administrative Burden Relating to the Use of Health IT and EHRs." Washington, DC: ONC, 2018. Available at https://www.healthit.gov/topic/usability-and-provider-burden/strategy-reducing-burden-relating-use-health-it-and-ehrs. Accessed January 2, 2019.

Overhage, M., and D. McCallie. "A Cross-Sectional, Nationwide, Multi-Institutional, Study of Ambulatory Medical Physician EMR Time Use." 2019 Under review.

Parsons, H. "What happened at Hawthorne." *Science*, vol. 183, no. 4128, 1974, pp. 922–932. Available at https://science.sciencemag.org/content/183/4128/922. Accessed February 14, 2019.

Payne, T. H., S. Corley, T. A. Cullen, T. K. Gandhi, L. Harrington, G. J. Kuperman, et al. "Report of the AMIA EHR-2020 Task Force on the Status and Future Direction of EHRs." *Journal of the American Medical Informatics Association*, vol. 22, no. 5, 2015, pp. 1102–1110. Available at https://doi.org/10.1093/jamia/ocv066. Accessed October 10, 2018.

Ratwani, R. M., E. Savage, A. Will, R. Arnold, S. Khairat, K. Miller, R. J. Fairbanks, M. Hodgkins, and A. Z. Hettinger. "A Usability and Safety Analysis of Electronic Health Records: A Multi-Center Study." *Journal of the American Medical Informatics Association,* vol. 25, no. 9, 2019, pp. 1197–1201. Available at https://doi.org/10.1093/jamia/ocy088. Accessed March 24, 2019.

Schulte, F., and E. Fry. "Death by a Thousand Clicks: Where Electronic Health Records Went Wrong." *Kaiser Health News*, 2019. Available at https://khn.org/news/death-by-a-thousand-clicks/. Accessed April 5, 2019.

Shanafelt, T. D., L. N. Dyrbye, C. Sinsky, O. Hasan, D. Satele, J. Sloan, and C. P. West. "Relationship Between Clerical Burden and Characteristics of the Electronic Environment with Physician Burnout and Professional Satisfaction." *Mayo Clinic Proceedings,* vol. 91, no. 7, 2016, pp. 836–848. Available at https://doi.org/10.1016/j.mayocp.2016.05.007. Accessed October 10, 2018.

Sinsky, C., L. Colligan, L. Li, M. Prgomet, S. Reynolds, L. Goeders, et al. "Allocation of Physician Time in Ambulatory Practice: A Time and Motion Study in 4 Specialties." *Annals of Internal Medicine,* vol. 165, no. 11, 2016, pp. 753–760. Available at https://doi.org/10.7326/M16-0961. Accessed October 10, 2018.

Sittig, D. F. "Audit Logs. Medical Liability and Health Care Law." Houston, TX: University of Texas Health Science Center, 2017, pp. 1–11. Available at https://www.researchgate.net/profile/Dean_Sittig/publication/314280662_Audit_Logs/links/58bf00e592851cd83aa12538/Audit-Logs.pdf?origin=publication_detail. Accessed February 4, 2019.

Sockolow, P. S., J. P. Weiner, K. H. Bowles, and H. P. Lehmann. "A New Instrument for Measuring Clinician Satisfaction with Electronic Health Records." *CIN – Computers Informatics Nursing,* vol. 29, no. 10, 2011, pp. 574–585. Available at https://doi.org/10.1097/NCN.0b013e31821a1568. Accessed October 10, 2018.

Tai-Seale, M., E. C. Dillon, Y. Yang, R. Nordgren, R. L. Steinberg, T. Nauenberg, et al. "Physicians' Well-Being Linked To In-Basket Messages Generated By Algorithms In Electronic Health Records." *Health Affairs,* vol. 38, no. 7, 2019, pp. 1073–1078. Available at https://www.healthaffairs.org/doi/full/10.1377/hlthaff.2018.05509. Accessed July 2, 2019.

Tai-Seale, M., C. W. Olson, J. Li, A. S. Chan, C. Morikawa, M. Durbin, W. Wang, and H. S. Luft. "Electronic Health Record Logs Indicate That Physicians Split Time Evenly Between Seeing Patients and Desktop Medicine." *Health Affairs,* vol. 36, no. 4, 2017, pp. 655–662. Available at https://www.ncbi.nlm.nih.gov/pubmed/28373331. Accessed February 18, 2019.

Tutty, M. A., L. E. Carlasare, S. Lloyd, and C. A. Sinsky. "The Complex Case of EHRs: Examining the Factors Impacting the EHR User Experience." *Journal of the American Medical Informatics Association,* vol. 26, no. 7, 2019, pp. 673–677. Available at https://academic.oup.com/jamia/advance-article/doi/10.1093/jamia/ocz021/5426085. Accessed April 5, 2019.

Wu, D. T. Y., N. Smart, E. L. Ciemins, H. J. Lanham, C. Lindberg, and K. Zheng. "Using EHR Audit Trail Logs to Analyze Clinical Workflow: A Case Study From Community-based Ambulatory Clinics." *AMIA Annual Symposium Proceedings*, 2018, pp. 1820–1827. Available at https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5977693/. Accessed July 5, 2019.

Zheng, K., D. A. Hanauer, N. Weibel, and Z. Agha. "Computational Ethnography: Automated and Unobtrusive Means for Collecting Data In Situ for Human–Computer Interaction Evaluation Studies." In *Cognitive Informatics for Biomedicine,* edited by Vimla L. Patel and Thomas G. Kannampallil. New York, NY: Springer, 2015. Available at https://ubicomp.ucsd.edu/wp-content/papercite-data/pdf/zheng2015computational.pdf. Accessed February 18, 2019.

Appendix A: Discussion Guide

This page has been left blank for double-sided copying.

Introduction

Thank you for agreeing to speak with us today! I believe we planned about 45 minutes for this interview, are you still available until [end time]?

Great, thank you. Some quick housekeeping before we begin. As you may recall from our emails scheduling this interview, Mathematica is conducting a study for ASPE to "Quantitate the Burden of EHRs" by examining the feasibility of using EHR audit log data to construct national measures of burden for outpatient clinicians. By burden, we mean the non-value-added effort for clinicians, as distinguished from work that adds value.

We will keep our discussion today confidential. This means that, although we will share aggregate findings with ASPE and may include quotes in reports and presentations, we would never attribute those quotes to you as an individual. For example, we might say "an expert on EHR design said X," or something of the sort. Is that OK with you?

Thank you! As I mentioned, I'm joined by my colleague Megan Fitzgerald, who will be taking notes. Is it OK with you if we also record the conversation as a backup to guard against technical difficulties and the like? Thank you.

\<start recording\>

And would you please repeat that we have your consent to record the conversation?

Thank you.

Do you have any questions for us before we begin?

Warm-Up

1. Please tell me about your experience studying burden, using EHRs, extracting and/or analyzing audit log data, and/or constructing measures related to EHRs.

2. What do you consider to be the key activities that clinicians complete in the EHR related to patient care?

If not mentioned, ask about importance of:

    a. Administrative tasks,
    b. Clinical documentation tasks, and
    c. Clinical review tasks

Questions for Clinicians and Experts on Clinician Experiences

3. We want to make sure our definition of burden is comprehensive and relevant. What do you consider to be the key dimensions of burden related to using the EHR in an outpatient setting?

   - Probe as necessary:

    a. Time spent entering information
    b. Complexity of entering information (e.g., too many clicks)
    c. Challenges finding information
    d. Challenges interpreting information (e.g., too much copy and paste) and cognitive burden
    e. Other

   - How would you differentiate between EHR-attributable burden and EHR use patterns that are driven by factors extrinsic to the EHR like your organization's workflow, the tempo of day-to-day operations, and your documentation preferences and those of other clinicians?

4. How, if at all, do you think the type of EHR task affects the burden of use?

   - What is most burdensome about completing administrative tasks in the EHR?

   - What is most burdensome about completing clinical documentation tasks in the EHR? [Probe as necessary about entering orders, alerts, modifications to orders, time searching for the right order.]

   - What is most burdensome about completing clinical review tasks in the EHR?

5. We want to learn more about factors that drive burden. How, if at all, do you think EHR design affects burden of use?

[Probe as necessary regarding templates, integration of voice recognition, alerts and embedded decision-support functions, and interoperability with other systems that support care coordination and care transitions, quality reporting, and population health management.]

6. How, if at all, does clinician specialty affect EHR burden? Is EHR burden different for primary care than it is for specialists?

7. Which tasks or elements of burden do you think are most valuable and feasible to measure (either for policymakers, provider organizations, vendors, or other stakeholders)? [Probe as necessary for rationale.]

    - *For each measurement category:* At what level of precision/accuracy can this be measured by audit log data?

Questions for Vendors and Experts on EHR Infrastructure

8. What, if any, measures of burden does [your system/EHR vendor] provide to practices? *[Probe as necessary on how measures are constructed, face validity with users.]*

9. What might make it difficult to accurately measure provider documentation tasks in an EHR, or across EHRs, using audit log data? *[Probe as necessary regarding data retention, data retrieval, data processing.]*

    - What strategies would you recommend to overcome these difficulties?

10. How, if at all, does the structure or content of [your system's/EHR vendor's] audit log data change over time? [*Probe as necessary regarding system updates from the vendor.]*

11. What are some key similarities between the different vendors' audit log files? Key differences?

12. How, if at all, could a practice's installation choices affect the structure or content of its audit log data? *[Probe as necessary regarding on-site decisions about what information to store, which optional EHR features to activate, the software version in use, and unintended consequences of on-site customization by practice staff.]*

13. We recognize that not all EHR use is burdensome and further seek to focus on burden that is *attributable to the EHR* and not driven by extrinsic factors *[if necessary, give example from earlier in the discussion, or suggest 'for example, messages or notifications from external sources (e.g., pharmacy) with requests that have already been fulfilled'].* How, if at all, can EHR audit logs be used for this work? Which elements of clinician burden are particularly well-suited to being measured by audit log data? Particularly ill-suited?

Wrap-Up

14. Is there anything else that we haven't talked about that may be important to address or that is relevant to our conversation today?

15. Do you have any (other) suggestions or recommendations for ASPE as it tries to address this issue?

Conclusion

Thank you for taking the time to speak with us today! We may be in touch over the coming months to clarify something from this conversation and will plan to reach out in the fall or winter if you are interested in the final report on our findings.

This page has been left blank for double-sided copying.