Methods to Facilitate the Capture, Use, and Reuse
of Structured and Unstructured Clinical Data

by

Tzu-Yu Wu

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Information)
in the University of Michigan
2016

Doctoral Committee:
    Associate Professor Kai Zheng, Co-Chair, University of California, Irvine
    Associate Professor Qiaozhu Mei, Co-Chair
    Associate Professor David Alan Hanauer
    Associate Professor Joyce M. Lee
    Associate Professor Mark W. Newman

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AMIA   American Medical Informatics Association

CBD   Computer-based Documentation

CPOE   Computerized Provider Order Entry

CPR   Computerized Patient Record

EHR   Electronic Health Record

HCI   Human-Computer Interaction

Health IT   Health Information Technology

HITECH   Health Information Technology for Economic and Clinical Health

ICD   International Classification of Disease

IOM   Institute of Medicine

IR   Information Retrieval

NLP   Natural Language Processing

PD   Participatory Design

PDA   Personal Digital Assistant

PROMIS   Problem-Oriented Medical Information System

SBQR   Semantically-based Query Recommendation

SDE   Structured Data Entry

SNOMED-CT   Systematized Nomenclature of Medicine - Clinical Terms

XML   Extensible Markup Language

ABSTRACT

Electronic health records (EHRs) have great potential to improve quality of care and to support clinical and translational research. While EHRs are being increasingly implemented in U.S. hospitals and clinics, their anticipated benefits have been largely unachieved or underachieved. Among many factors, tedious documentation requirements and the lack of effective information retrieval tools to access and reuse data are two key reasons accounting for this deficiency. In this dissertation, I describe my research on developing novel methods to facilitate the capture, use, and reuse of both structured and unstructured clinical data.

Specifically, I develop a framework to investigate potential issues in this research topic, with a focus on three significant challenges. The first challenge is structured data entry (SDE), which can be facilitated by four effective strategies based on my systematic review. I further propose a multi-strategy model to guide the development of future SDE applications. In the follow-up study, I focus on workflow integration and evaluate the feasibility of using EHR audit trail logs for clinical workflow analysis. The second challenge is the use of clinical narratives, which can be supported by my innovative information retrieval (IR) technique called "semantically-based query recommendation (SBQR)". My user experiment shows that SBQR can help improve the perceived performance of a medical IR system, and may work better on search tasks with average difficulty. The third challenge involves reusing EHR data as a reference standard to benchmark the quality of other health-related information. My study assesses the readability of trial descriptions on ClinicalTrials.gov and found that trial descriptions are very hard to read, even harder than clinical notes.

My dissertation has several contributions. First, it conducts pioneer studies with innovative methods to improve the capture, use, and reuse of clinical data. Second, my dissertation provides successful examples for investigators who would like to conduct interdisciplinary research in the field of health informatics. Third, the framework of my research can be a great tool to generate

future research agenda in clinical documentation and EHRs. I will continue exploring innovative and effective methods to maximize the value of EHRs.

CHAPTER 1

Introduction

Clinical data in general and patient records in particular are keystones of health care. They are fundamental to not only patient care but also clinical and translational research, as well as hospital administration, billing, and medical education. Clinical data are primarily captured and documented through the course of patient care, and in recent years supported by computer systems. The idea of using computer systems to document clinical data, or so-called computer-based documentation (CBD), is not new. It had been attempted by pilot and visionary projects such as the Problem-Oriented Medical Information System (PROMIS) since 1970's (1–3). With the advancing of computer and information technology (IT) over the past 30 years, CBD and computerized patient records (CPRs) have been drawing significant attention in clinical informatics research communities. In 1997, the Institute of Medicine (IOM) advocated CPRs as an essential technique for health care (4). This IOM report highlighted the need for standardization, knowledge representation, and direct clinical data entry by physicians in a structured fashion to improve data accuracy.

The IOM report triggered much research in the 1990's to model patient records, predict data entry behaviors, and develop structured data entry (SDE) applications (5–10). Unfortunately, a majority of these SDE prototype systems was not generalizable and only showed success in a small-scale, specific medical environment. Very few of them achieved promising evaluation results in clinical practice. One major challenge was the low user acceptance, partly because of the suboptimal design of the computer user interfaces. Unfortunately, there was limited room for improving such user interfaces because they were strictly guided by the design of patient record models. For example, Los et al. pointed out that the inconsistent representation of data elements on the user interface of OpenSDE (a novel and popular SDE prototype) was "a result of a

1

fundamental design principle of OpenSDE" (11). These lessons suggested that there must be careful considerations of and a balance between predictive data entry and user interface design when developing an SDE application.

While the main stream of research focused on modeling patient records and developing SDE applications, a group of researchers started to challenge this exclusive goal. These researchers, many of whom were from the Vanderbilt University, thought clinical data entry could be supported by multiple documentation tools, rather than exclusively by computerized structured data entry forms (12). Clinicians should freely choose a documentation tool based on their goals of documentation and patient care (13). To solve the challenge of integrating clinical data from multiple documentation tools, these researchers suggested using interface terminology, i.e. clinical terms locally agreed upon and frequently used, rather than using standardized terms such as International Classification of Disease (ICD) and Systematized Nomenclature of Medicine - Clinical Terms (SNOMED-CT), to improve data quality and consistency (14). Based on these research findings, Rosenbloom et al. further pointed out an inherent tension between structured and unstructured clinical data: "While structured entry emphasizes data standardization and structure, human adoption of CBD systems requires an emphasis on expressivity, efficiency, flexibility, and being well adapted to a typical workflow (15)". This important perspective led a paradigm shift from a dominant focus on structure data to a balance between both structured and unstructured data.

The historical trajectory and the paradigm shift in informatics research on clinical documentation and data capture inspired my dissertation and formed the framework for it (Table 1). This framework has two dimensions. First, clinical data type can be categorized into structured or unstructured. Second, clinical data lifecycle can have at least three stages: 1) captured or collected, 2) used or analyzed for patient care, and 3) reused or repurposed in multiple ways (16). In my dissertation, I focus on three challenging problems in this framework, which segues into my three areas of research. In the first research area (R1), I study 'best' practices of CBD to

inform optimal strategies with a focus on usability and workflow integration to capture structured clinical data more efficiently and with higher quality. In the second research area (R2), I explore ways to extract valuable information locked in unstructured clinical narratives to improve their value of use and reuse, for purposes such as decision support, quality improvement, and clinical and translational research. In the third research area (R3), I investigate potential communication issues, especially document readability and reading comprehension, when sharing patient records and other clinical information with patients and the general public.

Table 1.1: The framework of this dissertation.

	Capture	Use	Reuse
Structured	R1. Structured Data Entry (Ch.2-3)	-	-
Unstructured	-	R2. Medical Information Retrieval (Ch. 4)	R3. Comm. issues when sharing medical information (Ch. 5)

My dissertation is organized as follows. In Chapters 2-3, I focus on my first research area. Specifically, in Chapter 2, I systematically review the literature and synthesize four effective strategies to facilitate SDE; in Chapter 3, I conduct a novel study to examine the feasibility of using EHR audit trail log data for clinical workflow analysis, which is one of the gaps identified in my previous literature review. In Chapter 4, I switch the focus to my second research area. I analyze the query log of a semantically-based query recommendation feature and generate design implications based on the findings. In Chapter 5, which focuses on my third research area, I used clinical notes as a reference standard to benchmark the readability of trial descriptions on ClnicalTrials.gov, to further highlighted potential communication issues when sharing this trial information with general publics. In Chapter 6, I summarize the contributions of my dissertation and describe my future directions and projects to continue advancing knowledge in my three research areas.

Chapter 2 Preview

(Facilitating Clinical Structure Data Entry: A 30-year Systematic Review)

I conduct a systematic literature review to synthesize the effective strategies that have been proposed and attempted in the past. I generate a set of keywords (e.g. structured data entry, clinical documentation) and query them against three online publication repositories, including PubMed, ACM Digital Library, and IEEE Xplore. I also search these keywords in Google Scholar to improve the recall of the dataset. I set up inclusion and exclusion criteria and follow the PRISMA statement to generate the final publication set for analysis (17). Using the Grounded theory, I synthesize four effective strategies to facilitate SDE from the selected papers. These effective strategies include: 1) Knowledge Base, 2) Interface Terminology, 3) Interface Usability, and 4) Workflow Integration. I further integrate these findings and propose a multi-strategy model to guide the development of future SDE applications.

Chapter 3 Preview

(Using EHR Audit Trail Logs for Clinical Workflow Analysis: An Example from Ambulatory Clinics)

Based on the previous literature review, I have found that an SDE application should consider and carefully integrate clinical workflow. In order to build such a workflow-integrated applicaiton, a critical first step is understanding and measuring clinical workflow. While Time & Motion studies have been regarded as the "gold standard" method for workflow analysis (18), they are costly and have limited capability of measurement. My goal of this study is to evaluate the feasibility of using EHR audit trail log data as a valid data source for workflow analysis.

I collect EHR audit trail log data from three clinics (two primary care and one specialty clinic) of an ambulatory healthcare organization in the western US. This organization launched an EHR enhancement in 2013, resulting in behavioral changes among the clinicians as indicated in a previous funded project that I participated (19), I select three previous findings to guide the analysis of the present study: 1) a workaround order-placing behavior, 2) a shift of tasks between

clinical roles, and 3) a phenomenon of deferred documentation. I explore the ability of the EHR audit trail log data to provide supportive evidence to these findings using various data analytic techniques.

Chapter 4 Preview

(Query Log Analysis and Design implications of Semantically-based Query Recommendation for an Electronic Health Record Search Engine)

While structured data entry is preferred due to its standardization and machine-readable format, clinical notes are still widely used due to their flexibility and expressivity. Information in clinical notes, however, is usually locked in sentences and cannot be utilized without help from special techniques, such as information retrieval (IR) (20,21). Retrieving clinical notes can be challenging because of the pervasive use of abbreviations, acronyms, and synonyms. To overcome this challenge, query expansion can be an effective method to assist users to find relevant documents.

I develop a novel IR technique called "semantically-based query recommendation (SBQR)" and deploy this feature on a prototype EHR search engine. This feature allows users to retrieve clinical documents with terms in a similar meaning to the search keywords. For example, when users search the term "hearing loss", they retrieve documents containing not only "hearing loss" but also its synonyms such as "difficulty of hearing". I co-conduct an evaluation study of this prototype system and its SBQR feature, which shows promising results. In my follow-up study included in this chapter, I analyze the query log automatically collected in the previous evaluation study and have found that SBQR brought users together with similar information needs but varying search techniques. This SBQR feature highly correlates to the participants' perceived positive performance of the prototype system, and may perform better on tasks of average difficulty.

Chapter 5 Preview

(Assessing the Readability of ClinicalTrials.gov)

ClinicalTrials.gov provides critical functions of disseminating trial information to the public and helping the trials recruit participants. In this study, I assess the readability of trial descriptions at ClinicalTrials.gov using multiple quantitative measures. The analysis includes all 165,988 trials registered at ClinicalTrials.gov as of April 30, 2014. To benchmark the readability of trial descriptions, I include in the analysis all 955 Health Topics articles from MedlinePlus and a random sample of 100,000 clinical notes from an EHR system as the reference standards. I apply five readability measures to characterize these medical corpora. The results show that the trial descriptions on ClinicalTrials.gov were the most difficult to read, even more challenging than clinicians' notes.

References

1. Weed LL. Medical records that guide and teach. N Engl J Med. 1968 Mar 14;278(11):593–600.

2. Weed LL. The problem oriented record as a basic tool in medical education, patient care and clinical research. Ann Clin Res. 1971 Jun;3(3):131–4.

3. Schultz JR, Davis L. The technology of PROMIS. Proceedings of the IEEE. 1979;67(9):1237–44.

4. Institute of Medicine (US) Committee on Improving the Patient Record. The Computer-Based Patient Record: Revised Edition: An Essential Technology for Health Care [Internet]. Dick RS, Steen EB, Detmer DE, editors. Washington (DC): National Academies Press (US); 1997 [cited 2015 Aug 8]. Available from: http://www.ncbi.nlm.nih.gov/books/NBK233047/

5. Rector AL, Nolan. Foundations for an Electronic Medical Record. Methods of Information in Medicine. 1991;30:179–86.

6. Kirby J. Predictive data entry in medical records. 1992;

7. Heathfield HA, Hardiker NR, Kirby J. Using the PEN&PAD information model to support hospital-based clinical care. Proc Annu Symp Comput Appl Med Care. 1994;452–6.

8. Moorman PW, van Ginneken AM, van der Lei J, van Bemmel JH. A model for structured data entry based on explicit descriptional knowledge. Methods Inf Med. 1994 Dec;33(5):454–63.

9. van Ginneken AM. Structured data entry in ORCA: the strengths of two models combined. Proc AMIA Annu Fall Symp. 1996 Nov;797–801.

10. Los RK, van Ginneken AM, de Wilde M, van der Lei J. OpenSDE: Row modeling applied to generic structured data entry. J Am Med Inform Assoc. 2004 Apr;11(2):162–5.

11. Los RK, van Ginneken AM, van der Lei J. Extracting data recorded with OpenSDE: possibilities and limitations. Int J Med Inform. 2005 Jul 2;74(6):473–80.

12.   Shultz E, Rosenbloom T, Kiepek W, Fitzhenry F, Adams P, Mahuli A, et al. Quill: a novel approach to structured reporting. AMIA Annu Symp Proc. 2003;1074.

13.   Rosenbloom ST, Crow AN, Blackford JU, Johnson KB. Cognitive factors influencing perceptions of clinical documentation tools. J Biomed Inform. 2007 Apr;40(2):106–13.

14.   Rosenbloom ST, Miller RA, Johnson KB, Elkin PL, Brown SH. Interface terminologies: facilitating direct entry of clinical data into electronic health record systems. J Am Med Inform Assoc. 2006 Jun;13(3):277–88.

15.   Rosenbloom ST, Denny JC, Xu H, Lorenzi N, Stead WW, Johnson KB. Data from clinical notes: a perspective on the tension between structure and flexible documentation. J Am Med Inform Assoc. 2011 Mar 1;18(2):181–6.

16.   Alex Ball. Review of Data Management Lifecycle Models [Internet]. University of Bath; 2012. Available from: http://opus.bath.ac.uk/28587/

17.   Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Int J Surg. 2010;8(5):336–41.

18.   Lopetegui M, Yen P-Y, Lai AM, Embi PJ, Payne PRO. Time Capture Tool (TimeCaT): development of a comprehensive application to support data capture for Time Motion Studies. AMIA Annu Symp Proc. 2012;2012:596–605.

19.   Zheng, K, Ciemins, E, Lanham, H, Lindberg, C. Examining the Relationship Between Health IT and Ambulatory Care Workflow Redesign. (Prepared by Billings Clinic under Contract No. 290-2010-0019I-1). AHRQ Publication No. 15-0058-EF. Rockville, MD: Agency for Healthcare Research and Quality; 2015.

20.   Hanauer DA. EMERSE: The Electronic Medical Record Search Engine. AMIA Annu Symp Proc. 2006;941.

21.   Hanauer DA, Mei Q, Law J, Khanna R, Zheng K. Supporting information retrieval from electronic health records: A report of University of Michigan's nine-year experience in developing and using the Electronic Medical Record Search Engine (EMERSE). J Biomed Inform. 2015 Jun;55:290–300.

CHAPTER 2

Facilitating Clinical Structured Data Entry: A 30-year Systematic Review

2.1 Introduction

Clinical data, especially when being computerized and stored in an Electronic Health Record (EHR) system as well as meaningfully used, have great potential to facilitate decision-making, improve quality of care and outcomes, support evidence-based medicine, and further constitute the basis of learning health systems (1). The meaningful use of EHRs has been promoted by the HITECH Act since 2009, which provides monetary incentives and imposes penalties to unleash the potential and maximize the benefits of EHRs (2–5). These benefits, however, cannot be achieved to a large extent without clinical data being recorded in a structured, computer-analyzable format.

The idea of recording clinical structured data to increase its utility was widely accepted in the 1990's. For example, Dick et al. in their Institute of Medicine report emphasized the benefits of standardized structured data and the importance of data being recorded electronically and directly by physicians in order to leverage the accuracy and quality of clinical data (6,7). This report triggered much research investigating the best practices of structured data entry (SDE) and its computer applications. Although many research projects and system prototypes have shown strong evidence of the benefits of SDE, recording structured data may remain a significant hurdle for clinicians for at least two reasons: 1) lack of usability and 2) low expressivity. The lack of usability can come from at least two sources. First, it can come from the inconsistency between the inherent (or manually defined) data structure and a clinician's mental model (8,9). When the inconsistency is large, clinicians may record data in free-text as a workaround or "exist strategy", even when coding options are available (10). Second, the lack of usability can come from poorly designed user interface, i.e. the presentation of structured data elements on a computerized form.

One example of such low-usability interface can be a lengthy dropdown, showing hundreds of diagnostic codes at once. Even if the structure of these diagnostic codes fits a clinician's mental model, they still can cause usability issues because clinicians can easily make mistakes during intensive scrolling and checking, especially under major time constraints. These usability issues can largely affect data quality and introduce medical errors, leading to patient safety hazards. (11). In addition, structured data are likely to have low expressivity compared to their free-text counterparts. This is because clinical observations are binned into a fixed number of predefined options and therefore unable to capture many nuance findings. To one extreme, structured data could be so rigid that it seriously hinders clinicians' reasoning of a patient's medical history (12,13).

To address the potential issues of SDE in usability and expressivity, researchers have been investigating alternatives to rapidly and flexibly capture clinical data while keeping their structure, or generating the structure afterwards, for computational use and reuse. One stream of research leverages Natural Language Processing (NLP) techniques to extract entities and relationships from clinical narratives (14). While utilizing NLP techniques can avoid potential usability issues of SDE and maintain a high level of expressivity in clinical data, they are not frequently adopted in clinical practice due to significant issues such as lack of user-centered design and scalability. (15) Another stream of research focuses on the automatic collection of structured data using techniques such as barcode scanning, optical character/mark recognition, and near field communication with medical devices. (16–18) While these techniques can remove much of the data collection burden from clinicians, more research is needed to understand how to integrate these data with EHRs, how to extract meaning information from them, and how to present the information in an effective manner. Still another stream of research involves using medical scribes to completely remove the documentation responsibility from clinicians, which has been shown to significantly improve the number of patients seen and the quality of physician-patient interactions, although with increased personnel cost. (19,20)

These alternative approaches, however, still require much more work to evaluate their efficacy in clinical practice. Moreover, as a portion of clinical data would need manual editing and verification, it is inevitable that clinicians spend a great deal of time entering computerized structured clinical data to maximize their level of use and reuse. This paper therefore systematically surveys the literature and aims to identify effective strategies that have been suggested and attempted by researchers to facilitate SDE. The rest of paper is organized as follows: The methods section describes the literature search strategy using the PRISMA statement (21), followed by the results section describing effective SDE-facilitating strategies synthesized from the literature. Next, the findings are discussed and future research is suggested to address the gap(s) identified.

2.2 Methods

2.2.1 Search Strategy

The search strategy followed the guidelines in the PRISMA statement (21). Multiple rounds of searches were conducted with the goal of iteratively improving the result quality. In the first round, relevant papers in PubMed were searched by the term "Structured Data Entry" (with quotation marks), with publication dates ranging from January 1986 to December 2015. This search returned a small set of 86 records. After examining the results, the query terms were expanded due to the low coverage of qualified papers known by this author. This expanded query was also searched in publication databases in non-medical fields including ACM Digital Library and IEEE Xplore. Note that the query against IEEE Xplore was customized to better reflect domain-specific variations (e.g. clinical vs. healthcare). Table 2.1 below lists the search queries and the data repositories.

Table 2.1: Search terms and data repositories.

Repository (# records retrieved)	Search Terms	Additional Criteria
PubMed (661)	"structured data entry" OR "clinical data entry" OR "clinical data capture" OR "computer-based documentation" OR "clinical documentation"	General Search; Publications dates 1986 and 2015; Results must have full text; Written in English;
ACM digital library (35)	"structured data entry" "clinical data entry" "clinical data capture" "computer-based documentation" "clinical documentation" (matches any)	Advanced Search > search ACM full-text collection; Published between 1986 and 2015; Results must have full text;
IEEE Xplore (60)	(PubMed search terms) OR (("structured" OR structure) AND (clinical OR health OR healthcare) AND ("data entry" OR "data capture" OR documentation))	Advanced Search > Command Search; Metadata Only; Published between 1986 and 2015;

2.2.2. Inclusive and Exclusive Criteria

Relevant papers were selected based on the following criteria. First, a paper was included if it focused on improving SDE methods rather than simply using SDE as a means of data collection. Second, data must be directly and manually recorded by end users. Papers utilizing medical scribes or devices such as barcode scanning were therefore excluded. Papers in a non-clinical context were included as long as the proposed methods can be applied to a clinical setting. In addition, papers focusing on collecting patient self-reported data were excluded. Third, data were expected to be recorded in a structured and coded fashion. Papers that involved extracting structured information from clinical narratives using NLP techniques were therefore excluded. On the other hand, papers learned from narratives to construct SDE template for future use were considered, e.g. extracting frequently used terms from notes to produce a more usable SDE form. Lastly, studies in the citation network were examined only in the first degree, i.e. references of a reference were not scanned. Figure 2.1 below illustrates the systematic review procedure with the number of publications selected.
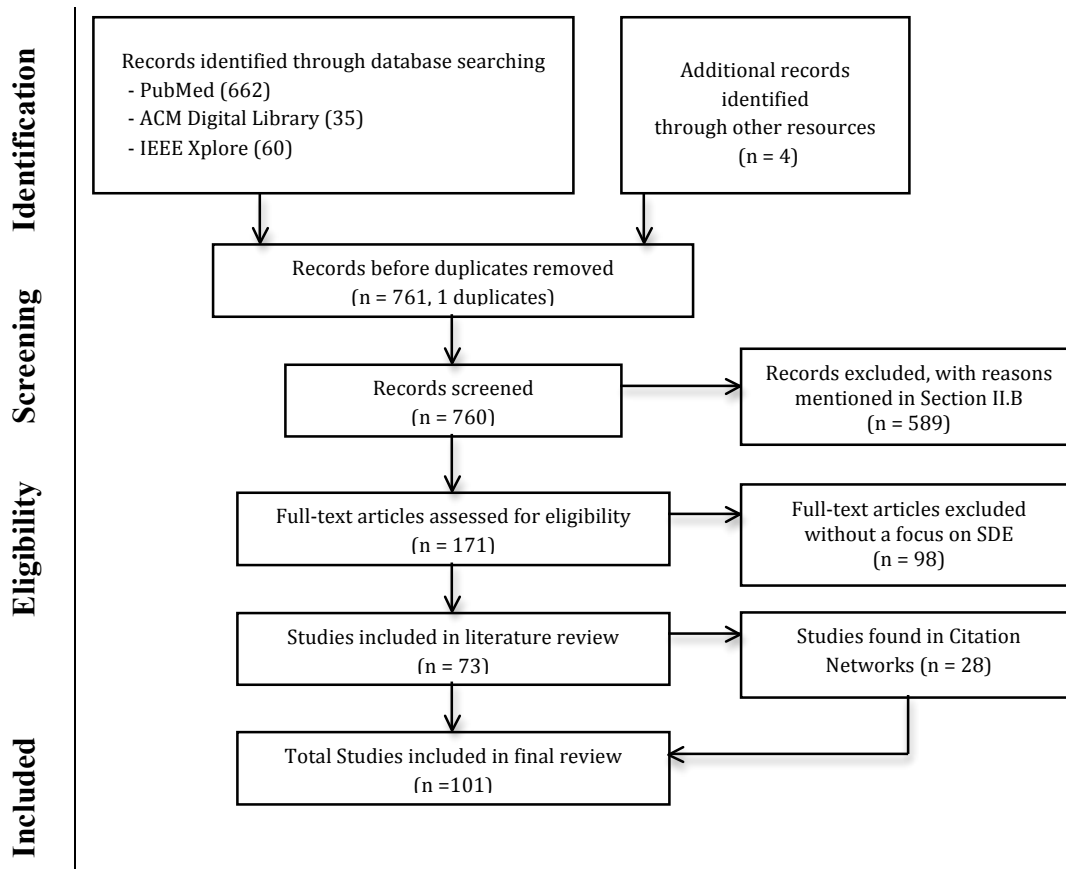
Figure 2.1: Systematic search process following PRISMA.

2.2.3 Data Analysis

The final selected publications were managed using Zotero, a cross-platform citation management software package (22). All selected papers were carefully read by this author to extract eighteen attributes from each, such as Objectives, Method, Specialty/Clinical Setting(s), System(s), Evaluation, and Findings (2.7 Appendix). The selected papers were repetitively coded, categorized, and thematically organized into effective SDE-facilitating strategies using the Grounded theory.

2.3 Results

Through the systematic search process, this author selected 101 papers and synthesized four effective strategies to facilitate SDE: 1) Knowledge Base, 2) Interface Terminology, 3) Interface Usability, and 4) Workflow Integration. This author further proposed a multi-strategy model to inform the design and implementation of future SDE applications.

2.3.1 Knowledge Base

A knowledge base contains a set of controlled vocabulary and sematic relations among them to comprehensively represent the knowledge about medical data in a specific domain. This strategy facilitates SDE because a knowledge base can predict data entry elements so that SDE forms could be adaptive to end-users. This strategy emerged in the early 90's from projects that aimed to prove the benefits of structured clinical data over free-text data in neurosurgery, gastroenterology, and pharmacology (23–27). The investigation of knowledge bases continued until 2008 based on the literature search (28–30). In the early 90's, researchers started to model structured data and support SDE using a knowledge base. One group of researchers focused on domain-specific SDE applications and achieved many successes. For example, the UltraSTAR project implemented an ultrasound reporting system with a hierarchical concept model (31,32), showing a higher level of report completeness than narratives and achieving good user satisfaction (33,34). Kahn et al. further expanded the project by introducing WWW techniques and standardized XML-based languages (35–37). This new version of UltraSTAR, however, was only evaluated in a lab setting and not generalized to other medical specialty. Another successful example was the SISCOP project in gastrointestinal endoscopy, which collected structured information and generated natural language reports. SISCOP-generated reports contained 60% more information with 50% less errors than their narrative counterparts (38–40). The success of these knowledge-based projects, unfortunately, was built upon their highly specialized domains, which made the knowledge bases relatively easy to design but less generalizable.

Another group of researchers was dedicated to developing a generalized approach for constructing a knowledge base. One such approach was the "Structured Meta-Knowledge" (SMK) formalism of the PEN & PED project. This SML formalism adopted a descriptive model view to collect data on what is clinically observed and believed, rather than on how care is supposed to be delivered (which is also called a prescriptive model). This formalism was constituted by three sub-models and two levels of observations, and written in a network-like language (41–43). By supporting predictive data entry, this formalism claimed to resolve the dilemma between expressivity and speed of use.

However, the generalizability of the SML formalism, and the PEN&PED project as a whole, was not very promising (41,44). The experiment in a nursing setting and an elderly care uncovered the inability of the system to carry out patient care with sufficient details (45–47). The deployment of this system in clinical routines further revealed the issues of suboptimal organization and presentation of clinical concepts in a computerized form, which led to the significant redesign of SMK formalism (48). One fundamental issue of the SMK formalism was that it could present too many data entry options to end-users. While these options may be conceptually related, they may not be clinically equally relevant and important, and therefore should be prioritized. To address this issue, Moorman et al. proposed a explicit descriptional knowledge base with refined semantic relations (49). A small-scale experiment in endoscopy successfully showed a success where structured forms generated by this knowledge base can effectively replace narratives and produce more diverse information (50).

This refined knowledge base was further extended by van Ginneken et al. as one component of a two-level computerized patient record framework (51,52), which is called Open Record for Care (ORCA). OCRA benefited SDE by balancing efficiency and flexibility as well as increasing retrievability of patient records. However, the evaluation of OCRA revealed several issues, including fail to consider domain-dependability (53), poor report quality due to the combination of structured and free-text data (54), lack of efficiency to support routine care (55), and

inconsistent retrieval results because of the occurrence of semantically equivalent expressions in its knowledge base (56).

The refinement of OCRA led to the emergence of a generic SDE application called "OpenSDE", which was featured on its innovative integration of a flexible data model, a knowledge base supporting data presentation, an tree-enhanced user interface, and a set of user-maintainable concepts (57–60). While OpenSDE can be easily deployed (61), it still was not very practically feasible, partly due to its inherent inconsistency in many dimensions (62–64). Moreover, studies showed that OpenSDE may lead to inefficient data capture and may be only suitable for well-defined datasets (65,66). This evidence suggests that using a generalized knowledge base as a primary approach to support SDE may not be very practical. These knowledge-based SDE applications often produced inflexible, inconsistent, or irrelevant data entry elements preventing high user acceptance.

2.3.2 Interface Terminology

This strategy can facilitate SDE because interface terminologies are easily understandable and highly differentiable to end users compared to standardized terminologies (or reference terminologies) such as ICD9 and SNOMED. Standardized terminologies may be problematic because of their potential threats to usability, which is one of the main reasons behind the low user acceptance of knowledge-based SDE application. (67,68) Researchers found that standardized terminologies have low coverage of clinical concepts and may need to be extended repeatedly (69–71). They also found that using standardized terminologies for data capture may lead to coding variations (9).

The definition of interface terminologies was coined by Rosenbloom et al. in 2006 as "a systematic collection of healthcare–related phrases (terms) that supports clinicians' entry of patient-related information into computer programs" (72). This group of researchers further examined the mapping between SNOMED and two interface terminologies, and developed an

evaluation model of interface terminologies (73,74). Following this definition, Daniel Le et al. in 2007 shared their experience implementing an interface terminology and indicated the importance of integrating such a terminology with the standardized ones. (75)

The idea of collecting daily-used clinical terms to support SDE was not new and has been researched since late 1990's, in parallel with the development of knowledge-based SDE applications. These projects involved using word frequency analysis and natural language processing (NLP) techniques to identify commonly used terms in clinical narratives. Their evaluation showed that these empirically identified terms can support existing SDE, and often had more granular concepts that were not fully covered by standardized terminologies (76–78).

Making these empirically identified terms more clinically usable largely depends on the balance between pre- and post-coordination of these terms (72). The following example shows the difference between these two types of coordination. In the pre-coordination, a problem "chest pain" is described exactly by a controlled vocabulary. In the post-coordination, this problem is represented as a combination of two concepts, "chest" and "pain". The pre-coordination provides "logic and intricacy" and post-coordination allows "expressivity" and more complete "domain coverage". Balancing pre- and post-composition therefore could reconcile the dilemma of efficiency and expressivity and have the ability to address the highly dynamic documentation language in routine care. Madani et al. took this viewpoint and refined their pre-coordination vocabulary set. They identified commonly combined concepts and all their modifiers. (79) Meanwhile, Kim et al. demonstrated that interface terminology required continuous effort of refinement. (80)

Rosenbloom and colleagues in 2013 further implemented and iteratively refined an interface terminology called "CHISL" (Categorical Health Information Structured Lexicon), and explained how CHISL can succeed in terms of usability. (81) However, CHISL still has room for improvement in generating natural language reports due to the need for a proper Part of Speech tagger. In addition, more research is needed to demonstrate how CHISL can effectively support

the variance of individual dictation styles and integrate with the pervasive hedge phrases in clinical documents. (82,83)

2.3.3 Interface Usability

The usability of an SDE application is key to user adoption. (84) This strategy specifically focuses on the usability of SDE computer interfaces. To achieve high usability, a computer interface should be "efficient, effective, learnable, memorable, highly satisfying, and has low error rate in supporting task execution in a specific context of use" (85,86). The usability of a computer interface can be improved in at least three areas: 1) single element, 2) data organization, and 3) input methods. In the first area, researchers have proposed and evaluated text prediction methods to improve the efficiency and the quality of structured data (87–89). For example, Canfield proposed priming intelligent split menus, which show frequently or possibly chosen options on the top, followed by the rest options in an alphabetical order. (87,90) Researchers also demonstrated that small changes in a single element can largely affect the data accuracy and even hospital financial bottom lines (91–93).

In the second area of improvement on data organization, researchers have proposed graphic-based method, especially body images, to enable efficient and intuitive SDE (94–96). A more advanced log-based method proposed by Zheng et al. (94,95) was able to reveal unexpected navigation behaviors and further optimize the data organization (96). In the third area of improvement on input method, one group of researchers promoted pen and paper input behaviors to help clinicians smoothly transition from paper-based documentation (97,98). Another group of researchers focused on providing multiple input methods so that clinicians can freely choose a suitable input method based on their documentation needs and personal habits (99–101). They also argued that there is no one-side-fit-all documentation tool and the selection of input methods should consider factors such as usability and clinical workflow (102).

It is worth noting that interface usability improvements can also be achieved by other methods, such as Participatory Design (PD). While PD considers the needs and reconciles the diverse objectives of stockholders, it is usually time consuming and may not capable of collecting actual usage patterns but ideal ones (103). Other methods to improve interface usability included applying human factor analysis to reduce human-machine errors, and integrating qualitative methods in the early stage of system design (104,105). The effectiveness of usability-improving methods should be examined through rigorous usability testing. The literature review identified a few studies conducting a usability test on an SDE application in the field of dentistry and oncology (106–109).

2.3.4 Workflow Integration

This strategy recommends that an SDE application should consider and integrate clinical processes and workflow. To achieve this goal, a critical first step is to understand the workflow. One early project identified in the literature search shared its experience integrating an SDE application with everyday clinical routine (110). This study designed a workflow-integrated SDE application based on experts' knowledge. More recent studies, on the other hand, used rigorous methods to make sense of workflow and evaluate the effectiveness of integration. These methods included observational time analysis, document analysis, literature search, expert interview, survey, workflow modeling, and feedback loop (111–114). These methods, however, only analyzed workflow at a surface level, i.e. they gauged the time efficiency, document quality, and user satisfaction of workflow integration as a whole without examining step-wised changes in details.

In addition to using rigorous research methods, another group of researchers considered portable devices as an effective strategy of workflow integration due to their mobility. Webster and Copenhaver advocated for a portable device with touch screen technology, which "optimally reduces inherent tradeoffs between information utility and system usability on one hand, and speed and accuracy of data entry on the other" (115). Other studies utilized Personal Digital

Assistant (PDA), tablet computer with stylus, and Apple iPad to replace existing (paper) documentation and automate as well as optimize the current workflow (116–118).

Still another group of researchers focused on the guidance of workflow integration, they highlighted more considerations and proposed an overall plan for designing and implementing such workflow-integrated SDE applications. Specifically, Embi et al. conducted a large-scale focused group study and found that documentation is "intertwined tightly" with clinical and administrative workflow (119). They suggested that future design of computerized patient documentation should address issues in the areas of communication, expressivity, information availability, workflow, and confidence. They also found that physicians, nurses, and administrators held different viewpoints on the above issues. Saleem et al. conducted ethnographic observations to identify efficiency strategies in ambulance care. They found that clinicians used paper artifacts, note templates, copying and pasting, and pre-populated notes to fit documentation activities into their workflow and to support their cognitive needs. (120) Knaup et al. proposed a systematic planning for the design of a workflow-integrated SDE application, which contains several steps. First, developers should analyze current documentation activities and infrastructure, and then they should design a terminology and a management system. Next, developers should fit these subsystems into workflow by planning the logical architecture of collaboration. Finally, developers should deploy all documentation tools and build a collaborative environment. (121)

2.3.5 A Multi-Strategy Model

This author combines the above four strategies into a multi-strategy model to inform the future design and implementation of an SDE application. Figure 2.2 shows the composition of such a model. As indicated by the first two strategies, this model contains a knowledge base that keeps the ontology of terms and a terminology service that handles the mapping between an interface terminology and standardized terminologies. This author suggests two additional services, one for workflow awareness and the other for usability optimization. The workflow awareness

service perceives documentation activities as a series of steps in a clinical workflow while considering clinical roles, timing, and physical locations. The usability optimization service not only fine tunes the presentation of information based on devices and tasks, but also allows the personalization and customization of end users to further minimize the gaps between a clinician's mental model and the system design.
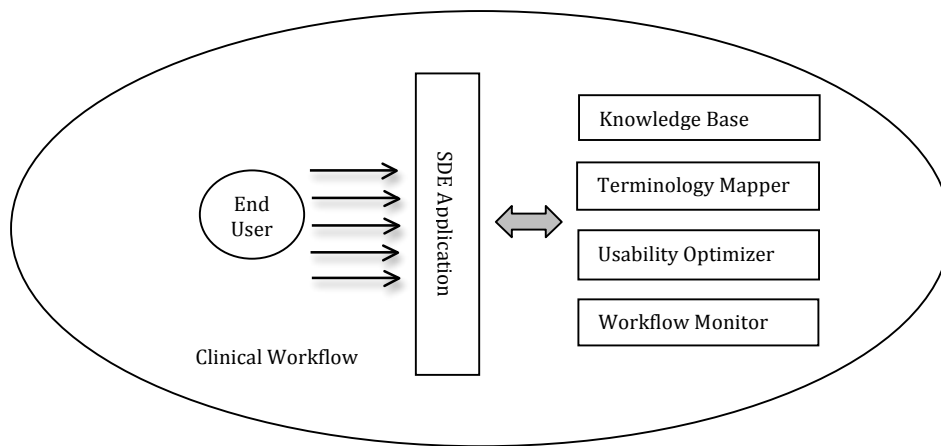
Figure 2.2: A multi-strategy model and its components.

One question about this model is how to coordinate the communications between these four services. This author argues that the coordination should be focused first on workflow awareness and usability optimization, followed by the terminology and knowledge base. Previous research has shown that a SDE application primarily guided by a knowledge base is likely to have limited generalizability and low user acceptance. Moreover, as recent studies suggested, the user acceptance and the practical feasibility of an SDE application largely depend on how well such an application is design in interface usability and workflow integration.
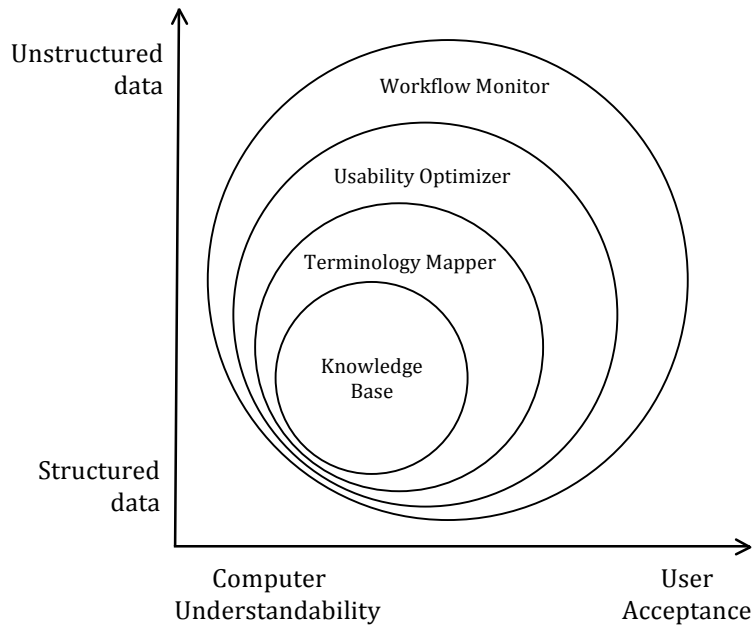
Figure 2.3: The coordination of four proposed services

2.4 Discussion

This study conducted a systematic literature review and synthesized four effective strategies to facilitate clinical SDE, including 1) Knowledge Base, 2) Interface Terminology, 3) Interface Usability, and 4) Workflow Integration. The results showed that the first strategy (Knowledge Base) tends to over-emphasize the benefits of structured data and the ability of such a knowledge base to comprehensively guide data collection processes. It is not stating that such the knowledge-based strategy is ineffective. Rather, it indicates that this strategy should be used in combination with other strategies to leverage user acceptance. On the other hand, the workflow integration strategy expands the scope of clinical SDE by bringing in considerations such as workflow analysis, mobility, and collaboration. In addition, the chorological order of these effective strategies in the literature suggests a paradigm shift of SDE from an expert focus to an end user focus, from a one-size-fits-all solution to customization, from a strong preference on

21

structured data to a balance between structure and expressivity, and from individuals to teamwork. Another observed trend is that the improvements of SDE applications were built upon the advancing of computer technology in the past 30 years. For example, as the computers became more decentralized and personal, it is reasonable to see the design of SDE applications move away from a centralized knowledge base to a distributed view on interface usability and workflow integration.

This paradigm shift is also reflected by government reports and policy views. Dick et al. in the 90's advocated the benefit of direct data entry from physicians in a structured format (6,7), leading to more than 15 years of efforts on improving the efficiency, effectiveness, and generalizability of knowledge-based SDE applications. More recent policy viewpoints seemed moving away from a strong emphasis on structured data and direct data entry by clinicians, and move toward a balance between structured and unstructured data as well as data collection from multiple sources. For example, the AMIA community in its 2011 policy meeting recognized that data could be collected "from multiple sources as appropriate and necessary, including nuanced medical discourse". (122) The report of the AMIA EHR 2020 task force further pointed out the need to simplify and speed EHR documentation and recommended automating data collection and entering data using "alterative approaches and media that could be more efficient of providers time". (123) This report also highlighted the equal value of information entered by physicians and other care team members as well as patients.

This study has at least two limitations. First, the quality of the systematic search largely depended on the design of the keywords. This issue has been mitigated by refining the query over multiple rounds to ensure high recall, i.e. known eligible papers were largely discovered. Second, the selected papers were solely examined, analyzed, and synthesized by this author, which may exhibit personal biases.

This study suggests several research opportunities to improve SDE in the future. First, since SDE should consider workflow, it is necessary to monitor and make sense of clinicians' current daily

activities to inform the design. While qualitative approaches such as interviews and observations may uncover some patterns, these approaches are limited due to their narrower scope, higher cost, and potential human biases. To supplement findings from qualitative approaches, researchers can utilize EHR audit trail logs to construct a more comprehensive and objective view of clinicians' documentation activities. These log data, however, are massive and have vendor-specific designs with varying data granularity, which introduces challenges in uniformly analyzing these data and drawing meaningful patterns from them. A study has indicated potential quality issues of EHR log data to support user behavioral analysis (124). More research is needed to extract and analyze high quality information from these EHR log data using state-of-art methods such as visual analytics on big data.

Second, it is urgent to improve the usability of SDE applications. This author suggests the need for more concrete guidelines for designing clinical SDE application with high usability, especially on mobile and tablet devices, as well as the need for best practices with promising evaluation results in daily routines. For example, the usability heuristics developed by Nielsen and Molich (125,126) may be modified and extended to evaluate clinical data capture systems. In addition, each clinician may have personal preferences and unique strategies of data entry to improve efficiency. Software personalization therefore could be another interesting research topic to further improve the usability of clinical SDE applications.

Third, the methods to effectively achieve the balance between structured and unstructured data should be researched. For example, Johnson et al. proposed a novel data entry method called "structured narrative", which achieves the balance by enabling structured data entry when composing free text notes (61). This data entry method records clinical data in an XML format, which introduces another set of challenges in maintaining, organizing, and retrieving information. Some research questions in this area include, but are not limited to, how to flexibly maintain data to support downstream uses such as decision support, research, and education, and how to simultaneously retrieve and rank cases with structured and unstructured data given their unique nature and characteristics.

2.5 Conclusions

This study systematically synthesized four SDE-facilitating strategies and further proposed a multi-strategy model to inform the design and implementation of future SDE applications. The study also suggests future research of SDE applications on workflow integration, usability, and the impact of the balance between data types.

2.6 References

1. Jha AK. Meaningful Use of Electronic Health Records<subtitle>The Road Ahead</subtitle> JAMA. 2010 Oct 20;304(15):1709.

2. Blumenthal D. Stimulating the Adoption of Health Information Technology. New England Journal of Medicine. 2009 Apr 9;360(15):1477–9.

3. Blumenthal D. Launching HITECH. New England Journal of Medicine. 2010 Feb 4;362(5):382–5.

4. Blumenthal D, Tavenner M. The "Meaningful Use" Regulation for Electronic Health Records. New England Journal of Medicine. 2010 Aug 5;363(6):501–4.

5. Powsner SM, Wyatt JC, Wright P. Opportunities for and challenges of computerisation. Lancet. 1998 Nov 14;352(9140):1617–22.

6. Institute of Medicine (U.S.). The computer-based patient record: an essential technology for health care. Washington, D.C: National Academy Press; 1991. 190 p.

7. Institute of Medicine (U.S.). The computer-based patient record: an essential technology for health care. Rev. ed. Washington, D.C: National Academy Press; 1997. 234 p.

8. McDonald CJ. The barriers to electronic medical record systems and how to overcome them. J Am Med Inform Assoc. 1997 Jun;4(3):213–21.

9. Patrick TB, Richesson R, Andrews JE, Folk LC. SNOMED CT coding variation and grouping for "other findings" in a longitudinal study on urea cycle disorders. AMIA Annu Symp Proc. 2008;11–5.

10. Zheng K, Hanauer DA, Padman R, Johnson MP, Hussain AA, Ye W, et al. Handling anticipated exceptions in clinical care: investigating clinician use of "exit strategies" in an electronic health records system. J Am Med Inform Assoc. 2011 Dec;18(6):883–9.

11. Middleton B, Bloomrosen M, Dente MA, Hashmat B, Koppel R, Overhage JM, et al. Enhancing patient safety and quality of care by improving the usability of electronic health record systems: recommendations from AMIA. Journal of the American Medical Informatics Association. 2013 Jan 25;20(e1):e2–8.

12. Walsh SH. The clinician's perspective on electronic health records and how they can affect patient care. BMJ. 2004 May 15;328(7449):1184–7.

13. Rosenbloom ST, Denny JC, Xu H, Lorenzi N, Stead WW, Johnson KB. Data from clinical notes: a perspective on the tension between structure and flexible documentation. J Am Med Inform Assoc. 2011 Mar 1;18(2):181–6.

14. Nadkarni PM, Ohno-Machado L, Chapman WW. Natural language processing: an introduction. J Am Med Inform Assoc. 2011 Oct;18(5):544–51.

15. Chapman WW, Nadkarni PM, Hirschman L, D'Avolio LW, Savova GK, Uzuner O. Overcoming barriers to NLP for clinical text: the role of shared tasks and the need for additional creative solutions. J Am Med Inform Assoc. 2011 Oct;18(5):540–3.

16. Nichols JH, Bartholomew C, Brunton M, Cintron C, Elliott S, McGirr J, et al. Reducing medical errors through barcoding at the point of care. Clin Leadersh Manag Rev. 2004 Dec;18(6):328–34.

17. Bergeron B. Clinical data capture: OMR and OCR and your flatbed scanner. MedGenMed. 2005;7(2):66.

18. Morak J, Schwetz V, Hayn D, Fruhwald F, Schreier G. Electronic data capture platform for clinical research based on mobile phones and near field communication technology. Conf Proc IEEE Eng Med Biol Soc. 2008;2008:5334–7.

19. Arya R, Salovich DM, Ohman-Strickland P, Merlin MA. Impact of scribes on performance indicators in the emergency department. Acad Emerg Med. 2010 May;17(5):490–4.

20. Bank AJ, Obetz C, Konrardy A, Khan A, Pillai KM, McKinley BJ, et al. Impact of scribes on patient interaction, productivity, and revenue in a cardiology clinic: a prospective study. Clinicoecon Outcomes Res. 2013;5:399–406.

21. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Int J Surg. 2010;8(5):336–41.

22. Center for History and New Media. Zotero Quick Start Guide [Internet]. Available from: http://zotero.org/support/quick_start_guide

23. Kondziolka D, Schwartz ML, Walters BC, McNeill I. The Sunnybrook Neurotrauma Assessment Record: improving trauma data collection. J Trauma. 1989 Jun;29(6):730–5.

24. Kuhn K, Swobodnik W, Johannes RS, Zemmler T, Stange EF, Ditschuneit H, et al. The quality of gastroenterological reports based on free text dictation: an evaluation in endoscopy and ultrasonography. Endoscopy. 1991 Sep;23(5):262–4.

25. Kuhn K, Gaus W, Wechsler JG, Janowitz P, Tudyka J, Kratzer W, et al. Structured reporting of medical findings: evaluation of a system in gastroenterology. Methods Inf Med. 1992 Nov;31(4):268–74.

26. Kuhn K, Zemmler T, Reichert M, Heinlein C, Roesner D. Structured data collection and knowledge-based user guidance for abdominal ultrasound reporting. Proc Annu Symp Comput Appl Med Care. 1993;311–5.

27. Zellner D, Schromm T, Frankewitsch T, Giehl M, Keller F. Structured data entry for reliable acquisition of pharmacokinetic data. Methods Inf Med. 1996 Sep;35(3):261–4.

28. Thurin A. A conceptual model of diagnostic findings in echocardiography. In IEEE; 1997 [cited 2013 Nov 1]. p. 637–40. Available from: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=648130

29. Bouamrane M-M, Rector A, Hurrell M. Gathering Precise Patient Medical History with an Ontology-Driven Adaptive Questionnaire. In IEEE; 2008 [cited 2013 Nov 1]. p. 539–41. Available from: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4562053

30. Ammon D, Hoffmann D, Jakob T, Finkeissen E. Developing an architecture of a knowledge-based electronic patient record. In ACM Press; 2008 [cited 2016 Jul 28]. p. 653. Available from: http://portal.acm.org/citation.cfm?doid=1368088.1368180

31. Greenes RA, Barnett GO, Klein SW, Robbins A, Prior RE. Recording, retrieval and review of medical data by physician-computer interaction. N Engl J Med. 1970 Feb 5;282(6):307–15.

32. Bell DS, Greenes RA, Doubilet P. Form-based clinical input from a structured vocabulary: initial application in ultrasound reporting. Proc Annu Symp Comput Appl Med Care. 1992;789–90.

33. Bell DS, Pattison-Gordon E, Greenes RA. Experiments in concept modeling for radiographic image reports. J Am Med Inform Assoc. 1994 Jun;1(3):249–62.

34. Bell DS, Greenes RA. Evaluation of UltraSTAR: performance of a collaborative structured data entry system. Proc Annu Symp Comput Appl Med Care. 1994;216–22.

35. Kahn CE Jr, Wang K, Bell DS. Structured entry of radiology reports using World Wide Web technology. Radiographics. 1996 May;16(3):683–91.

36. Kahn CE Jr. A generalized language for platform-independent structured reporting. Methods Inf Med. 1997 Aug;36(3):163–71.

37. Kahn CE Jr. Self-documenting structured reports using open information standards. Stud Health Technol Inform. 1998;52 Pt 1:403–7.

38. Gouveia-Oliveira A, Raposo VD, Azevedo AP, Salgado NC, Almeida I, Silva AM, et al. SISCOPE: a multiuser information system for gastrointestinal endoscopy. Endoscopy. 1991 Sep;23(5):272–7.

39. Gouveia-Oliveira A, Raposo VD, Salgado NC, Almeida I, Nobre-Leitão C, de Melo FG. Longitudinal comparative study on the influence of computers on reporting of clinical data. Endoscopy. 1991 Nov;23(6):334–7.

40. Gouveia-Oliveira A, Salgado NC, Azevedo AP, Lopes L, Raposo VD, Almeida I, et al. A unified approach to the design of clinical reporting systems. Methods Inf Med. 1994 Dec;33(5):479–87.

41. Nowlan WA, Rector AL, Goble CA, Horan B, Howkins TJ, Wilson A. PEN&PAD: A Doctors' Workstation with Intelligent Data Entry and Summaries. 1990 Nov 7;941–2.

42. Rector AL, Nolan. Foundations for an Electronic Medical Record. Methods of Information in Medicine. 1991;30:179–86.

43. Goble CA, Glowinski AJ, Nowlan WA, Rector AL. A descriptive semantic formalism for medicine. In IEEE Comput. Soc. Press; 1993 [cited 2013 Nov 1]. p. 624–31. Available from: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=344017

44. Kirby J. Predictive data entry in medical records. 1992;

45. Heathfield HA, Hardiker NR, Kirby J. Using the PEN&PAD information model to support hospital-based clinical care. Proc Annu Symp Comput Appl Med Care. 1994;452–6.

46. McDermott D, Heathfield H, Kirby J. A clerking tool for the patient record system. Medinfo. 1995;8 Pt 2:1664.

47. Heathfield HA, McDermott D, Kirby J. Computer-based medical clerking. Medinfo. 1995;8 Pt 1:377–81.

48. Kirby J, Rector AL. The PEN&PAD data entry system: from prototype to practical system. Proc AMIA Annu Fall Symp. 1996;709–13.

49. Moorman PW, van Ginneken AM, van der Lei J, van Bemmel JH. A model for structured data entry based on explicit descriptional knowledge. Methods Inf Med. 1994 Dec;33(5):454–63.

50. Moorman PW, van Ginneken AM, Siersema PD, van der Lei J, van Bemmel JH. Evaluation of Reporting Based on Descriptional Knowledge. Journal of the American Medical Informatics Association. 1995 Nov 1;2(6):365–73.

51. van Ginneken AM. Structured data entry in ORCA: the strengths of two models combined. Proc AMIA Annu Fall Symp. 1996 Nov;797–801.

52. van Ginneken AM, Stam H, Moorman PW. A multi-strategy approach for medical records of specialists. Int J Biomed Comput. 1996 Jul;42(1-2):21–6.

53. van Ginneken AM, de Wilde M, van Mulligen EM, Stam H. Can data representation and interface demands be reconciled? Approach in ORCA. Proc AMIA Annu Fall Symp. 1997;779–83.

54. van Mulligen EM, Stam H, van Ginneken AM. Clinical data entry. Proc AMIA Symp. 1998;81–5.

55. van Ginneken AM, Stam H, van Mulligen EM, de Wilde M, van Mastrigt R, van Bemmel JH. ORCA: the versatile CPR. Methods Inf Med. 1999 Dec;38(4-5):332–8.

56. Doupi P, van Ginneken AM. Structured physical examination data: a modeling challenge. Stud Health Technol Inform. 2001;84(Pt 1):614–8.

57. Los RK, van Ginneken AM, de Wilde M, van der Lei J. OpenSDE: Row modeling applied to generic structured data entry. J Am Med Inform Assoc. 2004 Apr;11(2):162–5.

58. van Ginneken AM. Considerations for the representation of meta-data for the support of structured data entry. Methods Inf Med. 2003;42(3):226–35.

59. Los RK, van Ginneken AM, van der Lei J. OpenSDE: a strategy for expressive and flexible structured data entry. Int J Med Inform. 2005 Jul 1;74(6):481–90.

60. van Ginneken AM, Verkoijen MJ. A multi-disciplinary approach to a user interface for structured data entry. Stud Health Technol Inform. 2001;84(Pt 1):693–7.

61. Venema AC, van Ginneken AM, de Wilde M, Bogers AJJC. Structured data entry for narrative data in a broad specialty: patient history and physical examination in pediatrics. BMC Med Inform Decis Mak. 2007;7:31.

62. Los RK, van Ginneken AM, van der Lei J. Extracting data recorded with OpenSDE: possibilities and limitations. Int J Med Inform. 2005 Jul 2;74(6):473–80.

63. Los RK, Roukema J, van Ginneken AM, de Wilde M, van der Lei J. Are structured data structured identically? Investigating the uniformity of pediatric patient data recorded using OpenSDE. Methods Inf Med. 2005 Aug;44(5):631–8.

64. Los RK, van Ginneken AM, Roukema J, Moll HA, van der Lei J. Why are structured data different? Relating differences in data representation to the rationale of OpenSDE. Med Inform Internet Med. 2005 Dec;30(4):267–76.

65. Roukema J, Los RK, Bleeker SE, van Ginneken AM, van der Lei J, Moll HA. Paper versus computer: feasibility of an electronic medical record in general pediatrics. Pediatrics. 2006 Jan;117(1):15–21.

66. Venema AC, van Ginneken AM, de Wilde M, Bogers AJJC. Is OpenSDE an alternative for dedicated medical research databases? An example in coronary surgery. BMC Med Inform Decis Mak. 2007;7:31.

67. Douglas K, Nubie M. System design challenges: the integration of controlled vocabulary use into daily practice. Stud Health Technol Inform. 1997;46:167–71.

68. DeFriece RJ. Design considerations for intelligent data entry: development of MedIO. Proc Annu Symp Comput Appl Med Care. 1995;91–5.

69. Logan JR, Klopfer KC. The use of a standardized terminology for comparison of free text and structured data entry. Proc AMIA Symp. 2000;512–6.

70. Marin HF. Comparing nursing terms from patient records to the ICNP, beta version. Stud Health Technol Inform. 2002;90:232–5.

71. Cheung NT, Fung V, Chow YY, Tung Y. Structured data entry of clinical information for documentation and data collection. Stud Health Technol Inform. 2001;84(Pt 1):609–13.

72. Rosenbloom ST, Miller RA, Johnson KB, Elkin PL, Brown SH. Interface terminologies: facilitating direct entry of clinical data into electronic health record systems. J Am Med Inform Assoc. 2006 Jun;13(3):277–88.

73. Rosenbloom ST, Brown SH, Froehling D, Bauer BA, Wahner-Roedler DL, Gregg WM, et al. Using SNOMED CT to represent two interface terminologies. J Am Med Inform Assoc. 2009 Feb;16(1):81–8.

74. Rosenbloom ST, Miller RA, Johnson KB, Elkin PL, Brown SH. A model for evaluating interface terminologies. J Am Med Inform Assoc. 2008 Feb;15(1):65–76.

75. Daniel-Le Bozec C, Steichen O, Dart T, Jaulent M-C. The role of local terminologies in electronic health records. The HEGP experience. Stud Health Technol Inform. 2007;129(Pt 1):780–4.

76. Hersh WR, Campbell EH, Evans DA, Brownlow ND. Empirical, automated vocabulary discovery using large text corpora and advanced natural language processing tools. Proc AMIA Annu Fall Symp. 1996;159–63.

77. Kreis C, Gorman P. Word frequency analysis of dictated clinical data: a user-centered approach to the design of a structured data entry interface. Proc AMIA Annu Fall Symp. 1997;724–8.

78. Kashyap V, Turchin A, Morin L, Chang F, Li Q, Hongsermeier T. Creation of structured documentation templates using Natural Language Processing techniques. AMIA Annu Symp Proc. 2006;977.

79. Madani S, Mirza J, Shultz E. Identification of pre-coordinated term candidates in a Cardiology Outpatient Service. AMIA Annu Symp Proc. 2008;1036.

80. Kim Y, Park H, Kim H-G, Kim YO. The Development of Medical Record Items: a User-centered, Bottom-up Approach. Healthc Inform Res. 2012 Mar;18(1):10–7.

81. Rosenbloom ST, Miller RA, Adams P, Madani S, Khan N, Shultz EK. Implementing an interface terminology for structured clinical documentation. J Am Med Inform Assoc. 2013 Jun;20(e1):e178–82.

82. Brown ML, Quiñonez LG, Schaff HV. A pilot study of electronic cardiovascular operative notes: qualitative assessment and challenges in implementation. J Am Coll Surg. 2010 Feb;210(2):178–84.

83. Hanauer DA, Liu Y, Mei Q, Manion FJ, Balis UJ, Zheng K. Hedging their mets: the use of uncertainty terms in clinical documents and its potential implications when sharing the documents with patients. AMIA Annu Symp Proc. 2012;2012:321–30.

84. Varma V, Varma S, Haq S, Haq MM, Raju N, Varma R. Structured data system for a breast cancer medical record. Stud Health Technol Inform. 2009;143:354–7.

85. Nielsen J. Usability engineering [Internet]. San Francisco, Calif.: Morgan Kaufmann Publishers; 1994 [cited 2013 Dec 29]. Available from: http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=582280

86. Abran A, Khelifi A, Suryn W, Seffah A. Usability Meanings and Interpretations in ISO Standards. Software Quality Journal. 2003;11(4):325–38.

87. Spenceley SE, Warren JR. The intelligent interface for online electronic medical records using temporal data mining. In IEEE Comput. Soc; 1998 [cited 2013 Oct 17]. p. 266–74. Available from: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=648321

88. Hua L, Wang S, Gong Y. Text prediction on structured data entry in healthcare: a two-group randomized usability study measuring the prediction impact on user performance. Appl Clin Inform. 2014;5(1):249–63.

89. Hua L, Gong Y. Toward User-Centered Patient Safety Event Reporting System: A Trial of Text Prediction in Clinical Data Entry. Stud Health Technol Inform. 2015;216:188–92.

90. Canfield K. Priming intelligent split menus with text corpora for computerized patient record data-entry. Int J Biomed Comput. 1995 May;39(2):263–73.

91. Hogan WR, Wagner MM. Free-text fields change the meaning of coded data. Proc AMIA Annu Fall Symp. 1996;517–21.

92. Ahlbrandt J, Henrich M, Hartmann BA, Bundschuh BB, Schwarz J, Klasen J, et al. Small cause - big effect: improvement in interface design results in improved data quality - a multicenter crossover study. Stud Health Technol Inform. 2012;180:393–7.

93. Barnes SL, Waterman M, Macintyre D, Coughenour J, Kessel J. Impact of standardized trauma documentation to the hospital's bottom line. Surgery. 2010 Oct;148(4):793–7; discussion 797–8.

94. McCullagh PJ, McGuigan J, Fegan M, Lowe-Strong A. Structure data entry using graphical input: recording symptoms for multiple sclerosis. Stud Health Technol Inform. 2003;95:673–8.

95. Lowe-Strong A, McCullagh PJ. Monitoring of symptoms and interventions associated with multiple sclerosis. Stud Health Technol Inform. 2005;117:223–8.

96. Zheng K, Padman R, Johnson MP, Diamond HS. An interface-driven analysis of user interactions with an electronic health records system. J Am Med Inform Assoc. 2009 Apr;16(2):228–37.

97. Lussier YA, Maksud M, Desruisseaux B, Yale PP, St-Arneault R. PureMD: a Computerized Patient Record software for direct data entry by physicians using a keyboard-free pen-based portable computer. Proc Annu Symp Comput Appl Med Care. 1992;261–4.

98. Poon AD, Fagan LM. PEN-Ivory: the design and evaluation of a pen-based computer system for structured data entry. Proc Annu Symp Comput Appl Med Care. 1994;447–51.

99. Shultz E, Rosenbloom T, Kiepek W, Fitzhenry F, Adams P, Mahuli A, et al. Quill: a novel approach to structured reporting. AMIA Annu Symp Proc. 2003;1074.

100. Rosenbloom ST, Kiepek W, Belletti J, Adams P, Shuxteau K, Johnson KB, et al. Generating complex clinical documents using structured entry and reporting. Stud Health Technol Inform. 2004;107(Pt 1):683–7.

101. Rosenbloom ST, Crow AN, Blackford JU, Johnson KB. Cognitive factors influencing perceptions of clinical documentation tools. J Biomed Inform. 2007 Apr;40(2):106–13.

102. Rosenbloom ST, Stead WW, Denny JC, Giuse D, Lorenzi NM, Brown SH, et al. Generating Clinical Notes for Electronic Health Record Systems. Appl Clin Inform. 2010 Jan 1;1(3):232–43.

103. Rasmussen SL, Lyng KM, Jensen S. Achieving IT-supported standardized nursing documentation through participatory design. Stud Health Technol Inform. 2012;180:1055–9.

104. Elrod J, Androwich IM. Applying human factors analysis to the design of the electronic health record. Stud Health Technol Inform. 2009;146:132–6.

105. Horsky J, McColgan K, Pang JE, Melnikas AJ, Linder JA, Schnipper JL, et al. Complementary methods of system usability evaluation: surveys and observations during software design and development cycles. J Biomed Inform. 2010 Oct;43(5):782–90.

106. Thyvalikakath TP, Schleyer TKL, Monaco V. Heuristic evaluation of clinical functions in four practice management systems: a pilot study. J Am Dent Assoc. 2007 Feb;138(2):209–10, 212–8.

107. Thyvalikakath TP, Monaco V, Thambuganipalle HB, Schleyer T. A usability evaluation of four commercial dental computer-based patient record systems. J Am Dent Assoc. 2008 Dec;139(12):1632–42.

108. Walji MF, Kalenderian E, Tran D, Kookal KK, Nguyen V, Tokede O, et al. Detection and characterization of usability problems in structured data entry interfaces in dentistry. Int J Med Inform. 2013 Feb;82(2):128–38.

109. Bostrom PJ, Toren PJ, Xi H, Chow R, Truong T, Liu J, et al. Point-of-care clinical documentation: assessment of a bladder cancer informatics tool (eCancerCareBladder): a randomized controlled study of efficacy, efficiency and user friendliness compared with standard electronic medical records. J Am Med Inform Assoc. 2011 Dec;18(6):835–41.

110. Vahl CF, Tochtermann U, Gams E, Hagl S. Efficiency of a computer network in the administrative and medical field of cardiac surgery. Concept of and experience with a departmental system. Eur J Cardiothorac Surg. 1990;4(12):632–8.

111. Apkon M, Singhaviranon P. Impact of an electronic information system on physician workflow and data collection in the intensive care unit. Intensive Care Med. 2001 Jan;27(1):122–30.

112. Bürkle T, Martin M, Schütz A, Starke K, Wagner S, Ries M, et al. Workflows in cancer treatment and their influence upon clinical documentation. Stud Health Technol Inform. 2013;192:1181.

113. Wagner S, Beckmann MW, Wullich B, Seggewies C, Ries M, Bürkle T, et al. Analysis and classification of oncology activities on the way to workflow based single source documentation in clinical information systems. BMC Med Inform Decis Mak. 2015;15:107.

114. Carlson KL, McFadden SE, Barkin S. Improving Documentation Timeliness: A "Brighter Future" for the Electronic Medical Record in Resident Clinics. Acad Med. 2015 Dec;90(12):1641–5.

115. Webster C, Copenhaver J. Structured data entry in a workflow-enabled electronic patient record. J Med Pract Manage. 2001 Dec;17(3):157–61.

116. Becker TK, Gries A, Martin E, Bernhard M. Experiences with a PDA-based documentation system in clinical research. J Med Syst. 2012 Apr;36(2):647–51.

117. Winkler C, Seifert J, Reinartz C, Krahmer P, Rukzio E. Penbook: bringing pen+paper interaction to a tablet device to facilitate paper-based workflows in the hospital domain. In ACM Press; 2013 [cited 2013 Oct 31]. p. 283–6. Available from: http://dl.acm.org/citation.cfm?doid=2512349.2512797

118. Kaka H, Ayearst R, Tran M, Touma Z, Bagovich M, Vinik O, et al. DEVELOPING AN IPAD® APPLICATION FOR DATA COLLECTION IN A RHEUMATOLOGY RESEARCH CLINIC. Int J Technol Assess Health Care. 2015 Jan;31(1-2):99–102.

119. Embi PJ, Weir C, Efthimiadis EN, Thielke SM, Hedeen AN, Hammond KW. Computerized provider documentation: findings and implications of a multisite study of clinicians and administrators. J Am Med Inform Assoc. 2013 Aug;20(4):718–26.

120. Saleem JJ, Adams S, Frankel RM, Doebbeling BN, Patterson ES. Efficiency strategies for facilitating computerized clinical documentation in ambulatory care. Stud Health Technol Inform. 2013;192:13–7.

121. Knaup P, Garde S, Haux R. Systematic planning of patient records for cooperative care and multicenter research. Int J Med Inform. 2007 Mar;76(2-3):109–17.

122. Cusack CM, Hripcsak G, Bloomrosen M, Rosenbloom ST, Weaver CA, Wright A, et al. The future state of clinical data capture and documentation: a report from AMIA's 2011 Policy Meeting. J Am Med Inform Assoc. 2013 Jan 1;20(1):134–40.

123. Payne TH, Corley S, Cullen TA, Gandhi TK, Harrington L, Kuperman GJ, et al. Report of the AMIA EHR 2020 task force on the status and future direction of EHRs. J Am Med Inform Assoc. 2015 May 28;

124. Cruz-Correia R, Boldt I, Lapão L, Santos-Pereira C, Rodrigues PP, Ferreira AM, et al. Analysis of the quality of hospital information systems Audit Trails. BMC Med Inform Decis Mak. 2013;13:84.

125. Nielsen J, Molich R. Heuristic evaluation of user interfaces. In ACM Press; 1990 [cited 2016 Aug 2]. p. 249–56. Available from: http://portal.acm.org/citation.cfm?doid=97243.97281

126. Nielsen J. Enhancing the explanatory power of usability heuristics. In ACM Press; 1994 [cited 2016 Aug 2]. p. 152–8. Available from: http://portal.acm.org/citation.cfm?doid=191666.191729

2.7 Appendix - Attributes of data analysis

Seq.	Description	Example
1	Document ID, e.g. PMID or DOI	pubmed_1778139
2	Number of Citations on Google Scholar	40
3	Study Type (Method or Evaluation or both)	Evaluation
4	Author(s)	Gouveia-Oliveira et al.
5	Title	Longitudinal comparative study on the influence of computers on reporting of clinical data
6	Year	1991
7	Objective / Problems	use structure data entry system result in 1. unnatural language style of report 2. loss of info however, narrative data also have high missing info and no controlled terms
8	Nature of Innovation	fully structured data entry system and natural language report generation
9	Method	NA
10	Specialty / Clinical Setting(s)	Endoscopy
11	Participants	1925 examinations in 4 types of lesions physicians
12	System(s)	SISCOPE
13	Device (Desktop, Laptop, Tablet, etc)	NA
14	Evaluation	comparison with free text t-5mo (=1mo) vs. t(=12mo) vs. t+2mo (=1mo)
15	Measures	1. physician compliance = computer-generated reports / actual # examinations 2. information efficiency = recorded # data items / total # data items 3. educational effect, whether free-text entry will be more accurate after the intervention
16	Performance / Findings	1. acceptance was constantly high 2. SISCOP 60% more info than free-text, 18% vs. 48% error rate 3. no educational effect, need to provide feedback to remind docs of important descriptive items
17	Limitation	NA
18	Notes	reference of pubmed_1743125

CHAPTER 3

Using EHR Audit Trail Logs for Clinical Workflow Analysis: An Example from Ambulatory Clinics

3.1 Introduction

In recent years, clinical data capture and documentation has increasingly been supported by computer systems, especially by Electronic Health Record (EHR) systems, as incentivized and regulated by the HITECH Act and the Meaningful Use (1,2). While computer-based documentation (CBD) may be preferred due to its improvements in availability, shareability, and reusability of clinical data (3,4), it may introduce negative impacts. For example, a poorly designed CBD system may fail to consider clinical workflow so that it contributes to cumbersome documentation activities as well as the capture of inaccurate and incomplete data (5).

In order to design a workflow-integrated CBD system, a critical first step is to understand and measure current workflow and documentation behaviors. Researchers and engineers have been using gold standard techniques such as Time and Motion (T&M) studies to achieve this goal (6). T&M studies, however, are often costly and incapable of collecting data in a large scale because of the challenges in recruiting numerous participants and observers as well as coordinating observation sessions. Meanwhile, T&M studies are inherently limited in accuracy due to human observers' cognitive overload and biases, resulting in suboptimal quality of behavioral data such as incompleteness and misinterpretation.

An alternative method to make sense of clinical workflow is through the secondary use of EHR data, which has drawn significant attention from researchers. For example, Ozkaynak et al. (2015) characterize the workflow in pediatric emergency rooms by analyzing encounter and

34

diagnosis data using Markov chain and visualization tools (7). Kirkendall et al. (2014) understood the workflow of medication order-dosing alerts by simulating order activities in a testing EHR environment (8). Redd et al. (2014) measured the efficiency of documentation by extracting the chart open and completion time in each patient encounter (9). These studies, although suitable for their purposes, only utilized EHR data on a surface level. The data and methods in these studies did not have the ability to uncover more nuanced behavioral patterns of a large amount of individual clinicians.

This issue can be addressed by using a special type of EHR data, i.e. EHR audit trail logs, which automatically and minutely record clinicians' behaviors for security and privacy purposes. These logs can be a cost-effective alternative to characterize individuals' documentation behaviors and have the potential to facilitate the understanding of clinical workflow and processes. For example, Hirsh et al. used audit files to understand time spending in primary care (10). However, to the best knowledge of this author, there is no study evaluating EHR audit trail logs as a valid source for clinical workflow analysis. The present study bridges this methodological gap by assessing the ability of EHR logs to provide supportive evidence to known workflow changes. Specifically, the study analyzes a sample of EHR audit trail log data extracted from an ambulatory healthcare organization. This organization launched an EHR enhancement project in mid 2013, resulting in subsequent workflow and behavioral changes among the clinicians as identified in a previous study (11). The present study chooses three findings from the previous work and evaluates the ability of the EHR audit trail log data to provide supportive evidence to these known behavioral changes. Then, the implications of utilizing such log data for clinical workflow analyses are discussed.

3.2 Methods

3.2.1 Study Site

The participating study site was a not-for-profit ambulatory healthcare organization located in an urban area in the western United States (Organization West). This organization contained more than 20 branches and affiliated locations including both primary care and specialty clinics. Organization West served several states over a widely rural area, with over 50% of patients being rural residents and around 50% of who are on Medicare/Medicaid. After nearly 10 years of using an EHR system (Cerner[1]), Organization West launched a significant update to enhance the system capability in 2013, such as an electronic patient homepage, a standardized message center, a new computerized provider order entry (CPOE), and an e-prescribing functionality. During the EHR enhancement, one of the clinics experimented with a "Core Team Model," in which a provider was teamed up with three clinical staff members to eliminate inefficiencies and patient wait times.

3.2.2 Data Collection

This study targeted three clinics in Organization West, including a primary care clinic in the main hospital (Primary Care 1), another primary clinic that experimented with the "Core Team Model" (Primary Care 2), and an Ear Nose Throat clinic (Specialty Care). A total of 24 clinicians were participated and equally distributed in each clinic. For each primary care clinic, the study participants included four providers, two nurses, one medical assistant, and one receptionist. For the specialty care clinic, the study participants included five providers, two nurses, one medical assistant, and no receptionist. The providers included in the study were physicians (MD), ophthalmologists (OD), nurse practitioners (NP), and physician assistants (PA). This sample data allowed the comparison of clinicians' behavioral changes in many dimensions, e.g. primary care vs. specialty care; with and without the "Core Team Model"; and providers vs. non-providers.

---

[1] Cerner Cooperation. http://www.cerner.com/

The participants' audit trail log data were extracted based on the timestamp. These data were de-identified prior to being presented to the research team.

3.2.3 Data Categorization

The log data were stored in a standalone SQLite database and categorized using Structured Query Language. The log data were categorized into three stages based on their timestamp: 1) PRE: log records comprised of a period of eight weeks prior to the launch date, 2) DURING: log records for a period of four weeks after the launch date, and 3) POST: log records of the eight weeks after the end of the DURING stage. The analysis only focused on the data in the PRE and POST stages to maximize the potential differences. The start and end of a week was set from Monday to Sunday. Moreover, the date and time information of the timestamps were parsed and stored separately, e.g. date, hour, and day of the week. The time information was further categorized into three types: 1) Regular hours (REG): time between 7am and 6pm from Monday to Friday; 2) off hours (OFF): time not in regular hours from Monday to Friday; and 3) weekend hours (WKN): anytime during Saturday and Sunday only.

In addition, the clinical roles in the log data were marked as Provider (P), Nurse (N), Medical Assistant (A), and Receptionist (R). Note that the study created a new event category in addition to the original event names and tasks defined by the EHR vendor. This new event category included seven groups: 1) Communication, 2) Entering, 3) Login/Logout, 4) Processing, 5) Reading, 6) Printing, and 7) Other. Each event was first mapped to a category separately by this author and a research assistant. Then, the mappings were discussed and an agreement was reached. If a consensus of an event mapping cannot be reached, this event will be categorized as "Other".

3.2.4 Hypotheses

The primary objective of this study was to evaluate the ability of EHR audit trail log data to provide supportive evidence to the known workflow changes. In particular, it aimed to reflect documentation-related behavioral changes among clinicians after the rollout of an EHR enhancement project. Three previous findings were chosen as reference standards of the evaluations: the first evaluation involved a workaround order-placing behavior in Primary Care 1, where a provider would have to prescribe a medication more than once in order to achieve the desired dose in the POST stage. For example, a provider would place two medication orders, one at 1.5 mg and the other at 0.5 mg, to equal the total intended dose of 2 mg. Since the order dosage was not captured in the log data, the providers in Primary Care 1 were alternatively hypothesized to have more number of order logs daily. Their patient volume in the PRE and POST stage was hypothesized to remain stable.

The second evaluation was a shift of tasks in Primary Care 2, where providers spent less time on computer-related tasks after the EHR enhancement project, while other clinicians consumed those tasks due to the implementation of the Core Team model. The providers in Primary Care 2 were hypothesized to have fewer computer-entering-related logs in the POST stage, compared to the PRE stage. On the other hand, non-providers in the POST stage were hypothesized to have more computer-entering-related logs.

The third evaluation was the phenomenon of "deferred documentation" in the Specialty Care clinic, highlighting differences in dictation. It is noted that providers in the PRE stage would dictate findings between patient visits, while in the POST stage they tended to document patient visits at the end of a day. It is hypothesized that providers in Specialty Care would have an increased number of logs in the off hours (and even weekends) in the POST stage.

3.2.5 Data Analysis

The data analysis was supported by Python, including libraries such as numpy, scipy, and xlrd. The analysis used the total number of logs and the percentage of logs as two primary measures. Specifically, the analysis began with a statistical summary of the logs, followed by three evaluations. For the hypotheses of the first evaluation, the log data related to "adding an order" were selected. The total number of orders placed per day was counted, and the days without any orders placed were dropped. The daily patient volume was estimated by the total number of patients accessed per day. For the second evaluation, the total number and the percentage of events in each coded event categories (e.g. Entering and Reading) were calculated. For the third evaluation, the same measures were used to quantify the off-hour and weekend logs. All the above measures in the PRE and POST stage were compared using a two-tailed, unequal variance t-test.

In addition, for the second evaluation, the fragmentation of the log events was assessed based on their categories. This measure, as proposed by Zheng et al. (12), calculated a daily fragmentation score to represent the level of cognitive overload. For each date, the log event categories were summarized based on their timestamp and sorted in an ascending manner. It is noted that logs with varying categories can have identical timestamp because multiple documentation behaviors may happen, or be triggered, at the same time. Fragmentation was detected if there was no overlap between the event categories of the current timestamp and its previous one. A daily fragmentation score was defined as the total number of detected fragmentation over the total number of timestamps on that day. The clinicians' daily fragmentation scores in the PRE and POST stage were compared using the same t-test technique in the other evaluations.

3.3 Results

3.3.1 Statistical Summary

A total of 2.65 million EHR audit trail logs were extracted from the participated organization. The EHR logs contained records in the PRE and POST stage, for a period of eight weeks each.

Overall, the total number of logs of Primary Care 1 & 2 was approximately the same, whereas the log volume of Specialty Care was half of the previous two. This is likely due to the differences of clinical practices between a primary and a specialty care clinic (Table 3.1). In terms of the log volume by clinical roles, the providers in Primary Care 2 had a much lower number than the nurses (87,600 vs. 141,090). In comparison, the average number of logs of the providers and nurses in Primary Care 1 were more comparable (111,351 vs. 116,078). A further examination of the log volume considering the PRE and POST stage is listed in Table 3.2. This table shows that the nurses in Primary Care 1 had a 20% decrease of their log volumes and the providers had a 20% increase and in the POST stage. On the other hand, the nurses in Primary Care 2 had a 10% increase in their log volumes while the providers maintain a similar number of logs. This observation might be attributed to the implantation of the Core Team model in Primary Care 2, which increased the nurses' documentation responsibility.

Table 3.1: Summary of the sample EHR audit trail log data.

Clinic	PRE (8 wks)	Launch Date	POST (8 wks)	Total	Provider (avg. total)	Nurse (avg. total)	Assistant (avg. total)	Receptionist (avg. total)
Primary Care 1	5/20/13 7/14/13	7/16/13	8/12/13 10/6/13	1,149,295	4 (111,351)	2 (116,078)	1 (215,347)	1 (256,386)
Primary Care 2	6/24/13 8/18/13	8/20/13	9/16/13 11/10/13	971,210	4 (87,600)	2 (141,090)	1 (143,758)	1 (194,868)
Specialty Care	8/5/13 9/29/13	10/1/13	10/28/13 12/22/13	529,999	5 (64,099)	2 (74,045)	1 (61,410)	-

Table 3.2: The average number of logs in the PRE and POST stage (providers and nurses).

Clinic/ Role-Stage	Provider (4)		Nurse (4)	
	PRE	POST	PRE	POST
Primary Care 1	50,591	60,759	64,664	51,414
Primary Care 2	43,452	44,148	66,782	74,308
Specialty Care	36,605	27,493	46,680	27,365

3.3.2 The First Evaluation

As shown in Table 3.3, the statistical test of the first evaluation indicated that the providers in Primary Care 1 placed significantly more orders in the POST stage (all $p<0.05$) and maintained the same number of patients seen, which is consistent with the hypotheses. The same statistical test was extended to the providers in Primary Care 2, which didn't show the same pattern. As shown in the second half of Table 3.3, two of the four providers in Primary Care 2 had no significant change in their order numbers, with one of them seeing more patients in the POST stage. This finding suggests that changes of EHR systems may impact clinicians' behaviors differently even though they operate within the same heath system and behave similarly in their clinical practice.

Table 3.3: Number of orders placed and patients seen for each provider in the two primary care clinics.

Clinic	Measure	Provider	PRE	POST	P-value
Primary Care 1	Number of Orders Placed	BC1-04	12.00	35.25	0.0 *
		BC1-05	42.35	53.15	0.0192 *
		BC1-09	44.68	53.53	0.0433 *
		BC1-10	15.94	30.14	0.0 *
	Number of Patients Accessed	BC1-04	42.50	34.75	0.3491
		BC1-05	42.57	42.58	0.9971
		BC1-09	31.45	35.32	0.1939
		BC1-10	40.47	41.94	0.6158
Primary Care 2	Number of Orders Placed	BC2-02	23.04	32.11	0.0066 *
		BC2-03	16.69	16.07	0.8386
		BC2-04	28.89	38.12	0.0342 *
		BC2-16	27.38	33.84	0.1826
	Number of Patients Accessed	BC2-02	31.77	28.34	0.1475
		BC2-03	29.47	35.11	0.0399 *
		BC2-04	36.39	34.46	0.5024
		BC2-16	23.14	22.04	0.6417

* p-value < 0.05

3.3.3 The Second Evaluation

The second evaluation began at coding the log events to the pre-defined high-level categories. This author and the research assistant achieved 81% agreement on coding the event categories initially, and resolved most of the discrepancies afterwards. Table 3.4 lists top ten frequent events that were assigned to "Computer Entering", and their vender defined names and types as well as their frequencies in Primary Care 2.

Table 3.4: Top 10 "Computer Entering" events

EVENT TYPE	EVENT NAME	COUNT	PCT
Maintain Person	Chart Access Log	141,616	42%
Maintain Person	Ensure	70,647	21%
Maintain Encounter	Patient-Provider Relations	22,068	7%
View Orders	Modify Details	16,719	5%
Maintain Encounter	Ensure	14,681	4%
Maintain Person	Patient	12,921	4%
Maintain Clinical Events	Write/Update Results	11,331	3%
Maintain Order	Add	9,770	3%
Maintain Order	Tasks	5,939	2%
Maintain Encounter	Prsnl Relationship	4,894	1%

The statistical test examined the differences of the computer-entering-related events between the PRE and POST stage. The results (Table 3.5) showed that the participated nurses and the medical assistant in Primary Care 2 had a significant increase in their percentage of computer entering logs as hypothesized. However, the receptionist had an unexpected drop in their percentage of entering-related logs (53% -> 49%). On the other hand, the providers had a slight but not significant decrease in their number and percentage of logs, which is not consistent with the hypothesis. The workflow fragmentation analysis of these providers (Table 3.6) indicated that three out of four providers had a lower level of fragmentation, although only one of them was statistically significant. The other providers had a slight but not significant increase in the level of fragmentation (0.398 -> 0.401). Together, the analysis of the second evaluation supports the hypotheses of shift clinical workflow, where the nurses and the medical assistant had an

increased percentage of computer-entry-related logs. While the providers didn't have a significant increase, they had an improved level of fragmentation in the POST stage, which may help reduce their cognitive overload, leading to their positive attitude toward the EHR change (i.e. more time with patients).

Table 3.5: The average number and percentage of "Computer Entering" logs in the PRE and POST stage.

Clinical Role	Participant ID	Number of Logs			Percentage of Logs		
		PRE	POST	p-value	PRE	POST	p-value
P	BC2-02	364.19	354.72	0.7983	24%	24%	0.6771
P	BC2-03	484.57	486.17	0.9743	29%	27%	0.0889
P	BC2-04	327.75	294.79	0.4302	22%	21%	0.4339
P	BC2-16	341.48	318.71	0.6085	27%	27%	0.9254
N	BC2-11	515.22	655.14	0.005*	27%	32%	0.0*
N	BC2-19	515.33	603.91	0.0516	24%	27%	0.0*
A	BC2-22	673.34	683.12	0.8806	28%	31%	0.0*
R	BC2-45	1375.36	1236.83	0.0737	53%	49%	0.0*

* 0.05 significance

Table 3.6: The workflow fragmentation analysis of the providers in the Primary Care 2.

Participant ID	PRE	POST	Difference	p-value
BC2-02	0.428	0.361	-0.067	0.0043*
BC2-03	0.388	0.365	-0.023	0.2669
BC2-04	0.358	0.306	-0.052	0.1305
BC2-16	0.398	0.401	0.003	0.8725

* 0.05 significance

3.3.4 The Third Evaluation

Table 3.7 summarizes the statistical test of the providers' weekly off-hour logs in Specialty Care. Three providers had a 2-6% increase while the other two providers had a 2% (or less) drop. Unfortunately, none of the changes were significant at the 0.05 level. The statistical tests of the weekend logs also found no significant differences. A further exploratory using visualization (line charts) suggested that providers in Specialty Care had increased number of logs right after 5pm (Figure 3.1). The statistical test was then performed again on the hourly summary of the log data. As shown in Table 3.8, four of five providers in Specialty Care had a significantly higher number of logs. They deferred documentation activities in at least one of the hours between 5pm and 10pm. The other provider (BC3-01) seemly had deferred the documentation activities in the PRE stage, which makes the EHR change less impactful in the POST stage.

Table 3.7: The average number and percentage of weekly off-hour logs in the PRE and POST stage.

Participant ID	Number of off-hour Logs			Percentage of off-hour Logs		
	PRE	POST	p-value	PRE	POST	p-value
BC3-01	761.43	1,093.0	0.3598	0.10	0.16	0.2120
BC3-04	73.50	283.5	0.0818	0.02	0.04	0.2749
BC3-05	462.88	688.2	0.3625	0.09	0.11	0.7078
BC3-10	303.29	167.4	0.2173	0.05	0.03	0.1476
BC3-14	0	9.6	0.1453	0	0	0.1357

Figure 3.1: The visualization of the number of off-hour logs of the provider BC3-04.
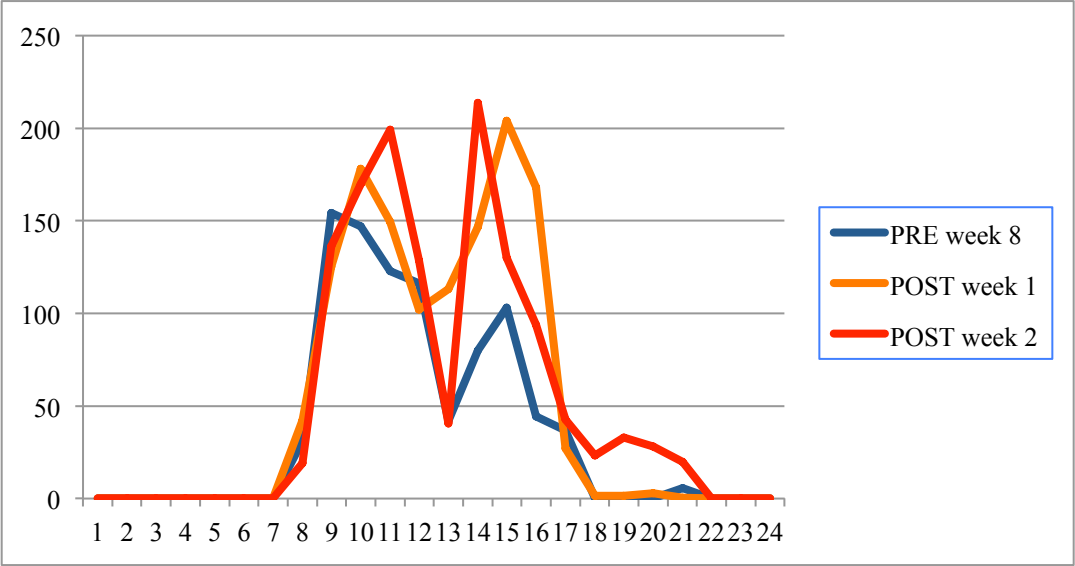
Table 3.8: The average number of daily off-hour log in the PRE and POST stage.

Participant ID	Hour	PRE	POST	p-value
BC3-01	5pm	173.26	147.7	0.7476
	6pm	118.96	186.5	0.3326
	7pm	61.93	86.25	0.6350
	8pm	11.07	0.50	0.1782
	9pm	0	0	-
BC3-04	5pm	0.91	16.65	0.0263 *
	6pm	0.36	19.10	0.0373 *
	7pm	0.33	12.00	0.064 **
	8pm	4.61	12.05	0.2927
	9pm	3.33	0	0.4416
BC3-05	5pm	53.14	64.42	0.6525
	6pm	17.05	13.50	0.8124
	7pm	31.97	0.04	0.1661
	8pm	12.11	0.29	0.3917
	9pm	0.14	26.75	0.0393 *
BC3-10	5pm	71.74	88.26	0.6039
	6pm	43.71	19.22	0.2146
	7pm	0	0.35	0.0978 **
	8pm	5.03	0.35	0.4277
	9pm	0	0.13	0.1174
BC3-14	5pm	0.72	6.78	0.0626 **
	6pm	0	2.00	0.1717
	7pm	0	0.09	0.2417
	8pm	0	0	-
	9pm	0	0	-

* 0.05 significance; ** 0.10 significance.

3.4 Discussion

This study evaluated the feasibility of using EHR audit trail logs as a valid data source for clinical workflow analysis. The results showed that EHR audit trail logs contained rich information of clinicians' documentation behaviors and workflow. The analyses base on the log data successfully produced supportive evidence to the three known behavioral and workflow changes. The study showed several benefits of using EHR audit trail logs to analyze clinical workflow. First if all, since the log data were recorded automatically and consistently, an analysis of one clinic can be easily applied to another clinic with almost no extra cost. Take the first evaluation for example, the changes of the order numbers and patient volumes was initially examined in Primary Care 1 and then directly applied to Primary Care 2 with nearly no extra effort. This benefit of easy extension makes this method more attractive than other workflow analysis tools, e.g. Time and Motion studies and work sampling.

Another benefit of using EHR logs for workflow analysis is to provide an objective view of clinicians' EHR behaviors and to verify the findings derived from other methods. Take the second evaluation for example, the previous study suggested that the providers in Primary Care 2 had more time for patient communication because they shifted documentation responsibilities to other clinicians as a result of the Core Team model. The analysis of the present study, however, showed that the providers' EHR log volume did not drop significantly although other clinicians indeed had a significant increase of the log volume. One possible explanation is that EHR audit trail logs may not fully capture the documentation process as they are not design for such intension. Another explanation of this phenomenon, as suggested by the workflow fragmentation analysis, is that the providers may have less fragmented documentation activities, leading to a smoother workflow and higher quality of patient interactions.

Still another benefit of using EHR logs for workflow analysis is the ability to compare and contrast results in multiple dimensions. In this study, the dimensions included clinical roles (providers vs. non-providers), care types (primary vs. specialty), care models (normal vs. the

47

core team model), time (regular vs. off-hour), and event categories (Entering vs. Other). Moreover, EHR audit trail logs enable detailed comparison showing the differences of individual clinician's behaviors. Take the third evaluation for example, the analysis not merely supported hypothesis of deferred documentation in Specialty Clinic, but further indicated which hour(s) the documentation activities were deferred to, which were varying among the five providers.

Although the rich information in EHR audit trail logs has several benefits, it can introduce analytical challenges. One such challenge is the high complexity and high level of noise in the log data. It could be very difficult to identify meaningful patterns solely based on EHR log data. The present study voided this problem by utilizing the findings from the previous study to shed light on the analysis. This suggests that using EHR logs for clinical workflow analysis may require mixed methods, e.g. integrating the qualitative results to the quantitative analysis. It may also require extensive effort on data exploratory, which can be supported by visual analytics techniques. Another challenge is that EHR logs may require high-level categorization to reduce their complexity and to make them more meaningful. For example, in the present study, the logs were manually categorized into seven groups in order to compare "Computer Entering" activities with other kinds of activities. Since each vendor has a different mechanism capturing audit trail logs, and since each study has different analysis goals, it seems that researchers would need to spend a significant amount of time labeling EHR logs. Data mining and machine learning techniques may help facilitate this process.

This study has several limitations. First, this study only analyzed EHR audit trail logs of one organization in an ambulatory setting. Since this is a proof-of-concept study, this author believes that the scope is adequate. Future research can consider expanding the scope by analyzing EHR logs from multiple organizations, from different EHR systems, and/or in an inpatient setting. Second, the study did not obtain the EHR audit trail logs of all clinicians in the targeted organization, but only the logs associated with the 24 participants in the three selected clinics. This limitation prevents the study from performing a comprehensive patient-centered analysis

48

because the selected participants may not work on the same patients in the study period. Next, the author did not have the opportunity to observe clinical activities and make sense of these activities with their EHR logs. Rather, the author understood the log data based on the recorded attributes such as event types, names, and the sequence and timestamp of the logs. To mitigate potential biases in the interpretation, a research assistant was hired to provide viewpoints. The research assistant also helped categorize the logs and resolve the discrepancy between the author's and the assistant's categorization. Last but not least, many results were not significant at the 0.05 level, and even not significant at the 0.1 level. One may conclude that these results were not promising. However, just because a difference is not statistically significant, doesn't mean that it has no clinical impact. These insignificant results still showed changes (increase or decrease of the means) aligned with the hypotheses.

3.5 Conclusions

This study evaluated the validity of using EHR audit trail logs for clinical workflow analysis. The three evaluations successfully provided supporting evidences and justify the use of the data. Future research can focus on developing a generalized analysis framework that utilizes this valid data source to monitor clinical workflow and uncover more nuance patterns of behavioral changes.

3.6 References

1.   Blumenthal D. Launching HITECH. N Engl J Med. 2010 Feb 4;362(5):382–5.

2.   Blumenthal D, Tavenner M. The "Meaningful Use" Regulation for Electronic Health Records. New England Journal of Medicine. 2010 Aug 5;363(6):501–4.

3.   Embi PJ, Yackel TR, Logan JR, Bowen JL, Cooney TG, Gorman PN. Impacts of computerized physician documentation in a teaching hospital: perceptions of faculty and resident physicians. J Am Med Inform Assoc. 2004 Aug;11(4):300–9.

4.   Johnson KB, Ravich WJ, Cowan JA. Brainstorming about next-generation computer-based documentation: an AMIA clinical working group survey. Int J Med Inform. 2004 Sep;73(9-10):665–74.

5. Ammenwerth E, Mansmann U, Iller C, Eichstädter R. Factors affecting and affected by user acceptance of computer-based nursing documentation: results of a two-year study. J Am Med Inform Assoc. 2003 Feb;10(1):69–84.

6. Lopetegui M, Yen P-Y, Lai AM, Embi PJ, Payne PRO. Time Capture Tool (TimeCaT): development of a comprehensive application to support data capture for Time Motion Studies. AMIA Annu Symp Proc. 2012;2012:596–605.

7. Ozkaynak M, Dziadkowiec O, Mistry R, Callahan T, He Z, Deakyne S, et al. Characterizing workflow for pediatric asthma patients in emergency departments using electronic health records. J Biomed Inform. 2015 Aug 30;

8. Kirkendall ES, Kouril M, Minich T, Spooner SA. Analysis of electronic medication orders with large overdoses: opportunities for mitigating dosing errors. Appl Clin Inform. 2014;5(1):25–45.

9. Redd TK, Read-Brown S, Choi D, Yackel TR, Tu DC, Chiang MF. Electronic health record impact on productivity and efficiency in an academic pediatric ophthalmology practice. J AAPOS. 2014 Dec;18(6):584–9.

10. Hirsch AG, Jones JB, Lerch VR, Tang X, Berger A, Clark DN, et al. The electronic health record audit file: the patient is waiting. J Am Med Inform Assoc. 2016 Jul 3;

11. Zheng, K, Ciemins, E, Lanham, H, Lindberg, C. Examining the Relationship Between Health IT and Ambulatory Care Workflow Redesign. (Prepared by Billings Clinic under Contract No. 290-2010-0019I-1). AHRQ Publication No. 15-0058-EF. Rockville, MD: Agency for Healthcare Research and Quality; 2015.

12. Zheng K, Haftel HM, Hirschl RB, O'Reilly M, Hanauer DA. Quantifying the impact of health IT implementations on clinical workflow: a new methodological perspective. J Am Med Inform Assoc. 2010 Aug;17(4):454–61.

CHAPTER 4

Query Log Analysis and Design implications of Semantically-based Query Recommendation for an Electronic Health Record Search Engine

4.1 Introduction

Clinical documentation is a central component of healthcare, which is fundamental to clinicians' ability to review patients' medical histories, make sense of their current medical problems, and decide proper treatments and care plans (1,2). Clinical documentation is often accomplished through the use of electronic health record (EHR) systems, which collects a massive amount of detailed patient data. In addition to the firsthand use of these patient data for clinical care, they can be re-used for a variety of purposes such as education, administration, and billing. Among these secondary uses, supporting clinical and translational research has been a primary focus due to its potential to enhance medical knowledge, support decision-making, and improve quality of care as well as patient outcomes.

Using patient data to support clinical and translational research, however, can sometimes be very difficult. One common issue is that a significant amount of these data is recorded in free-text form, partly because recording free-text data can align well with clinical workflow and practice(3). These free-text data, including progress notes and discharge summaries, contain valuable clinical information often unavailable elsewhere and can provide great insight to patients' medical situations. While utilizing these free-text clinical data in clinical research projects for purposes such as diagnosis and patient cohort identification is not uncommon, these data are often utilized manually (e.g. manual chart review) and therefore suboptimal, which cannot be scaled up without proper support of computer technology.

One such support is information retrieval (IR), which catches relevant documents from a corpus given users' information needs. The Retrieval of medical documents has at least one fundamental challenge. That is, clinicians and clinical researchers, even with many years of training, are frequently unable to formulate the queries that can most effectively describe their information needs to obtain the top relevant documents. This is attributable to the different kinds of uncertainty involved in a search process, including lack of knowledge about the corpus, inability to evaluate document relevance, and unfamiliarity with the variation of medical terms.

This fundamental uncertainty can be mitigated by a technique called "query expansion", which has been widely used in modern IR systems such as Google and Bing. Query expansion improves system performance by reformulating input queries to alternatives containing key terms that otherwise would have been missed. Studies have proposed several mechanisms to effectively expand query terms to achieve higher system performance, such as local/global document analysis (4) and lexical-semantic relations (5). Unfortunately, these traditional query expansion mechanisms may not successfully boost the performance of a medical IR system if not adequately considering the nature of these medical corpora, i.e. the pervasive use of abbreviations, acronyms, and synonyms. For example, the term "pat" can be a shortening of the word "patient" in many situations but can refer to "paroxysmal atrial tachycardia" in a specific context. For another example, when clinicians search notes with the term "car accident" in a medical IR system, they should also reasonably expect to see notes with the term "vehicle accident" or "motor accident" because of their semantically overlapped meanings.

Although query expansion techniques have been improved and applied to medical fields in recent years (6–9), few studies have investigated end users' perceptions of the performance gains in them. My previous study addressed this gap by designing a prototype system with a semantically-based query recommendation feature (SBQR) (10). This feature expanded queries to include their synonyms based on medical concepts defined in the Unified Medical Language System as well as collected empirically. The performance of this prototype system was evaluated in a lab setting in another study that I co-conducted (11). The results showed very positive

perceptions of the participants toward SBQE, suggesting the possibility of successful implementation of SBQR in medical IR systems.

The goal of this follow-up study is to conduct a deeper analysis to support the previous findings and generate potential explanations for the previous observations primarily derived from the self-reported data. The present study is focused on query log, a type of data representing an objective view of end users' information seeking behaviors. By utilizing these query log data automatically recorded during the previous user study, I aimed to investigate 1) if the query terms were formulated differently under the support of SBQR; and 2) in what ways SBQR may improve the search results and positively affect users' perceived performance of the prototype system.

4.2 Methods

A web-based prototype IR system was built with SBQR as illustrated in Figure 4.1. The workflow of SBQR started with parsing user-supplied queries to corresponding medical concepts using MetaMap, a tool for identifying concepts available from National Library of Medicine. These medical concepts were further expanded to include their synonyms based on two synonym sets. One was a pre-defined set in the Unified Medical Language System; the other was an empirical collection of a homegrown medical IR system called "Electronic Medical Record Search Engine (EMERSE)" (12). Next, the SBQR-expanded queries were matched to the indexes of the corpus containing about 100,000 medical documents. The index was built based on medical concepts and synonyms following the same process. With the support of SBQR, users of the prototype system were expected to find more relevant documents. For example, users can obtain documents containing both "hearing loss" and "difficulty of hearing" when querying either of the terms in the prototype system.
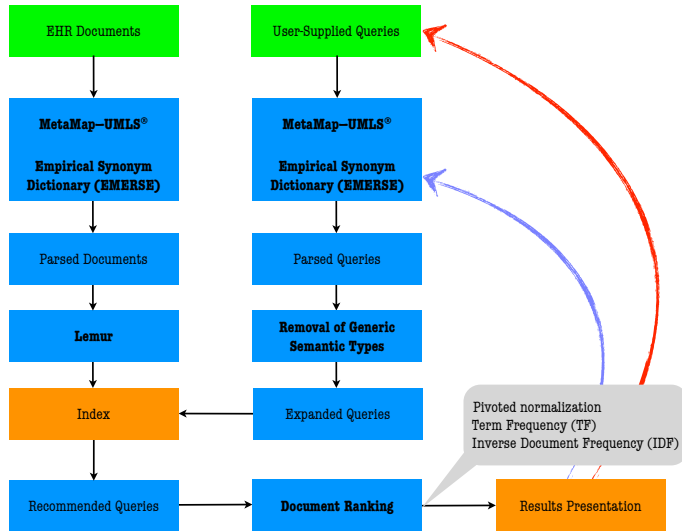
Figure 4.1: Workflow the Semantic-Base Query Recommendation (SBQR) in the prototype system (10).

This prototype medical IR system kept track of users' query behaviors during the previous user study, including the original user-supplied queries, the parsed concepts, the expanded terms when SBQR was activated, the top 30 retrieved documents, and the timestamps. Five pre-defined scenarios with various levels of difficulty were given to the 33 participants, who were active EMERSE users. In each scenario, the participants were told to turn SBQR off at the beginning and formulate as many queries as they needed until they retrieved satisfactory results. The participants then turned SBQR on, and conducted another round of searching until reaching another set of satisfactory results. Participants, however, were free to turn SBQR on and off several times during the search process. For comparative analysis, this author therefore chose to include only query log data related the first query when SBQR was turned off (Q_{A1}) and the last query when SBQR was turned on (Q_{Bn}) in each participant-scenario. Figure 4.2 illustrates a generalized search process with the highlighting of the selected queries for analysis.
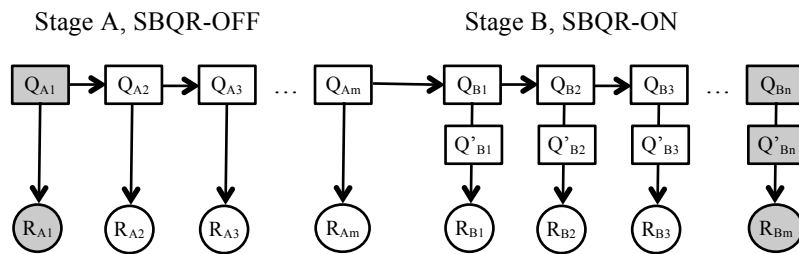
Figure 4.2: A generalized search process and the selected queries for analysis. Q_{A1} is the first user-supplied query when SBQR is turned off, which leads to the retrieved documents R_{A1}. Q_{Bn} is the last user-supplied query when SBQR is turned on, which is expanded to Q'_{Bn} by SBQR and leads to the retrieved documents R_{Bm}.

The analysis began by summarizing the participants' information seeking behaviors (query activities) in terms of number of queries per user, number of terms per query per user, and percentage of queries that were given when SBQR was turned on. Then, the analysis focused on answering the first research question, which is whether the participants formulated different queries with the support of SBQR. i.e. comparing Q_{A1} and Q_{Bn}. The difference between Q_{A1} and Q_{Bn} was measured by the degree of term overlapping. Specifically, in each scenario for each participant, the terms of Q_{A1} were extracted by a simple heuristic because the participants were asked to separate terms in a query with comma. These extracted terms were further folded to their lower case form and merged together to construct a term vector of Q_{A1}. The term vector of Q_{Bn} was extracted in the same process. The similarity between Q_{A1} and Q_{Bn} was measured by the

Jaccard similarity coefficient of their term vectors, defined as the ratio of the intersection of the terms of these two vectors over the union of them (F1). I hypothesized that Q_{A1} and Q_{Bn} would be similar so that the participants' perceived performance toward the system in the previous study could be largely attributable to the use of SBQR, rather than to the participants' own manipulation of queries.

$$J(Q_{A1}, Q_{Bn}) = \frac{|V_{A1} \cap V_{Bn}|}{|V_{A1} \cup V_{Bn}|} \qquad (F1)$$

Next, the analysis switched the focus to the second research query, which was to examine whether participants with the same information needs but using varying search strategies would retrieve a similar set of documents through SBQR. As illustrate in Figure 4.3, I hypothesized that the participants would be "brought together" with the help of SBQR given the semantic overlap of the concepts in their queries. This hypothesis was examined by comparing the Entropy of the result sets, i.e. comparing R_{A1} and R_{Bn}. A result set referred to the union of all top-10 documents retrieved in each participant-scenario. This Entropy measure gauged the level of uncertainty in a probability distribution, which is constructed by a set of probability scores of the retrieved documents in a result set. The probability score of a document was calculated by the ratio of the frequency of this document to the sum of all frequency of the documents in this result set. These probability scores were further turned into an Entropy score of a result set using the formula described in (F2) As can be seen, x represents the probability score of one document in a results set. If the hypothesis were true, SBQR would return a converged result set in each scenario, which can be measure by comparing the Entropy scores of R_{A1} and R_{Bn}. Otherwise, a positive gain would be observed.

$$H(x) = -\sum (x * \log(x)) \qquad (F2)$$
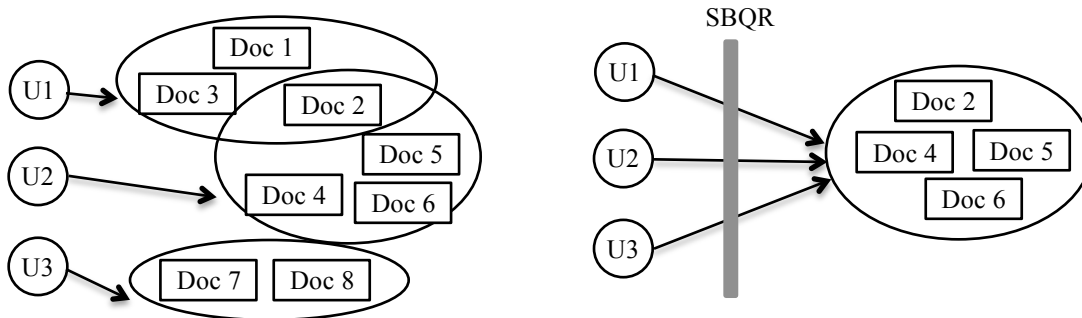
Figure 4.3: The hypothesized effectiveness of SBQR. Participants with the same information needs may use different search strategies (e.g. terms), resulting in lower overlap between the retrieved documents (left, high uncertainty); Under the support of SBQR, users may be "brought together" based on the semantic meaning of the terms and therefore retrieved a similar set of documents.

The difference of Jaccard similarity scores among the scenarios was further examined. First of all, the independence between scenarios and participants, the normality of the scores in each scenario, and the homogeneity of the variance among the scenarios were tested using the Chi-square test, the Kolmogorov-Smirnov test for goodness of fit, and the Bartlett and the Levene test, respectively. If the distribution of the Jaccard similarity scores is normal in each scenario, the Tukey HSD test will be employed to examine the difference of the group means. Otherwise, the kruskal test (nonparametric equivalent of ANOVA) will be used.

In terms of the analysis techniques, the query log data were stored and manipulated in a standalone SQLite database. Two data views were created to capture the first query when SBQR was turned on and the last query when SBQR was turned off in each participant-scenario. The functions to calculate Jaccard similarity and Entropy were implemented in Python 2.7 with existing libraries such as "pandas", "numpy", and "scipy". The use of the query log data was

permitted by the University of Michigan Health Sciences and Behavioral Sciences Institutional Review Boards

4.3 Results

The information seeking behaviors (query activities) of all 33 participants were automatically recorded in the previous user experiments, resulting in 10,451 records and 2,098 queries. Table 4.1 shows an example of one participant's queries in Scenario 2. In this example, User 005 submitted a total of six queries, with two of them being SBRQ-supported (the 4th and the 5th query). The user started with a single term query "dcis" without SBRQ, and then expanded the query manually to "non-invasive dcis". The user ended up with a query consisting of two concepts "dcis, breast cancer". The query activities of all participants are summarized in Table 4.2. As can be seen, participants formulated between 6 and 13 queries, with nearly half being submitted when SBQR was turned on, except for Scenario 1 where only one third of queries were submitted when SBRQ was turned on. The queries contained 2-4 terms and their length was between 27 to 44 characters. In the easy scenario, the participants tended to formulate shorter queries, both in terms of the average number of terms (2.56) and the average number of characters (27.12). On the other hand, in the scenario with high difficulty, the participants tended to formulate longer queries with more terms (3.28) and characters (44.28).

Table 4.1: A set of queries given by User 005 in Scenario 2 *"You are interested in identifying patients who have the non-invasive form of breast cancer known as DCIS."*

LOG_ID	USER_ID	ODR	QUERY	SBQR mode	TIMESTAMP
1000006702*	005	1	dcis	OFF	3/15/12 09:54:34
1000006712	005	2	non-invasive dcis	OFF	3/15/12 09:55:11
1000006717	005	3	non-invasive dcis breast cancer	OFF	3/15/12 09:55:30
1000006762	005	4	non-invasive dcis breast cancer	ON	3/15/12 09:57:39
1000006776*	005	5	dcis, breast cancer	ON	3/15/12 09:59:23
1000006795	005	6	dcis, breast cancer	OFF	3/15/12 10:00:30

* Queries selected for analysis: first query when SBQR was turned off (1000006702, Q_{A1}) and last query when SBQR was turned on. (1000006776, Q_{Bn})

Table 4.2: Summary of query activities.

Scenario	Estimated Difficulty	Avg. # of Queries per User	Avg. % of Query with SBQR	Avg. # of Terms per User-Query	Avg. Length (characters) per User-Query
1	Medium	9.33	36.1%	2.69	32.26
2	Low	8.24	45.2%	2.56	27.12
3	Medium	13.45	43.2%	3.54	36.21
4	High	8.24	46.5%	3.28	44.28
5	Medium	6.97	45.5%	3.02	34.03

The average Jaccard similarity coefficient of each scenario is reported in Table 4.3. Overall, the Jaccard similarity coefficient showed 77% similarity between the first SBRQ-turned-off query (Q_{A1}) and last SBRQ-turned-on query (Q_{Bn}). This high coefficient implies that the participants formulate similar queries even with the support of SBQR, suggesting that end users' positive perception of the prototype system likely resulted from the automatic expansion of query terms by SBQR, rather than from end users' own manual modification of query terms. Since the distribution of the Jaccard coefficient in the scenario groups was likely to be non-normal, the Kruskal test was used to check any significant difference among the group means, which is shown to be not significant (p=0.26). Although the differences were not statistically significant,

Scenario 1 (first in the order) and Scenario 4 (high difficulty) seemed to have lower Jaccard similarity coefficient than others. On the other hand, Scenario 2 (low difficulty) seemed to have a higher Jaccard similarity coefficient. Together with the query analysis in Table 4.2, it seems that the participants in an easy scenario submitted shorter and similar queries, while the participants in a challenging scenario submitted longer and varying queries.

Table 4.3: Summary of query similarities.

Test of Independence	Test of Normality	Test of Variance	Scenario	Estimated Difficulty	Avg. Jaccard Coefficient
p=1.0 (independent)	All p < 0.01 (non-normal)	p = 0.15 (equal variance between groups)	1	Medium	0.69 ± 0.35
			2	Low	0.86 ± 0.22
			3	Medium	0.80 ± 0.32
			4	High	0.71 ± 0.37
			5	Medium	0.79 ± 0.33

The Entropy analysis also showed evidence supporting the hypothesis. As listed in Table 4.4, a negative gain in the Entropy scores was observed for the scenarios 1, 3, and 5, indicating that the result sets were converged. This suggests that SBQR effectively brought the participants together and led to a more condensed result set. However, the positive gain in the Entropy scores in the scenarios 2 and 4 indicates that the result sets were even more diverse when SBQR was turned on in these scenarios. The additional analysis showed that the differences in the Entropy scores were highly correlated to the perceived system performance (Pearson correlation coefficient: -0.85). This high negative correlation could mean that the more condensed a result set is when SBQR was turned-on, the more likely the participants would perceive higher system performance. Considering the estimated difficulty, SBQR did not perform well when the estimated difficulty was on the extremes (Scenario 2 and 4). This author suspects that in a high difficulty scenario, participants could formulate varying queries preventing the retrieved results from converging no matter how much help SBQR provided. In a low difficulty scenario, since the information needs could be so specific and clear, SBQR introduces "noise" into the result set by adding too many semantically related, but less relevant terms

Table 4.4: The analysis of Entropy of top-10 result sets. The Pearson correlation between the Entropy difference (%) and the perceived performance is -0.85.

Scenario	Entropy of SBQR-Off	Entropy of SBQR-On	Entropy Difference	Entropy Diff. (%)	Perceived Performance
1	4.4384	3.6747	-0.7637	-17.21	4.24
2	3.0688	3.3858	0.3170	10.33	3.94
3	3.8411	3.5537	-0.2874	-7.48	4.42
4	3.9398	4.0709	0.1311	3.33	4.09
5	3.6617	2.9248	-0.7369	-20.12	4.55

4.4 Discussion

Utilizing unstructured clinical data for firsthand and secondary uses, such as decision support and clinical and translational research, often requires information retrieval techniques that can assist clinicians in finding relevant information locked in massive free-text sentences. This author developed such a technique (SBQR) that expanded query terms to their semantically overlapped forms in a prototype medical IR system, and evaluated the perceived system performance in a user experiment. In the present study, the author further analyzed the query log recorded in the user experiment to seek potential explanations for participants' positive perceptions of the prototype system. The results showed that participants formulate very similar queries with SBQR's assistance, indicating that the perceived positive system performance was likely attributable to the effectiveness of SBQR, rather than to the manual modifications of queries when a search progressed. Moreover, the participants' information seeking behaviors were consistent with the estimated difficulty of the scenarios. That is, the participants in an easy scenario tended to formulate shorter and more similar queries, while the participants in a challenging scenario tended to formulate longer and less similar queries.

The Entropy analysis revealed that the estimated difficulty of a scenario could be a contingent factor of participants' positive perception of the system performance. SBQR achieved higher performance in scenarios with medium difficulty as opposed to those at the extremes. One possible explanation could be that in the extremes cases SBQR either introduces noise or provides limited help. This finding suggests at least two design ideas of SBQR in modern medical IR systems: 1) SBQR can be designed as a user-controllable option, in which it would be necessary to educate users as to the best timing to turn on SBQR. 2) SBQR can also be designed as a semi-automated feature, which is activated based upon the observed and inferred difficulty of users' information needs, possibly through the analysis of query terms and retrieved documents (13,14).

The strength of this study lies in the use of a combination of objective query log and self-report survey data to uncover participants' complicated behaviors when using a novel IR feature, namely SBRQ. The study has several limitations, one of which is that the use of SBQR was not strictly controlled in the user experiment, although participants were asked to turn SBQR off first and then turn it on. In fact, participants can and did alternate between turning SBQR on and off when a search progressed. Although this allowed participants to conduct searches more naturally, it imposed extra complexity in the analysis. To resolve this complexity, this author targeted the first query when SBQR was turned off and the last query when SBQR was turned on to maximize potential term differences, and ignored all other queries in between. Another limitation is that neither the participants were asked to provide relevance feedback on the retrieved documents, nor there is a gold standard of document relevance for each scenario. Although the Entropy analysis supported that the perceived system performance was highly correlated with the degree of convergence of the retrieved results among participants, without gold standards, the relationship between the convergence and the quality of search results cannot be verified.

4.5 Conclusions

This study analyzed query log data regarding the use of SBQR to investigate reasons behind the participants' positive perception of system performance reported in the previous study. The results showed that this positive perception was likely attributable to the effectiveness of SBRQ, and was likely contingent upon the difficulty of a search scenario. Modern medical IR systems should consider the design of SBQR as a user-controllable option or a semi-automated feature that is triggered when the difficulty of user's information needs can be inferred. This study confirms that SBQR has a great potential to overcome the challenges in retrieving medical documents, which primarily results from the pervasive use of acronyms and synonyms in this type of corpus.

4.6 References

1. Institute of Medicine (U.S.). The computer-based patient record: an essential technology for health care. Washington, D.C: National Academy Press; 1991. 190 p.

2. Institute of Medicine (US) Committee on Improving the Patient Record. The Computer-Based Patient Record: Revised Edition: An Essential Technology for Health Care [Internet]. Dick RS, Steen EB, Detmer DE, editors. Washington (DC): National Academies Press (US); 1997 [cited 2015 Aug 8]. Available from: http://www.ncbi.nlm.nih.gov/books/NBK233047/

3. Rosenbloom ST, Denny JC, Xu H, Lorenzi N, Stead WW, Johnson KB. Data from clinical notes: a perspective on the tension between structure and flexible documentation. J Am Med Inform Assoc. 2011 Mar 1;18(2):181–6.

4. Xu J, Croft WB. Query expansion using local and global document analysis. In ACM Press; 1996 [cited 2015 Feb 13]. p. 4–11. Available from: http://portal.acm.org/citation.cfm?doid=243199.243202

5. Voorhees EM. Query Expansion using Lexical-Semantic Relations. In: Croft BW, van Rijsbergen CJ, editors. SIGIR '94 [Internet]. London: Springer London; 1994 [cited 2015 Feb 13]. p. 61–9. Available from: http://link.springer.com/10.1007/978-1-4471-2099-5_7

6. Hersh W, Price S, Donohoe L. Assessing thesaurus-based query expansion using the UMLS Metathesaurus. Proc AMIA Symp. 2000;344–8.

7. Liu Z, Chu WW. Knowledge-based query expansion to support scenario-specific retrieval of medical free text. Information Retrieval. 2007 Feb 9;10(2):173–202.

8. Díaz-Galiano MC, Martín-Valdivia M., Ureña-López LA. Query expansion with a medical ontology to improve a multimodal information retrieval system. Computers in Biology and Medicine. 2009 Apr;39(4):396–403.

9.   Jain H, Thao C, Zhao H. Enhancing electronic medical record retrieval through semantic query expansion. Information Systems and e-Business Management. 2012 Jun;10(2):165–81.

10. Danny T.Y. Wu, Lei Yang, Mei Qiaozhu, David A. Hanauer, Kai Zheng. Towards Intelligent and Socially Oriented Query Recommendation for Electronic Health Records Retrieval. In Health Search and Discovery (HSD) workshop at the Association Computer Machinery (ACM) Special Interest Group on Information Retrieval (SIGIR), Dublin, Ireland; 2013.

11. David A. Hanauer, Danny T.Y. Wu, Lei Yang, Qiaozhu Mei, Katherine B. Murkowski-Steffy, VGVinod Vydiswaran, et al. Development and Empirical Evaluation of Semantically-based Query Recommendation for an Electronic Health Record Search Engine. Journal of Biomedical Informatics (in review).

12. Hanauer DA. EMERSE: The Electronic Medical Record Search Engine. AMIA Annu Symp Proc. 2006;941.

13. David Carmel, Elad Yom-Tov, Adam Darlow, Dan Pelleg. What makes a query difficult. In The Association Computer Machinery (ACM) Special Interest Group on Information Retrieval (SIGIR); 2006.

14. Carmel D, Yom-Tov E. Estimating the Query Difficulty for Information Retrieval. Synthesis Lectures on Information Concepts, Retrieval, and Services. 2010 Jan;2(1):1–89.

CHAPTER 5

Assessing the Readability of *ClinicalTrials.gov*

5.1 Introduction

5.1.1 Background and Significance

Clinical trials are the bedrock of research for a variety of medical interventions including drugs, devices, and therapies intended to improve treatment efficacy and patient outcomes. Today, many clinical trials must be registered in a US National Institutes of Health repository (http://ClinicalTrials.gov) as a means to improve their accessibility to the public and to enhance participant recruitment (1,2). Although ClinicalTrials.gov does not include all trials ever conducted, recent regulatory requirements have led to the exponential growth of the number of studies registered, with a ten-fold increase occurring over the past decade (3).

Each study registered at ClinicalTrials.gov is accompanied with a detailed description covering all aspects of the trial protocol including the disease(s) targeted, intervention under evaluation, and requirements for participant recruitment. The registry hence serves not only as a mechanism for ensuring the ethics and integrity of the trials through increased transparency, but also a credible source of information for patients who are interested in participating or in learning about the results of the trials. As described in the ClinicalTrials.gov mission statement, the website was established in part to fulfill the goal of "providing patients, their family members, health care professionals, researchers, and the public with easy access to information on publicly and privately supported clinical studies on a wide range of diseases and conditions (4)." Additionally, the US Congress Food and Drug Modernization Act, which led to the creation of ClinicalTrials.gov, requires that the details about all clinical trials registered must be "in a form that could be readily understood by the public (5)."

However, the registry's potential for facilitating information dissemination and participant recruitment could be limited if the public, with varying literacy levels, are unable to read and properly understand the descriptions of the trials. Poor readability can also be a source of self-selection bias undermining the broad applicability of study findings, as those who are able to better comprehend the trial protocols may be more likely to volunteer for study participation (6). Thus, it is important to investigate the readability of trial descriptions available at ClinicalTrials.gov to ensure that the study information can be effectively conveyed to a wide audience with varying literacy.

Readability is known to affect the comprehensibility and communication effectiveness of text (7). Developing readability measures and validating/applying them in different empirical settings have thus been of great interest to researchers and educators in a wide range of domains (8-11). In this study, this author evaluated the readability of ClinicalTrials.gov trial descriptions using four general-purpose readability scoring algorithms (12-16) in addition to a measure specifically developed to work with medical text (17). The evaluation was conducted by comparing the readability of trial descriptions to the readability of two other related but distinct corpora: (1) Health Topics articles from MedlinePlus—a website created and maintained by the U.S. National Library of Medicine to provide the general public high-quality information about diseases, conditions, and wellness (18), and (2) clinician notes retrieved from the electronic health records (EHR) system used at our institution that were created for conveying internal communication among medical professionals. It was hypothesized that clinician notes would be most difficult to read, followed by clinical trial descriptions and then MedlinePlus Health Topics articles.

5.1.2 Previous Work

While this author is not aware of any prior studies that have specifically looked at the readability of clinical trial descriptions, there has been empirical work to assess lay persons' ability to retell the trial descriptions they read, revealing considerable comprehension errors (19,20). In addition, there have been studies investigating the readability of patient handouts and health education

pamphlets (21-23), online health content (24-26), and informed consent forms (11,27). These studies consistently found that patient and consumer health information resources tend to be difficult to read and require a literacy level higher than their intended audiences. For example, several studies demonstrated that patient consent forms for both patient care as well as clinical research were often written in very complex language (28-32), with one study suggesting that surgical consent forms were written at the level of scientific journals (33). Even Institutional Review Board consent form templates, which are intended to serve as the model for easy-to-understand text for laypersons, were deemed too complex for their proposed benchmarks (i.e., 5th and 10th grade) (34).

Nevertheless, prior readability studies conducted in healthcare have several notable limitations. First, the sample size employed was often small (no more than a few hundred documents). Second, most of the studies applied readability scoring algorithms developed for general purposes that do not take into account the unique characteristics of healthcare text (35). In this study, this author attempted to address these limitations by analyzing a much larger dataset, consisting of all trials registered at CliniclTrials.gov, all health articles from MedlinePlus, and 100,000 randomly selected clinician notes retrieved from an EHR system, using both general-purpose and medical specific readability assessment measures.

5.2 Methods
5.2.1 Corpora and Text Features
Three corpora were comparatively analyzed in this study. The first corpus contained all 165,988 clinical trial studies available at CliniclTrials.gov as of April 30, 2014. Each trial provided a detailed description on the website about its study objectives, target patient population(s), and approaches in the following four structured sections: *Purpose*, *Eligibility*, *Contacts and Locations*, and *More Information*. Among them, the Purpose section often begins with a narrative introduction and ends with a "detailed description" subsection orienting readers to the most important facts about the study setting(s) and the overall research design (a sample is

provided in Figure 1). Because these narratives serve as the entry point for readers to skim and decide whether the trial is of potential interest and worth exploring further, their readability is crucial. The analysis therefore was focused on these narratives extracted from the Purpose section. For convenience, this corpus is referred to as "Trial Description" in this paper. It contained approximately 1.5 million sentences and 33 million words.

Figure 5.1: A sample *Purpose* section from ClinicalTrials.gov.

PURPOSE
We hypothesize that individuals with Alpha-1 Antitrypsin (AAT) deficiency have ongoing liver injury which is not detected by the usual blood tests used to look at liver function. This ongoing liver injury leads to cirrhosis in a significant number of adults with AAT deficiency.

Detailed Description:
Our overarching hypothesis is that liver disease in adults with AAT deficiency is the result of the accumulation of the abnormally folded protein within the endoplasmic reticulum of the hepatocyte. In some individuals, the intrinsic cellular mechanisms of the hepatocyte are sufficient to clear adequate amounts of the abnormally folded protein such that liver disease does not occur. In AAT deficient individuals who develop liver disease, environmental and other genetic factors stress the hepatocyte, and the normal cellular mechanisms that maintain homeostasis are disrupted, leading to liver disease.

For this proposal, our hypothesis is that the prevalence of liver disease in adults with AAT is higher than previously reported because liver injury and fibrosis is not accurately detected by available routine liver testing. Testing this hypothesis will require an initial evaluation for liver disease with liver function testing and imaging, and then histologic confirmation by liver biopsy.

In addition to "Trial Description," two other corpora were analyzed in order to obtain benchmarks to better interpret the results generated by readability scoring algorithms. The second corpus we analyzed consisted of all 955 "Health Topics" articles in English available at MedlinePlus as of April 30, 2014 (a sample is provided in Figure 2). Because these Health Topics articles are carefully curated by the US National Library of Medicine with the goal of disseminating high quality, easy-to-understand health information to the general public, they should be highly comprehensible by laypersons and thus should receive the best readability evaluation scores. This corpus is referred to as "MedlinePlus" in this paper. It contained a total of 13,630 sentences and 136,032 words. Note that three other types of consumer-oriented materials

available at MedlinePlus (a Medical Encyclopedia, Drug & Supplements information, and "Video & Cool Tools") were not included in our MedlinePlus corpus. This is because the Encyclopedia and the Drug & Supplements information are highly structured (i.e., a majority of this content is expressed via bullet points), and the "Video & Cool Tools" are mostly multimedia resources with little text for analysis.

Figure 5.2: A sample MedlinePlus Health Topics article on "Aortic Aneurysm".

Aortic Aneurysm

An aneurysm is a bulge or "ballooning" in the wall of an artery. Arteries are blood vessels that carry oxygen-rich blood from the heart to other parts of the body. If an aneurysm grows large, it can burst and cause dangerous bleeding or even death.

Most aneurysms are in the aorta, the main artery that runs from the heart through the chest and abdomen.

There are two types of aortic aneurysm:

- Thoracic aortic aneurysms - these occur in the part of the aorta running through the chest
- Abdominal aortic aneurysms - these occur in the part of the aorta running through the abdomen

Most aneurysms are found during tests done for other reasons. Some people are at high risk for aneurysms. It is important for them to get screening, because aneurysms can develop and become large before causing any symptoms Screening is recommended for people between the ages of 65 and 75 if they have a family history, or if they are men who have smoked. Doctors use imaging tests to find aneurysms. Medicines and surgery are the two main treatments.

The third corpus analyzed in this study consisted of 100,000 free-text narrative clinician notes (a sample is provided in Figure 3) randomly retrieved from the EHR system in use at the University of Michigan Health System (UMHS), a tertiary care academic medical center with over 45,000 inpatient admissions and 1.9 million outpatient visits annually (36). The homegrown EHR system, called CareWeb, allows clinicians to create notes via dictation/transcription or via typing (37). These notes are generally unstructured, but clinicians could use simple, customizable text-based templates if desired. The corpus contained multiple document types retrieved from CareWeb generated in both inpatient and outpatient areas including admission notes, progress notes, radiology reports, and narrative assessments and plans. Because these clinician notes were

composed by medical professionals and intended to be read by other medical professionals, they were hypothesized to be most difficult to read across the three study corpora. This "EHR" corpus contained over 5 million sentences with about 56 million words.

For patient privacy protection reasons, all documents contained in the EHR corpus were first de-identified before they were used in this study. The identification was performed using the MITRE Identification Scrubber Toolkit (MIST) (38), and was based on a well-performing, locally developed model that we previously evaluated and reported in the literature (39). Identifiable information including names, ages, and dates was replaced with standardized placeholders such as *[NAME]*, *[AGE]*, and *[DATE]*, with the majority of the clinical text remaining in its original form.

Figure 5.3: A sample clinical note from UMHS.

[NAME] is a [AGE] old female who underwent a composite resection of her mandible and mouth floor with local advancement flap and tracheostomy on [DATE] transferred to the SICU overnight for respiratory distress. She was initially diagnosed with an invasive squamous cell carcinoma of the mandibular alveolus. She was staged at a T3 N1 squamous cell carcinoma of the right mandibular alveolus. She underwent an uneventful surgery and post-operatively, she was doing well with a medicine team following for her multiple comorbidities including diabetes, COPD, HTN and GERD. She was though to be fluid positive over the last several days. She was receiving 200cc of fluid an hour over the 14 hrs prior to her decompensation last evening. Yesterday evening, she began to have increasing respiratory distress with increasing oxygen requirements on trach mask. She was placed on 80% and transferred to the SICU after receiving 60mg of Lasix. Upon arrival, she notes some difficulty with breathing, denies any chest pain, abdominal pain, cough.

5.2.2 Surface Metrics

First of all, each corpus was characterized using four surface metrics: average document length, average sentence length, vocabulary size, and vocabulary coverage. Vocabulary size is defined as the number of distinct words contained in the corpus. Vocabulary coverage is the percentage of distinct words that can be found in known dictionaries.

In the empirical study, a medical English dictionary was developed by combining entries extracted from an open-source English spell-check tool, GNU Aspell (40), with medical terms

extracted from an open-source medical spelling checker, OpenMedSpel (41). In addition, a comprehensive dictionary of medical terminologies was created based on the content of the Unified Medical Language System (UMLS) 2013AB Metathesaurus which includes more than 2.9 million concepts and 11.4 million unique concept names from over 160 source vocabularies (42). The first dictionary was referred to as the Basic Medical English Dictionary ("Med-Dict") and the second was referred to as "UMLS."

5.2.3 Readability Measures

In this study, four general-purpose readability scoring algorithms and one medical specific algorithm were applied to measure the readability of each corpus. The four general-purpose measures have been popularly used in a wide range of domains, including healthcare, all of which produce a readability score in the form of the number of years of education required to comprehend the material under evaluation (43). Below, the mechanism underlying each of these general-purpose readability measures was briefly summarized.

1. New Dale-Chall (NDC), computed based on the average number of words per sentence and the percentage of 'unfamiliar' words not covered in a pre-defined dictionary (12,13);

2. Flesch-Kincaid Grade Level (FKGL), computed based on the weighted sum of the average number of words per sentence and the average number of syllables per word, and then adjusted by a baseline score (14);

3. Simple Measure of Gobbedygook (SMOG), computed based on the square root of the average number of syllables per word of words that have three or more syllables (15);

4. Gunning-Fog Index (GFI), computed based on the average number of words per sentence and the complexity-syntax patterns of the words (16).

The fifth readability scoring algorithm, called the Medical-Specific Readability Measure (MSRM), was developed by co-authors of this paper (QTZ and JP) specifically for assessing the readability of medical text (17). Besides average sentence length and average word length, MSRM makes use of several additional text features such as average number of sentences per paragraph and parts of speech. Rather than producing an absolute score, MSRM estimates the relative distance of the features of the text being evaluated to those of a set of *easy-to-read* text samples and a set of *difficult-to-read* text samples. The *easy-to-read* samples consist of content extracted from online health education materials whereas the *difficult-to-read* samples consist of text extracted from scientific journal articles and medical textbooks. The scores produced by the algorithm range between -1 and 1, wherein 1 indicates the best readability. The mathematical underpinnings of MSRM can be found in the original publication (17).

5.2.4 Analysis Procedures

The readability of each of the study corpora was independently evaluated using the five scoring algorithms. Also, a composite score was produced by averaging the grade level metrics generated by the four general-purpose measures. No stop words were removed prior to the analysis as it might change the text features and subsequently affect the readability scoring. Pairwise differences among the readability scores of the three corpora were conservatively tested using ANOVA with Tukey's Honestly Significant Difference (HSD). All statistical analyses were performed in *R* version 3.0.2. The Institutional Review Board at the University of Michigan reviewed and approved the research protocol of this study.

5.3 Results

5.3.1 Surface Metrics

The surface metrics of the three study corpora are reported in Table 5.1. Consistent with the hypothesis, MedlinePlus Health Topics articles appear to the easiest to read as they have the shortest average sentence length, smallest vocabulary size, and highest vocabulary coverage (97.1% by Med-Dict alone and 99.4% by Med-Dict and UMLS combined). In comparison, the average sentence length of the Trial Description corpus is more than two-fold longer than that of MedlinePlus, and its vocabulary size is more than twenty times larger. Further, only one third of the words used in ClinicalTrials.gov descriptions are covered by Med-Dict. The EHR corpus, not surprisingly, has the largest vocabulary size and the least vocabulary coverage by known dictionaries, and is therefore likely most difficult to read.

Table 5.1: Surface metrics.

Surface metrics	Trial Description	MedlinePlus	EHR
Average sentence length (number of words)	26.1👎	10.2👍	12.3
Vocabulary size	147,978	6,939👍	307,750👎
Vocabulary covered by Med-Dict	34.8%	97.1%👍	15.3%👎
Vocabulary covered by UMLS	38.0%	66.9%👍	17.6%👎
Vocabulary covered by UMLS + Med-Dict	53.7%	99.4%👍	24.6%👎

👍 Likely associated with the best readability; 👎 Likely associated with the poorest readability.

5.3.2 Readability Scores

Table 5.2 reports the readability scores. All five measures consistently rated ClinicalTrials.gov trial descriptions as the most difficult corpus, which requires 15.8 to 21.1 years of education on average to be able to proficiently read and understand. The MedlinePlus Health Topics articles were consistently rated as the corpus that is easiest to read, requiring no more than a high school level of education (8.0 to 11.3 years). The scores of the EHR corpus always fell in the middle range. The Tukey's HSD test results showed that the differences across the mean readability scores of the three corpora are all statistically significant regardless of the readability measure used. Table 5.3 provides some sample narratives illustrating text that was rated easy to read vs. text that was rated difficult to read.

Table 5.2: Readability scores.

Scoring algorithm	Trial Description	MedlinePlus	EHR
NDC	15.8 ± 0.8☞	11.3 ± 2.3☜	15.1 ± 1.7
FKGL	17.2 ± 4.2☞	8.0 ± 1.4☜	9.1 ± 1.9
SMOG	17.9 ± 3.1☞	10.7 ± 1.3☜	11.9 ± 1.5
GFI	21.1 ± 4.6☞	10.9 ± 1.9☜	12.9 ± 2.2
Average of the four above	18.0 ± 3.0☞	10.2 ± 1.5☜	12.2 ± 1.5
MSRM	-0.44 ± 0.28☞	-0.10 ± 0.23☜	-0.36 ± 0.18

☜ Best readability; ☞ Worst readability.

Table 5.3: Sample text from each of the study corpora.

Corpus	Readability	Average Number of Years of Eduaction Requried for Proficient Reading*	Sample text
Trial Description	Easy	7.2	"This study plans to learn more about the immune system's response to breast cancer in young women."
	Hard	58.9	"The primary objectives of this study are: To evaluate the safety and tolerability of TH-302 monotherapy and in combination with bortezomib in subjects with relapsed/refractory multiple myeloma. To identify the dose-limiting toxicities (DLTs) and determine the maximum tolerated dose (MTD) of TH-302 monotherapy and in combination with bortezomib in subjects with relapsed/refractory multiple myeloma. To identify a recommended Phase 2 dose for TH-302 and dexamethasone with or without bortezomib in subjects with relapsed/refractory multiple myeloma."
MedlinePlus	Easy	5.2	"Did you know that the average person has 5 million hairs? Hair grows all over your body except on your lips, palms and the soles of your feet. It takes about a month for healthy hair to grow half an inch. Most hairs grow for up to six years and then fall out. New hairs grow in their place."
	Hard	17.2	"Dupuytren's contracture: a hereditary thickening of the tough tissue that lies just below the skin of your palm, which causes the fingers to stiffen and bend."
EHR	Easy	4.6	"Will continue to follow and assess identified deficits and goals."
	Hard	28.6	"Right hydroureter confirmed by retrograde pyelogram prior to stent placement." "CHF with ischemic cardiac myopathy and ejection and an ejection fraction of 35%. PVOD with bilateral carotid stenosis."

 * Average scores pooling the results generated by the four conventional measures.

Figure 5.4 depicts the distribution of readability scores among the documents in each corpus. Figure 5.5 illustrates the variation among the scores. As shown in Figures 5.4 and 5.5, a majority of ClinicalTrials.gov trial descriptions were rated more difficult to read compared to the EHR and MedlinePlus corpora, which suggests that the findings of this study were robust and were not caused by a few outliners. Among the three corpora studied, the readability scores of MedlinePlus Health Topics articles have the least amount of variation, whereas the readability scores of ClinicalTrials.gov trial descriptions are most widely spread out.
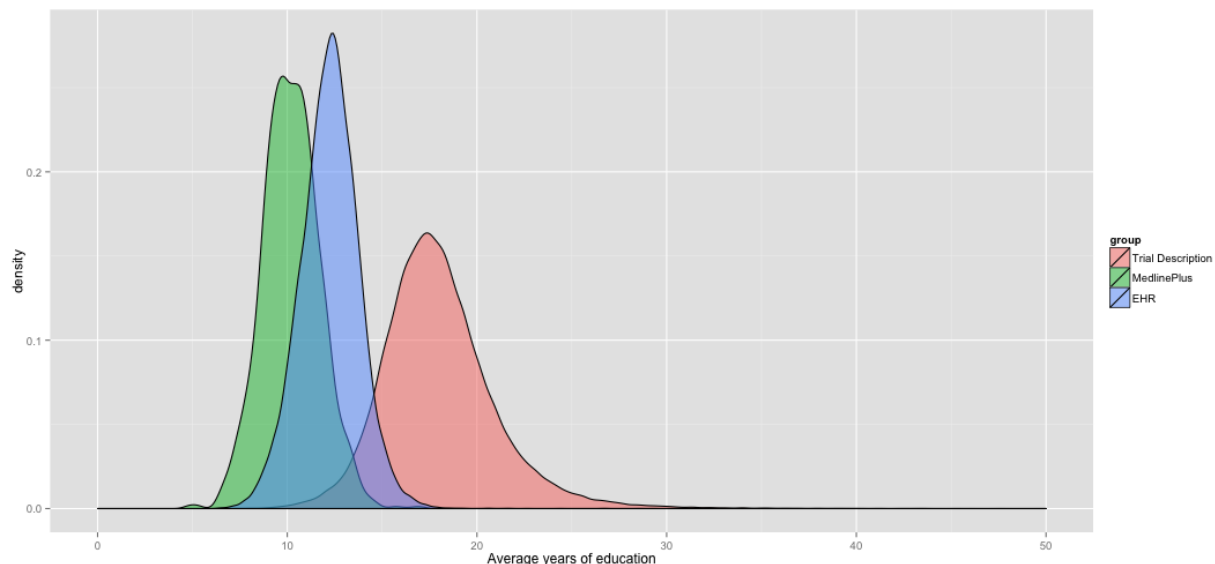

Figure 5.4: Distributions of average readability scores of the three study corpora (average scores were computed by pooling the results generated by the four conventional readability measures).
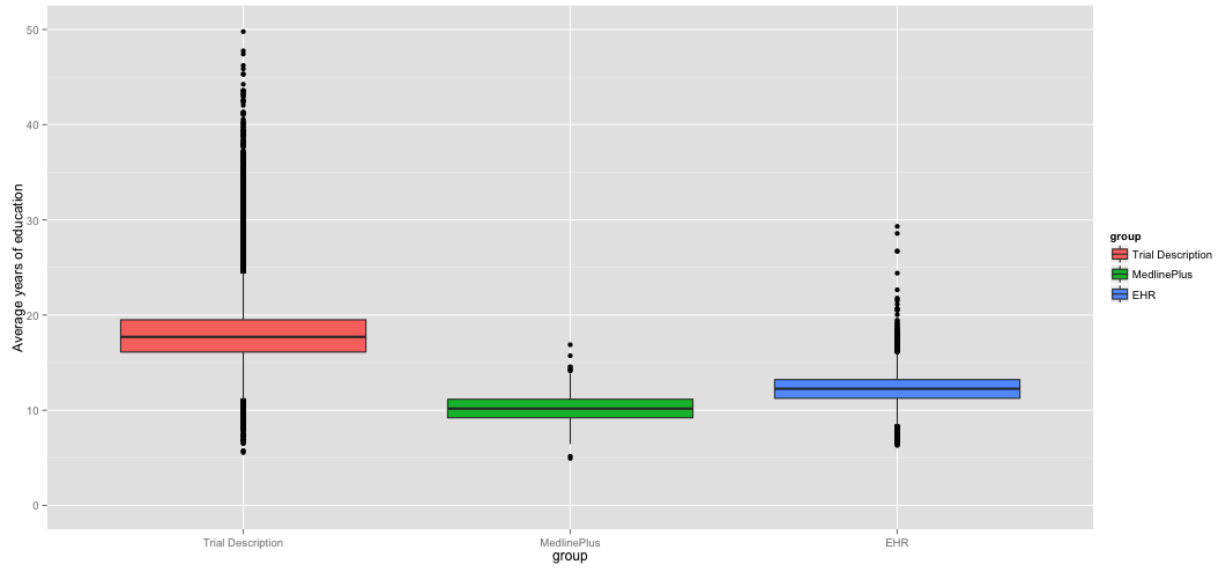
Figure 5.5: Variability of average readability scores (average scores were computed by pooling the results generated by the four conventional readability measures).

5.4 Discussion

While there have been studies assessing the readability of healthcare text such as patient education materials and informed consent forms (11,21-23,27-33), to the best of this author's knowledge, no research has been conducted to date to evaluate the readability of trial descriptions at ClinicalTrials.gov. As this federal registry plays an important role in informing the general public about clinical trial studies, information published at the website should be prepared in a manner that can be easily read and understand by laypersons. The evaluation of the present study however suggests concerning results. Every readability scoring algorithm employed in the study rated ClinicalTrials.gov trial descriptions as the most difficult corpus, on average requiring 18 years of education, or a postgraduate level, to proficiently read and comprehend. While these readability algorithms do not provide precise measures of the number of years of education required to comprehend the material, the results strongly suggest that ClinicalTrials.gov trial descriptions have severe readability issues.

This study was not designed to provide concrete guidelines as to how to improve the readability of ClinicalTrials.gov. That said, the results do suggest several areas where potential improvements could be made. For example, the very long average sentence length of the Trial Description corpus adversely affected its readability scores. Breaking down long sentences into shorter ones can thus be a quick way to improve the readability of many trial descriptions. Further, about two third of the words used in ClinicalTrials.gov trial descriptions are not found in the basic medical English dictionary. Those submitting these descriptions could find the process of changing all these terms to be a significant burden, or they may view trial descriptions without complex medical nomenclatures to be less scientifically rigorous. Thus, an alternative strategy might be to provide a consumer-oriented companion version of the description that uses plain layperson English; e.g., "chickenpox virus" instead of "varicella zoster virus," and "removal of kidney stone" instead of "nephrolithotomy."

This study applied multiple measures to ensure the reliability of readability scoring. The fact that MedlinePlus Health Topics articles were consistently rated as the easiest-to-read corpus to a certain degree validates the selected readability measures. The finding that clinician notes are generally easier to read than ClinicalTrials.gov trial descriptions is however surprising. This might be due to the fact that the available readability scoring algorithms are not best suited to evaluate the readability of clinician notes due to their unique characteristics. Prior research does show that machine-rated readability of clinical notes often generates convoluted results: when evaluated at the lexical level, the readability of clinical notes tends to be comparable to that of easy-to-read documents (e.g. consumer-oriented education materials); while when evaluated at the syntactic and semantic level, it tends to be comparable to that of difficult-to-read documents (e.g., publications in scientific journals) (44). Another possible explanation is that clinician notes 'speed-written' in time-sensitive patient care environments are often succinct, and complex medical words are often abbreviated or 'acronymized,' resulting in shorter sentences and shorter words with fewer syllables that may work in their favor when rated by readability scoring algorithms. This however could make the document much more difficult to read by patients (45). With the US healthcare system now becoming increasingly 'wired,' this situation might improve, or it might deteriorate. On the one hand, modern and "meaningful use" certified EHRs discourage clinicians from writing or dictating unstructured, free-text notes in favor of generating such notes from structured templates or with text generated by computer algorithms. This change has the potential to improve readability as it may reduce the amount of abbreviations, acronyms, and non-standard medical language used in clinician notes. However, on the other hand, EHR-generated clinician notes populated from templates and structured data may appear absurd to human readers, and may lack sufficient context explaining the medical conditions described. Therefore, laypersons, perhaps also clinicians themselves, may find computer-generated notes more challenging to read and understand. To the best of this author's knowledge, the MSRM method used in this paper developed by Kim et al. (2007) was the first and only attempt to develop custom readability scoring algorithms for clinician notes (17). Future work to develop better algorithms for assessing the readability of medical text, and for

understanding and improving the readability of computer-generated clinician notes populated from templates and structured data in modern EHRs, is therefore critically needed.

This study has several limitations. First, even though we included a readability scoring algorithm tailored to evaluating medical documents, the general-purpose readability measures used in the study were not specifically designed to work with healthcare text. Therefore they may not be able to generate highly accurate results. Second, clinician notes that we analyzed were retrieved from a single EHR system. The idiosyncrasies of the system, and of the local culture of clinical documentation, may also introduce biases into the study findings. Lastly, we only used computational methods to quantify the readability of the three study corpora, without engaging human readers who should ideally be drawn from a representative patient panel. Thus, the results of this study are only suggestive, not conclusive.

5.5 Conclusions

This study used five different scoring algorithms to evaluate the readability of clinical trial descriptions available at ClinicalTrials.gov. The evaluation was conducted in comparison with MedlinePlus Health Topics articles and clinician notes retrieved from an EHR system. The results show that ClinicalTrials.gov trial descriptions are the most difficult corpus, on average requiring 18 years of education in order to proficiently read and comprehend. Because ClinicalTrials.gov serves critical functions of disseminating trial information to the general public and helping the trials recruit patient participants, there is a critical need to develop guidelines and strategies to improve the readability of ClinicalTrials.gov trial descriptions so they can be understood by laypersons.

5.6 References

1. Food and Drug Administration Amendments Act of 2007 (FDAAA), Section 801. 2007; PUBLIC LAW 110–85—SEPT. 27. http://clinicaltrials.gov/ct2/manage-recs/fdaaa; accessed November 15, 2014.

2. Angelis CD, Drazen JM, Frizelle FA, et al. Clinical trial registration: a statement from the International Committee of Medical Journal Editors. N Engl J Med. 2004. 351(12):1250–1.

3. Trends, Charts, and Maps. 2014. http://www.clinicaltrials.gov/ct2/resources/trends; accessed November 15, 2014.

4. Background of ClincialTrials.gov. http://www.clinicaltrials.gov/ct2/about-site/background; accessed November 15, 2014.

5. Historical perspective on the development of ClinicalTrials.gov, and an overview of FDA's role in supporting the success of the database, and accessibility to clinical trials information by the public. http://www.fda.gov/ForConsumers/ByAudience/ForPatientAdvocates/ParticipatinginClinicalTrials/ucm143647.htm; accessed November 15, 2014.

6. Jüni P, Altman DG, Egger M. Systematic reviews in health care: assessing the quality of controlled clinical trial. BMJ. 2001;323(7303):42–6.

7. Garner M, Ning Z, Francis J. A framework for the evaluation of patient information leaflets. Health Expect. 2012;15(3):283–94.

8. Rosemblat G, Logan R, Tse T, et al. Text features and readability: expert evaluation of consumer health text. Mednet 2006. http://www.mednetcongress.org/fullpapers/MEDNET-192_RosemblatGracielaA_e.pdf; accessed November 15, 2014.

9. Zeng-Treitler Q, Kim H, Goryachev S, et al. Text characteristics of clinical reports and their implications on the readability of personal health records. Stud Health Technol Inform. 2007;129(Pt 2):1117–21.

10. Adnan M, Warren J, Orr M. Assessing text characteristics of electronic discharge summaries and their implications for patient readability. In: HIKM '10 Proceedings of the Fourth Australasian Workshop on Health Informatics and Knowledge Management. 2010;108:77–84.

11. Terblanche M, Burgess L. Examining the readability of patient-informed consent forms. Open Access Journal of Clinical Trials. 2010;2:157–62.

12. Dale E, Chall JS. A formula for predicting readability: instructions. Educ Res Bull. 1948;27(2):37–54.

13. Chall JS, Dale E. *Manual for Use of the New Dale-Chall Readability Formula*. Brookline, MA: Brookline Books. 1995.

14. Kincaid JP, Fishburne RP, Rogers RL, et al. *Derivation of New Readability Formulas (Automated Readability Index, Fog Count, and Flesch Reading Ease Formula) for Navy Enlisted Personnel*. Research Branch Report 8–75. Chief of Naval Technical Training: Naval Air Station Memphis, 1975.

15. McLaughlin GH. SMOG grading: A new readability formula. J Reading 1969;12:639–46.

16. Gunning R. *The Technique of Clear Writing*. New York, NY: McGraw-Hill International Book Co., 1952.

17. Kim H, Goryachev S, Rosemblat G, et al. Beyond surface characteristics: a new health text-specific readability measurement. AMIA Annu Symp Proc. 2007;11:418–22.

18. About MedlinePlus. http://www.nlm.nih.gov/medlineplus/aboutmedlineplus.html; accessed November 15, 2014.

19. Keselman A, Smith CA. A classification of errors in lay comprehension of medical documents. J Biomed Inform. 2012;45(6):1151–63.

20. Smith CA, Hetzel S, Dalrymple P, et al. Beyond readability: investigating coherence of clinical text for consumers. J Med Internet Res. 2011;13(4):e104.

21. Yin HS, Gupta RS, Tomopoulos S, et al. Readability, suitability, and characteristics of asthma action plans: examination of factors that may impair understanding. Pediatrics. 2013;131(1):e116–26.

22. Davis TC, Mayeaux EJ, Fredrickson D, et al. Reading ability of parents compared with reading level of pediatric patient education materials. Pediatrics. 1994;93(3);460–8.

23. Agarwal N, Hansberry DR, Sabourin V, et al. A comparative analysis of the quality of patient education materials from medical specialties. JAMA Internal Medicine. 2013;173(13);1257–9.

24. D'Alessandro DM, Kingsley P, Johnson-West J. The readability of pediatric patient education materials on the world wide web. Arch Pediatr Adolesc Med. 2001;155(7):807–12.

25. Risoldi Cochrane Z, Gregory P, Wilson A. Readability of consumer health information on the internet: a comparison of U.S. government-funded and commercially funded websites. Journal of Health Communication, 2012;17(9);1003–10.

26. Tian C, Champlin S, Mackert M, et al. Readability, suitability, and health content assessment of web-based patient education materials on colorectal cancer screening. Gastrointestinal Endoscopy. 2014;80(2);284–90.e2.

27. Langford AT, Resnicow K, Dimond EP, et al. Racial/ethnic differences in clinical trial enrollment, refusal rates, ineligibility, and reasons for decline among patients at sites in the National Cancer Institute's Community Cancer Centers Program. Cancer. 2014;120(6);877–84.

28. Baker MT, Taub HA. Readability of informed consent forms for research a veteran administration medical center. JAMA. 1983;250(19):2646–8.

29. Tarnowski KJ, Allen DM, Mayhall C, et al. Readability of pediatric biomedical research informed consent forms. Pediatrics. 1990;85(1):58–62.

30. Meade CD, Howser DM. Consent forms: how to determine and improve their readability. Oncol Nurs Forum. 1992;19(10):1523–8.

31. Hopper KD, TenHave TR, Hartzel J. How much do patients understand? Informed consent forms for clinical and research. AJR Am J Roentgenol. 1995;164(2):493–6.

32. Grossman SA, Piantadosi S, Covahey C. Are informed consent forms that describe clinical oncology research protocols readable by most patients and their families? J Clin Oncol. 1996;12(10):2211–5.

33. Grundner TM. On the readability of surgical consent forms. N Engl J Med. 1980;302(16):900–2.

34. Paasche-Orlow MK, Taylor HA, Brancati FL. Readability standards for informed-consent forms as compared with actual readability. N Engl J Med. 2003;348(8):721–6.

35. Wu TY, Hanauer DA, Mei Q, et al. Applying multiple methods to assess the readability of a large corpus of medical documents. Stud Health Technol Inform. 2013;192:647–51.

36. Facts & Figures, University of Michigan Health System. http://www.uofmhealth.org/about%20umhs/facts-figures; accessed November 15, 2014.

37. Zheng K, Mei Q, Yang L, et al. Voice-dictated versus typed-in clinician notes: linguistic properties and the potential implications on natural language processing. AMIA Annu Symp Proc. 2011:1630–8.

38. Aberdeen J, Bayer S, Yeniterzi R, et al. The MITRE Identification Scrubber Toolkit: design, training, and assessment. Int J Med Inform. 2010;79(12):849–59.

39. Hanauer D, Aberdeen J, Bayer S, et al. Bootstrapping a de-identification system for narrative patient records: cost-performance tradeoffs. Int J Med Inform. 2013;82(9):821–31.

40. Atkinson K. GNU Aspell. http://aspell.net/; accessed November 15, 2014.

41. e-MedTools. OpenMedSpel - Opensource Medical Spelling. http://e-medtools.com/openmedspel.html; accessed November 15, 2014.

42. Unified Medical Language System (UMLS) 2013AB Release Information. http://www.nlm.nih.gov/archive/20140415/research/umls/knowledge_sources/metathesaurus/release/index.html ; accessed November 15, 2014.

43. Ley P, Florio T. The use of readability formulas in health care. Psychology, Health & Medicine. 1996;1(1);7–28.

44. Zeng-Treitler Q, Kim H, Goryachev S, et al. Text characteristics of clinical reports and their implications on the readability of personal health records. Stud Health Technol Inform. 2007;129(Pt 2):1117–21.

45. Adnan M, Warren J, Orr M. Assessing text characteristics of electronic discharge summaries and their implications for patient readability. In: *HIKM '10 Proceedings of the Fourth Australasian Workshop on Health Informatics and Knowledge Management*. 2010;108:77–84.

CHAPTER 6

Conclusions

Clinical data are keystones to health care. I am dedicated to improving the quality of clinical data and maximizing the value of them. To achieve this goal, I developed a framework as described in Chapter 1 (Table 1.1) to organize potential issues that would affect the quality and value of clinical data. This framework, based on clinical data types and lifecycles, informed my three research areas and generated research ideas and projects. I included four projects (Chapters 2 to 4) in my dissertation; two of them are in my first research area and the other two are in my second and third research area, respectively.

In my first research area, I focused on 'best' practices for capturing structured clinical data more efficiently, comprehensively and accurately. I surveyed the literature and synthesized four SDE-facilitating strategies (Chapter 2), one being workflow-integration. To design a workflow-integrated SDE application, one has to understand and measure the current clinical workflow, which is usually done by ethnographic observations, interviews, work samples, and time motion studies. These types of workflow analyses, unfortunately, are limited in their capacity and cannot collect detailed data on a large scale. To address these issues, I conducted a novel study to evaluate the feasibility of using EHR audit trail log data as a cost-effective and scalable alternative to make sense of clinical workflow. (Chapter 3) The results showed that the log data have great potential to serve this purpose. However, analyzing log data usually requires input from a variety of sources to reduce the potential uncertainties found in highly complex and multi dimensional log data so that meaningful patterns in the data may be identified.

My contribution in my first research area includes proposing a comprehensive multi-strategy model to guide the design and development of future SDE application, and providing supportive evidence to the feasibility and validity of using EHR audit trail logs for clinical workflow analysis. My follow-up work involves developing a universal EHR log data schema that can support a wide range of purposes in addition to security checks, such as quality improvements and education. Also, using a universal EHR log data schema can make multi-site comparison more accurate and with less effort. This development should begin at demonstrating the benefits and limitations of using the current schema of EHR logs, and highlighting the inconsistency between EHR systems. This would motivate stakeholders such as policy makers, hospital administrators, and EHR venders, to promote the use of a universal EHR log data schema.

Another follow-up work in the first research area is the interface usability of SDE applications, which is another strategy identified in my literature review. I have been conducting a study that targets a particular dimension of usability in a particular type of SDE application, i.e. utilizing software personalization to adjust EHR systems for individual needs. This EHR personalization study, which is in its initial stage of investigation, involves semi-structured interviews and qualitative coding to learn why and how doctors personalize their EHR systems. I have completed 11 interviews, six male doctors and five female doctors, as of my dissertation defense. I have summarized the participants' demographics and clinical expertise, and will focus on synthesizing the types of personalization features, the management of these features, and the reasons for personalizing them. The findings from this study could provide insights into how to best develop future EHR personalization features to enhance the usability of EHR systems.

In my second research area, I focused on retrieving information locked in unstructured clinical data, such as discharge notes, for the purpose of supporting firsthand clinical use and secondary down-streaming uses. I participated in an NLM-funded project that developed a prototype medical information retrieval (IR) system with a semantically-based query recommendation (SBQR) feature. The analysis of the user experiment of this feature showed promising results.

My study presented in this dissertation further analyzed the query log data derived from the user evaluation. (Chapter 4) The results showed that the participants did not tend to formulate varying queries under the support of SBQR, suggesting that their perceived positive performance toward the system likely came from the effectiveness of this feature. Moreover, the SBQR seemed to work better on tasks (information needs) of average difficulty. Modern medical IR systems could consider implementing SBQR and educate users on the best timing for its use. Modern medical IR systems could also semi-automatically activate this feature if the difficulty level of users' information needs can be determined or inferred. This SBQR feature has been implemented and integrated in an institutional medical IR engine called EMERSE at the University of Michigan Health System (UMHS). EMERSE has been proved to be a highly effective and user-acceptable medical IR system. (1)

My contribution in my second research area includes developing a novel IR technique to support clinicians' search activities, and evaluating the system performance based on self-reported as well as objective data. My follow-up work will be twofold. First, I intend to migrate EMERSE to another research university's health system. While EMERSE is designed and constructed to be highly portable, whether this is the case remains an open question. EMERSE offers a free-of-charge license for an academic use so there is a low financial barrier to migration. On challenge, however, is that limited guidance is available to institutions for how to best migrate EMERSE with regard to policy, data management, system efficiency, and server configuration. My goal in this first focus, therefore, is to generate a technical report to share the experience in migrating EMERSE to another research university's health system, and to provide evidence to support its portability and generalizability. The second focus of my follow-up study will be continuing the development of potentially useful features to improve the performance of EMERSE, and conducting user studies to evaluate the effectiveness of these features. In addition, I plan to interpret clinicians' search behaviors by existing theories and models in the field of information seeking behaviors. The results can contribute to the theoretical foundation of medical information seeking.

In my third research area, I focused on identifying communication issues when sharing clinical information with patients and the general public, and developing innovative computer techniques to mitigate these issues. I published a study in this area on the readability of trial descriptions on ClinicalTrials.gov (Chapter 5). The results of this study showed that these trial descriptions were very difficult to read, even more problematic than the sample clinical notes. The findings suggest that the readability of trial descriptions on ClinicalTrials.gov should be improved. This is needed to ensure the success of this web-based registry, whose aim is to disseminate public-understandable information, and to avoid a potential selection bias that can occur if less readable trial descriptions are used for participant recruitment.

My contribution in my third research area includes presenting a pioneer study that triggers the attention of research communities to the readability of ClinicalTrials.gov (2). My follow-up work will focus on developing a special 'translator' for clinical and health information. This translator can semi-automatically generate a consumer-oriented companion version of health information deemed to be more readable for patients and the general public. This translator would also allow clinicians to keep their current practice of generating clinical data without worrying about patients being able to see information that they should not see. To design and implement such a translator, one fundamental aspect is to map professional terms to consumer (layperson) terms, e.g. mapping "ibuprofen" to "painkillers". This mapping can be derived from machine learning techniques, which often require user-labeled data to start with. I have co-conducted a research project that invited medical students to label medical concepts and modifiers in free-text posts in an online health forum. The goal of this project is to demonstrate the collection of high-quality annotation data from participants through a rigorous process, and the application of machine learning techniques to learn the mappings from the annotation data.

Another follow-up work in my third research area is to move beyond readability and focus on document comprehensibility, quality, and utility. Document comprehensibility can be approximated by the information flow between sentences and paragraphs. Document quality has

at least two dimensions: completeness and accuracy. Document utility gauges the extent to which a document is usable to patient care, billing, and other purposes. I will conduct studies to design and validate these computational measures, and demonstrate how these measures can leverage the value of EHRs. I will also pay significant attention to the user acceptance when deploying these measures in clinical routines.

In summary, in my dissertation I developed a framework to identify issues in clinical documentation and data use, and generated my three research areas: 1) structured data entry, 2) medical information retrieval, and 3) patient communication. I demonstrated the work that I have done in my PhD study and described my follow-up projects that are either ongoing or in a planning stage. The scope of my three research areas may be expanded in the future. For example, Johnson et al. (2008) proposes a novel data entry method that allows clinicians to capture structured data while typing free-text notes (4). This "structured narrative" blurs the boundary in my framework between the capture of structured data and the capture of unstructured ones. Moreover, the structured narrative adds complexity and new challenges in retrieving clinical documents that are related to my second research area. Another example is that structured data can be captured not just by clinicians, but also by devices and apps. Little is known about how to efficiently integrate these device data into a data capture process and how to effectively present them to facilitate the communication between clinicians and patients. I will consider these dimensions, expand my research areas accordingly, and investigate potential issues as well as the best strategies to deal with them.

I would like to conclude by re-emphasizing my three accomplishments in this dissertation. First, My novel methods in the areas of human-computer interaction and data mining can help clinicians capture structured clinical data more efficiently and to achieve a higher quality. Second, my methods can improve and promote the use and reuse of unstructured clinical data through information retrieval techniques considering the unique characteristics of clinical narratives. Third, my innovative approaches in natural language processing can produce

communication-effective clinical information for patients and the general public. With the ultimate goal of maximizing the quality and the value of clinical data, I will continue advancing my three research areas so I may contribute knowledge based on the findings of my work to academic and professional health informatics communities.

References

1. Hanauer DA, Mei Q, Law J, Khanna R, Zheng K. Supporting information retrieval from electronic health records: A report of University of Michigan's nine-year experience in developing and using the Electronic Medical Record Search Engine (EMERSE). J Biomed Inform. 2015 Jun;55:290–300.

2. Kang T, Elhadad N, Weng C. Initial Readability Assessment of Clinical Trial Eligibility Criteria. AMIA Annu Symp Proc. 2015;2015:687–96.

3. Garner M, Ning Z, Francis J. A framework for the evaluation of patient information leaflets. Health Expect. 2012 Sep;15(3):283–94.

4. Johnson SB, Bakken S, Dine D, Hyun S, Mendonca E, Morrison F, et al. An Electronic Health Record Based on Structured Narrative. Journal of the American Medical Informatics Association. 2008 Feb 1;15(1):54–64.