CONSOLIDATION OF CDA-BASED DOCUMENTS FROM MULTIPLE SOURCES: A MODULAR APPROACH

Seyed Masoud Hosseini Asanjan

Submitted to the faculty of the University Graduate School

in partial fulfillment of the requirements

for the degree

Doctor of Philosophy

in the School of Informatics and Computing,

Indiana University

September 2016

Accepted by the Graduate Faculty, Indiana University, in partial

fulfillment of the requirements for the degree of Doctor of Philosophy.

_____

Josette F. Jones, Ph.D., Chair

_____

Brain E. Dixon, M.P.A., Ph.D.

Doctoral Committee

_____

Daniel J. Vreeman, PT, DPT, MSc

May 17, 2016

_____

Anthony Faiola, Ph.D., M.F.A.

_____

Huanmei Wu, Ph.D.

DEDICATION

I dedicate this work to my parents, whose affection, love, and encouragements make me able to achieve such success and honor.

I also dedicate my dissertation to my loving wife, Zahra Derafshi. Without her support, this journey may have had a different outcome.

ACKNOWLEDGEMENTS

Over the past three years I have received support and encouragement from a great number of individuals. First, I thank Dr. Dixon who has been my mentor. His guidance and support has made this a thoughtful and rewarding journey. I would like to thank my dissertation committee for their mentorship as I moved from an idea to a completed study. I also thank Dr. Jones, my committee chair, for her advice and support.

Jonathan Meade and Jamie Schnitzius both provided valuable help in developing the CDA consolidation system. During data collection and writing Uzay Kirbiyik and Tim McFarlane from public health lab meetings, guided me to analyze the results. I acknowledge your kind help. I thank Zuoyi Zhang from Regenstrief institute for his assistance in the statistical analysis part. I especially thank Jennifer Williams for her help in submitting IRB forms and preparing the computer systems for developing the system. I would also like to thank the teachers who took part in this study for generously sharing their time and ideas. I have learned much though our conversations. I acknowledge Indiana Health Information Exchange (IHIE) for providing me with access to the real-world CCDs. Finally, thanks to physicians, researchers, and nurse practitioners at Regenstrief institute and IU school of nursing because of participating in this study as interviewees.

Seyed Masoud Hosseini Asanjan

Consolidation of CDA-Based Documents from Multiple Sources: A Modular Approach

Physicians receive multiple CCDs for a single patient encompassing various encounters and medical history recorded in different information systems. It is cumbersome for providers to explore different pages of CCDs to find specific data which can be duplicated or even conflicted. This study describes the steps towards a system that integrates multiple CCDs into one consolidated document for viewing or processing patient-level data. Also, the impact of the system on healthcare providers' perceived workload is evaluated.

A modular system is developed to consolidate and de-duplicate CDA-based documents. The system is engineered to be scalable, extensible and open source. The system's performance and output has evaluated first based on synthesized data and later based on real-world CCDs obtained from INPC database. The accuracy of the consolidation system along with the gaps in identification of the duplications were assessed. Finally, the impact of the system on healthcare providers' workload is evaluated using NASA TLX tool.

All of the synthesized CCDs were successfully consolidated, and no data were lost. The de-duplication accuracy was 100% based on synthesized data and the processing time for each document was 1.12 seconds. For real-world CCDs, our system de-duplicated 99.1% of the problems, 87.0% of allergies, and 91.7% of medications. Although the accuracy of the system is still very promising, however, there is a minor inaccuracy. Due to system improvements, the processing time for each document is reduced to average 0.38 seconds for each CCD.

The result of NASA TLX evaluation shows that the system significantly decreases healthcare providers' perceived workload. Also, it is observed that information reconciliation reduces the medical errors. The time for review of medical documents review time is significantly reduced after CCD consolidation.

Given increasing adoption and use of Health Information Exchange (HIE) to share data and information across the care continuum, duplication of information is inevitable. A novel

system designed to support automated consolidation and de-duplication of information across clinical documents as they are exchanged shows promise. Future work is needed to expand the capabilities of the system and further test it using heterogeneous vocabularies across multiple HIE scenarios.

Josette F. Jones, Ph.D., Chair

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBRIVIATIONS

Application Programming Interface (API)

Centers for Medicare and Medicaid Services (CMS)

Clinical Document Architecture (CDA)

Consolidated CDA (C-CDA)

Continuity of Care Document (CCD)

Electronic Health Record (EHR)

Health Information Exchange (HIE)

Health Information Exchange Organization (HIO)

Health Level Seven Standard (HL7)

Hoosier Healthcare Innovation Challenge (HHIC)

Indiana Health Information Exchange (IHIE)

Indiana network for patient care (INPC)

Integrating the Healthcare Enterprise (IHE)

Master Patient Index (MPI)

Meaningful Use (MU)

Model-view-controller (MVC)

National Aeronautics and Space Administration Task Load Index (NASA-TLX)

National Patient Safety Goal (NPSG)

Natural Language Processing (NLP)

Reconciliation of Clinical Content and Care Providers (RECON)

Systematized Nomenclature of Medicine - Clinical Terms (SNOMED CT)

The Joint Commission (TJC)

University of Mississippi Medical Center (UMMC)

Virtual Lifetime Electronic Record (VLER)

World Health Organization (WHO)

Extensible Markup Language (XML)

DEFINITIONS

Health Information Exchange (HIE): HIE has been defined as the communication of clinical data such as problem lists, clinicians' notes, or other critical medical information from one provider organization (e.g., doctor's office) to another (e.g., hospital) (Jha, Doolan, Grandt, Scott, & Bates, 2008).

Health Information Exchange Organization (HIO): The term HIE is used as broad as faxing a patient's medical record from provider to provider and as specific as entities and specialized organizations facilitating electronic exchange of health information among healthcare system stakeholders, such as hospitals and physician practices (B. E. Dixon, Zafar, & Overhage, 2010). In this dissertation, HIO is used as a type of organization that governs health information exchange among health care stakeholders within a defined geographic area.

Health Level Seven Standard (HL7): HL7 is one of the leading Standards Development Organizations (SDO) for exchange of clinical and administrative data among healthcare information systems and is recognized as the premier international SDO in the health care domain (Health Level Seven web site).

Clinical Document Architecture (CDA): "The HL7 CDA is a document markup standard that specifies the structure and semantics of a clinical document (such as a discharge summary, progress note, procedure report) for the purpose of exchange. A CDA document is a defined and complete information object that can include text, images, sounds, and other multimedia content. It can be transferred within a message, and can exist independently, outside the transferring message (Robert H Dolin et al., 2006)."

Continuity of Care Document (CCD): The CCD is an electronic document exchange standard developed by HL7 for sharing patient summary information between providers and organizations. CCD constrains the CDA and is provided as a template to improve the interoperability between clinical information systems (Dolin, Giannone, & Schadow, 2007).

Modular System: The main idea behind modular system design is to break a complex system down to a set of totally or partially independent components called modules and then plug these components together to work as a system. The advantage of this design is that these modules can be used in different systems independently.

CHAPTER 1

INTRODUCTION

Receiving care from multiple providers who practice in different locations and facilities leads to creation of medical records that are fragmented among multiple locations. Health Information Exchange (HIE) is a technique to electronically exchange patients' medical information across multiple organizations in order to improve the access to timely and longitudinal health information (Brian E. Dixon, 2016). Healthcare providers make clinical decisions based on the information they obtain from different sources, including medical records, reference books and colleagues (Dawes & Sampson, 2003); however, a national study shows they prefer to use electronic exchange of health information compared to paper reports since they believe different aspects of care and clinical workflow efficiency can be improved through using HIE (Patel, Abramson, Edwards, Malhotra, & Kaushal, 2011). Also, it is estimated that HIE can save billions of dollars through reduction of redundant tests or costs associated with paper-based ordering and reporting of results (Walker et al., 2005).

The Clinical Document Architecture (CDA) is a document markup format standard that can be used to represent a typical medical document such as discharge summary, progress note, or operative note (Robert H. Dolin et al., 2006). This document is designed by Health Level Seven (HL7) an international standards development organization that creates standards for exchanging clinical and administrative data among health information systems (Hosseini, Ahmadi, & Dixon, 2014). Many templates can be applied on CDA structure in order to create new medical documents for specific purposes. Continuity of Care Document (CCD) is one of the CDA templates that is not intended to be a complete medical history, but to provide the most critical information that is necessary for "the continuation of care". The CCD is broken across 17 sections that cover summary of administrative, demographic, and clinical information facts. The structure and semantics of CDA is originally designed by HL7 for the purpose of data exchange.

Being a template of CDA, CCD is planned to be used for exchange of "critical clinical information" among healthcare providers (Healthcare Information Technology Standards Panel (HITSP), 2009). Exchange of CCDs is required for HIE activities under the Meaningful Use criteria in order to increase the adoption of interoperable clinical information systems ("Medicare and Medicaid programs; electronic health record incentive program--stage 2. Final rule," 2012).

Whereas HIE significantly benefits the health sector, overloading physicians with a lot of medical records including duplicative and conflicting information can disrupt clinical workflow and lead to low quality of care delivery. Presence of too much information creates "information overload" that hinders physicians' understanding about patients' problems and makes the decision making process challenging (Hall & Walton, 2004). Without an efficient and effective method for de-duplicating redundant patient data, providers everywhere will be forced to review lengthy and redundant information spread across multiple documents. This will create further information overload and usability challenges for EHR systems, which are already criticized for not supporting clinical workflow (McDonald et al., 2014; Ratwani, Fairbanks, Hettinger, & Benda, 2015).

The CCD has a potential to carry duplicative and conflicting information if it is generated multiple times for a patient. Health Information Exchange Organizations (HIO) receive multiple CCDs for the same patient from different sources, but typically do not integrate them into a unified clinical summary. This challenge is another picture of the information duplication problem that faces healthcare organizations, specifically HIOs. Hence, the Indiana Health Information Exchange (IHIE) proposed CCD consolidation and de-duplication as a 'grand challenge' facing the HIE marketplace ("Hoosier Healthcare Innovation Challenge ", 2013). Clinical information reconciliation is a Certification Criteria for Electronic Health Record Technology that is proposed by Meaningful Use to address this challenge (Office of the National Coordinator (ONC), 2015). Clinical information reconciliation is a process of comparing and

harmonizing the information from multiple sources in order to remove redundant data and create a consistent and coherent medical record.

Currently, there is no proven solution for reconciliation and consolidation of multiple CCDs. The causes of duplication and the areas in which duplication happens in CCDs are also not clearly identified. In order to improve the interoperability among information systems and reduce information overload, a novel approach is needed to address challenge.

1.1. **Background**

Evidence Based Medicine (EBM) aims to transfer the lessons learned from research into routine practice and help clinicians to make clinical decisions based on the evidence from well designed and conducted research (Deshpande, Publicover, Gee, & Khan, 2003). To be compliant with EBM, physicians have to efficiently use research-based evidence along with patients' medical records in their clinical practice (Halligan & Donaldson, 2001); however, fast growing production of healthcare information and increasing number of published medical research leads to information overload that makes it difficult for practitioner to adopt the principles of the EBM (Klerings, Weinhandl, & Thaler, 2015; Noone, Warren, & Brittain, 1998).

Overflow of information in medical records creates similar problem when physicians receive several medical records for a patient gathered from multiple sources. This requires the physicians to sift through multiple documents that include potentially duplicative or conflicting information. Identifying the status of problems, allergies and medications among multiple documents needs a great deal of time and precision. This task becomes more difficult and time consuming if the patient has a complex medical history with dozens of problems, medications, and allergies.

To combat this challenge, information reconciliation and data de-duplication is addressed in the national and international level. The World Health Organization (WHO) launched the Collaborating Centre on Patient Safety initiative to identify, gather and disseminate existing

patient safety solutions to prevent harm to patients. One of the suggested solution areas in this initiative is focusing on actions for information reconciliation accuracy at transition of care (World Health Organization, 2007). According to the Meaningful Use (MU) rules, clinical information reconciliation is needed in the U.S. as a component of the Health IT Certification Criteria for Electronic Health Record Technology (Office of the National Coordinator (ONC), 2015). Medication reconciliation is also considered as a key patient safety factor in the Safety and Quality Improvement Guide Standard in Australia (Australian Commission on Safety and Quality in Health Care, 2012)

Much research also has focused on information reconciliation in different healthcare domains. For example, researchers developed different computer-based medication reconciliation solutions to reduce medication errors and overcome the barriers of manual reconciliation (Bassi, Lau, & Bardal, 2010; Plaisant et al., 2013; Pronovost et al., 2003; Silva et al., 2011; Ziminski, De la Rosa Algarin, Saripalle, Demurjian, & Jackson, 2012). Clinical information reconciliation is not restricted to medication domain, but various works has been done in other domains such a medical imaging (Reiner, 2011a, 2011b, 2011c), cancer screening (Wagholikar et al., 2013), and immunization information de-duplication (Hosseini et al., 2014; Miller, Frawley, & Sayward, 2000). Recently, Integrating the Healthcare Enterprise (IHE) published two RECON profiles to address clinical information reconciliation: (1) Reconciliation of Diagnoses, Allergies and Medications (RECON) (Integrating the Healthcare Enterprise (IHE), 2011), and (2) Reconciliation of Clinical Content and Care Providers (RECON) (Integrating the Healthcare Enterprise (IHE), 2015). IHE is a nonprofit organization that brings together key stakeholders in healthcare domain to coordinate interoperability activities. Upon its common framework, IHE develops implementation guidelines called “IHE Profiles” across a range of use cases in healthcare as well as public health. The RECON profiles describe some heuristics and considerations that can be helpful in the automation of reconciliation process. These heuristics can guide the designers of reconciliation systems to be aware of the issues and opportunities

available within medical records that can cause or prevent redundancy. The profiles are limited to three specific domains (diagnoses, allergies and medications), in which many domain specific points are described to assist the process of identifying duplications and conflicts.

While there have been extensive work regarding clinical information reconciliation, duplication of information in CDA-based documents is rarely or not addressed in research-based publications. For instance, reconciliation of medications has been automated through different techniques or immunization information de-duplication is addressed in many research studies; however, there is no optimal solution for reconciliation of clinical information in CDA-based documents. Also, data de-duplication in the essential sections of medical records such as allergies, diagnosis, and problems are not addressed neither independently nor as part of CDA or CCD documents. Although IHE addressed the reconciliation of these sections in the RECON profiles; however, they differ considerably from this work. First, the RECON profiles are not scientific publications nor based on original research studies - IHE profiles describe clinical information management use cases and specify how to use existing standards to improve health information interoperability. Second, these profiles only describe some considerations and heuristics for automating clinical information reconciliation and do not specifically focus on de-duplication of information in the CDA-based documents.

It is just a few years that Meaningful Use encourages the healthcare organizations to exchange Consolidated CDA (C-CDA) and CDA reconciliation is still a challenge facing HIOs and healthcare facilities. Many CCDs are collected in HIOs without being reconciled due to lack of any proven solution for CCD consolidation. The duplicative information is hard to be reconciled by physicians since they are often larger than the capacity of most peoples' working memory. Therefore, physicians deal with large amount information in their clinical practice. A novel technique is needed for detecting duplications and conflicts in CDA-based documents and generating a consolidated document. This technique should be able to minimize the burden of information overload and improve the quality of information.

1.2. **Theoretical Background: Information Overload**

The performance of an individual is positively correlated with the amount of the information she receives up to a certain point. In the other words, if an individual receives information beyond a certain point, her performance will decline (Chewning & Harrell, 1990). Once the individual is positioned in this condition called "information overload", providing further information will no longer be used by her for the purpose of decision making (O'Reilly, 1980). Figure 1, presents the decline in the performance after passing the information overload threshold (Schroder, Driver, & Streufert, 1967).

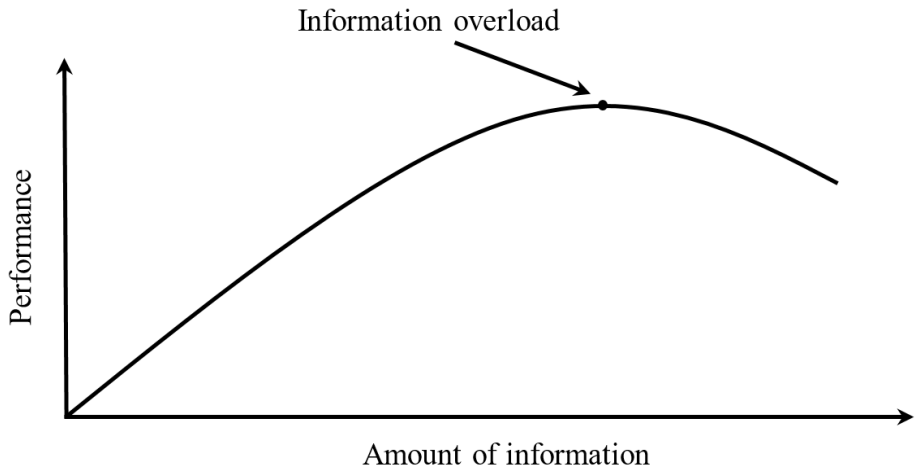

Figure 1. Decline in the performance after passing the information overload threshold.

Eppler and Mengis examined the theoretical basis of the information overload based on the literature in different domains to identify the main elements and situations that cause information overload. They consolidated the result of their research in a conceptual framework that shows the cause and effect of information overload in an organization (Eppler & Mengis, 2003). The main focus of this framework is on the relation of an individual's performance with the amount of information she receives. In this framework, they identified the relationship between main causes of information overload, their symptoms in organizations as well as suitable countermeasures to prevent the effects of information overload.

According to this conceptual framework, there are five main causes for information overload that are mutually interdependent: "Information itself (its quantity, frequency or intensity, and quality or general characteristics), the person receiving, processing or communicating information, the tasks or processes which need to be completed by a person, team or organization, the organizational design (i.e., the formal and informal work structures), and the information technology that is used (and how it is used)."

Among the aforementioned causes, the use or misuse of information technology (IT) is considered as a critical issue for creating information overload in the recent years (Bawden, 2001; Bawden & Robinson, 2009). Although information retrieval time is reduced through implementation of IT systems; however, it can also increase the amount of potentially useless and redundant information that a person should deal with (Edmunds & Morris, 2000).

HIOs are health-related organizations that mainly deal with information technologies. Gathering multiple CCDs for a patient by HIOs and retaining this information without reconciliation, leads to accumulation of duplicative and conflicting information. Providing this information to physicians is an example of information overload problem that is happening in the healthcare sector.

1.3. **Overview of the Dissertation**

The main goal of this work is to address the problem of information overload through consolidation of CDA-based documents and removing duplicative and conflicting information. This dissertation is comprised of three distinct scientific articles focusing on three aims of the dissertation. The aims are tailored to address this challenge in three steps: 1) development of CDA consolidation system, 2) evaluation of the system, and 3) assessing the impact of system on healthcare providers. The research aims and the questions addressed in each aim are as follows:

1.3.1. Aim 1. Developing a modular system to de-duplicate and consolidate CDA-based documents received from multiple sources.

Question 1. How accurate is the CDA Consolidation system compared to manual consolidation?

Question 2. What is the performance of system (the average time to process a document as well as the minimum, maximum, and average file sizes)?

Question 3. Is the Consolidated CCD valid based on HL7 standard?

1.3.2. Aim 2. Characterization of data representation among multiple documents.

Question 4. What are the challenges for identifying duplications?

Question 5. What are the differences in data representation among multiple documents?

Question 6. How much the real-world CCDs are compliant with HL7 standard?

1.3.3. Aim 3. Conducting a scenario-based study to evaluate the impact of the system on healthcare providers' perceived workload.

Question 7. What is the impact of system on participants' workload in the following sub-classes? Mental demands, Physical demands, Temporal demands, Performance, Effort and Frustration.

Question 8. What is the impact of system on participants' accuracy and efficiency in reconciling CCDs?

CHAPTER 2

STUDY 1 - CONSOLIDATING CCDS FROM MULTIPLE DATA SOURCES: A MODULAR APPROACH

2.1. **Background and Significance**

To make accurate care and treatment decisions, clinicians require complete, timely and relevant information about a patient. Unfortunately clinical information is fragmented across many different organizations and heterogeneous computer systems. Factors such as variations in insurance coverage, reliance on multiple providers, increasing specialty care, and disparate information systems contribute to information fragmentation (Vest & Gamm, 2010). Fragmentation creates clinical workflow inefficiencies, leads to over-utilization of health care through duplicative testing or imaging, and prevents coordination of care across healthcare systems (Frisse et al., 2012; Schoen et al., 2012). Health Information Exchange (HIE) can ameliorate fragmentation by facilitating communication between disparate health information systems (Acker et al., 2007; Brailer, 2005; Hripcsak et al., 2007).

Independent information systems providing clinical information are usually syntactically or semantically incompatible, making HIE challenging (B. E. Dixon, Vreeman, & Grannis, 2014). Standards such as Health Level Seven (HL7) support HIE amongst disparate information systems. HL7 is one of the leading standards for exchange of clinical and administrative data among healthcare information systems and is the most widely recognized international standards development organization in the health care domain(Hosseini et al., 2014; Office of the National Coordinator (ONC), 2014).

The CCD is an electronic document exchange standard developed by HL7 for sharing patient summary information between providers and organizations. The CCD constrains the HL7 Clinical Document Architecture (CDA) and is provided as a template to improve the interoperability between clinical information systems (Dolin et al., 2007). Further, the CCD

standard is required for HIE activities under the Meaningful Use criteria, regulations from the CMS that incentivize the adoption and use of interoperable clinical information systems ("Medicare and Medicaid programs; electronic health record incentive program--stage 2. Final rule," 2012).

Given pressure from health reform and market forces, clinical provider organizations and their information system vendors are required to generate and process CCD documents that contain current and past medical information about a patient. Sometimes a clinical information system or HIE organization must aggregate multiple CCDs received from a variety of sources and generate a consolidated CCD to facilitate delivery of complete information to a clinician. Care providers prefer to review consolidated information that represents a single, comprehensive picture of a patient's medical history and current condition, rather than multiple CCDs that may include duplicate or conflicting information. Providing too many documents or too much information could require providers to review many pages of a medical record to find a proverbial 'needle in the haystack', which could affect the quality of care delivered to patients.

HIE organizations, which facilitate the electronic transfer of clinical information among a group of health care organizations, seek efficient methods for consolidating multiple CCDs for a given patient. Consolidation of CCDs, however, is challenging and requires processes that can interpret, merge, de-duplicate, and resolve conflicts across complex documents involving complex data types. Given that meaningful use regulations and many HIE organizations are less than three years old, there do not exist a plethora of proven solutions for consolidating CCDs. HIOs need novel, innovative ideas and products to assist them in managing the growing volume of CCDs available for a patient in support of health care reform goals that seek to provide higher quality, less costly health care to a greater number of people.

In this article, we describe a modular approach to the consolidation and de-duplication of CCDs. The approach is designed to support instances of HIE where there is the need to consolidate multiple clinical documents prior to presentation of information to clinicians or

consumption into electronic health record (EHR) systems. We further describe an initial prototype system that embodies the approach. Although the prototype system is limited in its focus on consolidating three sections of a CCD (e.g., allergies, medications, problem list) and cannot resolve disparate information representations, the approach is nonetheless an important first step. The rest of the paper is organized as follows. Section 2 describes the design of the approach, prototype system and the methods used to pilot data acquisition and analysis. The prototype system was evaluated using synthetically generated CCDs from a single HIE with a common local vocabulary and did not consider disparate CCDs from multiple vendor systems with heterogeneous vocabularies. Section 3 is dedicated to the results of the pilot system evaluation and feasibility testing. Finally, Section 4 summarizes our conclusions and proposes future research directions, which include evaluation using real-world documents from multiple vendor systems and expansion to resolve semantic meaning from disparate information representations.

2.2. **Materials and Methods**

In order to aggregate multiple CCDs received from different sources and HIOs, we developed a modular solution to combine the CCDs into a single document and resolve the conflict and duplication of multiple documents for each patient.

2.2.1. Study Context

The design and development of our approach originated as a component of the 2013 Hoosier Healthcare Innovation Challenge (HHIC). HHIC is a state-based software development competition that brings together healthcare and technology professionals to provide creative solutions for some of the most challenging problems in healthcare ("Hoosier Healthcare Innovation Challenge ", 2013). As a part of the 2013 HHIC, the Indiana Health Information Exchange (IHIE) proposed CCD consolidation and de-duplication as a ‘grand challenge’ facing

the HIE marketplace. IHIE is one of the largest HIE organizations in the U.S., serving 25,000 physicians and over 10 million patients (Biondich et al., 2014; Biondich & Grannis, 2004). The HIE includes multiple integrated delivery networks, hospitals, physician practices, laboratories, radiology centers, long term post-acute care facilities, and public health agencies.

In recent years, IHIE observed that multiple HIE partners, including other HIE networks in Indiana, members of the eHealth Exchange national HIE, and private enterprise health systems within its own network, began implementing CCD interfaces to send and receive information about individual encounters as well as a patient's past medical history (Byrne et al., 2014; B. E. Dixon, Colvard, & Tierney, 2014). A total of 58 facilities within the IHIE network regularly send CCDs within the exchange. The growth in CCD interfaces was driven by meaningful use incentives, which emphasize the CCD as a preferred message type for providing a summary of care as patients' transition in the health system.

In order to meet customer demands, IHIE needed to gather a wide range of CCDs 'just-in-time,' consolidate the information across documents, and provide a single record of summarized medical information to an application or end user. For example, between July 2014 and April 2015, a total of 767,758 CCDs were exchanged for 359,503 unique patients. While many patients only had one CCD sent, the number of CCDs per patient ranged from 1 to 17 with the number of facilities ranging from 1 to 13. Over 12,000 patients had at least 2 CCDs exchanged during the 9 months. Given the variety of CCD documents potentially available for a given patient, IHIE perceived the need to implement a solution to consolidate and de-duplicate CCDs as they are exchanged across its network. To address this problem, the authors formed a team for the HHIC competition and developed an approach which was ultimately accepted by IHIE and selected as a best solution in the 2013 HHIC competition. The proposed solution was required to be open source. The original version of the system was published in GitHub ("CCD deduplication system - Source Code," 2013).

2.2.2. System Description

We designed the system to meet three paradigms in computer science: scalability, extensibility, and open source. IHIE desired the solution to provide for evolving needs as the HIE market and health IT standards evolved. This required the solution to be scalable, or expandable to greater numbers of interfaces and volumes of CCDs over time, and configurable as IHIE's customer needs or CDA document types evolved over time. Currently IHIE and its participants are exchanging the HITSP/C32 - Summary Documents Using HL7 Continuity of Care Document ("C 32 - HITSP Summary Documents Using HL7 Continuity of Care Document (CCD) Component,"), a variety of CDA document as required for Stage 1 of the U.S. Centers for Medicare and Medicaid Services (CMS) "meaningful use" program ("Health information technology: initial set of standards, implementation specifications, and certification criteria for electronic health record technology. Final rule," 2010). However, this is expected to evolve towards the Consolidated CDA (C-CDA) with later stages of meaningful use adoption by IHIE and its participants. Furthermore, Regenstrief, IHIE and their partners believe firmly in an open source philosophy or that the system should be available and extensible by not only the original developers but the entire HIE community. Therefore, we chose a modular approach implemented on an open source stack.

The system architecture (Figure 2) consists of three primary components: 1) a CCD Consolidation Engine that receives multiple CCD documents through an Application Programming Interface (API) and executes a series of rules to consolidate and de-duplicate data; 2) an Audit component that identifies and logs system events; and 3) a configuration component that enables implementers to configure the rules executed by the system. Currently, the system is designed with a RESTful API and a simple web application to enable users to configure their desired rule set but this can easily be swapped out to meet the needs of the application calling the system. Multiple CCDs are fed into the Consolidation Engine and the system returns a merged

CCD after executing variety of rules defined and configured in Rules engine. During the execution of processes, audit information is captured into a NoSQL database.

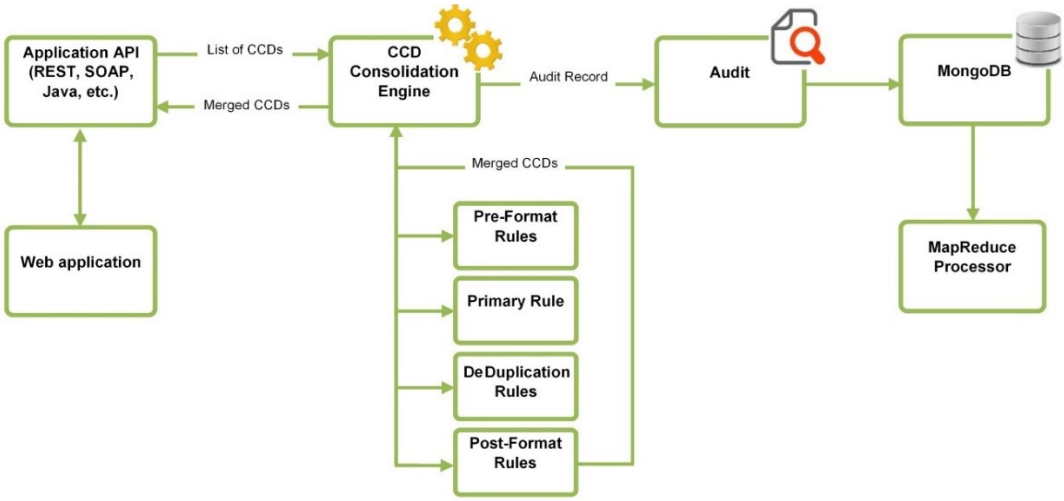

Figure 2. CCD consolidation application model.

2.2.2.1. *CCD Consolidation Engine*

The heart of the solution is the CCD Consolidation Engine. The engine is designed to execute a set of rules to consolidate and de-duplicate information across a set of CCDs into a single CCD. Four classes of rules were developed:

Pre-Format Rules examine incoming CCDs and inspect them for quality. Experience has shown that electronic clinical messages can vary in their interpretation of HL7 standards (Barnes, 2007), leading to imperfect messages that contain important data. Pre-Format rules can normalize incoming messages and enable imperfect messages to be processed (Zafar & Dixon, 2007). The Pre-Format rules can execute functions such as validating CCDs (e.g., applying criteria established by HITSP and HL7) (Healthcare Information Technology Standards Panel (HITSP), 2009; "HITSP Clinical Document and Message Terminology Component," 2010; "HITSP Summary Documents Using HL7 Continuity of Care Document (CCD) Component," 2009), dropping invalid CCDs (if this is desired), or strip unwanted sections of the CCD out.

Primary Rule ensures that the program returns something to the calling application. The idea is that one record in the list of CCDs is chosen as the "master CCD." Selection of the master CCD is based on a simple rule: choose the one with the highest number of entries (e.g., the biggest). The other CCDs are merged into the master CCD. If the merger fails for any reason, then the master CCD along with the non-merged CCDs will be returned. This rule prevents the system from failing to return anything and potentially leaving a clinician without any information to view, something that can lead to user frustration and abandonment of health IT systems.

De-duplication Rules merge, de-duplicate and handle conflicting information. The initial prototype contains nine main classes of rules. Some classes focus on examining and consolidating specific sections: e.g., Medications, Problems, Vital Signs, and Immunization. Other classes focus on resolving conflicts for certain attributes of CCD entries: e.g., Assigned Authors, Confidentiality. Each class is configurable and expandable to meet local needs. For example, in Medication class, functions extract data from the entry level of each CCD. The extracted information includes Medication date, Medication Status Code (e.g., active), administrationUnitCode (e.g., Tablets), manufacturedMaterial (e.g., Raloxifene 60 MG Oral Tablet), Medication Code (e.g., 312768), Medication CodeSystem (e.g., 2.16.840.1.113883.6.88), Medication OriginalText (e.g., Evista), Medication Name (e.g., Evista), and healthcare provider information. Then through firing multiple rules (usually comparison conditions), duplicated entries are discarded and the remaining items are merged into master CCD. To enable comparison, the class requires a minimum of three fields (code, code system, and healthcare provider name). If an entry does not possess at least these fields, then it is automatically retained or merged into the master CCD.

Post-Format Rules check the consolidated document for internal validity and ensure the format of the document conforms to applicable HL7 CDA and CCD constraints.

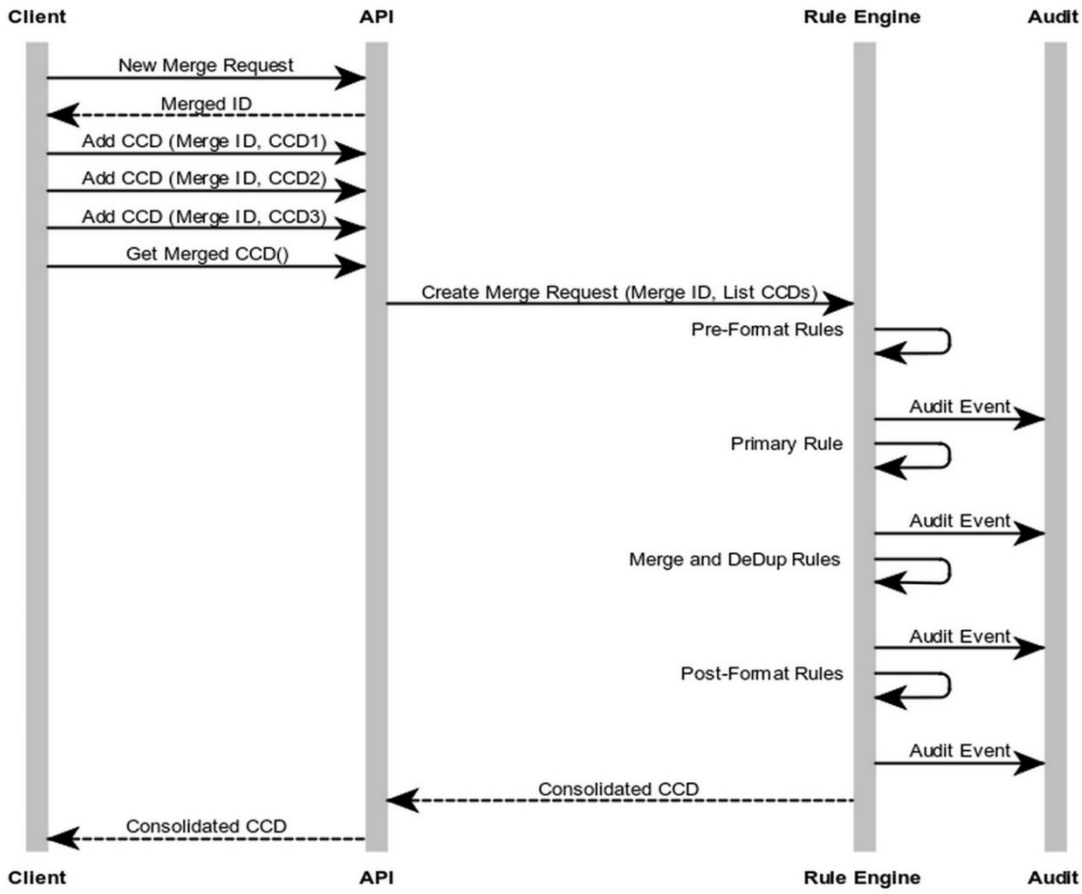

Figure 3. Depicts the sequence Diagram of consolidation of input CCDs through different modules of the system. This diagram shows that three CCDs are passed through an API to the Rules Engine. Rules are executed and, for every rule, one audit event is recorded (the list of input CCDs, the merged CCD, discarded data elements, and information on the executed rule) into a NOSQL database. Finally, the consolidated CCDs are returned to the calling application or service.

2.2.2.2. *Audit*

After each rule in the Consolidation Engine is executed, the engine sends the list of input CCDs, the merged CCD, discarded data elements, and information on the executed rule to the audit system. The audit system stores this information in a NoSQL database, such as MongoDB. The NoSQL database is used because of the large volume of data being sent by the audit system. NoSQL databases are designed to deal with large amounts of data for long term storage. Figure 3

presents a sequence diagram for consolidation of three CCDs through our application including feeding, merging and auditing processes.

2.2.2.3. *User Configuration*

An application layer was developed on top of the Consolidation Engine to enable user configuration. A simple demo web application (Figure 4) presents how user can select a set of rules to be fired against CCDs.

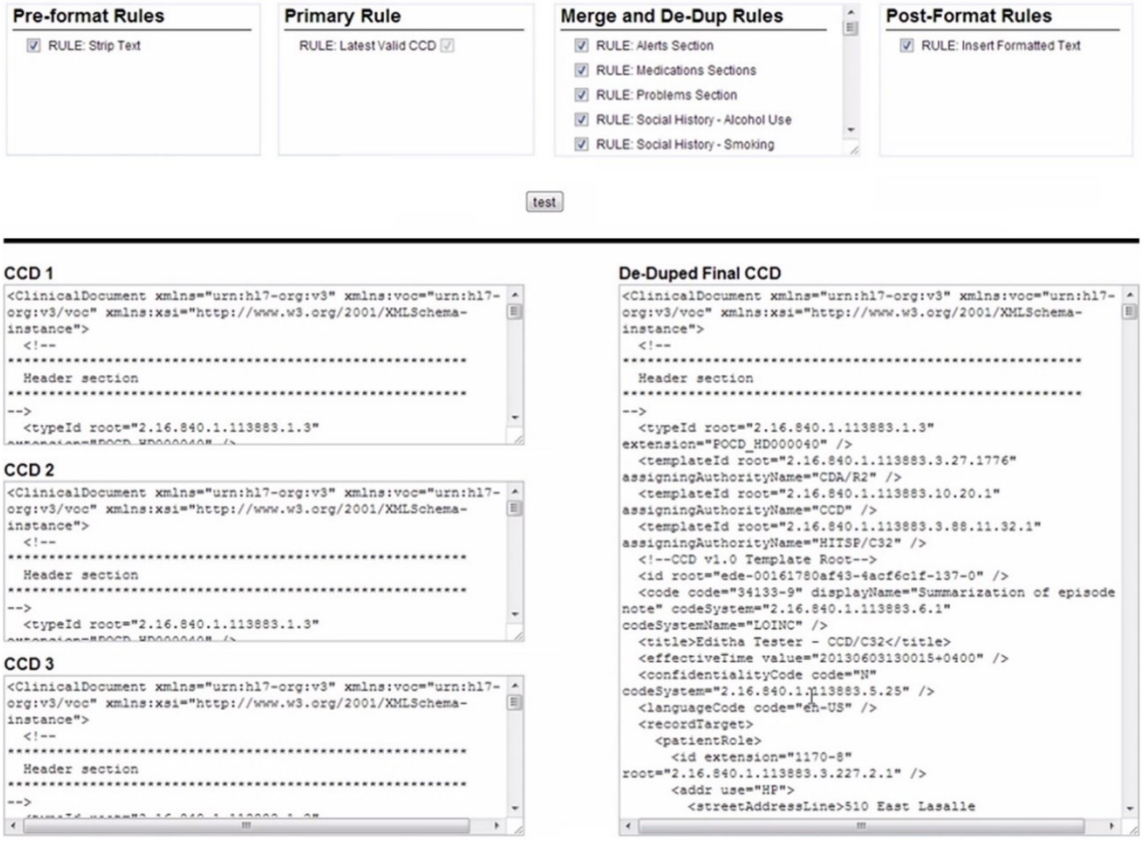

Figure 4. A Web-based user interface that enables selection of the desired rules to be fired against input CCDs. The three input CCDs are loaded into the text boxes at the left side of the figure and the consolidated, de-duplicated CCD in the right side of the figure appears after pushing the test button. The user can change the rule options from the top side of the interface as many time that is needed and the system will generate a new, consolidated CCD based on the selected rules.

2.2.3. Study Design

We evaluated the system using a corpus of 150 CCDs representing 50 unique patients sampled from IHIE's over 10 million patients (Biondich & Grannis, 2004; McDonald et al., 2005). Although a variety of rules have been developed in the Consolidation Engine, we focused the evaluation on three important sections of the CCD: Medications, Problems List, and Allergies. These sections contain key information necessary to perform important clinical tasks, such as medication reconciliation, drug-drug interaction checking, and drug-allergy checking.

The CCDs were input into the system for consolidation and de-duplication in batches of three documents where the three documents simulated disparate institutional data for the same patient. The system output was a single, consolidated CCD. We examined system performance as well as the difference between the input documents and the output document. Our analysis focused on whether the system correctly identified and de-duplicated the same information spread across three disparate CCDs.

The sample size is sufficient for detecting a difference in the size or number of entries for 50 (N) patients with a moderate effect size (0.4) at 80% power. For a corpus of 150 CCDs (3N), the largest decrease possible when the documents are exactly the same is 67%; but if there are duplicates within each report, the largest decrease would be >=67%. While one could choose to evaluate a range (2N, 3N, 4N, 5N) of options with respect to the number of CCDs, 3N has some basis in reality: an emergency encounter with resulting hospital stay followed by primary care follow-up; or a primary care visit followed by specialty visit followed by primary care follow-up. In these scenarios, a CCD would be generated when there is a transfer of care. While a larger sample of patients would be necessary to detect smaller effect sizes, we anticipated at least moderate effect sizes given our methods for generating the synthetic CCDs.

2.2.4. Methods for Data Acquisition

To generate the corpus of CCDs for testing, we first used the production instance of the Regenstrief CCD generator – a web service that generates a CCD for a given patient – to create de-identified CCDs for 50 unique patients selected from a population of 12 million. We randomly selected patients with a minimum of seven active problems, five active medications, and three documented allergies. We then twice cloned the CCDs to create multiple input documents for each patient. Next we manipulated the cloned CCDs to simulate disparate clinical information drawn from various EHR systems. An algorithm manipulated the medication, problem list, and allergy sections of the CCDs, randomly deleting existing entries and/or adding new entries to the documents.

After the clones were manipulated, the corpus contained 150 unique, de-identified documents for 50 unique patients. Manipulated CCDs were fed into the CCD consolidation service, in sets of 3, and merged CCDs were generated for each patient. When input into the system for testing, the system could not distinguish between the original or manipulated clone CCDs.

2.2.5. Methods for Data Analysis

System performance was determined based on the total time necessary to process the corpus of 150 input CCDs. We examined the average time to process a CCD as well as the minimum, maximum, and average file sizes. To examine accuracy, we manually de-duplicated the CCDs and compared the manually merged document with the consolidated CCD generated by the system.

2.3. **Results**

All 150 input CCDs were successfully consolidated into 50 output CCDs. Table 1 summarizes the number of entries within each section of the CCDs before and after consolidation.

The consolidation process significantly reduced the number of entries due to duplicate values in each section that the engine identified and removed during the consolidation process. The consolidation process also reduced the size of the documents, as well as the number of lines in each document. Manual review of the results confirm that the Consolidation Engine correctly identified all (100%) of the duplicate entries.

	A: Number of Entries in Input CCDs (N=150)	*B: Number of Entries in Consolidated CCDs (N=50)*	*Percentage Decrease from A to B*
Problems Section	7,511	3830	49%
Medication Section	4,822	1900	60.6%
Allergy Section	1,309	275	79%
File size on disc (KB)	92,205	39,199	57.5%
Number of lines in CCDs	1,636,385	687,192	58%

Table 1. Comparison of CCD properties before and after consolidation.

The current prototype is executing consolidation and de-duplication rules against the test dataset at a rate of approximately .009 to .03 seconds per rule depending on the complexity of the rule. The overall de-duplication time for 150 CCDs is 169 seconds which covers the time for firing rules on our focused sections (Problems, Medications, and Alert) and other rules such as Primary Merge Rule, Vital Signs, and Confidentiality.

The audit write method takes longer than the actual consolidation but the auditing methods can be assigned to a separate processor thread to run, which could improve performance. According to the Audit logs, the largest file size among the 150 CCDs is 6912 KB and the smallest one is 90 KB. The overall file size of 150 CCDs is 90.04 MB. For this size of data, the number of de-duplication occurred in Problems section is 3686 and this number is 2923 for Medications and 1035 for Alerts section. The overall number of entries in Problems section is 7511, in Medications 4822, and in Alerts section is 1309 (Table 1).

2.4. **Discussion**

Consolidating and de-duplicating information in clinical summary documents is a requirement of HIE systems, because data are fragmented and often duplicated across providers. In this study, we performed preliminary evaluation of a CCD Consolidation Engine designed to merge and de-duplicate CCDs as they are exchanged for consumption into an EHR or immediate presentation to a clinical user. Preliminary results are favorable, showing that with simulated data the consolidation engine was able to identify all (100%) of the duplicated entries for a limited set of document sections across groups of CCDs for the same patient.

Decreasing the number of entries (49% in Problems entries, 60.6% in Medications entries, and 79% in Allergies entries) may help providers to read medical information without confronting duplicate entries which may cause frustration or safety concerns. For example, when reconciling medication data to determine potential changes in a patient's drug regimen, duplicate entries may require additional time to review the list leading to clinician frustration or they may be hard to detect leading to unintended consequences. Such a scenario would only exacerbate clinician fatigue and usability challenges presented by HIE systems (B. E. Dixon et al., 2013; Gadd et al., 2011). Furthermore, consolidation significantly reduces the size of CCDs, which has an impact on the time required to process and present information to users in real-world systems. Even if there is no duplication among a group of CCDs, consolidation of documents reduces the overall size of the information because, at minimum, patient demographics are repeated across CCDs. This may speed processing of the information and potentially aid in the display of the information to end users. Therefore the consolidation may be a useful component of operational HIE systems, or implemented within a hospital or health system to process incoming or outgoing CCDs for a given patient before consuming the potentially duplicative information or rendering it to a clinical end user.

System performance is adequate for many use cases but insufficient for others. Total time to process 150 CCDs is 169 seconds, just over one second per document. This is likely reasonable

in an automated use case where input CCDs are considered for consumption into an EHR system, say, during an overnight batch process. However, this performance may not be adequate in a scenario where a clinician is waiting to view documents. In the Department of Veterans Affairs, providers can access multiple CCDs from non-VA providers through an initiative called the Virtual Lifetime Electronic Record (VLER) (Byrne et al., 2014). Clinicians initiate requests for external information by opening the VA's VistaWeb application, the component within its enterprise EHR that enables access to fragmented information captured by other VA facilities, as well as the Department of Defense. Currently the end user receives a potentially long list of CCDs for the patient from various VLER partners – usually private HIE organizations but could be another federal agency. Appling the CCD Consolidation Engine might be useful to reduce the number of documents to be rendered to the end user. However, if the process takes more than one second to complete, the end user may become frustrated. Past research at Regenstrief and others has demonstrated that clinicians expect sub-second performance from clinical information systems (Paterno et al., 2012). Furthermore, the results from this pilot using similarly structured CCDs may not hold up in a larger evaluation, therefore additional testing and evolution will be necessary to make the system robust under various conditions and contexts.

Although the CCDs generated by the Regenstrief Institute are conformant with HL7 standards, we have observed that many real-world CCDs being exchanged within HIOs as well as health systems are not always conformant. The Indiana HIE supported by Regenstrief utilizes a wide range of pre-processors to fix incoming messages from over 1000 source systems (Zafar & Dixon, 2007). These pre-processors help to add missing document structure components or move data from say a <text> block to a <observation> block if one does not exist. The CCD Consolidation Engine assumes that the received documents for consolidation are valid. Yet it is possible that non-conformant CCDs may get inserted into system at some point. So, instead of ignoring or deleting invalid documents, we designed our solution to be lenient in what it accepts with the intention of creating a robust solution for the heterogeneity that exists in real-world

clinical messages. For instance, when there is no valid template ID for the Medications section in a CCD and the values for "code" or "code system" are not valid, the application tries to parse the title of that section and if there is a word "Medications" in title, it assumes that section as Medications section. However, if there is no such useful information, the application skips that section.

2.4.1. Limitations and Future Development

Although the results of this evaluation are promising, the system has several limitations. First, because of the competition time constraints, we used synthetic data output from a single CCD generation engine. Second, while the system can detect conflicting information across documents there are few rules for how to adjudicate discrepancies. Third, the engine is not yet sophisticated enough to detect semantic overlap when varying code systems are present or free text is used.

This evaluation focused on synthetic data output from a single CCD generator, one developed by Regenstrief and used by IHIE. Although the 50 base CCDs were derived from real patients, the resulting sample was nonetheless small and derived from a single source. Future work will evaluate the approach and system using real-world CCDs exchanged as part of meaningful use from multiple sources. We will test the system using multiple CCDs exchanged among 60 facilities representing five health systems and more than 10 independent community hospitals.

The engine currently possesses several rules to detect conflicting information. For instance, one CCD may show that a patient is allergic to penicillin but the other CCD reports the patient has no known allergies. Another example is where one CCD documents a patient response to a question about Cigarette-Use as "Never Smoked Cigarette" but in another document the value is "Smoker." Currently the consolidation engine detects these conflicts, but the system is limited in what it does with such information. Presenting these conflicts after detection is

challenging because there does not exist a standard representation of this kind of information in CCD documents. In addition, while the current engine contains several such rules, there are many more varieties of conflicts. Future work will be necessary to expand the list of conflicts examined by the system. Moreover, once discovered there is no clear logic for adjudicating these kinds of conflicts. When to keep such conflicts or whether it is possible to make *a priori* determinations about which values might take precedence over others (e.g., always keep presumed allergies even if one document explicitly reports no known allergies) need to be addressed in future work and in a systematic way.

While the engine performed well detecting duplicates, the kind of duplicate entries it can detect is limited to cases where the same semantic encoding is used in multiple CCDs. The engine cannot currently detect semantic relationships between documents where different coding systems are used (e.g., SNOMED vs. ICD) or between encoded entries and free text entries (e.g., SNOMED vs. "Type 2 diabetes"). Future work will be necessary to enhance the engine to detect true semantic relationships we know exist in the real-world (Lin, Vreeman, McDonald, & Huff, 2011). Approaches that leverage the UMLS Metathesuarus may be explored to take advantage of already cross-linked ontologies. Natural language processing (NLP) approaches may also be necessary to link non-encoded observations in one CCD with encoded entries in other CCDs.

2.5. **Conclusion**

The use and exchange of CCDs are expected to increase during the next phase of meaningful use. It is helpful that providers can access patients' medical records from multiple locations through HIE, however, duplication and conflict of information is inevitable. The potential of having a document consolidation engine as a service, is a promising solution for HIOs to provide information that is easier to consume for clinical providers and more findable because of integration of homogenous information in one section. Our effort in this study is to present a novel prototype for de-duplication and consolidation of CCDs, which can be considered

a first step towards improving interoperability among information systems and information access among providers.

CHAPTER 3

STUDY 2 - RECONCILING DISPARATE INFORMATION IN CONTINUITY OF CARE DOCUMENTS: PILOTING A SYSTEM TO CONSOLIDATE STRUCTURED CLINICAL DOCUMENTS

3.1. **Background and Significance**

Many factors spread health information among multiple heterogeneous information systems, including variations in insurance coverage, reliance on multiple providers, and increase in specialty care systems (Vest & Gamm, 2010). Information fragmentation makes clinical workflow inefficient, creates barriers to care coordination, and leads to over-utilization of healthcare through duplicative testing or imaging (Frisse et al., 2012; Schoen et al., 2012). Health Information Exchange (HIE) is a promising solution to connect these fragmented and disparate health information systems and provide timely and relevant information to providers (Acker et al., 2007; Brailer, 2005; Hripcsak et al., 2007). Meaningful Use regulations from the U.S. Centers for Medicare and Medicaid Services (CMS) underscore the importance of HIE to support better care across the nation. Exchanging key clinical data, transmitting a summary care record, e-prescribing, integrating laboratory results, and immunization reporting are all different forms of HIE required in Meaningful Use Stage 2 and 3 (Kuperman, 2011). Using clinical data exchange standards is the key to enabling HIE among disparate information systems

Health Level Seven (HL7) is an international standards development organization that creates standards for exchanging clinical and administrative data among healthcare information systems (Hosseini et al., 2014; Office of the National Coordinator (ONC), 2014).The Meaningful Use rules have adopted one of HL7's standards, the Consolidated CDA (C-CDA) Implementation Guide, as “the sole standard for exchanging summary care records” ("Medicare and Medicaid programs; electronic health record incentive program--stage 2. Final rule," 2012). The C-CDA standard (HL7 Health Level Seven, 2012) actually contains templates for nine different clinical

documents, including Continuity of Care Document (CCD), Consultation Notes, and Discharge Summary. Of these, the CCD is one of the popular documents that is been generated and exchanged in healthcare facilities. CCD's popularity stems in part from its inclusion of the most critical information for making clinical decision such as summary of administrative, demographic, and clinical information facts.

Exchanging CCDs among healthcare organizations is an effective method for providing a clinical summary of patients to providers. Although increased information sharing via CCDs is an improvement over data silos, the technical limits of most current HIE infrastructure has created another problem. HIOs receive multiple CCDs for the same patient from different sources, but typically do not decompose them into a unified clinical summary.

Providers who receive these CCDs must sift through multiple documents that include potentially duplicative or conflicting information. Without an efficient and effective method for de-duplicating the redundant patient data, providers are forced to review the lengthy and redundant information spread across multiple documents.

At the national level, the certification criteria of Meaningful Use stages 2 and 3 recognize the necessity of clinical information reconciliation. One specific certification criterion requires that information systems incorporate problems, medications, and allergies sections from multiple documents to improve the overall clinical effectiveness ("Medicare and Medicaid programs; electronic health record incentive program--stage 2. Final rule," 2012; Office of the National Coordinator (ONC), 2012).

We developed a prototype system to consolidate and de-duplicate multiple CCDs for a single patient. The system was designed for use by an HIE, and we evaluated its performance on simulated data (Hosseini, Meade, Schnitzius, & Dixon, 2015). The result shows that all of the input CCDs were successfully consolidated, and no data were lost. We were encouraged by the preliminary results of the system, but recognized the limitations in that evaluation due to the time constraints and use of synthetic data. In the present study, we have further refined the

consolidation system and evaluated its performance on real-world CCDs sampled from the Indiana Network for Patient Care (INPC) database. The INPC created by investigators at the Regenstrief Institute and operated by the Indiana Health Information Exchange (IHIE) is the nation's largest and longest tenured HIE. IHIE now serves over 25,000 physicians and over 10 million patients (Biondich et al., 2014; Biondich & Grannis, 2004).

Specifically, the purpose of this study in the first place is to develop or change rules in the system's rules engine component for better identification of duplications in real-world CCDs, while also improving the whole system performance. Our second purpose is to evaluate the system based on de-duplication of real-world data instead of synthetic CCDs. And finally, our last aim is to identify the main causes of inaccurate consolidation in our system and investigate the differences in data representation in the test data set.

3.2. **Materials and Methods**

In this section, the improvement of prototype system to address limitations of our previous study is explained (Hosseini et al., 2015). Later, we will describe the study design for consolidation of real world CCDs both manually and automatically. We will also describe the methods used for the evaluation of system accuracy, along with other findings through this process.

3.2.1. System Description and Improvements

The CCD consolidation system (which is now called CDA consolidation system) was originally designed and developed to align with three paradigms in computer science: scalability, extensibility, and open source. Also, the system architecture consists of three primary components:

1. A CCD Consolidation Engine that receives multiple CCDs through an Application Programming Interface (API) and executes a series of rules to reconcile and de-duplicate data.
2. An Audit component that identifies and logs system events.
3. A Configuration component that enables implementers to configure the rules executed by the system.

Our previous system designed in a way to only accept CCD documents. One of the objectives of the system improvement in this study was redesigning the system to accept every document that is derived from CDA. In other words, in the current system all of the CDA-based documents are accepted as an input. Hence, our focus in this study is to consolidate three important sections of medical record (problems, medications, and allergies). Since, these sections are covered in CCD template of C-CDA, we designed the system to generate the final consolidated document in CCD format. Although we didn't restrict the format of input data (i.e., all CDA-based documents), the test data set from the INPC for system evaluation was in CCD format.

Based on our findings, we improved the rules engine component of the system by adding new rules and modifying the previous rules. Also, some new techniques in software development were applied to reduce de-duplication time and improve overall performance. We randomly selected and reviewed several CCDs (around 20) from INPC as our development data set in order to secure a more accurate understanding of the CCDs and in general CDA-based documents that are generated in real world. We also participated in HL7 working group meetings and discussed about the important consideration for CDA reconciliation with experts in this domain.

For instance, the attributes of all Extensible Markup Language (XML) elements on CDA (i.e., nullflavor or negationInd) play an important role in presenting information. We gave close attention to these data elements and redesigned several rules that effect the attributes during de-duplication. The translation code of the CDA entries is another element that we addressed in the

current version of CDA consolidation. During the process of reviewing development set of CCDs we realized that they might include multiple codes for a single entry that can be used in identifying duplications. This, however, was contingent upon whether different codes between CCDs were present for one concept. Another improvement to the system was the generating of a narrative block (human readable or free- text section) of consolidated CCD. In the current version, the system identifies free-text information related to every entry level data in the CCDs and temporarily keeps them in the memory as entry-narrative pairs. After reconciliation of the entry data, the related narrative blocks are merged in the narrative block section. Further free-text processing is applied to ensure the narrative block is a valid HTML. This feature in the previous version of the system was not accurate enough to create free-text data for real world documents.

3.2.2. Study Design

The study was designed to address two objectives. First, we wanted to know how accurately a consolidation system could de-duplicate input data set, and consolidate them into one document. In order to evaluate the accuracy of automatic consolidation, we reviewed the CCDs manually and compared the result of automatic consolidation with manual consolidation. We consider the results of the manual consolidation as a plausible means to identify duplications. The differences between the number of deduplications in manual and automatic methods were identified.

Our second objective was to identify the cause of inaccuracy in the consolidation and investigate the differences in data representation among test set of CCDs. This process is done primarily during the manual consolidation process.

3.2.2.1. *Automatic Deduplication and Consolidation*

We evaluated the system using a corpus of 522 CCDs representing 50 unique patients randomly sampled from INPC database. INPC contains CCDs exchanged among 60 facilities

representing five health systems and more than 10 independent community hospitals (Biondich & Grannis, 2004; McDonald et al., 2005). The methods for data sampling is explained later in this article. As noted, we designed the system to generate the final consolidated document in CCD format.

In our sample data for 50 patients, the number of CCDs for each patient varied from 2 to 36. Subsequently, each batch input to the system and consolidated CCDs are generated and saved in a specified folder for each patient. Along with the consolidated CCDs, other metadata about the CCDs was generated and saved within the same folder. Metadata types included: CCD receiving date, number of entries, custodians, and the elapsed time for deduplication.

3.2.2.2. *Manual Deduplication and Consolidation*

The 522 CCDs for each patient was manually reviewed entry by entry for each section (problems, allergies, and medication) and duplications were identified. For patients with a higher number of CCDs (>5 or 6) we used Microsoft Excel and VBA Scripts to facilitate the manual review.

3.2.3. Methods for Data Acquisition

We sampled our data from one of INPC databases in a PostgreSQL format which includes 176,169 CCDs received from the healthcare facilities, from 07/12/2014 to 11/06/2015 (484 days). The CCDs received from 24 facilities in the State of Indiana during this period.

The process of patient matching is outside of the scope of this study. In designing the system, we assumed that the CDA-based documents belonging to a patient would be identified by HIE in advance to input to the CDA consolidation system. In this study, however, CCDs in the database are not grouped based on each patient's identifier. As such, we had to extract CCDs for each patient and then group them within their specified folder. Patient identification and matching were conducted before inputting the CCDs into the system, i.e., based on the family and birth date

fields of CCDs. Since there was a chance of incorrect patient matching, the CCD identification of each patient was double checked during the manual review.

After all patients were identified, patients with two or more CCDs were selected and grouped in each of their respective folder. A total of 30,370 patient folders including 2 to 36 CCDs (overall 86,184 CCDs) were created. In order to sample data among 30,370 patients during the evaluation phase, we first categorized the patients based on their number of CCDs (2 to 36) and then randomly selected patients from each category (stratified sampling). Stratified sampling was applied to make sure there is at least one CCD from each category that might be lost in a more generalized random sampling. To balance the patient selection process, more patients were selected from categories that had a higher number of patients. In sum, 50 patients (including 522 CCDs) were randomly selected for the purpose of CCD consolidation.

3.2.4. Methods for Data Analysis

A comparison between manual and automatic de-duplication suggests how much automation is able to correctly identify duplications in three area: problem, allergy, and medication. For instance, if there were seven duplications from the manual review, and 5 in automatic method, we would conclude that the accuracy of the system is 71%. The same calculation was performed for all of the entries in each section after consolidation.

System performance was evaluated based on the total time necessary to process the input CCDs. We examined the average time to process a CCD as well as the minimum, maximum, and average file sizes. Further investigation was conducted to identify the gaps in system de-duplication process and different representations of data in CCDs.

3.3. **Results**

Table 2 shows the total number of entries along with the decrease in number of entries in 522 CCDs that were consolidated manually and automatically.

	Problems (n=6,531)	*Allergies (n=1,845)*	*Medications (n=3,255)*	*Total (n=11,631)*
Decrease in number of entries in manual deduplication (ideal)	89.9%	91.9%	57.3%	81.1%
Decrease in number of entries through system deduplication	89.1%	79.9%	52.5%	77.4%

Table 2. Total number of entries and decrease in number of entries through manual and system deduplication

Manual consolidation of 11,631 entries was done in approximately 150 hours, however, the same data set was automatically consolidated in 3.3 minutes (average 0.38 seconds for each CCD) using the CDA consolidation system. We considered the outcome of manual consolidation as an ideal de-duplication, which suggestions that 100% of the duplications were correctly detected in three CCD sections.

System consolidation had lower accuracy in the duplication detection compared to ideal consolidation. The system was able to de-duplicate 99.1% of problems, 87.0% of allergies, and 91.7% of medications. The decrease in number of entries after system deduplication was lower than ideal consolidation because of the lower deduplication percentage.

The size on disc for the input of 522 CCDs was 265 MB. Also, 522 CCDs were consolidated into 50 CCDs belonging to 50 patients with an overall size disc of 26.2 MB (90.1% decrease). Overall, there was an average of 41.5 days difference between all the 522 CCDs receiving dates for each patient. Figure 5 shows a sample of the receiving time differences for a patient with 15 CCDs. The CCDs for a patient (15 CCDs) are received in INPC with different time difference (days) from the first CCD receive day (day 0).

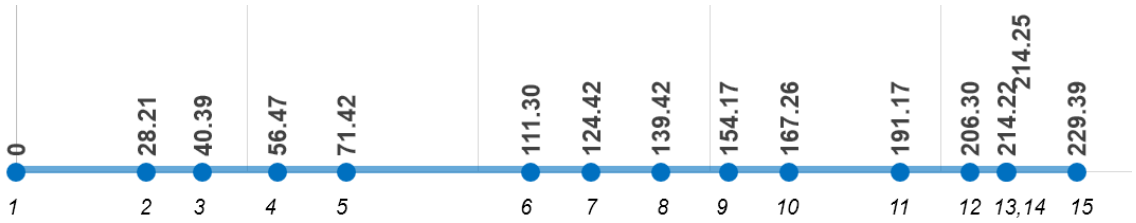

Figure 5. The CCDs for a patient (15 CCDs) are received in INPC with time differences (days) from the first CCD receive day (day 0).

3.3.1. Inaccuracy in system deduplication

As explained previously, our system de-duplicated 99.1% of the problems, 87.0% of allergies, and 91.7% of medications. This means there was some degree of inaccuracy in deduplication (problem 0.9%, allergy 13%, and medication 8.3%).

All of the inaccuracies in the problems and allergies were due to the reference to the free-text section of the CCD, rather than terminology code that identifies a specific problem or allergy. Our system identified duplications only based on the coding value—such as the SNOMED, RxNorm, or ICD9 codes. As shown in Figure 6, an allergy entry in a patient's CCD does not include any code for the particular allergenic substance. That is, a reference is only made to a free-text section of CCD. Figure 7 on the other hand, presents another allergy entry that is accurately processed by our system, because it includes two different codes for "Codeine," to which a patient is allergic.

```
<observation classCode="OBS" moodCode="EVN">
  <templateId root="2.16.840.1.113883.10.20.22.4.7" />
  <id root="                                    " />
  <code code="ASSERTION" codeSystem="2.16.840.1.113883.5.4" codeSystemName="HL7 ActCode" displayName="Assertion" />
  <statusCode code="completed" />
  <effectiveTime>...</effectiveTime>
  <value xsi:type="CD" codeSystem="2.16.840.1.113883.6.96" codeSystemName="SNOMED CT" code="419199007" displayName="Allergy to substance" />
  <author>...</author>
  <participant typeCode="CSM" contextControlCode="OP">
    <participantRole classCode="MANU">
      <playingEntity classCode="MMAT">
        <code nullFlavor="UNK">
          <originalText>
            <reference value="#ALLERGEN       " />
          </originalText>
        </code>
      </playingEntity>
    </participantRole>
  </participant>
  <entryRelationship typeCode="SUBJ" inversionInd="true">...</entryRelationship>
  <entryRelationship typeCode="MFST" inversionInd="true">...</entryRelationship>
</observation>
```

Figure 6. An allergy entry without code for substance to which patient is allergic to.

```
<observation classCode="OBS" moodCode="EVN">
  <templateId root="                                    " />
  <id root="                                   ' />
  <code code="ASSERTION" codeSystem="2.16.840.1.113883.5.4" codeSystemName="HL7 ActCode" displayName="Assertion" />
  <statusCode code="completed" />
  <effectiveTime>...</effectiveTime>
  <value xsi:type="CD" codeSystem="2.16.840.1.113883.6.96" codeSystemName="SNOMED CT" code="416098002" displayName="Drug allergy" />
  <author>...</author>
  <participant typeCode="CSM" contextControlCode="OP">
    <participantRole classCode="MANU">
      <playingEntity classCode="MMAT">
        <code code="2670" codeSystem="2.16.840.1.113883.6.88" codeSystemName="RxNorm" displayName="Codeine">
          <originalText>
            <reference value="#ALLERGEN       " />
          </originalText>
          <translation code="d00012" codeSystem="2.16.840.1.113883.6.314" codeSystemName="multum-drug-id" displayName="codeine" />
        </code>
      </playingEntity>
    </participantRole>
  </participant>
  <entryRelationship typeCode="SUBJ" inversionInd="true">...</entryRelationship>
  <entryRelationship typeCode="MFST" inversionInd="true">...</entryRelationship>
</observation>
```

Figure 7. An allergy entry with two codes for Codeine to which patient is allergic to.

As outlined in the medication section, reference to free-text rarely happened (with inaccuracy at .6%). At the same time, the majority of issues in identifying the duplications in medication was directly related to not using the RxNorm code. RxNorm codes are the permitted data values in the constrained value domain for the medications section. However, alternatives can be used by changing the value of the nullflavor attribute in the CCD to OTH (other) and using non-RxNorm code in the translation element of the entry. For instance, in our sample data, Multum Drug IDs are sometimes used in the translation elements of CCDs. However, it is also important to note that Multum Drug IDs are a much broader and less specific category that corresponds to the generic names of drugs. Therefore, there is a chance of having the same code for two different drugs in Multum Drug IDs.

For instance, the mediation “24 HR diltiazem Hydrochloride Extended Release Capsule” has two dosage of 240 MG and 180 MG. the Multum Drug ID for both dosages of this medication is the same (D00045), while, the RxNorm codes are different (830837 and 830845). Although these medications can be correctly de-duplicated based on their RxNorm code, however, RxNorms codes are not always provided and if these medications are received with only Multum Drug IDs the de-duplication might be difficult.

This potential point of confusion, reduced the accuracy of our system by 8.3%. Figure 8 is a real sample code section for a medication that has no RxNorm code and the translation section only carries Multum Drug ID.

```
<code nullFlavor="OTH">
  <originalText>
    <reference value="#MEDPROD6400065107" />
  </originalText>
  <translation code="d03768" codeSystem="2.16.840.1.113883.6.314" codeSystemName="multum-drug-id" />
</code>
```

Figure 8. A medication without RxNorm code.

The free-text part of CCD provides the information that is human readable and healthcare providers review this information, while, the structured data is consumed by the computer systems. Sometimes the free-text section of CCD was the same for two codes in the problems section of the test dataset. For instance, “Chronic diastolic heart failure (disorder)” has the SNOMED Code of 441530006 in the entry level of CCD and the SNOMED Code for “Diastolic heart failure (disorder)” is 418304008. However, these problems are presented as “Diastolic CHF” in the free-text section of CCD.

3.4. **Discussion**

Reconciliation of clinical information particularly among CDA-based documents is one of the challenges that face HIOs as they seek an efficient method for consolidating multiple medical documents for a given patient. In this study, we improved the previously developed CCD consolidation system, followed by an evaluation of system accuracy. This was done by means of

comparing automated system results to manual consolidation. Our primary findings are promising, as evidenced by the system's capability to de-duplicate CCDs with a significant degree of accuracy.

In sum, our findings suggest that if input CDA-based documents are compliant with the HL7 standard, the system is able to identify the majority of the duplications. Conversely, any reference to the free-text section of the CCD without providing coded values, makes the system incapable of identifying any duplications. This is because the system can only de-duplicate based on existing codes. Other key factors that have been problematic to identifying duplications is providing alternative information instead of permitted data values in CCDs. For instance, RxNorm codes are permitted for medications in CCD according to the HL7 standard, the lack of this code (in the medication section) will cause an inaccuracy in de-duplication.

Another issue regarding the quality of CCDs is presenting the same free-text section for two different entries. This problem is primarily shown when the entries have a similar name, while also having different codes for existing entries. Since our system de-duplicates the entries based on codes, both entries remain in the consolidated CCD. We consider this de-duplication as a valid and necessary consolidation. However, in the "human readable" section of the CCD, it is not possible for providers to discriminate between two entries. This is because the free-text is the same as in the original document.

Overall, if the quality of the dataset is acceptable and the CCDs are compliant to the standard, we can infer that the system consolidation is significantly accurate in identifying duplications. Moreover, this suggests that our high performance system demonstrates that the automatic consolidation of medical records is feasible. At the same time, however, our work also generated several concerns. For example, one of our greatest concerns during system development was its incapability to de-duplicate the entries if they are coded based on different terminologies. In other words, the concept mapping between different terminologies is not considered in the current version of the system design. These concerns, however, diminished over

time. In the sample data from real world CCDs, translations of the main codes are provided based on different terminologies. Since, we improved our system to use translation codes in de-duplication process the code mapping limitation was not considered as a big challenge.

We assumed that if CCDs are generated according to HL7 standard implementation guidelines and translation codes are provided (with the main codes), then a huge number of de-duplications can be done without the need for concept mapping among terminologies. Our hope, however, is to focus on this limitation in our forthcoming iteration of this research. This is because some facilities generate CCDs with local codes without translation. Nevertheless, an ideal system should be able to map these code to standard terminologies and identify duplications.

According to our findings, duplications in allergy (91.9%) and problems (89.9%) were higher than medication (57.3%). The reason for this difference is that usually medications are added or modified during treatment, which includes an updating of the record in every new CCD. Therefore, the number of duplications in medications are often less than problems and allergies.

CCDs are usually generated based on real-world triggers in the healthcare domain, which often includes a meaningful time difference between CCDs' generation dates. Although there were very few duplicate CCDs (received) in INPC because of system or user error, the majority of CCDs were unique and represented different aspects of a patients' medical history. This suggests that the value of consolidating multiple CCDs be measure relative to the increasing burden placed on clinicians during their review of multiple patient document.

The total time to process and de-duplicate each CCD was 1.13 seconds in previous version of our system. We reduced this number to 0.38 seconds in the current version because of system enhancements. However, we assume adding new rules for better CCD de-duplication will increase CCD de-duplication performance in the future. We also believe, however, that implementing good software developing techniques will also give rise to increased system performance.

3.4.1. Limitations and Future Development

Although the results of this study are promising, the system has several limitations. First, while the system can detect conflicting information across documents, there are few rules for how to adjudicate these discrepancies. Second, although we used CCD translation codes (along with the main codes to identify duplications), the de-duplication engine is not yet sophisticated enough to detect semantic overlaps when varying code systems are present or free text is used. For example, approaches that leverage the UMLS Metathesuarus may be explored to take advantage of already cross-linked ontologies. Natural Language Processing (NLP) approaches may also be necessary to link non-encoded observations in one CCD with encoded entries in other CCDs.

3.5. **Conclusion**

The use and exchange of CCDs are expected to increase over the next decade. As such, it is helpful that providers have access to patient medical records from multiple locations through their HIE. As the same time, however, duplication and conflict of patient information is inevitable. The potential of having a document consolidation engine as a HIE service will provide clinicians information that is (1) easier to use and understand and (2) more searchable/findable due to its integration of homogenous information into one section. This study represents a novel prototype for de-duplication and consolidation of CCDs, which we believe is a major step toward improving information access and the interoperability among information systems.

CHAPTER 4

STUDY 3 - IMPACT OF CDA CONSOLIDATION SYSTEM ON HEALTHCARE PROVIDERS' WORKLOAD

4.1. **Background**

The current state of healthcare dictates that providers' responsibilities are woven seamlessly with information technology, particularly, clinical information systems that they use to perform tasks. Fatigue and mental workload are associated with complexity and difficulty of task. In general, complex tasks lead to a lack of focus and less control on new and unexpected events—both of which increase the likelihood of error. In sum, higher cognitive and physical workload result in lower accuracy and performance and reduced efficacy in the accomplishment of tasks (DiDomenico & Nussbaum, 2008).

Information reconciliation is one of the most common tasks healthcare providers perform on a daily basis. Each time a patient transitions from one clinic or setting to another, clinicians need to review previous medications and medical history, followed by a reconciliation of any information differences across disparate records of the prior clinic or facility. This process takes considerable time and effort and providing multiple medical documents for a patient might be a potential for a considerable number of inaccuracies and oversights during the information reconciliation process. Studies have demonstrated (Institute for Healthcare Improvement, 2011; Redmond et al., 2016) that such problems especially in medication duplication lead to medical error, with a direct impact on patient care.

Automating information reconciliation has the potential to reduce cognitive and physical workload, key barriers to medication reconciliation (Boockvar, Santos, Kushniruk, Johnson, & Nebeker, 2011). Some health care systems and facilities participate in health information exchange (HIE), in which data and information are shared across organizational boundaries. Increasingly HIE involves the exchange of a documents encoded using the Health Level Seven

(HL7) Clinical Document Architecture (CDA) standard, including Continuity of Care Documents or CCDs. Exchanging CCDs and incorporating the CDA-based problems, medications, and allergies sections from multiple documents into electronic health records is required by the "meaningful use" incentive program to improve overall clinical effectiveness ("Medicare and Medicaid programs; electronic health record incentive program--stage 2. Final rule," 2012; Office of the National Coordinator (ONC), 2012).

Although increased information sharing via CCDs is an improvement over data silos, the technical limits of many HIE infrastructures has created another problem. HIOs receive multiple CCDs for the same patient from different sources, but typically do not decompose them into a unified clinical summary. Instead, providers who access an HIE interface must sift through multiple documents that include potentially duplicative or conflicting information. Without efficient and effective methods for de-duplicating data across documents, providers are forced to review lengthy and redundant information. Cumbersome review and reconciliation of information have been observed in prior studies on HIE (Byrne et al., 2014; Strauss et al., 2015), establishing the need for better tools.

The challenge of reconciling clinical content in CCDs has attracted both local and national attention. The Indiana Network for Patient Care (INPC), created by investigators at the Regenstrief Institute and operated by the Indiana Health Information Exchange (IHIE), is the nation's largest and longest tenured HIE. IHIE now serves over 25,000 physicians and over 10 million patients (Biondich et al., 2014; Biondich & Grannis, 2004). De-duplication of CCDs was named by IHIE as a "grand challenge" in the 2013 Hoosier Healthcare Innovation Challenge (HHIC). HHIC is a state-based software development competition that brings together healthcare and technology professionals to provide creative solutions for some of the most challenging problems in healthcare ("Hoosier Healthcare Innovation Challenge ", 2013).

As a part of the HHIC competition, we developed a prototype system to consolidate and de-duplicate multiple CCDs for a single patient (Hosseini et al., 2015). We further refined this

system and tested its real-world performance in CCD consolidation compared to manual review (Hosseini et al., 2016). We are encouraged by the preliminary results of the system's performance. Yet, we also wanted to characterize how the system's performance influenced the perceived workload of providers. Our novel system is able to consolidate multiple CDA-based documents into a single view. As such, our hypothesis is that this system has potential to significantly reduce providers' workload.

The purpose of this study is to evaluate the impact of the proposed CDA consolidation system on healthcare providers' perceived workload while reviewing patient medical documents. We also sought to determine the system's effect on provider time to review and reconcile medical information review.

4.2. **Methods**

In this study we recruited primary care providers and conducted three scenario-based tasks designed for reviewing CCDs. To evaluate the impact of the system on participants' perceived workload each scenario was conducted two times, one time for multiple CCDs and one time for a single consolidated CCD generated by the system. NASA TLX was used to evaluate providers' perceived workload. At the end of the study, three open-ended questions were asked in a face-to-face interview. Figure 9 presents the overall structure of the study.

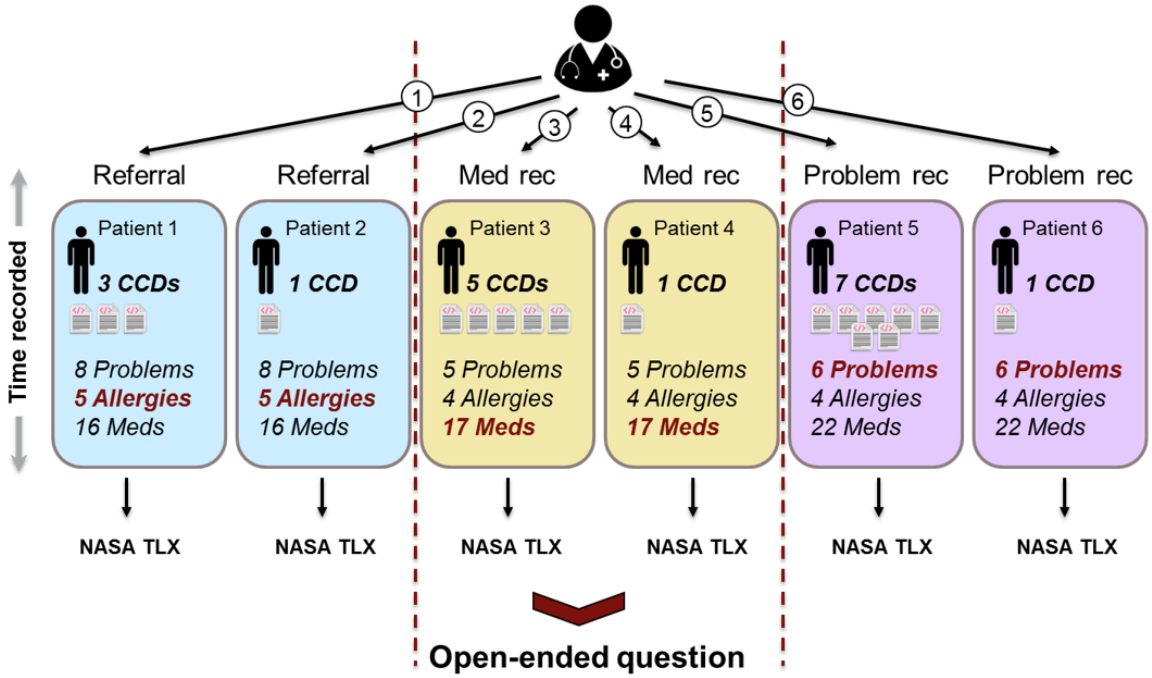

Figure 9. Overall structure of the study.

4.2.1. Study Design and Scenarios

This study evaluated how providers reviewed three sections in medical documents that are important for making clinical decisions: problems, allergies, and medications. We designed three transitions of care scenraios in which we could examine cognitive tasks associated with the review of these document sections: 1) referral, 2) medication reconciliation, and 3) problems reconciliation. Five primary care physicians and four nurse practitioners were recruited (9 total participants) to conduct the scenario-based tasks. Based on prior research (Gleason et al., 2004; Vira, Colquhoun, & Etchells, 2006) and our team's clinical experience, we obtained 8 as the minimum sample size required for this study to identify at least 22% reduction of time (mean value) needed for medical information reconciliation. We set alpha at 0.05 and power at 0.80. The participants were asked to read the scenarios and conduct specific tasks. Table 3 contains the scenarios were provided to the participants.

Scenario	*Description*
Referral	As a health provider you are reviewing electronic medical records (CCDs) of a patient that you want to refer to a specialist. To refer the patient to a specialist, you need to identify all of important allergies in the patient's medical record that are necessary for the specialist to know. You may identify these allergies based on reviewing the patient's medications and problems list. Your task is to reconcile all of the patient's allergies and then write a note to the specialist identifying those identified allergies.
Medication Reconciliation	Patient-X is a new patient who arrived at your clinic today. Your office has requested and received clinical records from Patient-X's previous provider. These records are ready for review prior to the visit. Your task is to review the CCDs in the documents and reconcile the patient's medications prior to the clinical encounter.
Problems Reconciliation	A new patient (Patient 5 or 6) is coming to your clinic for an initial visit. He/she's medical record is complex with multiple conditions, which your office received from the prior physician(s). Your task is to reconcile the complex problem list as part of preparation for seeing the new patient.

Table 3. Three scenarios for reviewing CCDs

The purpose of this study was to evaluate participants' perceived workload while reviewing two types of medical records: (1) Multiple CCDs for a patient versus (2) One consolidated CCD generated by CDA-consolidation system. Therefore, each scenario included two tasks: 1) One before consolidation and 2) One after consolidation task. No training information was provided to the participants. Also, participants were unaware that the CCD in the second task of each scenario was already consolidated and reconciled. Therefore, they reviewed the reconciled CCDs with the intention of finding duplications. Figure 10 is a screenshot of a sample CCD that is presented to the participants in a web browser.

Overall, each participant conducted six tasks within three scenarios. In each scenario the first task was to review multiple CCDs for the patient and the second task had only one consolidated CCD. All of the participants conducted a referral scenario first, then medication reconciliation, and finally the problems reconciliation scenario. The time to complete each task was recorded. Also, each participant's accuracy on performing the tasks were evaluated after the interview session.

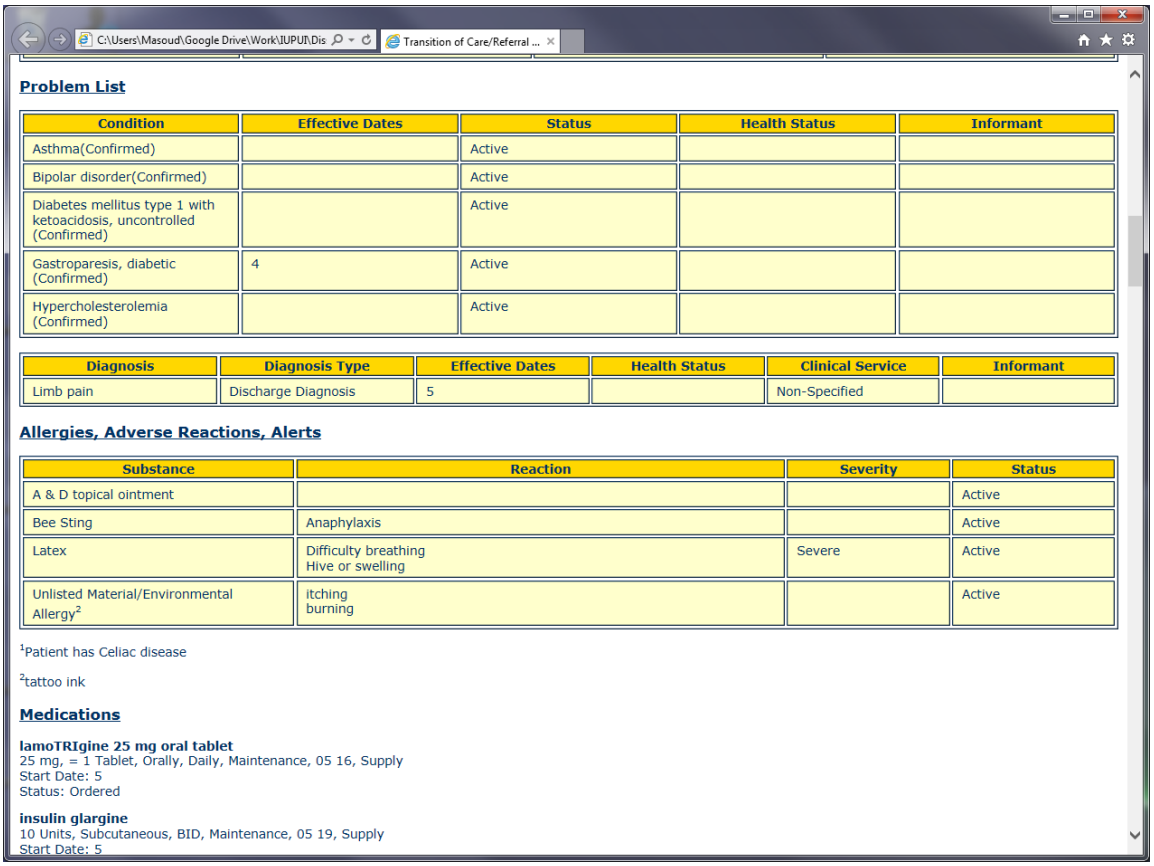

Figure 10. Screenshot of a sample CCD that is presented to the participants in a web browser

4.2.1.1. *NASA-TLX*

The National Aeronautics and Space Administration Task Load Index (NASA-TLX) was used for evaluating workload. NASA TLX is the most widely used research measure to assess task workload and effectiveness in humans (Hart, 2006). The NASATLX includes overall index

of mental workload as well as the relative contributions of six subscales: mental, physical and temporal task demands, and effort, frustration, and perceived performance (Hart & Staveland, 1988).

At the end of each task, participants completed the NASA-TLX. The NASA-TLX captures self-rating on a scale of 100 points for mental, physical and temporal demands as well as effort, performance, and frustration associated with each task. In total, each participant completed the NASA TLX six times.

An overall NASA-TLX score was defined based on the formula $\sum_{i=1}^{6}(r_i \times w_i)/\binom{6}{2}$. In this formula, 'r' represents the self-reported rate for each of the six measures in the NASA TLX and 'w' is the weight of the measure defined based on the fifteen paired-wise selection of measures. An open source HTML and JavaScript version of the NASA TLX application was used in this study (Vertanen, 2012).

4.2.1.2. *CCD Sampling*

The CCDs used in this study were sampled from the INPC database. To collect CCDs for each scenario, we queried the database for patients with a minimum of five problems, fifteen medications, and four documented allergies.

The CCDs in the database did not contain a global patient identifier. In order to select the relevant CCDs for a particular patient, we identified and matched patients based on the first name, last name and birth date fields of CCDs. Since there was a chance of incorrect patient matching, the CCD identification of each patient was double checked manually. To assess the effect of our consolidation system, we then selected patients with three or more CCDs.

For each scenario, two patients were randomly selected from our sample after applying the following filter. We ran the sample of CCDs through the consolidation system and then after reconciliation, we chose two patients for each scenario who had the same number of problems,

medications, and allergies. For instance, for the referral scenario two patients were randomly selected with the total of eight reconciled problems, five reconciled allergies, and sixteen reconciled medications. These two patients had a different number of entries in each of the three sections before reconciliation, however, after reconciliation, the number of entries were the same.

Because each scenario included two tasks, we randomly selected six patients with the aforementioned criteria. In each scenario, the second task was always to review the consolidated CCD.

4.2.2. Open-Ended Questions

After the six tasks were conducted and NASA TLX results were collected, the open-ended questions presented in table 4 were asked in a face-to-face interviews.

question 1	What were the challenges you faced in finding information from the patient's medical records in general? Inconsistencies (conflicts in different documents)
question 2	If you had a tool that could bring together patient information from medical documents into a single view, would this be helpful to you? If yes or no, why?
question 3	How would you use such a tool in your practice? Would you use it in advance of the encounter or during the visit?

Table 4. Open-ended questions

The interview was recorded in each session, followed by selected transcription and content analysis.

4.2.3. Statistical Analysis

Results from our scenarios were tabulated into a spreadsheet and we compared variables (time and NASA TLX subscales) using a paired t-test analysis. A p-value of less than 0.05 was considered statistically significant.

4.3. **Results**

The data is compared based on reconciling information in multiple versus consolidated CCDs and the results of the study are categorized in three main domains: 1) accuracy, 2) healthcare providers' perceived workload, and 3) efficiency.

4.3.1. Accuracy

In the referrals and problems scenarios all of the participants correctly reconciled the allergies and problems in multiple CCDs. The consolidated CCD were already reconciled and there was no need for de-duplication of information. However, since the participants were unaware this factor, they still attempted to locate duplications in consolidated CCDs. It is also important to note that in the medication reconciliation scenario, participants made a few mistakes while attempting to reconcile multiple CCDs. Two of the participants missed one medication in to the list of reconciled medications at the end of the task. Also, three of the participants were not able to reconcile the medication "Lorazepam" with its common brand name "Ativan," but rather included both names in the final medication list. The medication reconciliation scenario was conducted correctly for the consolidated version of CCDs without any mistake.

4.3.2. Perceived Workload

Based on the results of the paired t-test analysis, the overall perceived workload of reviewing consolidated CCDs are significantly lower than multiple CCDs. However, some of the NASA TLX measures didn't change significantly in each scenario. In Table 5, the underlined p-values that are greater than .05 indicate the non-significant measures in each scenario.

4.3.3. Efficiency

The results show that reviewing consolidated CCDs is faster than reviewing multiple CCDs. Table 5 and 6 demonstrate the result of statistical analysis and time reduction percentage for each scenario.

	Referral Scenario			*Med Rec Scenario*			*Problems Rec Scenario*		
	R1 Mean (SD)	*R2 Mean (SD)*	*p*	*M1 Mean (SD)*	*M2 Mean (SD)*	*p*	*P1 Mean (SD)*	*P2 Mean (SD)*	*P*
Time	131.2 (89.6)	54.1 (36.4)	0.008	297.2 (132)	183.9 (107)	0.000	145.2 (68.2)	50.7 (24.0)	0.000
MD	39.4 (25.1)	24.4 (13.6)	0.041	47.8 (26.4)	38.3 (20.3)	0.196	38.3 (26.2)	21.1 (17.6)	0.005
PD	16.7 (10.6)	12.2 (7.5)	0.026	31.1 (25.7)	22.8 (24.4)	0.033	18.9 (15.8)	13.3 (7.9)	0.048
TD	31.1 (27.0)	16.1 (10.8)	0.020	53.9 (23.0)	34.4 (19.8)	0.017	35.6 (17.4)	25.0 (15.8)	0.005
PE	27.2 (16.8)	20.6 (15.5)	0.052	26.1 (14.1)	23.3 (15.0)	0.048	21.1 (11.9)	15.6 (7.3)	0.037
EF	28.9 (22.7)	22.8 (15.2)	0.128	52.8 (26.1)	43.9 (24.1)	0.032	31.7 (20.6)	23.9 (16.9)	0.038
FR	22.2 (26.5)	12.8 (12.3)	0.179	33.9 (22.0)	27.2 (20.2)	0.044	22.8 (21.7)	17.8 (14.8)	0.156
OW	32.4 (21.7)	20.3 (10.2)	0.032	46.6 (20.4)	38.0 (17.1)	0.021	31.0 (14.1)	21.2 (11.4)	0.004

Table 5. Paired t-test analysis results for a multiple versus consolidated CCD review. Underlined numbers are p-values that are not considered as significant. The lower value for performance indicates the higher performance reported by the participant. (MD= Mental Demands, PD=Physical demands, TD=Temporal Demands, EF=Effort, PE=Performance, FR=Frustration, OW= Overall Workload).

Measure	*Referral Scenario*	*Med Rec Scenario*	*Problem Rec Scenario*
Time	58.8%	38.1%	65.1%
MD	38.0%	19.8%	44.9%
PD	26.7%	26.8%	29.4%
TD	48.2%	36.1%	29.7%
PE	24.5%	10.6%	26.3%
EF	21.2%	16.8%	24.6%
FR	42.5%	19.7%	22.0%
OW	37.2%	18.4%	31.5%

Table 6. Reduction percentage for each measure.

4.3.4. Provider Interviews

The open-ended interviews at the conclusion of study took approximately 20 minutes for each participant. Table 7 summarizes the responses from all participants into main topics.

	Topics
Topics in question 1	• Reconciliation is a huge time commitment. • Too much information (I need relevant information) • Uncertainty (missing, inaccurate or not complete data) • Inconsistencies (conflicts in different documents) • Finding information (Where to look) • Complex UI (I am not sure where the information is)
Topics in question 2	• This system helps finding different types of information easily. • Less time • Less searching • Less clicking
Topics in question 3	• Before visit (5 people) • During visit (2 people) - Data already consolidated and can be used during visit • Before and during visit (2 people)
Final discussion	• All of the participants agree that consolidation system is very useful • All of the participants agree that hyper-link is helpful • This is not an ultimate solution - good start point • I need the data in my EMR. • Categorize into type of visit note(admission, discharge, office visit), active medications, categorization of problems (heart problems) • Time stamped information (Active medication).

Table 7. Main topics that are discussed during open-ended discussion.

4.4. **Discussion**

Many previous studies addressed reconciliation of medication (Mueller, Sponsler, Kripalani, & Schnipper, 2012), however, reconciling other medical information such as allergies or problems is very rare or not exist. Also, there is no published study that addresses the duplicative information in CDA-based documents. In addition, developing electronic systems for medical information reconciliation is mostly limited to medication reconciliation which didn't evaluate the impact of the system on healthcare provider's workload or reduction of required time

for reviewing reconciled medication list (Agrawal & Wu, 2009; Boockvar, Blum, et al., 2011; Murphy, Oxencis, Klauck, Meyer, & Zimmerman, 2009; Poole, Chainakul, Pearson, & Graham, 2006; Schnipper et al., 2009; Showalter, Rafferty, Swallow, Dasilva, & Chuang, 2011).

High workload is considered a serious problem for healthcare providers, resulting in potential adverse events or miss-treatment during patient care(Vincent, Taylor-Adams, & Stanhope, 1998). Multiple studies suggest that the application of technology can provide a workload incentive to deliver better healthcare (Bardsley, Steventon, & Doll, 2013; Meng, Fallavollita, Habert, Weidert, & Navab, 2016). In sum, our findings suggest that a CDA consolidation system can significantly decrease the providers' workload by reducing in half the time to review multiple, duplicated medical records. Furthermore, when using this system, participants' self-reported a significantly reduced overall workload.

Participants perceived less overall workload while doing referral, medication reconciliation, and problems reconciliation scenarios when the CCDs are reconciled and consolidated using our CDA consolidation system. However, some aspects of provider perceptions about workload did not differ before and after consolidation. For instance, in referral scenario performance, effort, and frustration measures did not change significantly. We suspect that this was due to the lower number of CCDs in the first task of the referral scenario (3 CCDs), as well as the low number of allergies in the CCDs (5 allergies).

Among these three measures (performance, effort, and frustration), only frustration was not significantly changed in last scenario (problems reconciliation). The first task of the problem reconciliation scenario involved reviewing seven CCDs, which imposed higher workload and time demands. We were surprised that participants did not report a higher level of frustration with this task. One possible explanation is that the participants became familiar with the process of the scenarios. After conducting four tasks in the previous scenarios (referral and medication reconciliation) they may have become unintentionally primed and therefore were faster in executing the final scenario. This explanation is further supported by the similar mean times

between the referral scenario (with three CCDs) and the problem reconciliation scenario. Thus, "multiple-treatment interference" may have been an unintended flaw in the method design.

The participants believed reconciliation was a huge time commitment. Overall, our CDA-consolidation system reduced the time to review medical records by approximately 50% in all the scenarios. The time reduction in medication reconciliation (38.1%) was lower than referral and problems reconciliation scenarios (58.8% and 65.1% respectively). We suspect that this was because we didn't inform participants that the CCD in the second task of the medication reconciliation scenario was already reconciled. Since there was a high number of medication in the CCD (17 CCDs), participants searched for duplications in consolidating CCDs, even though there were none. In their open-ended responses, all participants mentioned that our CDA-consolidation system would be very helpful in their practice due to a time reduction for reviewing medical records, which includes searching and clicking.

Three participants did not successfully reconcile the medications from multiple CCDs. In one case they missed a medication and in another they included both 'Lorazepam' and 'Ativan," which are the same drugs with different names. Their oversights were not surprising. The mental task of reconciliation is more difficult when switching between five CCDs trying to synthesize while removing redundancies. Automating parts of the reconciliation process has the potential for reducing errors and avoiding unintended harm.

Participants noted several challenges are related to finding information in the patients' medical record. All believed that collecting information electronically is very helpful in capturing a more comprehensive history of patient medical records. However, overloading providers with too much information can make it difficult to find the information needed at a particular moment. The participants suggested that providers need only relevant information in their clinical practice and prefer not to review several pages of medical records even if it is already consolidated by our system. They also complained about missing, inaccurate, incomplete, and conflicting information in the medical records, which creates uncertainly while making clinical decisions. Although

addressing the problem of missing or inaccurate information is beyond the scope of this study, our system currently possesses several rules to detect conflicting information.

For instance, one CCD may show that a patient is allergic to penicillin, but other CCDs report that the patient has no known allergies. Another example is where one CCD documents a patient's response to a question about Cigarette-Use as "Never Smoked Cigarette," but in another document the value is "Smoker." Currently, our system detects these conflicts, but it is limited in providing the provider a technical mechanism to fix this conflict in the information.

Exposing and resolving these conflicts in CCD documents after detection is a challenge for future work. In addition, while the current engine contains several rules to address these issues, there are many more conflict varieties that need technical solutions. In addition, once discovered the system needs logic for adjudicating these kinds of conflicts. It is difficult make a priori determinations about which values might take precedence over others. For example, should the system always keep presumed allergies even if one document explicitly reports no known allergies? Such conflict resolution algorithms will be important areas of future work.

Another interview topic was the use of complex or poorly designed user interfaces as another barrier in finding information. Our small contribution in this regard was to use hyperlinks in the outline section of each CCD. This way, the participants were able to click on the section of interest for review (allergies, medications or problems list) instead of searching for that section in the whole CCD. All of the participants found the hyperlinks very helpful for finding information and actually used it while executing the tasks during the study. It is important to note, however, that a detailed evaluation of interface design was outside the scope of this study.

Participants indicated that they typically reviewed medical records in advance of their encounter or during a visit with a patient. Most would not change their approach if they could use a CDA consolidation tool. Only two of the participants suggested that if the data can be reconciled and consolidated before their encounter with the patient, they would rather review the information during the visit rather than in advance of it.

In our final discussion with the participants, we received some recommendations about data categorization. The participants recommended to categorize the information into types of patient visit notes (admission, discharge, or office visit), time stamp data (i.e. active medications), and categorization of problems (i.e. heart problems). Currently, the underlined structure of CCDs is not designed in a way to categorize the data based on these types. However, for generating a HTML version of a CCD, it is possible to improve the style sheet that transforms XML-based CCD to HTML. Through improving the style sheet, it is possible to select specific information from CCDs and present them in a categorized fashion.

4.5. **Conclusion**

In this study we sought to assess the impact of a CDA consolidation system on providers' perceived workload while executing a medical document review. Our findings demonstrated that consolidated CCDs can be reviewed in significantly less time compared to multiple CCDs (approximately 50% less for three scenarios). Also, participants reported they perceived less workload while reviewing consolidated CCDs. Although providers considered it challenging to review relevant information even in consolidated CCDs, participants found our CDA consolidation system very helpful in making accurate clinical decisions and unanimously agreed that they would like to use such a system in their clinical practice.

CHAPTER 5

CONCLUSION AND DISCUSSION

As policies such as meaningful use continue to push health information systems to generate and exchange CDA-based documents, the need for consolidation tools like the one tested in this dissertation will increase. CDA documents represent snapshots in time of care delivered to a patient, and as patients traverse multiple providers for acute and chronic care the number of CDA-based documents containing fragments of their complete, longitudinal medical record will grow. Both HIE networks and EHR systems need efficient methods for consolidating information to support use by providers who utilize various clinical workflows or to capture "new" or "updated" information into a repository for storage. In this dissertation, a novel approach is developed and evaluated to address this need. This section concludes the dissertation by discussing the research aims and questions and how this work addressed them. At the end, limitations and future works are discussed.

5.1. **Aim 1. Developing a modular system**

The first aim of the dissertation is to develop a modular CDA consolidation system that meets three paradigms in computer science: scalability, extensibility, and open source. This section is addressing the following research questions:

Question 1. How accurate is the CDA Consolidation system compared to manual consolidation?

Question 2. What is the performance of system (the average time to process a document as well as the minimum, maximum, and average file sizes)?

Question 3. Is the Consolidated CCD valid based on HL7 standard?

HIOs receive high volume of information from different facilities and provide services to many organizations. Therefore, they require software solutions that can stay functional and efficient against large number of requests (software scalability). Although hardware improvement

usually is the first solution that comes into mind, however, software development approaches has a significant impact on the system scalability. Using good software development techniques, the proposed solution performed a fast consolidation process (0.38 second per CCD) which guarantees a reasonable efficiency even if the system receives large load of input data. Such a solution should also be able to grow without being totally rewritten and allow the users to add new features and modify current functionalities (software extensibility). We addressed this paradigm through designing independent modules and applying Model-View-Controller (MVC) software architectural pattern (Krasner & Pope, 1988). The modularity of the system enables each component to function and be maintained independently and work in a loosely-coupled fashion. Also, the system is developed as an open source solution. The original version of the system was published in GitHub ("CCD deduplication system - Source Code," 2013). Using RESTful APIs and NoSQL database indicates that the CDA consolidation system is tailored and designed to meet the technical needs of HIOs. Developing RESTful APIs the system functionalities are exposed to be consumed through HTTP requests. HTTP is the most ubiquitous protocol for communication over the internet. The NoSQL database is used in the audit component to retain large volume of data to be further analyzed in the future.

Using a random sample of synthesized and real-world CDA documents exchanged within a large HIE network, we assessed the accuracy and performance of the system designed to consolidate and de-duplicate CDA documents generated by EHR systems for the same patient. The accuracy of the system in identifying duplications and generating standard CCD was promising using two data sets (synthesized and real-world). Although, CCD consolidation based on the real-world data had a minor inaccuracy, however, the system can still provide much better view of patients' medical history compared to multiple documents. The performance of the system is favorable and the file size is also reduced significantly in both methods. The results of the system evaluation based on the intentions of aim 1, indicate that the system was successful in

achieving the planned objectives. Nevertheless, there are some limitations that are explained in the next section of this document.

5.2. Aim 2. Characterization of data representation among multiple documents

Aim 2 focused on evaluating the system and characterization of data representation among multiple CCDs. In this aim, the research questions are designed to investigate the challenges for identifying duplications, the differences in data representation, and the compatibility of the input CCDs with HL7 standard. The research questions are:

Question 4. What are the challenges for identifying duplications?

Question 5. What are the differences in data representation among multiple documents?

Question 6. How much the real-world CCDs are compliant with HL7 standard?

We observed that nearly all of the issues encountered by the system were due to missing semantic coding within a structured data element. When a data element lacked semantic encoding, it was challenging to compare that element to the others included in the set of CCDs under examination by the system. In these cases, the system attempted to perform string comparisons on concept names and local identifiers. However, the lack of semantic encoding for those elements left duplicates in the final, output CCD.

While inaccuracies in de-duplicating concepts were caused primarily by a lack of semantic interoperability, overall we observed adherence to both syntactic and semantic standards. None of the CCDs were rejected by the system for failing to conform to the HL7 CDA specifications. Furthermore, the majority of structured data elements within the CCDs contained appropriately encoded concepts using terminologies such as SNOMED, LOINC and RxNorm. These are the terminologies required for ‘meaningful use’ certification ("2015 Edition Health Information Technology (Health IT) Certification Criteria, 2015 Edition Base Electronic Health Record (EHR) Definition, and ONC Health IT Certification Program Modifications. Final rule," 2015), and it was pleasant to find their use among CDA-based documents generated by a real-

world, enterprise EHR system. A prior analysis by Dixon et al. (B. E. Dixon, Vreeman, et al., 2014) of real-world messages in the INPC demonstrated low adoption of semantic standards.

There was high consistency in the data representation among multiple CCDs even though the documents were generated in different healthcare facilities. For example the date and time of the entries in different section (allergies, problems, or medication) were based on the same format and the system didn't face any problem in time comparison processes. This was because the CCDs syntactically and semantically adhered to the HL7standard. Being compliant with HL7 constraints and using standard data types minimized the differences in the data representation among documents.

5.3. **Aim 3. Impact of the system on healthcare providers' workload**

A scenario-based study was conducted to evaluate the impact of the system on healthcare providers' perceived workload. The research questions in the aim 3 are:

Question 7. What is the impact of system on participants' workload in the 6 sub-classes (mental demands, physical demands, temporal demands, performance, effort, and frustration)?

Question 8. What is the impact of system on participants' accuracy and efficiency in reconciling CCDs?

The study results show that the overall perceived workload of the participants was significantly reduced after reviewing consolidated CCDs. Medical information reconciliation is a complex and time consuming task. This study demonstrated automatic consolidation of medical documents has a potential to expedite clinical workflows. Although in some scenarios (i.e. performance, effort, and frustration in the referral scenario or mental demand in the medication scenario) providers' perception about workload did not differ before and after consolidation; however, using NASA TLX tool we collected adequate statistics to show that the CDA Consolidation system has a significant impact on reducing healthcare providers' workload.

The system also can assist providers to be more efficient in their clinical practice. Participants completed the scenarios based on the consolidated CCDs significantly faster than the tasks related to multiple documents (almost 50% in average for three scenarios). Also, they made no mistake in conducting the scenarios using consolidated documents versus multiple documents in which some of the participants couldn't reconcile the medication list correctly. This is another example of impact of information overload on clinical decision making. Information overload was decreased through removing duplications and consolidating multiple CCDs which led to reduction in medication error. According to the interview results in the third study, the participants found CDA consolidation system very helpful in making accurate clinical decisions and are interested in using this system in their practice.

5.4. **Limitations and Future Works**

The CDA consolidation system can be used or improved in the future research studies to address the challenge of information overload and data duplication. We explained the limitations of the current study and categorized the expected future works in two themes: a) Further development of the tool since it is open source, b) Additional studies to evaluate the tool.

5.4.1. Further development of the tool since it is open source

While it is expected the CDA consolidation system detects conflicting information across documents, however, the study's main focus is on de-duplication of information. The current rule sets are configured to allow, for example, an entry for the presence of an allergy and the notice of "no known allergies" as two distinct entries in the final allergy section. More sophisticated approaches are necessary to develop for resolving complex conflicts in the data observed across multiple CCDs. This is an area for future research by our team as well as others developing tools for consolidating and presenting data to providers drawn from multiple sources.

The current version of the system performs string comparison when there is no semantic encoding for the elements inside CCDs. However, the lack of semantic encoding for those elements left duplicates in the final, output CCD. One approach for the future might be to explore the use of fuzzy matching algorithms and word stemming, techniques used in natural language processing (Chaudhuri, Ganjam, Ganti, & Motwani, 2003; Yang & Chute, 1994).

The Meaningful Use rules refer to the "adoption of solely the Consolidated CDA (C-CDA) standard for summary care records. C-CDA implementation guide (HL7 Health Level Seven, 2012) includes 9 different clinical documents including CCD. Our focus in this study is on consolidation of three important sections of a medical record (problems, medications, and allergies). Since, these sections are covered in CCD, the consolidated document is generated in CCD format. However, the system is designed in a way to be extensible in the future to generate different templates of CDA. Since the original version of the system was published in GitHub as an open source solution ("CCD deduplication system - Source Code," 2013); therefore, further extensions can be applied on the system by other contributors. There are more opportunities to enhance the system from technical perspective such as consolidating other sections of CCD - CCD document contains 17 different sections among which we consolidated only three sections in this work - or improving the system to consolidate HL7 FHIR based documents.

5.4.2. Additional studies to evaluate the tool

There are some areas that this system can be piloted to perform other kinds of information reconciliation. For example, the performance, accuracy, and efficiency of the system can be assessed through consolidation of other templates of CDA such as Consultation Notes or and Discharge Summary. The result of system evaluation was promising using the CCDs (one of the popular templates of CDA) that we obtained from INPC; however, the system was never tested based on other CDA templates. Evaluation of the system based on CCDs that are not within IHIE network would be another use case for future studies.

Initially the system was designed to address the challenges in HIE Networks. HIOs facilitate the exchange of information among healthcare centers, however, there are some organizations that communicate without existence of HIOs and directly exchange the data among each other. Another opportunity for further research is to evaluate the functionality of the system in the situations that HIOs are not involved. For example, embedding the system in an EHR system to consolidate the CCDs transferred from other EHRs would be one of the use cases. This idea was originated from the interview session of the study 3. Although the participants in the interview were interested in using CDA consolidation system in their practice, however, they hypothesized that integration of the tool with EHR systems would be more beneficial. The future research needs to be done to validate this hypothesis.

REFERENCES

2015 Edition Health Information Technology (Health IT) Certification Criteria, 2015 Edition Base Electronic Health Record (EHR) Definition, and ONC Health IT Certification Program Modifications. Final rule. (2015). *Fed Regist, 80*(200), 62601-62759.

Acker, B., Birnbaum, C., Braden, J., Carlisle, V., Claybrook, D., Demster, B., . . . Grant, K. (2007). HIM principles in health information exchange. *Journal of AHIMA/American Health Information Management Association, 78*, 69-74.

Agrawal, A., & Wu, W. Y. (2009). Reducing medication errors and improving systems reliability using an electronic medication reconciliation system. *Jt Comm J Qual Patient Saf, 35*(2), 106-114.

Australian Commission on Safety and Quality in Health Care. (2012, October 2012). Safety and Quality Improvement Guide Standard 4: Medication Safety Retrieved from http://www.safetyandquality.gov.au/wp-content/uploads/2012/10/Standard4_Oct_2012_WEB.pdf

Bardsley, M., Steventon, A., & Doll, H. (2013). Impact of telehealth on general practice contacts: findings from the whole systems demonstrator cluster randomised trial. *BMC health services research, 13*(1), 1.

Barnes, M. (2007). Lessons learned from the implementation of clinical messaging systems. *AMIA Annu Symp Proc*, 36-40. Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=18693793

Bassi, J., Lau, F., & Bardal, S. (2010). Use of information technology in medication reconciliation: a scoping review. *Annals of Pharmacotherapy, 44*(5), 885-897.

Bawden, D. (2001). Information overload. *Library & information briefings*(92), 1-15.

Bawden, D., & Robinson, L. (2009). The dark side of information: overload, anxiety and other paradoxes and pathologies. *Journal of information science, 35*(2), 180-191.

Biondich, P. G., Dixon, B. E., Duke, J., Mamlin, B., Grannis, S., Takesue, B. Y., . . . Tierney, W. M. (2014). Regenstrief Medical Informatics: Experiences with Clinical Decision Support Systems. In R. A. Greenes (Ed.), *Clinical Decision Support: The Road to Broad Adoption* (2 ed., pp. 165-187). Burlington, MA: Elsevier, Inc.

Biondich, P. G., & Grannis, S. J. (2004). The Indiana network for patient care: an integrated clinical information system informed by over thirty years of experience. *J Public Health Manag Pract, Suppl*, S81-86. doi:00124784-200411001-00013 [pii]

Boockvar, K. S., Blum, S., Kugler, A., Livote, E., Mergenhagen, K. A., Nebeker, J. R., . . . Yeh, J. (2011). Effect of admission medication reconciliation on adverse drug events from admission medication changes. *Arch Intern Med, 171*(9), 860-861. doi:10.1001/archinternmed.2011.163

Boockvar, K. S., Santos, S. L., Kushniruk, A., Johnson, C., & Nebeker, J. R. (2011). Medication reconciliation: barriers and facilitators from the perspectives of resident physicians and pharmacists. *J Hosp Med, 6*(6), 329-337. doi:10.1002/jhm.891

Brailer, D. J. (2005). Interoperability: the key to the future health care system. *Health Affairs-Millwood Va Then Bethesda MA, 24*, W5.

Byrne, C. M., Mercincavage, L. M., Bouhaddou, O., Bennett, J. R., Pan, E. C., Botts, N. E., . . . Cromwell, T. (2014). The Department of Veterans Affairs' (VA) implementation of the Virtual Lifetime Electronic Record (VLER): findings and lessons learned from Health Information Exchange at 12 sites. *Int J Med Inform, 83*(8), 537-547. doi:10.1016/j.ijmedinf.2014.04.005

C 32 - HITSP Summary Documents Using HL7 Continuity of Care Document (CCD) Component. Retrieved from

http://www.hitsp.org/ConstructSet_Details.aspx?&PrefixAlpha=4&PrefixNumeric=32

CCD deduplication system - Source Code. (2013). Retrieved from https://github.com/jdmeade/ccd-dedup

Chaudhuri, S., Ganjam, K., Ganti, V., & Motwani, R. (2003). *Robust and efficient fuzzy match for online data cleaning.* Paper presented at the Proceedings of the 2003 ACM SIGMOD international conference on Management of data.

Chewning, E. G., & Harrell, A. M. (1990). The effect of information load on decision makers' cue utilization levels and decision quality in a financial distress decision task. *Accounting, Organizations and Society, 15*(6), 527-542.

Dawes, M., & Sampson, U. (2003). Knowledge management in clinical practice: a systematic review of information seeking behavior in physicians. *Int J Med Inform, 71*(1), 9-15.

Deshpande, N., Publicover, M., Gee, H., & Khan, K. S. (2003). Incorporating the views of obstetric clinicians in implementing evidence-supported labour and delivery suite ward rounds: a case study. *Health Info Libr J, 20*(2), 86-94.

DiDomenico, A., & Nussbaum, M. A. (2008). Interactive effects of physical and mental workload on subjective workload assessment. *International Journal of Industrial Ergonomics, 38*(11), 977-983.

Dixon, B. E. (2016). What is Health Information Exchange? In B. E. Dixon (Ed.), *Health Information Exchange: Navigating and Managing a Network of Health Information Systems* (pp. 3-20). Waltham, MA: Academic Press.

Dixon, B. E., Colvard, C., & Tierney, W. M. (2014). Identifying health facilities outside the enterprise: challenges and strategies for supporting health reform and meaningful use. *Inform Health Soc Care*, 1-15. doi:10.3109/17538157.2014.924949

Dixon, B. E., Simonaitis, L., Goldberg, H. S., Paterno, M. D., Schaeffer, M., Hongsermeier, T., . . . Middleton, B. (2013). A pilot study of distributed knowledge management and clinical decision support in the cloud. *Artif Intell Med, 59*(1), 45-53. doi:10.1016/j.artmed.2013.03.004

Dixon, B. E., Vreeman, D. J., & Grannis, S. J. (2014). The long road to semantic interoperability in support of public health: experiences from two states. *J Biomed Inform, 49*, 3-8. doi:10.1016/j.jbi.2014.03.011

Dixon, B. E., Zafar, A., & Overhage, J. M. (2010). A Framework for evaluating the costs, effort, and value of nationwide health information exchange. *Journal of the American Medical Informatics Association : JAMIA, 17*(3), 295-301. doi:10.1136/jamia.2009.000570

Dolin, R. H., Alschuler, L., Boyer, S., Beebe, C., Behlen, F. M., Biron, P. V., & Shabo , A. (2006). HL7 Clinical Document Architecture, Release 2. *Journal of the American Medical Informatics Association, 13*(1), 30-39. doi:10.1197/jamia.M1888

Dolin, R. H., Alschuler, L., Boyer, S., Beebe, C., Behlen, F. M., Biron, P. V., & Shabo, A. (2006). HL7 clinical document architecture, release 2. *Journal of the American Medical Informatics Association, 13*(1), 30-39.

Dolin, R. H., Giannone, G., & Schadow, G. (2007). Enabling joint commission medication reconciliation objectives with the HL7 / ASTM Continuity of Care Document standard. *AMIA Annual Symposium Proceedings*, 186-190.

Edmunds, A., & Morris, A. (2000). The problem of information overload in business organisations: a review of the literature. *International journal of information management, 20*(1), 17-28.

Eppler, M. J., & Mengis, J. (2003). *A framework for information overload research in organizations*. Retrieved from

Frisse, M. E., Johnson, K. B., Nian, H., Davison, C. L., Gadd, C. S., Unertl, K. M., . . . Chen, Q. (2012). The financial impact of health information exchange on emergency department care. *Journal of the American Medical Informatics Association : JAMIA, 19*(3), 328-333. doi:10.1136/amiajnl-2011-000394

Gadd, C. S., Ho, Y. X., Cala, C. M., Blakemore, D., Chen, Q., Frisse, M. E., & Johnson, K. B. (2011). User perspectives on the usability of a regional health information exchange. *Journal of the American Medical Informatics Association : JAMIA, 18*(5), 711-716. doi:10.1136/amiajnl-2011-000281

Gleason, K. M., Groszek, J. M., Sullivan, C., Rooney, D., Barnard, C., & Noskin, G. A. (2004). Reconciliation of discrepancies in medication histories and admission orders of newly hospitalized patients. *American Journal of Health System Pharmacy, 61*(16), 1689-1694.

Hall, A., & Walton, G. (2004). Information overload within the health care system: a literature review. *Health Information & Libraries Journal, 21*(2), 102-108.

Halligan, A., & Donaldson, L. (2001). Implementing clinical governance: turning vision into reality. *BMJ, 322*(7299), 1413-1417.

Hart, S. G. (2006). *NASA-task load index (NASA-TLX); 20 years later.* Paper presented at the Proceedings of the human factors and ergonomics society annual meeting.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in psychology, 52*, 139-183.

Health information technology: initial set of standards, implementation specifications, and certification criteria for electronic health record technology. Final rule. (2010). *Fed Regist, 75*(144), 44589-44654.

Health Level Seven web site. "About HL7". Retrieved from http://www.hl7.org/about/index.cfm

Healthcare Information Technology Standards Panel (HITSP). (2009). HITSP CDA Content Modules Component. http://www.hitsp.org/ConstructSet_Details.aspx?&PrefixAlpha=4&PrefixNumeric=83 Retrieved from http://www.hitsp.org/ConstructSet_Details.aspx?&PrefixAlpha=4&PrefixNumeric=83

Healthcare Information Technology Standards Panel (HITSP). (2009). C 32 - HITSP Summary Documents Using HL7 Continuity of Care Document (CCD) Component. Retrieved from http://www.hitsp.org/ConstructSet_Details.aspx?&PrefixAlpha=4&PrefixNumeric=32

HITSP Clinical Document and Message Terminology Component. (2010). http://www.hitsp.org/ConstructSet_Details.aspx?&PrefixAlpha=4&PrefixNumeric=80 Retrieved from http://www.hitsp.org/ConstructSet_Details.aspx?&PrefixAlpha=4&PrefixNumeric=80

HITSP Summary Documents Using HL7 Continuity of Care Document (CCD) Component. (2009). http://www.hitsp.org/ConstructSet_Details.aspx?&PrefixAlpha=4&PrefixNumeric=32 Retrieved from http://www.hitsp.org/ConstructSet_Details.aspx?&PrefixAlpha=4&PrefixNumeric=32

HL7 Health Level Seven. (2012). HL7 Implementation Guide for CDA® Release 2: Consolidated CDA Templates for Clinical Notes. Retrieved from http://www.hl7.org/implement/standards/product_brief.cfm?product_id=379

Hoosier Healthcare Innovation Challenge (2013). Retrieved from http://www.hoosierchallenge.org/

Hosseini, M., Ahmadi, M., & Dixon, B. E. (2014). A Service Oriented Architecture Approach to Achieve Interoperability between Immunization Information Systems in Iran. *AMIA Annual Symposium Proceedings, 2014:1797-805*.

Hosseini, M., Jones, J. F., Faiola, A., Vreeman, D., Wu, H., & Dixon, B. E. (2016). Reconciling Real-World CCD Documents Using CDA Consolidation System. *To be submitted to JAMIA*.

Hosseini, M., Meade, J., Schnitzius, J., & Dixon, B. E. (2015). Consolidating CCDs from multiple data sources: A modular approach. *Journal of the American Medical Informatics Association, 2016;23(2):317-23*, ocv084.

Hripcsak, G., Kaushal, R., Johnson, K. B., Ash, J. S., Bates, D. W., Block, R., . . . Overhage, J. M. (2007). The United Hospital Fund meeting on evaluating health information exchange. *Journal of Biomedical Informatics, 40*(6), S3-S10.

Institute for Healthcare Improvement. (2011). How to Guide: Prevent Adverse Drug Events (Medication Reconciliation). Retrieved from http://www.ihi.org/explore/adesmedicationreconciliation/Pages/default.aspx

Integrating the Healthcare Enterprise (IHE). (2011). Reconciliation of Diagnoses, Allergies and Medications (RECON). Retrieved from http://www.ihe.net/Technical_Framework_Archives/

Integrating the Healthcare Enterprise (IHE). (2015). Reconciliation of Clinical Content and Care Providers (RECON). Retrieved from http://www.ihe.net/Technical_Framework_Archives/

Jha, A. K., Doolan, D., Grandt, D., Scott, T., & Bates, D. W. (2008). The use of health information technology in seven nations. *International Journal of Medical Informatics, 77*(12), 848-854.

Klerings, I., Weinhandl, A. S., & Thaler, K. J. (2015). Information overload in healthcare: too much of a good thing? *Z Evid Fortbild Qual Gesundhwes, 109*(4-5), 285-290. doi:10.1016/j.zefq.2015.06.005

Krasner, G. E., & Pope, S. T. (1988). A description of the model-view-controller user interface paradigm in the smalltalk-80 system. *Journal of object oriented programming, 1*(3), 26-49.

Kuperman, G. J. (2011). Health-information exchange: why are we doing it, and what are we doing? *Journal of the American Medical Informatics Association, 18*(5), 678-682.

Lin, M. C., Vreeman, D. J., McDonald, C. J., & Huff, S. M. (2011). A characterization of local LOINC mapping for laboratory tests in three large institutions. *Methods Inf Med, 50*(2), 105-114. doi:10.3414/ME09-01-0072

McDonald, C. J., Callaghan, F. M., Weissman, A., Goodwin, R. M., Mundkur, M., & Kuhn, T. (2014). Use of internist's free time by ambulatory care Electronic Medical Record systems. *JAMA Intern Med, 174*(11), 1860-1863. doi:10.1001/jamainternmed.2014.4506

McDonald, C. J., Overhage, J. M., Barnes, M., Schadow, G., Blevins, L., Dexter, P. R., & Mamlin, B. (2005). The Indiana network for patient care: a working local health information infrastructure. An example of a working infrastructure collaboration that links data from five health systems and hundreds of millions of entries. *Health Aff (Millwood), 24*(5), 1214-1220. doi:10.1377/hlthaff.24.5.1214

Medicare and Medicaid programs; electronic health record incentive program--stage 2. Final rule. (2012). *Fed Regist, 77*(171), 53967-54162.

Meng, M., Fallavollita, P., Habert, S., Weidert, S., & Navab, N. (2016). Device-and system-independent personal touchless user interface for operating rooms. *International Journal of Computer Assisted Radiology and Surgery*, 1-9.

Miller, P. L., Frawley, S. J., & Sayward, F. G. (2000). IMM/Scrub: a domain-specific tool for the deduplication of vaccination history records in childhood immunization registries. *Computers and Biomedical Research, 33*(2), 126-143.

Mueller, S. K., Sponsler, K. C., Kripalani, S., & Schnipper, J. L. (2012). Hospital-based medication reconciliation practices: a systematic review. *Arch Intern Med, 172*(14), 1057-1069.

Murphy, E. M., Oxencis, C. J., Klauck, J. A., Meyer, D. A., & Zimmerman, J. M. (2009). Medication reconciliation at an academic medical center: implementation of a comprehensive program from admission to discharge. *Am J Health Syst Pharm, 66*(23), 2126-2131. doi:10.2146/ajhp080552

Noone, J., Warren, J., & Brittain, M. (1998). Information overload: opportunities and challenges for the GP's desktop. *Stud Health Technol Inform, 52 Pt 2*, 1287-1291.

O'Reilly, C. A. (1980). Individuals and information overload in organizations: is more necessarily better? *Academy of management journal, 23*(4), 684-696.

Office of the National Coordinator (ONC). (2012). Health Information Technology: Standards, Implementation Specifications, and Certification Criteria for Electronic Health Record

Technology, 2014 Edition; Revisions to the Permanent Certification Program for Health Information Technology. *Fed Regist, 77*(171), 54163-54292.

2014 Edition Release 2 Electronic Health Record (EHR) Certification Criteria and the ONC HIT Certification Program; Regulatory Flexibilities, Improvements, and Enhanced Health Information Exchange, (2014).

Office of the National Coordinator (ONC). (2015). 2015 Edition Health Information Technology (Health IT) Certification Criteria, 2015 Edition Base Electronic Health Record (EHR) Definition, and ONC Health IT Certification Program Modifications. Retrieved from https://federalregister.gov/a/2015-25597

Patel, V., Abramson, E. L., Edwards, A., Malhotra, S., & Kaushal, R. (2011). Physicians' potential use and preferences related to health information exchange. *International Journal of Medical Informatics, 80*(3), 171-180.

Paterno, M. D., Goldberg, H. S., Simonaitis, L., Dixon, B. E., Wright, A., Rocha, B. H., . . . Middleton, B. (2012). Using a Service Oriented Architecture Approach to Clinical Decision Support: Performance Results from Two CDS Consortium Demonstrations. *AMIA Annu Symp Proc, 2012*, 690-698. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/23304342

Plaisant, C., Chao, T., Wu, J., Hettinger, A. Z., Herskovic, J. R., Johnson, T. R., . . . Shneiderman, B. (2013). *Twinlist: novel user interface designs for medication reconciliation.* Paper presented at the AMIA Annual Symposium Proceedings.

Poole, D. L., Chainakul, J. N., Pearson, M., & Graham, L. (2006). Medication reconciliation: a necessity in promoting a safe hospital discharge. *J Healthc Qual, 28*(3), 12-19.

Pronovost, P., Weast, B., Schwarz, M., Wyskiel, R. M., Prow, D., Milanovich, S. N., . . . Lipsett, P. (2003). Medication reconciliation: a practical tool to reduce the risk of medication errors. *Journal of critical care, 18*(4), 201-205.

Ratwani, R. M., Fairbanks, R. J., Hettinger, A. Z., & Benda, N. C. (2015). Electronic health record usability: analysis of the user-centered design processes of eleven electronic health record vendors. *Journal of the American Medical Informatics Association : JAMIA, 22*(6), 1179-1182. doi:10.1093/jamia/ocv050

Redmond, P., Carroll, H., Grimes, T., Galvin, R., McDonnell, R., Boland, F., . . . Fahey, T. (2016). GPs' and community pharmacists' opinions on medication management at transitions of care in Ireland. *Fam Pract, 33*(2), 172-178. doi:10.1093/fampra/cmw006

Reiner, B. I. (2011a). Medical imaging data reconciliation, part 1: innovation opportunity. *Journal of the American College of Radiology, 8*(9), 622-625.

Reiner, B. I. (2011b). Medical imaging data reconciliation, part 2: clinical order entry and imaging report data reconciliation. *Journal of the American College of Radiology, 8*(10), 720-724.

Reiner, B. I. (2011c). Medical imaging data reconciliation, part 3: reconciliation of historical and current radiology report data. *Journal of the American College of Radiology, 8*(11), 768-771.

Schnipper, J. L., Hamann, C., Ndumele, C. D., Liang, C. L., Carty, M. G., Karson, A. S., . . . Gandhi, T. K. (2009). Effect of an electronic medication reconciliation application and process redesign on potential adverse drug events: a cluster-randomized trial. *Arch Intern Med, 169*(8), 771-780. doi:10.1001/archinternmed.2009.51

Schoen, C., Osborn, R., Squires, D., Doty, M., Rasmussen, P., Pierson, R., & Applebaum, S. (2012). A Survey Of Primary Care Doctors In Ten Countries Shows Progress In Use Of Health Information Technology, Less In Other Areas. *Health Aff (Millwood)*. doi:10.1377/hlthaff.2012.0884

Schroder, H. M., Driver, M. J., & Streufert, S. (1967). *Human information processing: Individuals and groups functioning in complex social situations*: Holt, Rinehart and Winston.

Showalter, J. W., Rafferty, C. M., Swallow, N. A., Dasilva, K. O., & Chuang, C. H. (2011). Effect of standardized electronic discharge instructions on post-discharge hospital

utilization. *J Gen Intern Med, 26*(7), 718-723. doi:10.1007/s11606-011-1712-y

Silva, P. A. B., Bernstam, E. V., Markowitz, E., Johnson, T. R., Zhang, J., & Herskovic, J. R. (2011). *Automated medication reconciliation and complexity of care transitions.* Paper presented at the AMIA Annual Symposium Proceedings.

Strauss, A. T., Martinez, D. A., Garcia-Arce, A., Taylor, S., Mateja, C., Fabri, P. J., & Zayas-Castro, J. L. (2015). A user needs assessment to inform health information exchange design and implementation. *BMC Med Inform Decis Mak, 15*, 81. doi:10.1186/s12911-015-0207-x

Vertanen, K. (2012). NASA-TLX in HTML and JavaScript. Retrieved from https://www.keithv.com/software/nasatlx/

Vest, J. R., & Gamm, L. D. (2010). Health information exchange: persistent challenges and new strategies. *Journal of the American Medical Informatics Association, 17*(3), 288-294.

Vincent, C., Taylor-Adams, S., & Stanhope, N. (1998). Framework for analysing risk and safety in clinical medicine. *BMJ, 316*(7138), 1154-1157.

Vira, T., Colquhoun, M., & Etchells, E. (2006). Reconcilable differences: correcting medication errors at hospital admission and discharge. *Quality and Safety in Health Care, 15*(2), 122-126.

Wagholikar, K., Sohn, S., Wu, S., Kaggal, V., Buehler, S., Greenes, R. A., . . . Chaudhry, R. (2013). Workflow-based data reconciliation for clinical decision support: case of colorectal cancer screening and surveillance. *AMIA Summits on Translational Science Proceedings, 2013*, 269.

Walker, J., Pan, E., Johnston, D., Adler-Milstein, J., Bates, D. W., & Middleton, B. (2005). The value of health care information exchange and interoperability. *Health Aff (Millwood), Suppl Web Exclusives*, W5-10-w15-18. doi:10.1377/hlthaff.w5.10

World Health Organization. (2007). WHO Collaborating Centre on Patient Safety Solutions. *Control of Concentrated Electrolyte Solutions. Patient Safety Solutions, 1*(5), 1-3.

Yang, Y., & Chute, C. G. (1994). An example-based mapping method for text categorization and retrieval. *ACM Transactions on Information Systems (TOIS), 12*(3), 252-277.

Zafar, A., & Dixon, B. E. (2007). Pulling back the covers: technical lessons of a real-world health information exchange. *Stud Health Technol Inform, 129*(Pt 1), 488-492. Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=17911765

Ziminski, T. B., De la Rosa Algarin, A., Saripalle, R., Demurjian, S., & Jackson, E. (2012). *SMARTSync: Towards patient-driven medication reconciliation.* Paper presented at the Bioinformatics and Biomedicine Workshops (BIBMW), 2012 IEEE International Conference on.

CURRICULUM VITAE
Seyed Masoud Hosseini Asanjan

EDUCATION

2012 - 2016 • Ph.D., BioMedical and Health Informatics, School of Informatics and Computing, Indiana University, IN, USA
2009 – 2011 • M.Sc., Medical Informatics, Tehran University of Medical Science, Tehran, Iran
2004 – 2007 • B.Sc., Software Engineering, Azad University, Zanjan, Iran

PUBLICATIONS

- **Hosseini, M.,** & Dixon, B. E. (2016). Chapter 8 - Syntactic Interoperability and the Role of Standards In B. E. Dixon (Ed.), Health Information Exchange: Navigating and Managing a Network of Health Information Systems. Waltham, MA: Academic Press.
- **Masoud Hosseini**, M. H., Josette Jones. (2015). Hidden Aspects of Breast Cancer Patients' Life: A NLP Approach. Paper presented at the IndyBigData Conference.
- **Hosseini, M**., Meade, J., Schnitzius, J., & Dixon, B. E. (2015). Consolidating CCDs from multiple data sources: A modular approach. Journal of the American Medical Informatics Association, 2016;23(2):317-23, ocv084.
- **Hosseini, M**., Ahmadi, M., & Dixon, B. E. (2014). A Service Oriented Architecture Approach to Achieve Interoperability between Immunization Information Systems in Iran. AMIA Annual Symposium Proceedings, 2014:1797-805.
- **Hosseini, M**., Meade, J., Schnitzius, J., & Dixon, B. E. (2015). Consolidating CCDs from multiple data sources: A modular approach. Journal of the American Medical Informatics Association, 2016;23(2):317-23, ocv084.
- Zolnoori, M., Jones, J., Moin, M., Heidarnejad, H., Fazlollahi, M., & **Hosseini, M.** (2013). Evaluation of User Interface of Computer Application Developed for Screening Pediatric Asthma. In C. Stephanidis & M. Antona (Eds.), Universal Access in Human-Computer Interaction. Applications and Services for Quality of Life (Vol. 8011, pp. 563-570): Springer Berlin Heidelberg.
- **Hosseini, M.**, Ahmadi, M., Niakan, A. (2010). Control of User Access Level Based on Service Oriented Architecture. Tele-Medicine and Electronic Hospital Symposium, Tehran.

PUBLICATION UNDER PREPARATION

- **Hosseini, M.**, Jones. J., Faiola, A., Wu, H., Vreeman, D., Dixon, B. E. Impact of reconciled medical information on providers' perceived mental workload. Journal of the American Medical Informatics Association. (in final preparation to be submitted to JAMIA).
- Jones, J., Hosseini, M., Pradhan, M., Kulanthaivel, A., **Hosseini, M**. Modeling Patient Generated Data into Meaningful Topics: a Novel Approach. AMIA 2016 (in final preparation to be submitted to JBI).
- **Hosseini, M.**, Jones, J., Faiola, A., Wu, H., Vreeman, D., Dixon, B. E. Reconciliation of CDA-based clinical documents. Journal of the American Medical Informatics Association. (in final preparation to be submitted to JAMIA).

TEACHING AND MENTORING EXPERIENCES

2015 - 2016 • Co-instructor of course INFO B535 Clinical Information Systems, Indiana university.
2015 • Co-instructor of course INFO B642 Clinical Decision Support Systems, Indiana university.
2015 • Co-instructor of course INFO B582 Health Information Exchange, Indiana university.

2014 • Guest lecturer of course INFO B582 Health Information Exchange, Indiana university.
2013 - 2014 • Mentor and committee member of MS thesis in Health Informatics, Indiana University.
2010 • Speaker of Workshop "Microsoft SQL Server", Tehran University of Medical Science.
2006 • Lecturer of course Data Base Management System, Computer Science Department, Technical and Vocational College of Tabriz, Tabriz, Iran.

PROFESSIONAL WORKING EXPERIENCE

Intern - Roche Diabetes Care, Inc., Indianapolis, IN
I started my work at Roche Diabetes Care as an intern in May 2015 and since then I worked on various projects mostly focused on diabetes information exchange. Roche Diabetes Care is a pioneer in the development of blood glucose monitoring systems and a global leader for diabetes management systems and services. Roche is among 100 best companies at USA to work for and its internship acceptance rate is 4%. (http://www.roche.com)
2016 • Developing a mobile app for capturing information from devices (i.e. fitbit, glucose meter) and transferring to Roche endpoints through apigee API proxy server.
2016 • Designing a simplified proprietary messaging standard and an HL7 FHIR-based profile for diabetes and chronic disease domain.
2015 • Intercepting data traffic using proxy server and investigating message invalidities.
2015 • Developing EMR Simulator with capability of generating standard messages (HL7 C36, C-CDA, Continua WAN) to be exchanged with Roche endpoints.

Regenstrief Institute, Inc., Indianapolis, IN
Along with studying PhD in Health Informatics since 2012, I have being serving as Research Assistant at Regenstrief Institute. Regenstrief Institute, Inc. is an internationally respected informatics and healthcare research organization, recognized for OpenMRS and LOINC. (http://www.regenstrief.org)
2013 - Present • Developing a system for De-Duplication and consolidation of HL7 CCD documents received from multiple facilities.
2012 - 2014 • Developing a HL7 C62 generator to exchange messages between Regenstrief Institute and Indiana Veterans' Affairs (VA).

Software Engineer - EPD Co., Tehran, Iran
I worked at EPD Co. from 2007 to 2012 as a software engineer and during that 7 years I involved in various projects in three main themes: a. Software development and testing, b. System analysis and design c. working with HL7 standards. EPD Co. is a prestigious medical software developing company in Iran which produces software solutions for healthcare enterprises. (http://www.epd.ir/en)
2010 - 2012 • Analysis, Design and implementation of SOA-based EHR compatible with HL7 (V3).
2010 - 2011 • Development of interoperable immunization system based on SOA and HL7 (V3).
2009 • development and implementation of HL7 CDA generator as a new feature in EPD HIS.
2008 - 2009 • Developing Entity Identification Service (EIS) web service and exchanging HL7 (V3) messages for fulfilling interoperability between EPD-HIS and EIS.
2008 - 2009 • Investigating HL7 (V3) standards and customizing several HL7 (V3) messages in order to improve the compatibility of EPD systems with HL7 standard.

2007 - 2008 • Developing finger print recognition API used in EPD products such as time attendance and reservation systems.

COMPUTER SKILLS

Languages and Technologies
C#, CSS, CSS3, HTML, HTML5, Java, JavaScript, jQuery, JSP, Python, Swift, XAML

Data Management and Messaging
JSON, LINQ, MS SQL Server, MySQL, NHibernate, SQLite, XML, XSLT

System Analysis and Design
Classic RUP, RUP SOMA, IBM Rational Rose, Enterprise Architect, Visual Paradigm

Research and Statistical Software
MALLET, MATLAB, R, SAS, SPSS, WEKA

Project and Source Management
Microsoft Team Foundation System (TFS), MS Project, GitHub

Other Technologies
Apigee API management, Crystal Reports, MS WCF, MS WPF, SQL Server Reporting Service (SSRS)

PROFESSIONAL MEMBERSHIP AND ACTIVITIES

- Article reviewer of journal Health Policy and Technology.
- Article reviewer of journal Applied Clinical Informatics (ACI) Health Policy and Technology.
- Article reviewer of American Medical Informatics Association proceedings (AMIA).
- Official Member of ISIRI/ISO TC 215 – Health Informatics work Group of Iran Ministry of Health and Education, Tehran, Iran.
- Member of Heath Level Seven (HL7) organization.
- Member of ingenuity Society of Tehran University of Medical Science, Tehran, Iran.

HONORS AND AWARDS

2015 • Best Poster Award at IndyBigData Conference, 2015.
2013 • Winner of "Health Exchange Data Management" challenge at Hoosier Healthcare Innovation Challenge 2013 (A collaborative effort).
2011 • Ranked 5th in the Iranian nationwide Medical Informatics PhD entrance exam.
2011 • Awarded as Exceptional Talented student in Tehran University of Medical Science
2009 - 2011 • Ranked as top student during my M.Sc. course (highest GPA among students in all of the semesters).
2009 • Ranked 11rd in the Iranian nationwide Medical Informatics M.Sc. entrance exam.

LINGUAL SKILLS

Azeri (Native), Persian (Native), English (Professional), Arabic (Basic), Turkish (Basic).