A BIG DATA AUGMENTED ANALYTICS PLATFORM TO OPERATIONALIZE EFFICIENCIES AT COMMUNITY CLINICS

Kislaya Kunjan

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

June 2017

Accepted by the Graduate Faculty, Indiana University, in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Doctoral Committee

____

Josette F. Jones, PhD, Chair

____

Tammy Toscos, PhD

____

Huanmei Wu, PhD

August 15, 2016

____

Richard Holden, PhD

DEDICATION

Dedicated to my parents, my family (Meghana and Tasmin) and friends for their great love, support, and encouragement over the years

ACKNOWLEDGEMENTS

I would like to start by expressing my utmost gratitude to Dr. Tammy Toscos, for her relentless guidance and support as my primary advisor. Dr. Toscos has been a pleasure to work with, thanks to her scientific rigor, leadership, integrity, fairness, sense of humor, and pragmatism.

I would like to extend my sincere thanks to my committee members who have all been of great help to me: Dr. Jones, through her expansive knowledge of health informatics as well the nitty gritty of administration; Dr. Huanmei Wu, through her technical expertise and leadership skills; Dr. Richard Holden through his knowledge of what constitutes research in general and user-centered research in particular.

Special thanks to Dr. Brad Doebbeling. He had been my primary advisor for the initial years of this research after which he departed from Indiana University. He has been a great mentor to me through his vast yet unassuming knowledge of health services research and informatics as well as the vision for this research. He shepherded me through the design phases of this research and its initial implementation. I owe it to him and the co-PIs of this funded research project for providing me with this learning opportunity of a lifetime.

Special appreciation and thanks to our project manager, Amy Miller, for her agility, optimism and generosity that made it a pleasure being a member of the research team.

I also like to extend my gratitude to Dr. Hadi Kharrazi and Dr. Mathew Palakal, with whom I had the distinct pleasure of working with during the early years as a doctoral student. To the remaining faculty and staff of the School of Informatics and Computing at Indiana University, I have learned much from you all over the past 5 years. Thank you all for the support and guidance that has helped me to achieve this goal.

Kislaya Kunjan

A BIG DATA AGUMENTED ANALYTICS PLATFORM TO OPERATIONALIZE EFFICIENCIES AT COMMUNITY CLINICS

Community Health Centers (CHCs) play a pivotal role in delivery of primary healthcare to the underserved, yet have not benefited from a modern data analytics platform that can support clinical, operational and financial decision making across the continuum of care. This research is based on a systems redesign collaborative of seven CHC organizations spread across Indiana to improve efficiency and access to care.

Three research questions (RQs) formed the basis of this research, each of which seeks to address known knowledge gaps in the literature and identify areas for future research in health informatics. The first RQ seeks to understand the information needs to support operations at CHCs and implement an information architecture to support those needs. The second RQ leverages the implemented data infrastructure to evaluate how advanced analytics can guide open access scheduling – a specific use case of this research. Finally, the third RQ seeks to understand how the data can be visualized to support decision making among varying roles in CHCs.

Based on the unique work and information flow needs uncovered at these CHCs, an end-to-end analytics solution was designed, developed and validated within the framework of a rapid learning health system. The solution comprised of a novel heterogeneous longitudinal clinic data warehouse augmented with big data technologies and dashboard visualizations to inform CHCs regarding operational priorities and to support engagement in the systems redesign initiative. Application of predictive analytics on the health center data guided the implementation of open access scheduling and up to a 15% reduction in the missed appointment rates. Performance measures of importance to specific job profiles within the CHCs were uncovered. This was followed by a user-centered design of an online interactive dashboard to support rapid assessments of care delivery. The impact of the dashboard was assessed over time and formally validated through a usability study involving cognitive task analysis and a system usability scale questionnaire.

Wider scale implementation of the data aggregation and analytics platform through regional health information networks could better support a range of health system redesign initiatives in order to address the national ‘triple aim’ of healthcare.

Josette F. Jones, RN, Ph.D., Chair

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: GENERAL BACKGROUND

1.1 Introduction

The United States (US) healthcare system is beset with inadequate access, increasing costs, and low or inconsistent quality of care[1]. Average healthcare expenditure per capita is two-and-a-half times that of other developed countries even as the overall health care ranking remains low[2]. Primary health care is particularly challenged in terms of access, costs and outcomes relative to peer economies[3].

Concurrently, the healthcare landscape is entering a new era with the passage of the two important federal legislation: Health Information Technology for Economic and Clinical Health (HITECH) Act and the Patient Protection and Affordable Care Act (ACA). The HITECH Act passed in 2009[4], has been promoting the adoption and meaningful use of health information technology (IT), notably Electronic Health Record systems (EHRs)[5] and health information exchanges (HIEs)[6]. The ACA passed in 2012,[7] is greatly expanding health insurance access to citizens and creating new incentives to change clinical practice through better care coordination and quality. The provisions from both these legislations are making electronic healthcare data integral for clinical practice as well as making primary care health providers more accessible to patients[8]. The rapid adoption of Health IT is also catalyzing the premise of a highly participatory rapid learning health system (LHS), which seeks to derive actionable knowledge based on real-world experiences of millions of patients receiving care from health care providers, thereby hoping to reduce the 17-year average latency between discovery of new medical knowledge and its widespread use[9]. The Institute for Healthcare Improvement (IHI)[10] posits that new health system designs must be developed to simultaneously pursue the three dimensions (“triple aim”) of healthcare: improve the patient experience of care (including quality and satisfaction); improve the health of populations; and reduce the per capita cost of health care. More recently, the ‘triple aim’ has been expanded to a ‘quadruple aim’, adding the goal of improving the work life of health care providers[11]. Leveraging health IT has been recognized as key to achieving these “aims” of healthcare[12,13].

Within primary care, Community Health Centers (CHCs) function as ‘safety net’ ambulatory providers of acute, chronic and preventive care services to the low income and in medically underserved areas.[14] There are 1375 federally supported CHC organizations, also known as Federally Qualified Health Centers (FQHCs) or simply, health centers, that are spread over 9800 clinical service delivery sites serving over 23 million patients[15]. The passage of the key provisions of the ACA starting in 2014 aimed at reducing the uninsured population, has led to CHCs assuming a greater role in the primary care delivery landscape[16]. This is because the previously uninsured population now has coverage through the ACA and seeks health care predominantly at the CHCs[17]. The CHCs in turn have also received additional funding to meet this new demand for services through facility construction, renovation and the acquisition of health IT[18,19]. The expansion and investment in health IT serves to enhance the role of CHCs in the learning health system paradigm.

CHCs do share some health IT commonalities with other outpatient primary care clinics (many of which are affiliated with large hospital systems). In particular, they both rely on EHRs as a means to capture clinical information and manage their practice. However, CHCs differ significantly from other outpatient clinics in terms of their patient population. From a financial perspective, this creates challenging revenue streams due to a disadvantageous payer mix, with average operating margins at CHCs being less than 1%. This in turn limits their ability to make substantial capital investments (either in terms of facilities or health IT solutions)[20]. However, with federal funding support for health IT, the CHCs have seen great expansion in their implementation and use of EHRs since 2010[21]. Over 92% of all CHCs now use EHRs, which is higher than other office-based physician practices[22].

CHCs are keen to achieve the status of patient-centered medical homes (PCMH)[23] – a care delivery model which is patient-centered (responsive to individual patient preferences, needs and values) and coordinated through a primary care physician while being committed to quality and safety. Two-thirds (68%) of CHCs have thus far achieved PCMH recognition (a provision within the ACA). A critical factor for achieving PCMH recognition relates to improving patient appointment access defined in terms of providing same day appointments to patients (through more open scheduling), monitoring no-shows, and identifying opportunities to improve patient access to appointments[24]. CHCs are known to face unusually high no-show rates,[25] partially attributable to their unique clientele, difficulty in retaining providers needed for building patient relationships that in turn support visit adherence[26], and their inability (as safety-net providers) to penalize those who don't show for their scheduled appointments. This calls for a greater understanding of the clinic operations and in particular as it pertains to the appointment scheduling system and how it can be redesigned to deliver high quality care in terms of appointment access, outcomes, and costs[27].

This research builds upon efforts to better understand the unique work and information flow needs at CHCs from a health IT perspective. It then builds an analytics platform to support an operational redesign initiative aimed at delivering care that is efficient and accessible. Improving access to care through a redesign of the appointment scheduling system is a specific use case of this research, given its importance in light of CHCs achieving PCMH recognition. However, the data analytics supports measures that extend beyond the optimization of the scheduling system. Definitions of key terms used in this dissertation are included at the end of this chapter.

1.1.1 Importance of Topic

Despite having the most expensive health care delivery system, the US ranks last overall among 11 industrialized countries on measures of health system quality (effectiveness, safety, coordination, patient-centeredness), access to care (costs, timeliness of care), efficiency, equity, healthy lives, and health expenditures per capita[28]. The US Department of Health and Human Services (HHS) through the Office of the National Coordinator for Health IT (ONC) has been making unprecedented efforts to put information technology to work in health care[29]. More importantly, the ACA has established comprehensive healthcare insurance reforms to increase access to healthcare, improve quality, lower costs, provide comprehensive care through PCMH[30]. Not only does the ACA rely on CHCs to provide care for the millions of newly insured patients, but also ties increased funding to demonstrated quality and efficiency[31].

Various federally funded agencies, such as AHRQ (Agency for Healthcare Research and Quality), NIH (National Institutes of Health), and PCORI (Patient-Centered Outcomes Research Institute) are fostering health IT innovations[32], with focus on such initiatives as Health Information Exchanges (HIE), Usability and system design for integration with clinic workflow, Big Data to Knowledge (BD2K) Initiative[33], and PCORnet (National Patient-Centered Clinical Research Network)[34].

For the nation to remain competitive, it is important to assess the impact of federal legislation and associated initiatives on access to health care, and ultimately the quality of care for the underserved (especially in the areas of preventive care and chronic disease management). AHRQ's National Healthcare Disparities Report (2013) examined the differences in quality of care among racial and ethnic groups, and brought to light the lack of a comprehensive set of data elements that can produce national and state estimates for healthcare quality (in terms of effectiveness of care, patient safety, timeliness of care, patient centeredness, and efficiency of care).[35] Notably, Indiana ranks in the lowest quartile in the quality of ambulatory care (about 55% of ambulatory care services are with primary care providers[36] and CHCs account for a sizeable portion of those visits), yet the state ranked in the highest quartile in terms of hospital care. This report underscores the value of improving the various the dimensions of care (outcomes, access, efficiency, equity) in primary care and community health center settings, in particular, given their increasing importance in the era of the ACA. Furthermore, increasing healthcare spending on primary care has been shown to pay for itself via resulting reductions in overall health spending[37].

Sound data-driven clinical and operational decision making plays a crucial role in improving the quality of care, yet challenges remain in the design, development, implementation, adoption and utilization of health IT solutions (of which EHR systems are the most visible component)[38]. Leveraging health IT solutions for improving clinical and operational performance often requires the creation of data repositories[39] since they reduce barriers in assessing and analyzing data in an efficient and cost-effective manner[40]. A data warehouse that can efficiently aggregate structured and unstructured data from multiple disparate data sources, including patient records, clinical quality scores, and payer data can be used to improve clinical and financial outcomes across the CHC practice and the region at large, as well as foster informatics and operations research[41,42]. Actionable discoveries through aggregation, statistical analysis and creative slice-and-dice of data is at the foundation of a data-driven business[43].

1.1.2 Defining Healthcare Access and Efficiency

Access to care and efficiency are important concepts within the context of this research and hence are explained in greater detail. As noted, additional terms used in this document have been defined at the end of this chapter.

Access to health care means having "the timely use of personal health services to achieve the best health outcomes"[44]. Attaining satisfactory access to care requires three discrete steps: Gaining entry into the health care system, getting access to sites of care where patients can receive needed services and finally finding providers who meet the needs of individual patients and with whom patients can develop a relationship based on mutual communication and trust[45]. A common way to measure health care access is to assess the presence or absence of specific resources that facilitate health care, such as having health insurance or a usual source of care.

Within the context of healthcare operational research, however, access to care is defined more in terms of assessments of how easily patients can gain access to needed health care. A few common operational measures related to this include: patient demand (for appointment), available supply (of appointment slots), patient wait time for appointments, no-show rates, office visit cycle time, and patient satisfaction[46]. Operational measures on patient access are influenced by the scheduling system in place. In a system of open access scheduling (also known as advanced access and same-day scheduling), most, if not all patients can receive an appointment slot on the day they call, almost always with their personal physician[47]. This model replaces the traditional approach of differentiating between urgent and routine appointments, which results in the routine visits being put off until a later date. The traditional scheduling approach also fosters high no-show rates, demand variability,[48] and often diverts frustrated patients to urgent care/ER facilities where they receive less personalized care at a much higher cost to the healthcare system[49].

Efficiency: Efficiency is the ability to avoid waste in producing a desired result. Efficiency (η) is often measured as the ratio of useful work output to total work input or $\eta = \frac{W_{out}}{W_{in}}$. In ambulatory healthcare operations, a closely related term is that of production efficiency or productivity, where the work output is usually the number of patients seen, but can also be measured as billed revenue and Relative Value Units (units to quantify the provider work, practice expense, and malpractice costs for specific services to appropriately establish payment)[50]. Work input is usually a time interval defined in hours, days or months.

1.2 Review of the Literature and Knowledge Gaps

A review of the literature and the knowledge gaps within the area of healthcare data analytics is described. The area is broken down into 3 sections namely, information needs and challenges, advanced analytics, and finally user-centered dashboards, particularly as applied to community health centers.

1.2.1 Information Needs and Challenges

The need to increase patients' access to health care, make the care better coordinated, and more effective and efficient is well established.[51] While federal incentives from meeting 'Meaningful Use' requirements[52], has led to an increased adoption of EHR systems, many technological challenges have come to the fore, including interoperability challenges and difficulty leveraging the large data sets for meaningful analytics.[53] While health centers continue to make progress using health IT systems, key challenges have become evident in organizing and validating data from EHRs and other IT systems. The urgency of responding to these challenges can often cause organizations to be reactive and fragmented in their approach to analytics, without the opportunity to first build a solid infrastructure for data management and use[54]. Therefore, although CHCs have made progress in adopting EHR systems, analyzing data from these systems remains complex and time consuming, limiting the ability of clinic leaders to draw meaningful insights[55]. Even with rapid adoption of advanced health IT at certain health centers, the availability of functionalities supporting performance measurement and feedback has lagged even further[56].

An understanding of the operational domain through a work and information needs analysis is at the very foundation of a multi-source data integration and analytics platform[57]. Specifically, there is a need to understand the facts, procedures, and processes involved in generating the data. A lack of this understanding makes it difficult to identify the data elements and navigate through the explanations in the data dictionary. Understanding the “why” and “how” a particular data element is collected or measured ensures that important information about the processes and systems that generated the original data is not lost in the integration process and that data summaries do not lead to poor conclusions.

Recent studies have shown that the impact of health IT on efficiency are mixed and that is attributed to sluggish adoption of EHR systems system coupled with poor interoperability, and the failure of organizations to integrate the technology into existing workflow or reengineer care processes to take full advantage of efficiencies offered by the health IT[58]. Some of proposed strategies for enhancing the benefits of EHRs include: mandating vendors to open-up their APIs (application program interfaces), incentivizing development of low-cost consumer informatics tools, and promoting congressional review of HIPAA (Health Insurance Portability and Accountability Act) to optimize the balance between data privacy and reuse[59].

While much of the focus of the last decade, via meaningful use incentives has been to encourage health care providers to implement EHRs and capture and share data important to quality and cost, the work ahead is to ensure that systems are designed and implemented in a way that yields the promised benefits to efficiency and quality[60]. A study assessing the impact on CHCs efficiency following EHR implementation showed that there were efficiency improvements tied to clinical process (e.g. ordering labs), but the impact from an operational perspective was not as well studied[61]. Elsewhere, CHCs lagged behind in the use of health IT for performance monitoring likely due to resource constraints for training and ongoing IT support[62]. A study on using EHRs and data warehouses collaboratively in CHCs have highlighted implementation challenges along with the need for big data augmentation[63].

This research intends to fill in the knowledge gap of understanding the unique information needs of CHCs in the post-HITECH era and then designing a CHC-centric information architecture aimed at supporting data-driven systems redesign initiatives in these settings.

1.2.2 Advanced Analytics

The combination of data warehousing, data mining and evidence-based medicine has commenced an innovative application field of information technology in the health care industry [64]. Since the early 2000s, studies became increasingly common on appointment scheduling systems in health care, particularly as some large primary care clinic systems began implementing open access scheduling[65]. Research has identified several factors that affect the performance of appointment scheduling systems, including: arrival and service time variability, patient and provider preferences, available information technology and the experience level of the scheduling staff[66]. Methods from industrial engineering and operations research (such as agent-based discrete event simulation modeling and stochastic models) have been applied in health care to inform outpatient scheduling[67].

The area of appointment scheduling has greater relevance in CHCs due to their unusually high no-show rates[68]. Predictive modeling of no-shows based on logistic regression has been widely performed to improve clinic scheduling systems in a number of care settings. For example, a pediatric clinic performed no-show modeling to deal with overbooking effectively[69]. The study showed that appointment no-show modeling has negative impacts that include underutilized medical resources, increased healthcare costs, decreased access to care, and reduced clinic efficiency and provider productivity. A study in a Veteran Affairs Medical Center performed no-show modeling to improve capacity utilization[70]. Studies at a primary care center demonstrated that no-show behavior was independently correlated with the following four factors: the patient's age and race, the presence of any physician-identified psychosocial problems, and the percent of non-cancelled appointments that were kept during the prior 12 months[71]. Another study in an academic practice site illustrated that no-show analysis was used to assess the effect of moving to a carve-out model of an advanced access scheduling system on no-show rates[72]. This specific study found that the lead time for an available appointment was not a significant factor in appointment keeping behavior and that implementation of the concepts of advanced access scheduling does not directly result in a reduction in no-show rates. Instead, provider-patient continuity, wait time to an appointment, and other factors may be a more important measures. Previous study also showed that implementing open access requires an overhaul of scheduling infrastructure, including decreasing appointment types, simplifying operational processes, using past appointment requests to predict future demand and developing contingency plans for times when demand exceeds prediction[73].

Interestingly, no-show prediction has also been applied in airline bookings to maximize revenue, where travelers' no-show and standby behavior was predicted using passenger and directional itinerary information[74]. Passenger-based predictive modeling of airline no-show rates was implemented using three different predictive methods: a C4.5 decision-tree, a segmented Naive Bayes algorithm, and a new aggregation method for an ensemble of probabilistic models[75]. In addition, there are studies that have applied text mining techniques like co-location/co-occurrence identification and topic modeling for various healthcare applications. These include: redundancies in EHR corpora[76], translation of medical summaries[77], semantic classification of consumer health content[78], and automatically labeling topics in drug-drug reactions[79].

The literature review calls for a need to develop patient-centered scheduling system that most commonly involve some form of open access scheduling, yet it is not known how to best design open access scheduling interventions for CHCs and in particular, how advanced data analytics can help. The study applies statistical and machine learning models on large scale clinic appointment data from multiple clinics to identify patterns of no shows and cancellations necessary to guide implementation of open access scheduling as well as shed light on patient behavior and preferences. Additional analytics is performed to assess efficiencies and provider work output. Finally, the knowledge discovery process is used to guide operationalization of a patient-centered scheduling system in community health centers.

1.2.3 User-Centered Dashboards

The concept of user-centered design is rather simple – taking the user into account every step of the way as a product is developed, but the implications of this simple concept are quite complex[80]. Within the context of healthcare, the user can be a patient and/or some healthcare personnel. Interestingly, the term 'patient-centered' has been in use since the 1970s[81]. If health care organizations seek to become patient-centered, they must create and nurture an environment in which their most important asset—their workforce—is valued and patients are engaged in all efforts to improve care. An important way to achieve this commitment and engagement is to involve employees and patients directly in the design and implementation of patient-centered processes[82].

The premise of a learning healthcare system (LHS) is based on data capture, analysis, interpretation and then delivery of tailored messages to individual decision makers in the health care practice before implementing the recommended changes[83]. Despite the importance of tailoring messages for decision makers, there is a lack of research on understanding the nature of customization for individual roles (providers, management and support staff) in health centers. Alignment of information systems with organizational objectives and strategies is a key management objective in health care organizations[84]. The vast majority healthcare organizations experience conflicts varying from minor disagreements to fully disruptive behavior among nurses, physicians, physician assistants and administrators[85]. There often are inherent structural tensions or conflict between physicians, mid-level providers (nurse practitioner/physician assistant), clinical support staff (nurses, patient care assistants), and administrative staff[86]. By understanding the unique information needs of different stakeholders, Information systems can be designed to support the core missions of the health center[87].

A recent systematic review of the current state of evidence for the use of clinical and quality dashboards in health care environments found that there is considerable heterogeneity in the design of dashboards and users targeted[88]. The review called for detailed research studies to obtain evidence of the efficacy and establish guidelines for the dashboard design. Several standardized measures have been developed to capture patient reported outcomes (such as symptom status, physical function, mental health, social function, and wellbeing) with a focus on improving the quality of care from the patient's perspective [89]. Evidence-based practice dashboards for nursing have also been evaluated[90]. For providers, user-centered design principles have been applied to develop graphical dashboards of healthcare questionnaire responses from patients. Performance measurement often use scorecards, dashboards, and key performance indicators (KPIs) and to effectively measure the same, providers need more than just a balanced scorecard. They must also be able to access performance data from a central data warehouse[91]. In particular, data-driven decision making in healthcare is particularly valuable when incorporating predictive measures (as opposed to simply historical or descriptive data)[92]. Studies have called for rapid and low cost impact assessment of interventions through health care management dashboards and recommend use of experienced evaluators and standard methods[93]. Finally, research calls for applying the emerging area of visual analytics (the science of analytical reasoning facilitated by advanced interactive visual interfaces[94]) as a means to address many of the information overload challenges when making sense of large multisource healthcare data sets[95].

In summary, there is a gap in knowledge when it comes to understanding what are the various performance measures of interest to various roles within a community health center. This is at the foundation of a LHS platform that tailors analytics to stakeholders. Furthermore, existing

CHC information solutions from EHRs and analytics vendors are more focused on providing mandatory reporting solutions as opposed to delivering dashboard visualizations that are designed, developed and implemented using user-centered principles. There is also a need for appropriate methods to assess the usability and efficacy of dashboards on decision making. In this research study, we apply usability principles and agile development methods to engage with stakeholders at community health centers to design, implement and assess visual analytics based dashboards to support operational performance monitoring and decision making.

1.2.4 Rationale of the Proposed Research

To be successful in the evolving era of healthcare, CHCs that provide care for the underserved populations must not only skillfully adopt EHR systems, but also use data in new ways to manage patient care and their clinic operations to realize improvements[96]. Review of the literature has highlighted the need for advanced predictive and visual analytics that is supported by an optimal information architecture at the back end and user-centered design principles on the front end to delivery role-specific measures of interest. The specific gaps in knowledge addressed by the dissertation include: information needs assessments within health centers, advanced analytics to support optimize operations and finally, development of user-centered dashboards to support system redesign initiatives. This research has sought to understand the clinic work and information flow needs that will guide the development of an analytics and visualization platform supported by robust backend data systems. It intends to address the unmet need for leveraging health IT solutions within CHCs to improve appointment access, efficiencies, and quality of care. It is centered on building a clinic data architecture that is augmented with big data technologies and clinic-specific analytics in order to enhance the operational and financial health of the community health care system. Here, analytics are applied to the use case of implementing open access scheduling. The research will focus on aiding CHC managers, providers and staff better understand the issues of patient access to appointments and impact of local interventions through visualization of data elements that reflect key quality measures.

This work carefully considers the unique information needs and workflow of CHCs, with the goal of supporting thoughtful planning and effective decision-making to redesign care. Data-driven dashboards tailored for operational decision making can better meet the needs of management, when backed by use of predictive analytics, apply web-based interactive visualizations and apply standard usability principles to gather design input as well as rapidly assess the learnability and usefulness of such tools.

1.3 Research Questions

The overall hypothesis driving this research proposal is that a user-centered data analytics platform improves operational decision making at health centers. The impact on operational decision making will be evaluated through improvements in targeted clinic performance measures and qualitative feedback from participating health centers.

Successful implementation of organizational interventions depends on multiple factors (notably organizational engagement and culture), as well as how thoroughly the organization has followed the implementation program. In order to be considered an evidence-based organizational intervention, it was decided that CHC managers must find the analytics platform to have content validity (the extent to which the measures represent all facets of the given construct) and support organizational buy-in. This is a necessary step to support the implementation of any intervention aimed at improving clinic performance. Optimizing open access scheduling is a specific use case for this analytics.

The proposal seeks to understand the information needs within CHCs, build a multidimensional data architecture and analytics platform responsive to the expressed needs, and lastly incorporate role-specific analytics into a user-centered interactive dashboard to serve as a novel decision support tool geared towards improving clinic operational performance.

There are 3 research questions (RQs) governing this research and each is associated with 2 sub-RQs, as outlined below. Hypotheses, where applicable (i.e. when the study uses quantitative methods) are also noted under each of the sub-RQs.

- RQ1: What are the operational information needs at CHCs and the optimal data architecture to support those needs? – *The Information Architecture*
    - What are the information needs to support operations at CHCs?
    - What is the optimal data architecture to support those needs?
- RQ2: How can clinic-specific analytics guide appointment access? – *The Data Analytics*
    - How can predictive analytics be used to optimize appointment scheduling?
        - Specific Hypothesis: The time dimension of appointments is independent of appointment adherence
    - How can analytics be used to better evaluate provider productivity?
        - Specific Hypothesis: Productivity measured as Visits/hour and RVUs are directly correlated

- RQ3: How can clinic-specific data be harnessed to effectively promote reflection and decision making among varying roles in community healthcare delivery settings? – *The Dashboard*
    - What role-specific measures are effective for performance monitoring in health centers?
    - What are the design and assessment needs for interactive dashboards of user-defined performance metrics?

The underlined words are used as labels to refer to the RQs/sub-RQs in the next section illustrating their inter-relationship.

1.3.1 Inter-relationship of the RQs

RQ1 is the foundational component of this research in that it aims to understand the information needs at CHCs and supporting those needs through an enterprise data warehouse (EDW). The EDW serves as a data source for clinic-specific analytics (RQ2) focused on optimizing clinic operations. The two primary operational data analytics tied to the use case of implementing open access scheduling are: predictive analytics on appointment adherence; and analyzing provider productivity. The clinic-specific data analytics also informs the design of the visualizations within the dashboard. The EDW serves as the back-end to the dashboard design and implementation (RQ3). This interrelation between the hypothesis, the RQs and the associated sub-RQs is illustrated in Figure 1.

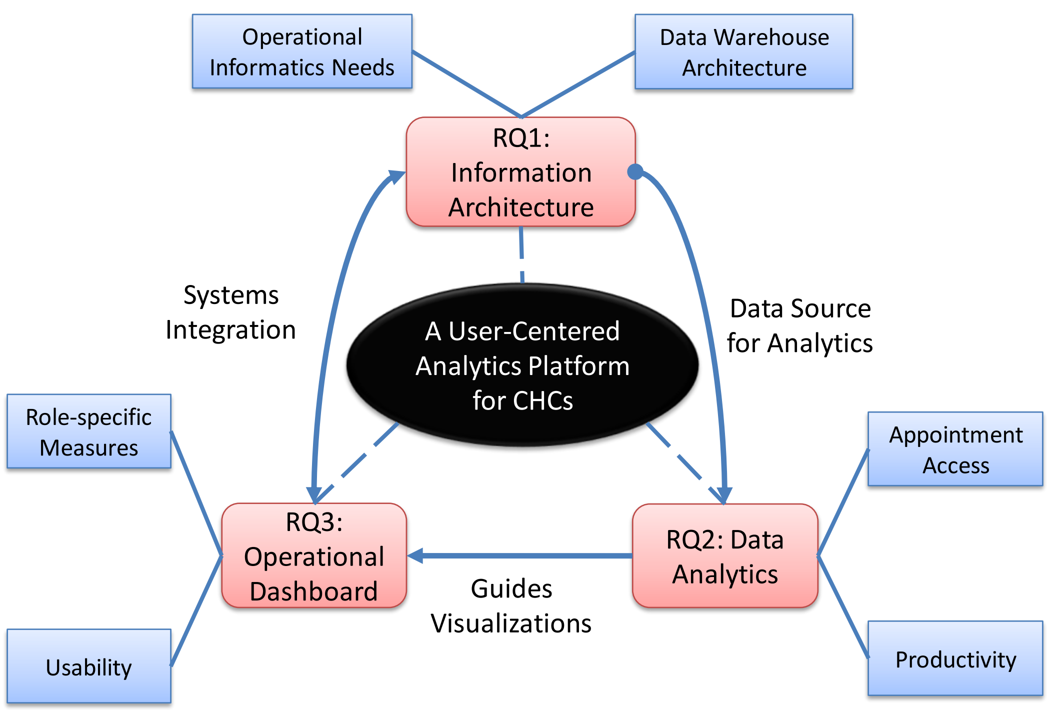

Figure 1: Connected Map of the RQs

1.3.2 Applying the Conceptual Framework of LHS

The Institute of Medicine (IoM)[97] has defined a 'Learning Healthcare System' (LHS) as a system in which, "science, informatics, incentives, and culture are aligned for continuous improvement and innovation, with best practices seamlessly embedded in the delivery process, and new knowledge captured as an integral by-product of the delivery experience"[98].

In this research, the LHS framework has been adapted for operationalizing a user-centered data analytics platform. The LHS has a virtuous cycle of 6 steps[99] that are shown in capitalized form: Plan/Gather data, Assemble Data, Analyze Data, Interpret Findings, Tailored Messaging, and Change Practice. The adaptation of the LHS cycle from a health IT driven operational optimization perspective has been referenced within brackets for each of the steps. Figure 2 illustrates the application of this conceptual framework for this research.

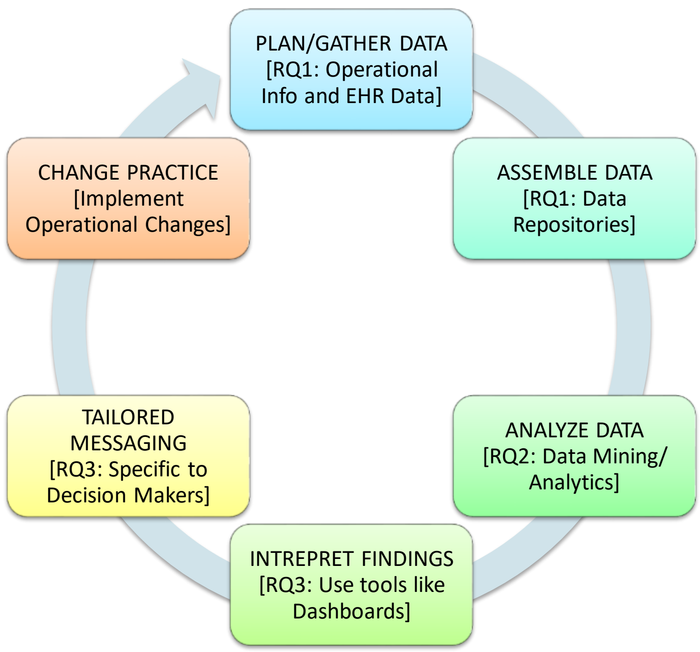

Figure 2: Framework of the Learning Healthcare System Cycle applied in this research

The first step in this research was to gather operational data from the CHCs on their information needs and challenges through questionnaires, interviews, process mapping, and data from the clinic EHRs and other operational/financial data stores. An efficient data repository/warehouse was then created to house collected data, followed by performing clinic-specific data mining/analytics and interpretation of the findings using tailored dashboard visualizations that expose the impact of implemented changes and need for additional change. The change in practice then feeds back into the cycle, creating a continuous learning system.

It should be noted here that the steps of gathering data through interpreting findings is made possible by recent IT innovations, but steps of tailoring messages specific to decisions makers and changing practice presents a challenge that requires the incorporation of methods from disciplines such as implementation science, human-computer interaction, and health economics[100].

1.4 Overview of the Subsequent Dissertation Chapters

The subsequent dissertation chapters 2, 3, 4 and 5, correspond to RQ 1, RQ 2, RQ 3, and closing remarks, respectively. Each of those chapters are written in the form an expanded journal article containing sections on introduction, literature review and identified gaps, methods used, results, discussion and conclusion. A brief summary of the research aim/hypothesis for each of the chapters along with the methods and results is now provided.

RQ 1/Chapter 2: The research question guiding this chapter is 'What are the operational information needs at CHCs and the optimal data architecture to support those needs?' referred to as the 'The Information Architecture'. The title of this chapter is: Information Needs at Community Clinics: Challenges and Solutions. The methods used to answer this RQ is primarily qualitative in nature and include: Use of a detailed CHC information questionnaire, rapid-ethnography-based workflow observations, and semi-structured interviews. The chapter maps the clinic workflows, describes the CHC information needs and challenges faced from a health IT perspective and designs and implements a data architecture tailored to the needs in CHCs that operate in a resource constrained environment. A data warehouse augmented with big data technologies has been populated with over 3.5 million visits records over six years from seven different CHC organizations using four different EHR systems. The chosen data infrastructure directly addressed the information challenges faced by the CHCs. This set the stage for clinic-specific data analytics to guide a particular intervention, aimed at improving access to appointments, described in Chapter 3.

RQ 2/Chapter 3: The research question guiding this chapter is 'What are key clinic-specific analytics needed to guide appointment access?' referred to as 'The Data Analytics'. The title of this chapter is: Large Scale Appointment Data Mining: Informing Patient-Centered Scheduling. The methods used to answer this RQ are quantitative in nature. There are two chosen use cases (the sub-RQs) that demonstrate the value of the analytics for appointment access, each having its own hypothesis. The first evaluates the value of predictive analytics to optimize appointment scheduling and the specific hypothesis is that the time dimension of appointments is independent of patient appointment adherence. The second sub-RQ is aimed at understanding the best approach to evaluate provider productivity and the specific hypothesis is that the productivity measured as visits/hour and RVUs (relative value units) are directly correlated. Following the steps of data extraction and scrubbing, statistical analysis, machine learning and visualization techniques are applied to evaluate the hypotheses. The results show that some aspects of the time dimension have a strong effect on patient appointment adherence, notably the lead time for appointment. The analysis also shows that the productivity assessments using RVUs are a better indicator of provider productivity. The information needs uncovered in Chapter 2 (RQ1) and the analytic results described in Chapter 3 (RQ2) lead to the need to find an effective means to present data analytics to a variety of user groups (RQ3), which is discussed in Chapter 4.

RQ 3/Chapter 4: The research question guiding this chapter is 'How can clinic-specific data be harnessed to effectively promote reflection and decision making among varying roles in community healthcare delivery settings?' referred to as 'The Dashboard'. The title of this chapter is: A User-Centered Dashboard for Care Delivery in Community Health Centers: Design, Development and Evaluation. The methods used to answer this RQ are both qualitative and quantitative in nature. The chapter describes the process of design, development, implementation and usability testing of an interactive web dashboard that is driven by the data warehouse created in RQ1. The design phase discovers role-specific measures of interest to the various CHC staff through the application of the card sorting technique and probing with a high-fidelity prototype dashboard. Following the development and deployment of the dashboard, a cognitive task analysis/walk-through and an enhanced system usability survey was conducted to assess the usability of the dashboard. Over 45 different measures are uncovered and dominant themes identified specific to various employee roles within CHCs. Survey results showed that users found the dashboard to be highly usable.

Closing Remarks/Chapter 5: The dissertation concludes with a summary of the results from the three RQs, scientific contribution to the field, future work, limitations, and final remarks.

1.5 Innovation

The process of healthcare delivery is creating volumes of data in large-scale organizational and clinical information systems. This calls for integrating informatics and health services research to understand the structure and function of information systems to use existing data effectively through advanced analytics as well develop new tools to collect data, with an aim to improve healthcare delivery in line with the LHS framework[101]. The novel aspects of this research are at the interface of informatics, health services research, implementation science, and systems redesign. In particular, the research lays out an innovative approach to foster a LHS-based framework in CHCs which have unique needs and barriers.

Multi-EHR Integrated Repository: This research has identified and validated a core set of clinical and practice management data elements from 4 different EHR data models, with one consisting of as many as 1900 tables and 31,000 column names. A method was developed that integrates the data model from these 4 EHR systems. This rich multi-CHC multi-EHR longitudinal data repository of over 3.5 million patient-visit records demonstrates a big data augmented community health information network to perform practice-based research. The warehouse architecture addresses the unique information needs of collaborating CHCs. This is the first time that a health center data warehouse has been architected based on aggregation of EHR and operational data to inform and assess the impact of iterative health systems redesign efforts. Notably, the Indiana University School of Informatics and Computing has invested in personnel to extend and support this resource to foster collaborative research and partnerships with the state-wide primary care association.

Advanced Clinic-specific Analytics: Implementing interventions that optimize appointment scheduling is a challenging system redesign proposition. This dissertation research makes use of advanced data analytic methods of relevance to CHC operations and applies a unique combination of machine learning, statistical analysis and visualization techniques to guide effective implementation and monitoring of appointment access.

Role-specific Measures of interest: While there is a well-documented need for business intelligence in healthcare and the benefit of dashboards (that include rich interactive visualizations customized to individual users) as a tool to meet this need, this research uncovers role-specific performance measures of interest within health centers and shows how user-centered interactive visual dashboards can be designed and implemented to support those measures.

Summarily, the combination of data warehousing, machine learning and visualization has commenced an innovative application field of information technology in the health care industry[102] and this research seeks to build an information infrastructure around that to offer support to each stage of the operational learning health system cycle as applied to CHCs.

1.6 Key Concepts and Definitions

Affordable Care Act[103]: The comprehensive health care reform law enacted in March, 2010. The law has 3 primary goals: Make affordable health insurance available to more people (The law provides consumers with subsidies ("premium tax credits") that lower costs for households with incomes between 100% and 400% of the federal poverty level); Expand the Medicaid program to cover all adults with income below 138% of the federal poverty level (Not all states have expanded their Medicaid programs); Support innovative medical care delivery methods designed to lower the costs of health care generally.

Acute Care[104]: Same as urgent care, is care for an illness, injury or condition serious enough that a reasonable person would seek care right away, but not so severe that it requires emergency room care.

Big Data[105]: Big data is a term for data sets that are so large or complex that traditional data processing applications are inadequate. Challenges include analysis, capture, data curation, search, sharing, storage, transfer, visualization, querying, updating and information privacy.

Chronic Disease Management[106]: An integrated care approach to managing illness which includes screenings, check-ups, monitoring and coordinating treatment, and patient education. It can improve the individual's quality of life while reducing health care costs associated with chronic diseases, by preventing or minimizing the effects of the disease.

Community Health Centers (CHCs)[107]: They are funded nonprofit health centers or clinics that serve medically underserved areas and populations. CHCs that received federal funding through Health Resources and Services Administration are called Federally Qualified Health Centers (FQHCs). FQHCs provide primary care services regardless of the individual's ability to pay.

Data Analytics[108] (also Data Mining): Health care analytics is a product category used in the marketing of business software and consulting services. It makes extensive use of data and statistical methods to perform explanatory and predictive modeling. Related fields are data mining, and knowledge discovery in databases (KDD). Data mining is the application of specific algorithms for extracting patterns from data and is a specific step within the larger KDD process[109].

Data Warehouse[110]: In computing, a data warehouse (DW), also known as an enterprise data warehouse (EDW), is a system used for reporting and data analysis, and is considered as a core component of the business intelligence environment. DWs are central repositories of integrated data from one or more disparate sources.

Decision Making [111] : Decision-making is the process of identifying and choosing alternatives based on the values and preferences of the decision-maker.

EHRs[112]: An electronic health record (EHR) is a digital version of a patient's paper chart. EHR systems (EHRs) are real-time, patient-centered records that make information available instantly and securely to authorized users. While EHRs do contain the medical and treatment histories of patients, they are built to go beyond standard clinical data collected in a provider's office and can be inclusive of a broader view of a patient's care. Modern EHR systems have native integration with Practice Management (PM) systems as well.

Federal Poverty Level (FPL)[113]: A measure of income issued every year by the Department of Health and Human Services. Federal poverty levels are used to determine an individual's eligibility for certain programs and benefits, including savings on Marketplace health insurance, and Medicaid and CHIP coverage. For e.g., per 2016 guidelines, it was $11,880 for individuals and $20,160 for a family of three.

Health IT[114]: Health information technology (health IT) makes it possible for health care providers to better manage patient care through secure use and sharing of health information. Health IT includes the use of electronic health records (EHRs) instead of paper medical records to maintain people's health information.

Information Architecture (IA): IA is a structural design of how all the components in information are integrated. It includes, the hardware, software, data marts and/or granular data warehouses[115]. The main components of IA are: plans, policies, principles, data models, standards, frameworks, technologies, organization, and processes. They are often equated with enterprise data architecture and data warehouse architecture. The Data Management Association (DAMA) posits a top-to-bottom hierarchy of knowledge, information, and data[116].

Information Needs: Information needs arise when people recognize a gap in their state of knowledge, that is, when they experience 'an anomalous state of knowledge' and wish to resolve that anomaly[117]. It is the need for information that individuals ought to have to do their job effectively or solve a problem satisfactorily. There is an implied value judgement in this – the meeting of need is beneficial or necessary to the person – and would be recognized as such. People do not have information needs per se, rather they experience a problem or difficulty or are under some pressure, and these cognitive and emotional needs may be met, by obtaining and then applying the appropriate information[118].

Medicaid[119]: Insurance program that provides free or low-cost health coverage to some low-income people, families and children, pregnant women, the elderly, and people with disabilities. Many states have expanded their Medicaid programs to cover all people below certain income levels.

Business informatics (BI) or Organizational Informatics[120]: Business Informatics (also known as Organizational Informatics) includes information technology, with a focus on the development of solutions for business problems. Information Systems has been said to have an "explanation-oriented" focus in contrast to the "solution-oriented" focus that dominates BI. BI researchers mainly focus on the creation of IT solutions for challenges they have observed or assumed.

Patient-Centered Outcomes Research[121]: Research that compares different medical treatments and interventions to provide evidence on which strategies are most effective in different populations and situations. The goal is to empower the patients and their doctors with additional information to make sound health care decisions.

Preventive Services[122]: Routine health care that includes screenings, check-ups, and patient counseling to prevent illnesses, disease, or other health problems.

Primary Care[123]: Health services that cover a range of prevention, wellness, and treatment for common illnesses. Primary care providers include doctors, nurses, nurse practitioners, and physician assistants. They often maintain long-term relationships with you and advise and treat you on a range of health related issues. They may also coordinate your care with specialists.

Primary Care Providers[124]: A physician (M.D. – Medical Doctor or D.O. – Doctor of Osteopathic Medicine), nurse practitioner, clinical nurse specialist or physician assistant, as allowed under state law, who provides, coordinates or helps a patient access a range of health care services.

Usability[125]: Usability is the ease of use and learnability of a human-made object such as a tool or device. In software engineering, usability is the degree to which a software can be used by specified consumers to achieve quantified objectives with effectiveness, efficiency, and satisfaction in a quantified context of use.

User-Centered Design[126]: The User-centered design (UCD) process outlines the phases throughout a design and development life-cycle all while focusing on gaining a deep understanding of who will be using the product. There are many variations of the UCD process. It can be incorporated into waterfall, agile, and other approaches. Depending on the needs, the user-centered design process is comprised of a number of methods and related tasks.

User Experience[127]: User experience (UX) focuses on having a deep understanding of users, what they need, what they value, their abilities, and also their limitations. It also takes into account the business goals and objectives of the group managing the project. UX best practices promote improving the quality of the user’s interaction with and perceptions of the product and any related services.

Workflow[128]: Workflow is the sequence of physical and mental tasks performed by various people within and between work environments. It can occur at several levels (one person, between people, across organizations) and can occur sequentially or simultaneously. Workflows may be viewed as one fundamental building block to be combined with other parts of an organization's structure such as information technology, teams, projects and hierarchies. For example, the workflow of ordering a medication includes communication between the provider and the patient, the provider's thought process, the physical action by the provider of writing a paper prescription or entering an electronic prescription into an electronic health record and transmitting the order electronically or having the patient take the prescription to the pharmacy to have the prescription filled.

CHAPTER 2: INFORMATION NEEDS AT COMMUNITY CLINICS: CHALLENGES AND SOLUTIONS

2.1 Introduction

A strong primary health care system can improve individual and population health outcomes while limiting the cost of health service provision[129]. Within the primary care sector, Community Health Centers (CHCs) function as 'safety net' providers of acute, chronic and preventive care services to low income and in medically underserved areas[130]. They have assumed greater importance with the passage of the Affordable Care Act (ACA), which has made healthcare more patient-centered, affordable and above all much more accessible to this newly insured patient population[131]. In 2015, there were 1375 federally qualified community health centers spread over 9800 neighborhood service sites that handle over 90 million patient-visits a year[132]. Effective and meaningful use of health information technology (IT) is vital to improving quality, efficiency, and reducing disparities in healthcare[133].

The National Healthcare Disparities Report examined the differences in quality of care among racial and ethnic groups, and brought to light the lack of a comprehensive set of data elements that can produce national and state estimates for healthcare quality (in terms of effectiveness of care, patient safety, timeliness of care, patient centeredness, and efficiency of care).[134] Notably, Indiana ranks in the lowest quartile in the quality of ambulatory care (about 55% of ambulatory care services are with primary care providers[135] and CHCs account for a sizeable portion of those visits), yet the state ranked in the highest quartile in terms of hospital care. This underscores the value of improving the various the dimensions of care (outcomes, access, efficiency, equity) in primary care delivered in CHC settings, in particular, given the large increase in patient volume accessing healthcare in the era of the ACA.

Sound data-driven clinical and operational decision making plays a crucial role in improving the quality of care, yet challenges remain in the design, development, adoption, implementation, and use of health IT solutions, of which electronic health records (EHRs) are the most visible component[136]. Leveraging health IT solutions for improving clinical and operational performance often requires the creation of data repositories[137] since they reduce the barriers to accessing and analyzing the data in an efficient and cost-effective manner[138]. A data warehouse that can efficiently aggregate structured and unstructured data from multiple disparate data sources – including patient records, clinical quality scores, and payer data – can be used to improve clinical and financial outcomes across the CHC practice as well as foster informatics and operations research[139,140]. Actionable discoveries through aggregation, statistical analysis and creative slicing-and-dicing of data is at the foundation of a data-driven business[141]. However, deriving knowledge from health care data requires an in-depth understanding of both the work and information flow within the source organizations[142].

Like all primary care clinics, CHCs also rely on EHRs as a means to capture clinical information and manage their practice, however, they differ in terms of their patient population. Figure 3 highlights the notable differences between the CHC patient population and the general US population in terms of the poverty level [e.g. 72% of CHC patients are at the Federal Poverty Level (FPL) vs. 15% for the US population], race [24% vs. 13% Black/African American], ethnicity [35% vs. 17% Hispanic], and insurance mix [42% vs. 16% Medicaid][143].

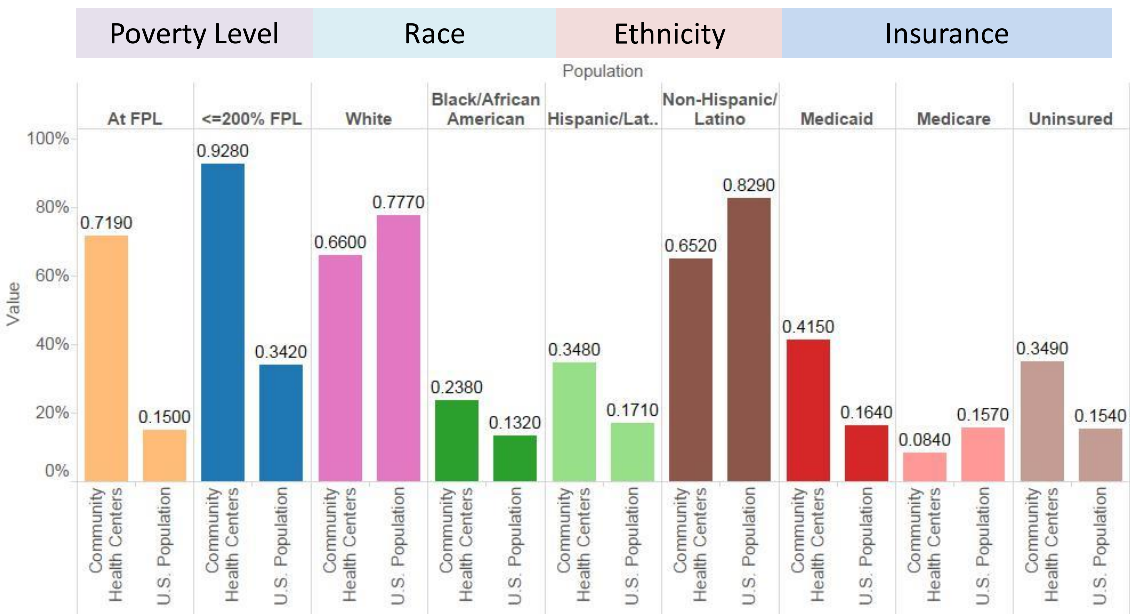

Figure 3: The CHC patient population in contrast to the general US population

The over representation of an indigent, uninsured and Medicaid patient population at CHCs affects their overall clinical operations, including their health IT. Notably, a disadvantageous payer mix at CHCs leads to average operating margins being less than 1%, which leaves them poorly equipped to make substantial capital investments (either in terms of facilities or health IT solutions)[144]. However, with federal funding support for health IT, the CHCs have seen great expansion in their implementation and use of EHRs since 2010[145], with over 92% of all CHCs now using EHRs.

Whereas federal incentives for meeting 'Meaningful Use requirements[146], has led to an increased adoption of EHRs across practices, many technological challenges have come to the fore. This is especially true when CHCs attempt to leverage large data sets for meaningful analytics, which requires organizing and validating data from EHRs and other health IT systems [147]. The urgency of responding to such a challenge can cause organizations to be reactive and fragmented in their approach to analytics, without the opportunity to first build a solid infrastructure for data management and use[148]. Even with rapid adoption of advanced health IT at certain health centers, the implementation has lagged for functionalities supporting performance measurement and reporting[149].

An understanding of the operational domain through a workflow and information needs analysis is at the very foundation of a multi-source data integration and analytics platform and in particular there is a need to understand the domain knowledge; the data model; and the data itself[150]. A lack of understanding of the facts, procedures, and processes involved in generating the data makes it difficult to identify the data elements and navigate through the explanations in the data dictionary (a collection of descriptions of the data objects or data models used in a particular database). Understanding the “why” and “how” a particular data element is collected or measured ensures that important information about the processes and systems that generated the original data is not lost in the integration process and that data summaries do not lead to faulty conclusions.

2.1.1 Knowledge Gaps

The need to improve access and efficiency in healthcare is well-established[151]. The effective utilization of health IT systems can support improvements in clinic operations, while yielding cost savings for health centers and payers alike[152]. Although CHCs have made progress in adopting EHR systems, the analysis of data from these systems remains complex and time consuming, limiting the ability of clinic leaders to draw meaningful insights[153]. CHCs, not unlike other health systems, are a potentially valuable source of clinical and administrative data. However, when the data is hard to extract or when it resides in multiple disparate silos, health systems cannot easily aggregate, analyze and assess the data in a cost-effective manner[154].

Systematic reviews have identified challenges faced by healthcare institutions to leverage their data for meaningful analytics. CHCs in particular face several resource barriers in this regard: general shortage of clinical and technologically skilled individuals in rural/underserved areas[155]; lack of integration into centralized health IT networks, or the scale necessary to justify set up of independent health IT infrastructure[156]; and shortage of CHC-specific vendor product offerings for analytics[157]. However, all these studies were done long before the implementation of the HITECH

Act in 2009 and in particular, the study [158] evaluating information needs for rural health 'professionals' was done in 2000.

The state of Indiana is among the few in the country that has not yet benefited from a Community Health Information Network, better known as Health Center Controlled Networks (HCCNs) – a network of primary care health centers that works towards achieving meaningful use of EHR systems and adopting technology-enabled quality improvement strategies, and engaging in health information exchange to strengthen the quality of care and improve patient health outcomes. This is in stark contrast to the fact that the state is home to the nation's largest health information exchange[159].

Lastly, while, there are multiple HCCNs in the nation[160], even those networks realize the need to augment their data warehouses with new technologies in the wake of the big data revolution[161]. It remains to be seen on how this can be implemented cost-effectively in resource constrained settings, such as CHCs. There is also lack of research identifying the resources needed at federally qualified health centers to benefit from this big data revolution[162].

The study intends to fill in the knowledge gap of understanding the unique information, analytic and resource needs of CHCs in the post-HITECH era and then designing a CHC-centric information architecture to address those needs for sound data-driven care delivery in these settings. The study also serves to propel forward the CHCs in Indiana, which is home to the largest HIE in the country, yet lacks an HCCN.

2.1.2 Understanding Information Needs and Information Architecture

Considering the importance of 'information needs' in this study, we start with defining what that is. The book, 'Assessing Information Needs: Tools, Techniques and Concepts for the Internet Age'[163] describes information needs as one that arises when a person recognizes a gap in his/her state of knowledge and wishes to resolve that gap. People do not have information needs per se, rather they experience a problem or difficulty, and these cognitive and emotional needs may be met, by obtaining and then applying that information. The authors point to four primary methods to yield the necessary high quality data to assess information needs: interviews, questionnaires, diaries, and observation. Of these, interviews are considered superior, particularly in-depth open-ended interviews and when combined with questionnaires and observations offer a fuller, richer and more trustworthy source of data.

Information Architecture (IA) is a structural design of how all the components in information are integrated. It includes, the hardware, software, data marts and/or granular data warehouses[164]. They are often equated with enterprise data architecture and data warehouse architecture, although it should be pointed out here that the Data Management Association (DAMA) posits a top-to-bottom hierarchy of knowledge, information, and data[165]. Specifically, data, "is the representation of facts as text, numbers, graphics, images, sound or video", while information is obtained by interpreting the context around data and finally, the resulting information then creates new knowledge that guides decision making.

2.1.3 Research Question

There are two research questions guiding this chapter. The first one is what are the information and analytic needs to support operations at CHCs, and secondly, what is the optimal data architecture to support those needs.

The present research study is an offshoot of a larger 3-year health systems redesign research project funded by the Patient Centered Outcomes Research Institute (PCORI)[166]. The larger project aims to improve efficiency and access to care for CHCs in the state by: identifying successful strategies to improve access; applying Delphi expert panel methods to enumerate and prioritize identify feasible and valid patient-centered, strategies for change; applying workflow and simulation modeling to develop clinic-centered recommendations for scheduling and access practice and policy changes; and finally, determining whether implementing patient-centered system of culture change and expanded open-access can improve access to care, relative to existing approaches. The focus of this chapter is on the health IT related needs of CHCs and how those needs are being met, which is critical to optimizing health care delivery.

2.2 Methods

2.2.1 Study Participants

Our research team partnered with 7 CHC systems representing 40 different sites geographically spread across the state of Indiana and serving over 150,000 patients each year. These CHCs were part of the larger collaborative project aimed at improving access to care. The CHCs operated in urban, semi-urban and rural settings and provide care to a diverse patient population in terms of race, ethnicity, insurance status, and income levels.

2.2.2 Specific Approach

The study took a novel approach to redesigning health systems based on clinic organizational information needs. In assessing the information needs for CHCs, the "positive deviance" framework – an established problem solving approach from complexity science[167] was applied. Positive deviants are individuals (or even organizations) who have an uncommon or unique approach to solving challenges and consistently demonstrate high performance in a given area of interest[168]. The positive deviant individuals are neither rebels nor do they disrupt their organizations, instead, they consider achieving the collective goal more valuable and important than conforming to organizational limitations and conventional wisdoms[169]. The 'positive deviant' individuals for this study were recommended by the clinic manager or the CHC-leadership.

The goal of this work was to understand the necessary information architecture at CHCs to support decisions about implementing change to improve clinic efficiency and access to care. Three complementary methods were used to accomplish this aim: 1) a clinic information questionnaire; 2) semi-structured interviews with key informants; and 3) workflow observations. Qualitative analysis of barriers and solutions in work and information flow was then performed to determine the appropriate information architecture.

2.2.2.1 Clinic Questionnaire

A 20-page CHC Information Questionnaire was developed that had detailed sets of questions on various aspects of the CHC operations. The questionnaire contained a mix of structured and open ended/descriptive questions and was adapted from multiple surveys compiled by AHRQ on workflow and IT assessments, and systems redesign[170]. The sections and major sub-sections within each questionnaire were: Clinic Information (name, site locations, services offered, health IT systems used, patient demographics); Service Information (appointment scheduling system, phone system, remote EHR access, policies for managing missed appointments, after hour calls/emails); Patient Visit Information (patient orientation, visit workflow, walk-ins, labs, shared/group visits, referrals, care-coordination, care teams, culturally appropriate services); Enrollment Practices (assistance with enrollment, best practices, problems, barriers); Quality and Performance Measures (measures tracked, how is it done and the role of IT). The instrument was drafted and refined by the study investigators. The questionnaire for each of the seven CHCs was filled out by the clinic site manager.

Data from the clinic questionnaire was collated across all seven CHCs and the notable characteristics were extracted and summarized based on an internal group consensus.

2.2.2.2 Clinic Staff Interviews

The next step was to understand the staff perspectives on the care delivery process, followed by workflow observations. There were two sets of interviews conducted. The first set of interviews was aimed at understanding the staff barriers to improving efficiency and access to care at CHCs and what specific strategies helped overcome some of those barriers. Within the context of these interview, there were challenges identified related to the use of health IT – the focus of the current study. The authors interviewed staff (N = 49) that included providers, administrators, nurses (RN and LPN), medical assistants, schedulers, front desk staff, social workers, care navigators and billing coordinators. All interviews were audio-recorded. The interview data were then transcribed and analyzed using a thematic and iterative approach. To increase the reliability of the coding process, a team of four researchers independently analyzed the same interview in order to categorize the responses, and repeated the process with new set of interviews until the codebook reached stability. Each of the interviews was then individually analyzed using the final codebook imported into a qualitative data analysis software, NVivo 10 from QSR International. For this particular study, only the codes labelled as 'health IT' were examined.

In addition to this first set of interviews, there was a second set of interviews specifically related to health IT. These key informant interviews of CHC staff were designed to characterize and enumerate barriers from a health IT perspective. The interviews were conducted face-to-face with at least two staff members ('the positive deviants') within each CHC. In-depth semi-structured interviews (N = 16) were carried with key informants (CEO, Medical Directors, Clinic Manager, IT director, Scheduler, QA manager) from at all seven CHCs. The topics explored in the interviews were guided by the aims of the study and implicitly focused on health IT-related needs, barriers, and facilitators. Unlike the first set of interviews wherein a formal thematic content analysis was performed, in this round of interviews, we reviewed and analyzed the collection of interview transcripts with the goal of identifying all issues centered around health IT.

2.2.2.3 Clinic Workflow Observations

Workflow observations was a multi-member effort, where individuals from the research team were assigned to observe and record various appointment related functions in a CHC. These included: front office, patient registration/check-in, phone triaging, scheduling, enrollment, referrals, labs, and check-out. The clinical aspects (e.g. 'intake' by Medical Assistant and 'exam' by physicians, physician assistants or nurse practitioners) were demonstrated by the staff, but not directly observed during the actual patient encounter to preserve confidentiality. AHRQ has developed a toolkit for 'Workflow Assessment in Health IT'[171], which was leveraged to capture the workflow information. This is a 5 step process: Step 1: Decide what processes to examine; Step 2: Create a preliminary flowchart; Step 3: Add detail to the flowchart and collect artifacts; Step 4: Determine who needs to be observed; Step 5: Observe and refine the flow charts.

A rapid ethnographic approach was used to probe into the work and information flow process[172]. Ethnography, which is 'on the ground' detailed observation and documentation, is considered rapid when there is a triangulation of the findings to ensure that they are accelerated, comprehensive and well developed. Typically, these data collection efforts occur over a few days, compared to the longer period for traditional ethnography. The workflow observations described above, as well as key informant interviews are considered part of the rapid ethnography. The following types of triangulation was done[173,174]: Methods triangulation (by using qualitative data from the workflow observations with the quantitative data included with the questionnaire); triangulation of sources (by comparing observations from different employees in the clinic); observer triangulation (using multiple observers for same processes). The goal was not to seek consensus, but to understand multiple ways of seeing the observational data.

The entire study, including the data collection from the clinics, was IRB approved by the Indiana University Office of Research Administration. Written documentation of these agreements, policies and approvals were shared with each partner prior to any data sharing.

2.3 Results

For this study, the results presented are focused on the health IT aspects tied to CHC operations. What follows are the related results from the clinic questionnaire, the themes from the semi-structured interviews and the findings from the work and information flow modeling. All of this informed the design of an information architecture, the core component of which is a multi-CHC data warehouse.

2.3.1 Clinic Survey Results

Four different EHRs were encountered at the seven CHCs based on different technology architectures: Client-Server based EHRs (GE Centricity, and NextGen) and Cloud-based EHRs (eClinicalWorks, and AthenaHealth). These four EHRs fall in the top 10 most popular ambulatory EHRs in the market. Two out of the seven CHCs were contemplating changing their EHR, while one made the switch in mid-2015. All seven CHCs had been with an EHR for at least 5 years.

Achieving PCMH status was one of the policy and financial initiatives that influenced both clinic operations and the adoption of health IT. One of the underlying principles for PCMH is enhanced access, i.e. making care available through systems such as open access scheduling and expanded hours. Table 1 provides information on the participating CHC systems, number of patients served, the number of physical service delivery sites, and the appointment scheduling method currently in place. We found that all of the CHC systems in our collaborative that are listed as using the traditional appointment scheduling method were actually in the midst of providing more hours to accommodate same-day/walk-in patients.

CHC System	Patients/Year	Sites	HER used	Scheduling Type
CHC-S	7,324	1	eClinicalWorks	Modified Open Access
CHC-H	59,286	12	eClinicalWorks	Modified Open Access
CHC-O	20,703	9	Athena Health	Modified Open Access
CHC-N	13,002	3	NextGen	Traditional
CHC-R	42,092	8	NextGen	Traditional
CHC-P	4,764	2	GE Centricity	Traditional
CHC-V	7,525	5	GE Centricity	Traditional
Totals	**154,696**	**42**		

Table 1: Select operational data on Participating Community Health Care Systems (2015)

The data from the EHR and scheduling systems were pulled using one of three methods depending on the reporting capabilities of the EHR used by each CHC and our needs. These data extraction methods included: Access to Reporting Module [CHC-S, CHC-O]; Direct Access to EHR Database [CHC-P]; and SQL-based Reporting [CHC-H, CHC-N, CHC-R, CHC-V].

The types of specialty services (beyond adult family medicine) offered at the CHCs is shown in Figure 4. All seven CHCs provided pediatrics and OB/GYN care, while only 3 of the seven provided pulmonary care, smoking cessation and podiatry care.

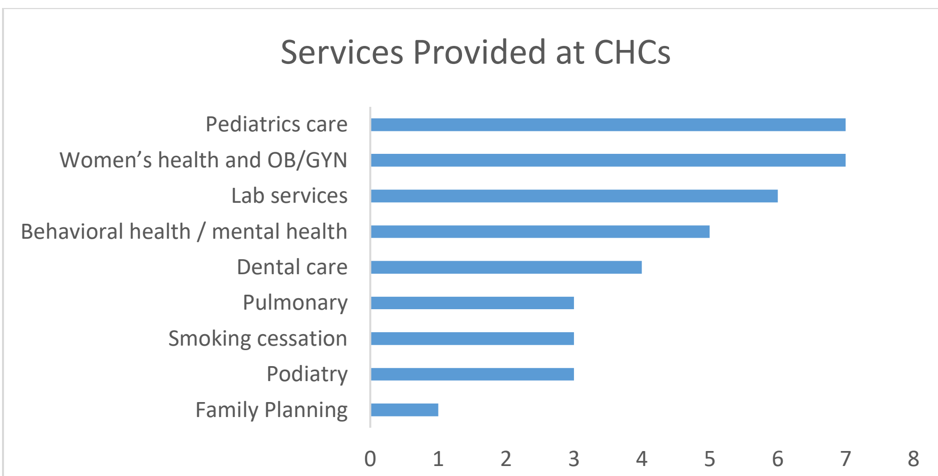

Figure 4: Availability of specialty care and services at the participating CHC Systems (N=7).

In terms of scheduling appointments, all seven CHC systems offered phone calls and in-person visits, referrals from hospitals and patient portals. All clinics provided advice over phone and have enabled provider access to EHR remotely. Access modalities are shown in Figure 5.

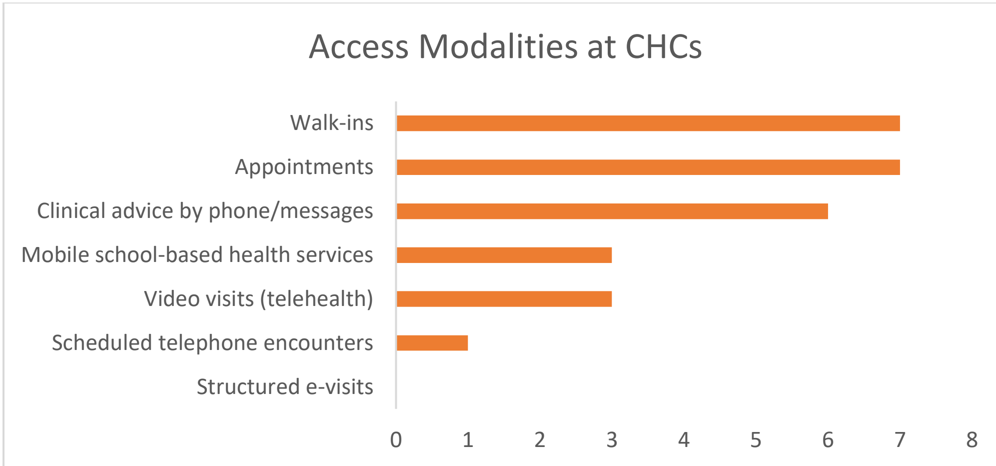

Figure 5. Modes of access to care at the seven CHC Systems in Indiana

2.3.2 Information Needs and Challenges

From the first set of interviews looking at the broader area of improving efficiency and access to care, there was a specific theme related to health IT. It was labeled as 'challenges with use of health IT'. There were other themes/sub-themes related to technology as well, but those were from the patient's perspective and have been in a separate research paper[175]. The issues uncovered from the specific theme on health IT were combined with the subsequent set of key informant interviews. The second set of interviews, were focused on the implementation and utilization of health IT in the CHCs. The following 5 themes and key associated sub-themes (in brackets) were identified from this combined set of interviews: 1) Personnel Constraints (training needs, time consumption); 2) Reporting Challenges (system deficiencies, inaccuracies); 3) Cost of Ownership (EHR purchase/set-up costs; ongoing maintenance/upgrade costs); 4) Workflow Management (tracking patient flow, resolving open/pending tasks); 5) System Limitations (poor interoperability, data migration).

Each of these themes are further described below with excerpts from the interview transcripts that best highlighted the issue. The names of the EHR systems used have been redacted.

Personnel Resource Constraints: All seven CHCs reported facing varying levels of personnel related resource constraints that was hampering their ability to leverage health IT for their operational and clinical quality improvement initiatives. Two of the largest sized CHCs had the economies of scale to sustain an internal IT department with at least 2 FTE openings. However, they too faced challenges in hiring, training, and retention of skilled employees. The following excerpts illustrate these challenges.

CHC COO: We've gotten so big so quick and so, high on our priority list is we are in search of an IT team…Unfortunately, we still do a lot of manual reporting work. And I know we can do canned reports on [EHR Name-1] and some we can build, but we don't have a person that has the experience or even has enough time that can focus on that. That's probably our frustration.
CHC IT Manager: We know we can run many reports in [EHR Name-3], but we need training on how to use them!
CHC QA Manager: None of us in the CHC know how to use the reporting system within [EHR Name-4]. We are only able to use it for the annual UDS [Uniform Data System] reports [for the Federal Bureau of Primary Healthcare]. [EHR Name-4] training is very expensive!

Reporting Challenges: Closely intertwined with resource constraints is the ability to run needed reports using the EHR data. This was one of the most important challenges at CHCs, considering how critical reporting is for CHCs, be it for mandatory federal grant funding purposes or making operational improvements. There are issues with reports that are not accurate or representative of the CHC needs. Additionally, a large amount of time is spent on using the EHR in a meaningful way.

CHC QA Manager: [EHR Name-4] has some canned reports [like one that reports productivity] but the calculation is not how we want.
CHC COO: When you're in [EHR Name-1], every day is so frustrating. When you're drilling down to know how many hypertensives we have, well it's in the EHR, but trying to pull a report that shows us this, umm, that's probably one of our biggest weaknesses right now.
CHC Operations Manager: We are not able to run the [appointment] no-show reports using the [EHR Name-3] report builder. Their canned productivity report is not correct and they have not been able to fix it for us. We are having to calculate that manually each month…. We got a low score in our children and teens BMI percentile quality measure [reported to HRSA/UDS]. Our population health management tool cannot pull that data since the percentile value is not stored as structured field within [EHR Name-3].
*CHC Clinic Supervisor on Time Spent on the EHR: Our EHR s***. I'll tell you that, I say it all the time. My boss told me that no matter how much it s*** we're not getting rid of it.….The biggest problem we have is with our time [we spend on it].*

Cost: The primary concern with CHCs here relates to the high upfront cost with the purchase and set-up of the new EHRs. Without federal funding these would have been hard to achieve, but the ongoing maintenance and upgrade costs (for newer modules or compliance with new federal requirements) takes a hit on the CHC budget. These issues have led to creation of new pricing models (with EHR Name-3], such as charging CHCs as a percentage of their billing volume, provided the CHC uses the EHR's claims processing services.

CHC QA Manager: We are paying a lot of money for [EHR Name-4]. We are evaluating alternatives. We like [EHR Name-3] because they charge us as a % of our practice revenue while using their billing services.

CHC COO on the EHR Costs: [EHR Name-1] will get you for every dime that you have! I don't even want to begin to tell you how much money we've had to spend on [EHR Name-1]….Everything in [EHR Name-1] is a cost. And so you can't get to the next step without purchasing an interface to get to the Meaningful Use requirements.

Workflow: Every step of the workflow needs to move seamlessly and in particular, the patient flow from arrival/check-in through departure/check-out flow, so that providers can operate efficiently and maximize time with patients. This is predicated upon automated tracking and reporting of the operational workflow. However, this remains a challenge with most EHRs and may often require the use of additional tools, leading to increased costs and redundancies. Interestingly, while EHR Name-3 is known to have this capability the CHC which uses that EHR is not able to process that particular report.

CHC-V Clinic Manager: There's no way in our [EHR Name-2] to track the in-between times [of the patient flow through the clinic]. One of the great things I have heard about [EHR Name-3], that we are looking at, other than being cloud-based which I kinda like, but also has its faults, is that they can get down in the weeds and track all this [patient flow]....so you can really wrap your brain around what's actually going on [in the clinic].

CHC-O Clinic Supervisor: With [EHR Name-3] I have to [manually] go through every single patient [to track their flow]. I do an audit on primary care to see the patient flow, particularly when I get complaints from patients.

System Limitations: Despite the rapid increase in EHR adoption, interoperability between various health IT vendors is very limited. The large EHR vendors in particular are resistant to implement efficient data interchanges due to competitive reasons and lack of federal mandates. Poor interoperability may also occur between subsystems of a same software product when those sub-systems are often brought into the vendor's product fold through acquisitions. Such is the case with EHR Name-2 vendor, which had acquired the practice management solution and later integrated that into their clinical solution. When CHCs change vendors (which is not uncommon), the issue of data migration becomes important.

CHC-V Clinic Manager: We use [EHR Name-2] for clinical and for practice management. They're, 2 separate systems [from the same company] and they're supposed to talk, sometimes they do, sometimes they don't. The [EHR Name-2] didn't make [the PM solution], they acquired it so it's got [EHR Name-2]'s name on it and they say that they work as one, but they don't!

CHC-O IT Manager: We moved from [EHR Name-2] to [the cloud-based] [EHR Name-3]….The version upgrade of [client-server based] [EHR Name-2] was causing many issues which prompted this change to [EHR Name-3]….Not all info was migrated into [EHR Name-3]. All of the structured data migration did not take happen as well as promised. It is only available as an 'archive'….[On the positive front] we like the national level bench marking available within [EHR Name-3].

The interview excerpts are representative of concerns we found throughout the seven CHCs participating in the study, irrespective of the EHR system they use, in terms of being able to leverage their data for efficient and cost effective analytics. Only one out of the seven CHCs in our study had a full time data analyst to support the organizational reporting/analytic needs. Not surprisingly, this clinic was also part of the largest CHC system in the state catering to over 60,000 patients across more than 10 sites in the Indianapolis metro area.

Most CHCs interviewed, expressed keen interest in being able to effectively draw insights from their data from their health IT systems, even as they encountered challenges in doing so using their existing EHR systems and reporting solutions. The key informant interview process highlighted several challenges in getting access to information and generating reports to support the care delivery process. One positive is that they have all benefited from investments in their ONC-certified EHRs using federal incentive dollars that started flowing in since the passage of the HITECH act of 2009 (that directly awarded health centers to support the purchase of health IT). The challenge that remains is leveraging the data in an effective and meaningful way to support improvements in clinical and operational outcomes.

2.3.3 Work and Information Flow Modeling

A generalized work and information flow map based on observations at each of the seven CHCs is shown in Figure 6. Referring to the figure, the process starts by either an incoming phone/email or a scheduled patient visit and the work and information flow path is dependent on the patient's reason for contacting the clinic [such as a request for information/medication refill, lab request or provider visit]. The flow path is different from established vs. new patients and acuity of the visit.

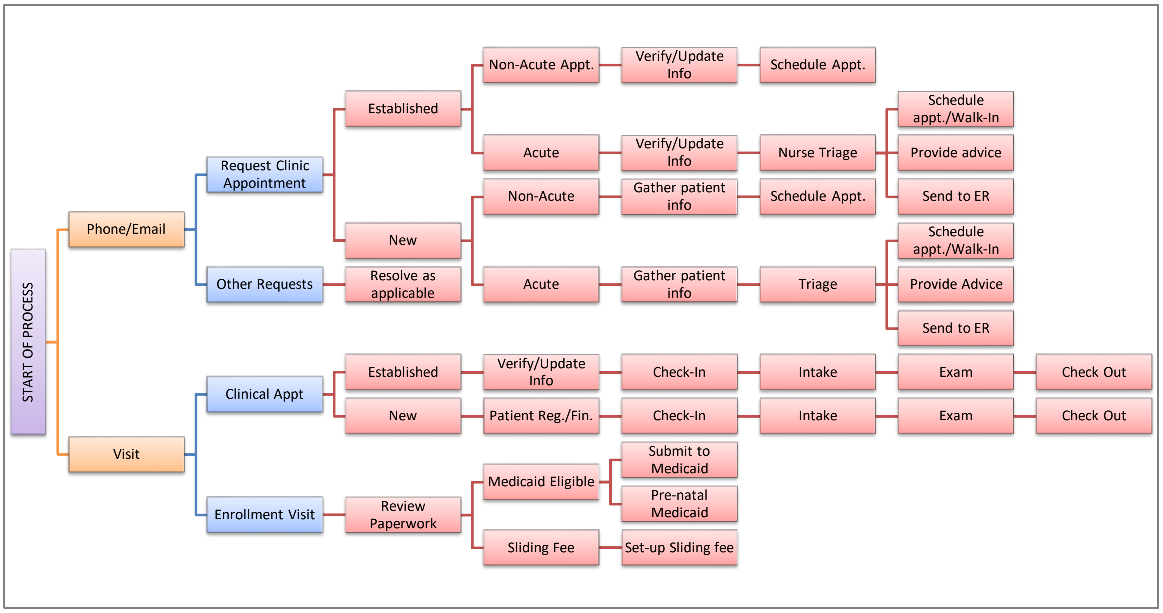

Figure 6: Generalized Work and Information Flow Possibilities within CHCs

The flow path provides an understanding of the workflow, which then makes it possible to associate that with the information needs. EHR data is directly associated with all of the activities shown in the workflow map for either information retrieval or documentation. This understanding of the domain guides the development of the information architecture described in the next section.

2.3.4 Summary of the Information Needs

The information needs assessment gathered through the clinic information tool, the key informant interviews and the workflow observations warranted the design and implementation of an appropriate information architecture. The information architecture refers to the manner in which the components of the information system are organized and integrated and is often synonymous with enterprise data architecture and data warehouse architecture.

There were various challenges uncovered through the key informant interviews, which were inhibiting effective utilization of health IT in the CHCs. These centered around: personnel constraints; cost of ownership of health IT systems; reporting challenges; workflow management; system limitations in terms of interoperability and data migration. Each of these challenges played a part in the design of the information architecture as described in the next section.

2.3.5 The Clinic Data Warehouse Architecture

A common and widely accepted technical definition of a data warehouse is that it is a ‘subject-oriented, integrated, non-volatile, and time-variant collection of data to support decision making'[176]. Data warehousing methodologies share a common set of tasks, including business requirements analysis, data design, architectural design, implementation and deployment[177]. A data warehouse contains only alpha-numeric data, excluding documents or other types of content. The data is stored separately from the organization's primary applications such as EHR/PM systems. The data is transformed to match a uniform data model, cleansed of duplicates and inaccuracies, and is extracted with business intelligence and reporting tools. A data warehouse contains the entire scope of data and can be used for both very general and very specific analysis and reporting.[178] Being a centralized store of data generated by multiple departments/sources, it is specially designed for data analysis, generating reports, and for other ad-hoc queries and turning them into meaningful data, so business decisions can be made based on this information. The major elements of developing any data warehouse are: planning; design of the data model; deployment of the technical infrastructure; and testing.

The general themes uncovered from the information needs assessments and its impact on the information architecture and specifically a centralized data warehouse platform for multiple participating health centers, is now described.

Personnel Constraints: A data warehousing platform for any organization requires multiple skill sets. Common roles include that of a project manager, database administrator, technical architect, ETL developer, front-end developer, trainer, data modeler, and QA. Since these are roles, a person may often take on multiple roles. While it is difficult to find people with extensive data warehousing experience, the situation is worse in non-metropolitan areas where most of the health centers are located. In such a situation there is a need to develop a data warehousing platform that is not constrained by such resources.

Reporting Challenges: The primary goal of a health center like all other organizations, is to make better decisions, be it financial, operational or clinical. A data warehouse enables better decision-making since the latter stores credible facts and statistics, and decision makers will be able to quickly retrieve that information based on their needs from a single point of truth as opposed to fragmented datasets. When data is standardized, there is also greater confidence in its accuracy. Without a data warehousing architecture, an information system that relies on using disparate data sources for analytics and combining them as necessary, would contain any number and types of data stores and analytics software. Such a system is complex and difficult to monitor, track or manage in an effective manner.

Cost of Ownership: Beyond the upfront cost of an EHR license and implementation and the maintenance costs, health centers dislike the idea of paying more for custom reporting and analytics. A multi-CHC centralized data warehouse containing normalized data for clinical and operational reporting is much more cost-effective than each individual health center having to work with their health IT vendor for custom reporting and analytics.

Workflow Management: Tracking and streamlining patient flow from check-in through check-out through a health center is critical both from an operational efficiency perspective as well as improving patient satisfaction. Additionally, from a workflow perspective it is important that users and consumers of health IT systems be able to use those systems effectively as well as identify and resolve any pending action items promptly. Appropriate data reporting, visualizations and analytic capabilities from a centralized data warehouse platform can support workflow management and make up for deficiencies of source health IT systems.

System Limitations: An appropriately designed information architecture can support information exchange between disparate software systems, be in different EHR systems or even external organizations like regional health information exchanges. When health centers change EHRs, data migration becomes a critical issue to be dealt with. A data warehouse platform presupposes knowledge of the EHR database structure and dictionary, which can support data migration related decision making and activities (such as data mapping/conversions, chart abstraction, etc.).

Given that CHC organizations have limited internal IT resources to dedicate to the creation and management of an effective warehousing system, they stand to gain from an architecture that benefits from economies of scale, yet is responsive to stakeholder information needs - as implemented in this study. The selected IA overcomes the challenges related to accessing and normalizing multi EHR systems data, generating operational performance reports, and performing associated analytics on data extracts. These challenges were magnified at smaller CHCs that lacked or could not afford the necessary IT resources for implementation, maintenance and upkeep. At the same time the data infrastructure had to be non-complex, efficient, adaptive, scalable, and benefited from a centralized resource allocation, yet provided the flexibility to perform onsite reporting and analytics by CHC staff. The organization and integration of data is achieved through working closely with the health centers in understanding their work flow and data modeling of their EHRs. The IA based on the information needs of the CHCs is now described.

2.3.6 Data Flow

We identified two primary categories of data that would directly be able to fulfill the health IT needs of the participating health centers. This includes: 1) Patient Level Data from integrated EHR/PM (Practice Management) Systems - the primary data source; and 2) Aggregated Clinic-level Data on Quality and Operations from Uniform Data Systems and IRS form 990. These two data categories are captured and integrated into a Microsoft SQL server based data warehouse augmented with Apache Hadoop based big data technologies. There is a data presentation layer where the data normalization occurs before any reporting and analytics is performed. This information flow diagram for the CDW is illustrated in Figure 7.

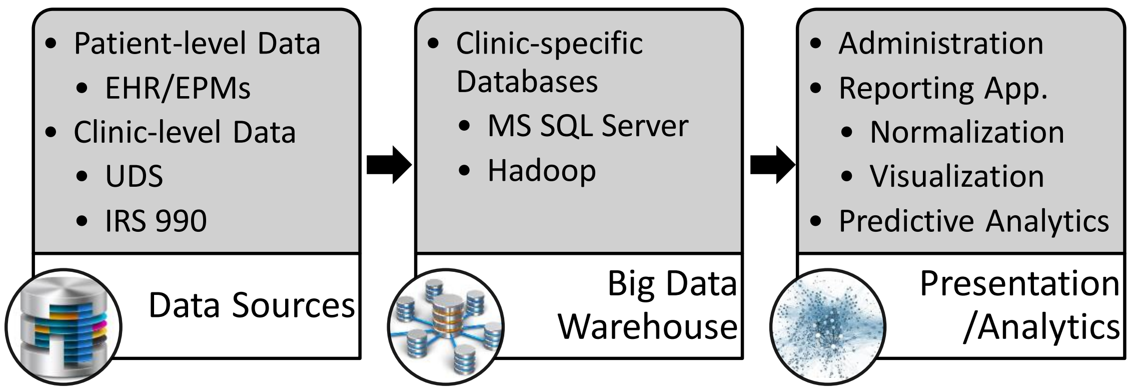

Figure 7: Information Flow Diagram for the Clinic Data Warehouse

The design and conceptualization of the CDW was based on the requirements gathered from all the stakeholders involved through an extensive multi-method needs assessment phase. While the initial choice of the data elements was driven by a narrow purpose to model and simulate the clinic's scheduling system, it soon expanded in scope to serve as a multidimensional CDW sourcing information from various domains. The objective was to understand the implications of the clinic operational characteristics on the clinic performance, particularly access to care, over time. For example, the CDW was used for evaluating the impact of specific scheduling methods on appointment adherence (no-shows, cancellations, reschedules), resource utilization, productivity, and various patient outcome measures.

The system architecture was defined by a subgroup of the research investigators in terms of: data sources, data model, interoperability requirements, data standards, data acquisition technologies, modes of acquisition, storage in a database server, privacy and security standards set up, data validity checks, data back-up performance and potential to build tools for data management and visualization. Roles and responsibilities were set for each of the above tasks, along with defining the business model. The policies and procedures were written, reviewed by CHC IT staff, leadership, legal counsel and revised. These policies and procedures were then developed into a written data sharing agreement and policies, which were signed by CHC leadership and Indiana University research leadership.

2.3.7 Obtaining Clinic and Patient Level EHR Data

The clinic financial and operational data categories include: operational data (clinic type, certification status, locations, size, patient mix, payer mix, service mix, staff utilization, scheduling system in use, enrollment system, quality measures and other dimensions of care, data from phone systems); and financial data (revenue, expense, margins, cash flow, payroll). The operational and financial data was obtained from the audited financial statements of the clinics, IRS Form 990 (annual returns for certain federally tax-exempt organizations) and Uniform Data Systems (UDS) filings. UDS reporting is done annually by all HRSA-funded health centers on a core set of information, including data on patient demographics, services provided, clinical indicators, utilization rates, costs, and revenues.

It was the patient level data that served as a backbone of the data warehouse. The data elements selected are those that are digitally captured as part of routine clinical practice in the EHR and reported to CMS. Collecting this data required the development of a core data model, given that information was being collected information from different healthcare facilities using different Patient Management and EHR systems. At a minimum the data elements were based off of the 'common Meaningful Use data set'. The data model incorporated elements and practices of those used by other large organizations, including: HL7's C-CDA[179], PCORnet[180], and FDA's Sentinel Initiative[181]. The data model helped promote efficient and streamlined collection of the data generated in each of the healthcare centers by organizing data into a standard structure.

The core data element categories included: *patient demographics, appointment data, date/time stamps of key events for patient and data flow, insurance information, provider information, medical problem list, encounter diagnoses, immunizations, medications, laboratory data, care plan, procedures, and healthcare referral information.*

While multiple ways to acquire patient data from the EHR were explored, the one ultimately selected was dependent on the EHR reporting capabilities, health IT resources in the health center, the data architecture of the CDW and the preferences of the CHC partners. The most efficient method involved obtaining the data dictionary/metadata (tables, column names, definition/meaning, data type, predefined values sets/descriptive text) of the relevant electronic datasets in place at each health center where possible, and identify the various data elements of our interest. The data elements were selected from four different EHR data models, with one consisting of as many as 1900 tables and 31,000 column names. SQL queries were then generated for each of the individual electronic datasets and given to the clinic staff to execute them on their database server. It was necessary to work closely with each CHC IT staff to extract the requested data, link and conduct de-identification procedures outlined in the project plan. De-identification was then verified in accordance with HIPAA safe-harbor provisions[182] by health center IT staff. The center then shared the data through a HIPAA compliant secure online healthcare data portal. The raw data was prepped programmatically before populating into the Microsoft SQL Database Server of the CDW making it available for ad-hoc querying and reporting. This eventual process required much discussion, revision, and assessment of risks with the various parties to ensure that all reasonable efforts were made to protect the confidentiality of the data. Figure 8 shows this process flow of data collection from the clinic EHRs.

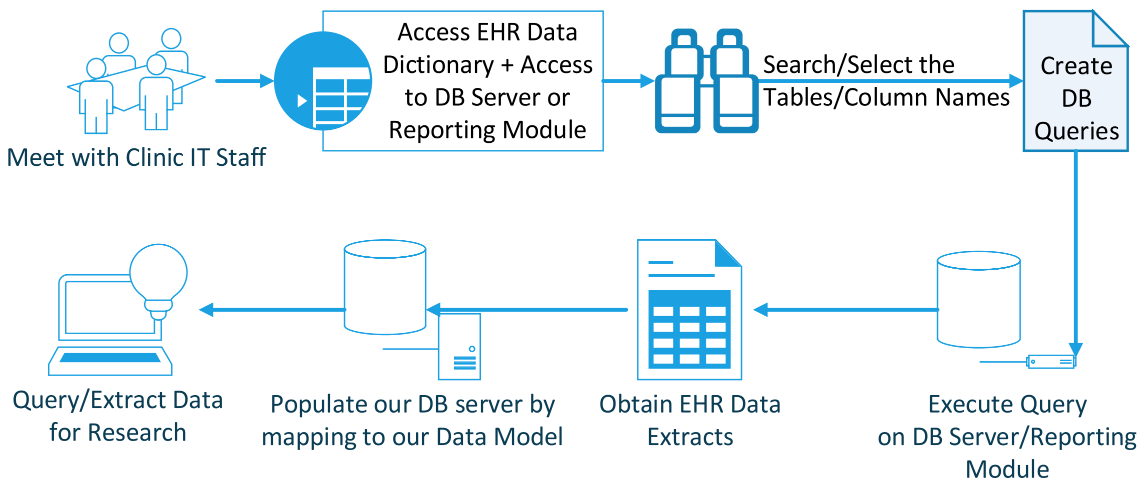

Figure 8: Process flow of data collection from the Clinic EHR Systems

In the current scenario, given the need for automated and rapid data ingestion, the data modeling technique was based on de-normalized tables. All table joins were performed at the presentation/analytics end of the information architecture. This has become possible with recent developments in the emerging area of visual analytic tools. It should be noted here that from a technical perspective, in a production environment involving nightly data refreshes, the process of loading the data into the warehouse would be automated. In such a case, we would have direct access to the database server to enable automated extraction of the data and transportation of the data from the EMR/EPM system(s) to the CDW. This automated process would involve the steps of: making a direct connection to the EMR/EPM database, execution of SQL queries, transfer of the data file to an SFTP server in the CDW environment, and finally the consumption of the incremental data extract by the data warehouse.

The data collection exercise, comprising over six years of EHR and UDS data (2010-16) from all participating clinics, has been completed. For all subsequent data requests (which are done on a monthly basis), the clinic only had to re-run the queries to append to the clinic-specific database. Data was mapped by the CDW staff to the same consistent format (e.g., with the same variable name, precision, and other metadata). This step was important in order to create a platform with much more rapid analytic capabilities.

Note that the data was already de-identified before populating into the database. A pseudo patient ID at the clinic level was used to map the data residing in different tables. In the first round of data collection, many of the raw data tables had to be prepared before loading them into the database. Populating free text information was particularly challenging given the existence of delimiters like commas, tabs and new lines. However, this was resolved programmatically and by specifying the input file format. Subsequent data loading was much more streamlined. Figure 9 shows the entity-relationship diagram for one of the clinic-specific database. There is a 'person' table at the center to which most of the other tables map. A person will have multiple encounters, diagnoses, payers and so on, while a 'provider master' table is mapped to tables like medication, procedures (performed on the 'person').

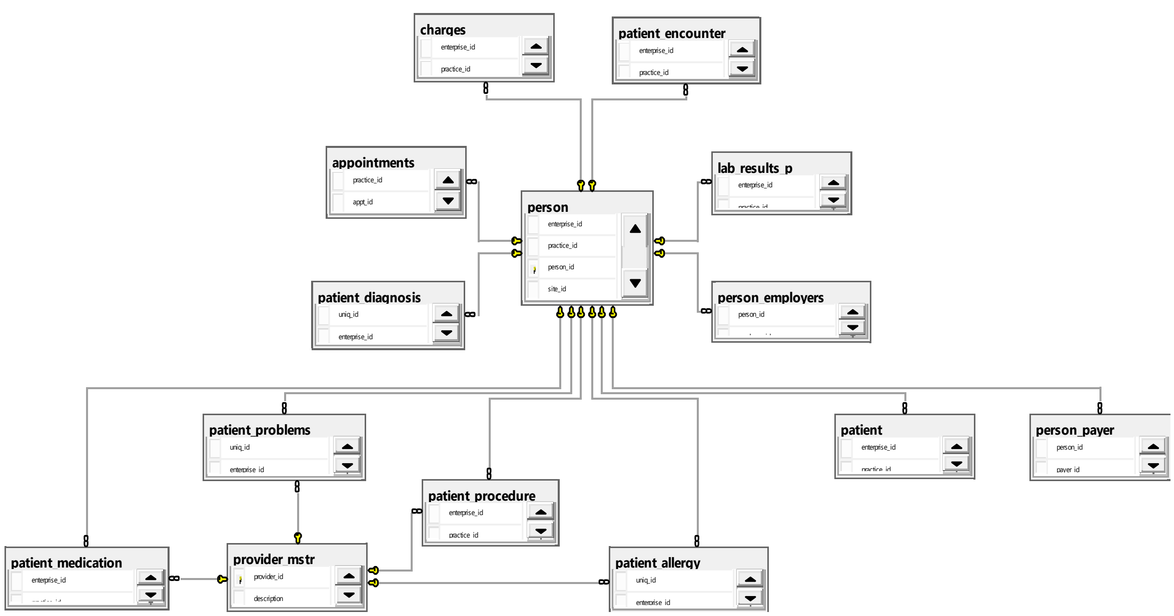

Figure 9: Snapshot of the truncated E-R/Database Diagram of the CDW

2.3.8 Pilot Big Data Augmentation

While there is a mix of structured and unstructured data in the CDW, the data size and ingestion rate of the current warehouse is not yet of the big data scale. However, there is a need to pilot the augmentation of big data technologies for specific use cases, within the context of traditional data warehousing and analytics for CHCs. The augmentation happened through deploying an open-source software framework, called Apache Hadoop – which is used for distributed storage and processing of very large data sets in a fault tolerant manner using commodity hardware. The primary use case for the CDW was data archival. The combination of *low-cost, scalable and query-able* storage offered by Hadoop's file storage subsystem was used to archive historical data from the data warehouse, which otherwise would have to be stored offline on relatively high cost tape/optical media.

Figure 10 illustrates the advantage in terms of cost, scalability, and query ability when using the big data/Hadoop platform for data archival compared to traditional data archival on tapes and optical media.

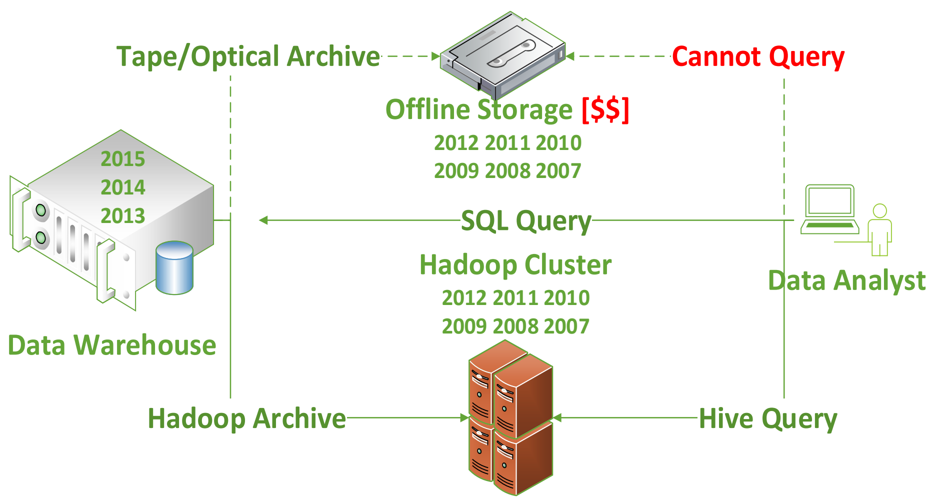

Figure 10: Big Data use case illustration for data warehouse archival

To illustrate this big data augmentation, two tasks where performed: 1) data archival to Hadoop and 2) querying from Hadoop for data visualization. To accomplish this, Apache Sqoop was utilized which is a command-line interface application for transferring data between relational databases and Hadoop. The Sqoop script was written to directly connect to the CDW using a JDBC driver and then move the tables directly into HDFS (Hadoop Distributed File System). Secondly, to show that querying and visual analytics can be performed on the data, a connection to the archived table was made and the data visualization was performed. The visualization served to validate the aforementioned tasks.

2.3.9 Technology Infrastructure

The CDW was administered through a new dedicated server that is housed within the Indiana University School of Informatics and Computing. The server specifications selected were such that it allowed for scalability in processing power and storage. The base hardware specs include: Intel Xeon CPU E-5 3.3GHz processor, 128 GB RAM, and 4x1 TB Hard Drives in a single Raid 1 array. It is running Windows Server 2012 with two partitions: One for the operating system/software applications and the other for data storage. The advantage of using dual partitions is in being able to perform the additional security step of data encryption. The server had remote desktop access for select authorized users and can also be accessed remotely through a SSL-VPN client. The server has an internal IP address and as such, does not allow internet access for data security reasons. Data was copied to the server by mapping a network drive using the UNC path.

The core software applications installed were: MS SQL Server (for database management), SAS and R (for statistical analysis); RapidMiner Studio (for data mining); and Tableau (for data visualization/dashboards). For the pilot big data augmentation piece, the Hortonworks' Hadoop data platform was deployed on an Oracle VM Virtualbox to create a single node (pseudo-distributed) Linux cluster. The sandbox VM mode allowed for full off network control of the data environment for data protection and security. The components of particular use within the Hadoop Data Platform were: Sqoop (to move data between relational databases and Hadoop File System) and Hive (a data warehousing solution to transform MapReduce jobs in Hadoop and one that leverages SQL and is suited for batch analytics).

The multidimensional Clinic Data Warehouse continues to capture clinical and operational data from CHCs in a uniform format to facilitate comparisons and analyses over time. The clinic data warehouse currently stores over **3.5 million** patient-visit records on about **0.5 million** unique patients from 2010-16 from all seven CHCs. It includes detailed information on appointments, scheduling, prescriptions, lab tests, screening/preventive measures, immunization, and quality measures under Meaningful Use Stage 2 criteria[183]. This volume of rich, longitudinal patient data is lending itself to large-scale data analytics tied to improving clinical and operational performance.

For the purposes of illustration of a use case of the CDW, a sample visualization of the patient flow time distribution at one of the sites in a participating CHCs is shown in

Figure 11. The patient flow stages are [seen on the top]: arrival → check-in → intake → exam → sign-off → check-out and the associated wait times between the stages. This particular histogram view shows the exam wait time for providers (names hidden) who were taking longer for their examinations. The filters on the right of the image allow for the real-time interaction with the visualization.

Figure 11: Visualization depicting the patient flow from arrival through check-out

2.4 Discussion

The discussion section provides a summary of the key findings, the study challenges and strategies employed, the study limitations, implications in the context of the totality of evidence, and future research directions.

2.4.1 Summary of Key Findings

This study has taken a novel approach towards systems redesign initiatives leveraging best practices and based on CHC staff-centered information needs. The needs assessment phase utilized multiple methods that complimented each other: administering a survey questionnaire to the CHC primary site manager, performing semi-structured interviews with key informants at all seven CHCs, and finally workflow observations of the key processes at the clinic sites. These methods informed the health IT related needs and uncovered wide ranging challenges related to EHR procurement, implementation and the ability to leverage data from health IT systems (EHR, operational, financial and quality) to support clinical and operational performance improvement. These challenges included: resource constraints (training needs, time consumption); reporting challenges (system deficiencies, inaccuracies); cost of ownership (EHR purchase/set-up costs; ongoing maintenance/upgrade costs); workflow management (tracking patient flow, resolving open/pending tasks); and system limitations (poor interoperability, data migration). A collaborative data warehousing solution has been designed and implemented as a means to overcome the challenges CHCs face in being able to benefit from analytics driven by a robust, centralized, and scalable data infrastructure and one that is augmented with big data technologies. Furthermore, the CDW architecture is such that it is straightforward, easy to use, has low maintenance cost, and the analytic capabilities complements the existing functionality of the various EHR systems in use at the CHCs.

2.4.2 Study Challenges and Strategies

There were various challenges encountered in acquiring the data from the clinics. For most clinics, this was the first time they received such a request to extract EHR data. Thus far, the clinics primarily ran reports to fulfill government requirements, in tandem with manual charting. Even though the clinics lacked adequate time or resources for learning and drawing insights from their data, they all readily recognized its importance to support their evidence-based quality improvement efforts. Adhering to Meaningful Use Stage 3 requirements and securing Patient-Centered Medical Home certifications with an adequate reporting infrastructure were prime concerns.

In terms of strategies, the study authors engaged meaningfully with all the clinics and presented a value proposition to the clinic leadership. The researchers benefited from a high level of participation and trust from the participating health centers, which was crucial to the success of this endeavor. The clinic staff were interviewed to understand their challenges and learn from their innovations, coupled with onsite workflow observations of key processes. The EHR data dictionary from two different EHR systems (covering five out of the seven CHCs) was obtained. These EHRs were based on client-server architecture. The data dictionary allowed the identification of all the data elements of interest and creation of the SQL queries to be executed on their database servers. This made it less onerous on the clinics as they simply had to run the provided queries and share the extracted data with the researchers. The remaining two CHCs were using a cloud-based EHR with no direct access to the database server. For those, the data was pulled using a web-based reporting tool that was integrated with the EHR.

There are many industry surveys that rate EHR systems on different dimensions. The researchers do not feel it appropriate to deliver an opinion piece on any particular EHR vendor, although we did identify many challenges from the user perspectives. Technology has been rapidly evolving in this space with an increasing shift from the traditional client-server architecture towards cloud based applications (particularly for small to medium sized ambulatory health centers), which allows for easy integration with external applications. Centralized data warehouse and analytics solutions such as the one developed in this research has been received enthusiastically by almost all CHCs in the study. They clearly want to benefit from economies of scale and assess their performance internally and relatively to peer health centers. As the state primary care association takes steps towards establishing a HCCN, the work demonstrated here is poised to support that effort.

2.4.3 Limitations of the Study

The findings from the study reflect the unique characteristics and information needs of CHCs in the state of Indiana and therefore may not entirely reflect other CHCs in the US. The data warehouse architecture which in turn was influenced by the information needs of these CHCs will have similar limits on the generalizability. Secondly, the data model may not reflect the universe of EHRs in the ambulatory care space. The data dictionaries (i.e. information describing the contents, format, structure and relationships within the database) were from four different commercial EHR vendors and other administrative systems. The data model may be less representative of other primary care practices who are not dedicated to meeting needs of the underserved. Finally, the design of the architecture and resource constraints imposed certain practical limitations on the frequency of data collected from the clinics, which in turn defined the type of analytics and decision support that can be performed. Further research is needed to investigate strategies for more efficient data sharing and use of standard data formats and ontologies, to expand this research. A collaboration with the state primary care association and their health center controlled network, will overcome some of these limitations.

2.4.4 Implications in the Context of the Totality of Evidence

The ability for clinic leaders to draw meaningful insights requires simplifying the process of data access and knowledge discover[184,185] while augmenting data warehouses with new technologies in the wake of the big data revolution[186]. Enabling structured accessible secure data captured through EHR systems provide mechanisms where clinical care informs research, and insights derived from research are reinvested in care. But challenges related to standardization, access, interoperability, and development of real-time analytics must be overcome to realize these benefits[187]. The potential quality and efficiency benefits derived from EHR usage has been shown to be particularly important for resource-constrained organizations, such as CHCs, that are constantly challenged to ‘do more with less’[188,189,190].

Studies have shown that data warehouses with appropriate analytic capabilities are generally most useful to a CHC within the context of participation in a larger organization such as a CHIN/HCCN or a large multisite health center, because of the economies of scale [191]. The information architecture developed as part of this study is an important step towards realizing the goal of a first state-wide multi-CHC information network. The architecture implemented is unique in that is tailored to the information needs of CHCs and addresses the challenges faced by them when it comes to leveraging their health IT system. Studies can be planned that leverage the CDW for system redesign initiatives and research in primary care. Providing the health centers with easy access to their data and enabling performance benchmarking against their peers can be powerful, creating opportunities for testing new technology solutions and positive change.

The CDW provides for the first time in the state of Indiana, a foundation for healthcare data analytics across the community health center operations and research enterprise. A modern analytics platform supported by such CDWs when widely implemented can better support health system redesign initiatives to address the 'quadruple aim' (better care experience for individuals, better health for populations, lower per capita cost of care and better work life of providers)[192].

2.4.5 Future Research Directions

Future work in this area can benefit from near real-time data warehousing[193], where incremental data extracts from the source systems are populated into the data warehouse for use cases that warrant such a need. From an operational perspective, these would be of particular value when monitoring patient flow through the clinic or responding on scheduling changes caused by patient non-adherence to their appointments.

From an analytics perspective, there are several research opportunities based on leveraging the CDW. Analytics can be broadly classified as being, descriptive, predictive, and prescriptive analytics. A few research examples under, 'descriptive analytics' are: development of enriched process models based on EHR time-stamps; ranking clinics based on adherence to evidence based best practices; and peer comparisons at the state and national level. Under 'predictive analytics' machine-learning models are built for various clinical applications including predicting treatment compliance and performing longitudinal network analysis and risk stratification of disease. Note that although patient data in the CDW is de-identified, a pseudo patient ID will allow longitudinal patient-centered studies across care settings. Assessing the impact of the affordable care act (ACA) on access to care will be of particular and timely relevance, considering that the CDW contains multidimensional and longitudinal (2010-16) data that is both pre and post implementation of the ACA.

Under ‘prescriptive analytics’ findings from the predictive analytics can be incorporated into standalone software applications or embedded within patient management systems to provide personalized decision support. The data can be further enriched by including external variables, such as data from community and geo-spatial information systems. There are opportunities for research and development of new technologies for patient engagement and management in the community, as well as integration of disparate care providers and care sites (inpatient, outpatient, community, “minute” clinics, prisons, other) as part of a care management team. The community clinics stand to benefit from such analytics since they lack the infrastructure that is usually available to academic medical centers. Enabling access to this dataset fosters community informatics research and development to improve care delivery, patient engagement and efficiency. A novel multidimensional clinic data warehouse of CHCs has been developed that supports data analytics and dashboards for improving access to care. If this effort is sustained and expanded, it should enable data-driven systems redesign, research, education and training, technology development and testing, and informed decision making over the years.

2.5 Conclusions

The study has developed a detailed understanding of the CHC information needs using multiple complimentary methods: questionnaires, key informant interviews, and workflow observations. The information needs called for design of a multidimensional clinic data warehouse - one that integrates patient, operational, financial, and quality data. We are encouraged that both the Indiana University and the state community health center primary care association have decided to enter into a partnership to foster further collaboration. A centralized data architecture was chosen that is relatively easy to implement and sustain, yet is rich in scope and content. Most importantly though it holds great meaning and relevance to the participating health centers, since the information architecture directly addresses their information needs and challenges. The goal is to enable free standing CHCs to participate in this data warehouse, in the same way that large hospitals participate in HIEs and benefit from large scale healthcare data analytics to improve overall clinic operational performance and achieve better scores on the ever evolving clinical quality measures in the value-based payment era. Developing strategies for interacting with technology developers and entrepreneurs, in order to test the feasibility of new tools and prototypes should speed innovation and dissemination of effective IT. Ultimately, expanding on the model of a CDW should help CHCs better meet the needs of their patients in continually learning, evaluating, and redesigning care.

CHAPTER 3: LARGE SCALE APPOINTMENT DATA MINING: INFORMING PATIENT-CENTERED SCHEDULING

3.1 Introduction

The Institute of Medicine's (IOM) landmark 2001 report 'Crossing the Quality Chasm'[194] defined high-quality health care as care that is safe, effective, patient-centered, efficient, equitable, and timely. IOM had released another report[195] in 2015 on transforming health care scheduling and access, which noted that timeliness was not as well-studied or understood as the other dimensions. Specifically, considerable evidence exists that poor access and long wait times negatively impact patients and families, as well as providers, in terms of health outcomes, patient satisfaction with care, health care utilization, and organizational reputation[196,197].

Health care access is a significant challenge in the US for many individuals for a variety of reasons, including inadequate health insurance coverage and an insufficient number of health care providers. The Affordable Care Act (ACA)[198] legislation has sought to address access problems by expanding Medicaid coverage and availability of insurance through the health insurance exchanges[199]. At the forefront of dealing with this challenge are Community Health Centers (CHCs) - safety net providers that are federally supported to provide primary care to socio-economically disadvantaged patient populations with or without health insurance and often living in medically underserved areas[200].

Access to health care means having "the timely use of personal health services to achieve the best health outcomes"[201] and is assessed by the presence or absence of specific resources that facilitate health care, such as having health insurance or a usual source of care. Within the context of healthcare operational research, however, access to care is assessed more in terms of how easily patients can gain access to needed health care. Common operational measures include: patient demand (for appointment), available supply (of appointment slots), patient wait time for appointments, no-show rates, office visit cycle time, and patient satisfaction[202]. Not surprisingly, these operational measures are going to be influenced by the appointment scheduling system in place.

Efficient and effective scheduling is critical for optimal utilization of scarce medical staff resources, especially in CHCs. Generally, two main classes of health care appointment scheduling are used[203]. One is the traditional scheduling system, where most if not all the appointments are booked in advance, and any urgent care needs are either deflected or added to existing appointments through double booking. The other is open access (also known as advanced access or same day scheduling), where a majority of the appointment slots may be held open for patients who call that day for routine, urgent or preventive visits, while the remainder are booked appointments.

Some of the problems with traditional scheduling systems include: delays in ability to schedule an appointment, high rates of missed appointments, high variability in staff workload;[204] poor patient satisfaction due to the long appointment lead times and lack of continuity of care[205]. The traditional scheduling approach also often diverts frustrated patients to urgent care facilities or Emergency Rooms (ERs) where they receive less personalized care at a much higher cost to the healthcare system[206,207]. Open access scheduling in general has been shown to address these issues by significantly reducing no-show rates and patient wait times[208]. However, implementing open access has its own challenges and failures, and requires a redesign of the clinic operations[209]. Key to implementing an open access scheduling system is to simplify both the appointment types, as well as the duration. Open access also requires a closer monitoring of the supply (provider availability) and demand (patient needs and appointment adherence), and matching these two variables across the time dimension[210].

3.1.1 Motivation

Effective appointment scheduling is of great importance in light of recent health care legislation, notably as health centers strive to achieve Patient-Centered Medical Home (PCMH)[211] recognition – a provision within the ACA. PCMH is a care delivery model which is patient-centered (responsive to individual patient preferences, needs and values) and coordinated through a primary care physician while being committed to quality and safety.

A critical factor in achieving PCMH recognition is the patient access in a health center. The PCMH model of care has a set of standards and elements to be met and the very first standard calls to '*Enhance Access and Continuity*'. The first element within this standard pertains to provisioning for same day appointments. Specifically, the CHCs aiming for PCMH recognitions are expected to implement more same-day appointments through open access scheduling, monitor their no-shows, and identify opportunities to improve patient access to appointments[212]. This calls for a greater understanding of the CHC patient scheduling systems and potentially leveraging their data infrastructure to design and implement patient-centered scheduling systems.

This chapter is focused on guiding the implementation of an optimized open access scheduling system through advanced analytic methods. It is part of a larger systems-based research project funded by the Patient-Centered Outcomes Research Institute (PCORI)[213] on improving efficiency and access to care in seven CHCs with over 40 clinical sites across the state of Indiana. The CHCs were selected such that they represented a broad range of urban, semi-urban and rural settings and provided care to a diverse patient population in terms of race, ethnicity, insurance status, and income levels.

3.1.2 Research Question

The primary research question (RQs) for this chapter is: How can clinic-specific analytics guide appointment access? There are two related sub-RQs each with its own hypothesis. The first being how can predictive analytics be used to optimize appointment scheduling, with the specific hypothesis being: The time dimension of appointments is independent of appointment adherence. The second sub-RQ is: How can analytics be used to better evaluate provider productivity? The specific hypothesis associated with that is the productivity measured as visits/hour and RVUs are directly correlated. Within the context of appointment scheduling and in particular, when a clinic includes open access hours in their schedule, it is important to understand the predictors of appointment adherence and also optimize the selection of the open access hours. In addition, accurate assessment of productivity is important in an open access scheduling system since high appointment adherence accompanied by a low productivity can adversely impact health center operations.

3.1.3 Related Work

The combination of data warehousing, data mining and evidence-based medicine has commenced an innovative application field of information technology in the health care industry[214]. Studies have shown that CHCs are known to face unusually high no-show rates,[215] partially attributable to their unique patient groups, difficulty in retaining providers, and their inability (as safety-net providers) to penalize those who no-show[216].

Studies on appointment scheduling systems in health care became increasingly common since the early 2000s, when some large primary care clinic systems began implementing open access scheduling[217]. Research has identified several factors that affect the performance of appointment scheduling systems, including: arrival and service time variability, patient and provider preferences, available information technology and the experience level of the scheduling staff[218]. Methods from industrial engineering and operations research (such as agent-based discrete event simulation modeling and stochastic models) have been applied in health care to inform outpatient scheduling[219, 220]. Related work on predictive analytics and monitoring productivity are now described.

3.1.3.1 Predictive Analytics

Predictive modeling of no-shows based on logistic regression has been widely performed to improve clinic scheduling systems in a number of care settings. For example, a pediatric clinic performed no-show modeling to effectively deal with overbooking[221]. The study showed that appointment no-shows have negative impacts that include underutilized medical resources, increased healthcare costs, decreased access to care, and reduced clinic efficiency and provider productivity. A study in a Veteran Affairs Medical Center performed no-show modeling to improve capacity utilization[222]. Prior studies at a primary care center demonstrated that no-show behavior was independently correlated with the following four factors: the patient's age and race, the presence of any physician-identified psychosocial problems, and the percent of non-cancelled appointments that were kept during the prior 12 months[223]. Another study in an academic practice site illustrated that no-show analysis was used to assess the effect of moving to a carve-out model - traditional scheduling with some open access included (carved out) on no-show rates[224]. This specific study found that the lead time for an available appointment was not a significant factor in appointment keeping behavior and that implementation of the concepts of advanced access scheduling does not directly result in a reduction in no-show rates. Instead, provider-patient continuity, wait time to an appointment, and other factors may be a more important measures. Previous study also showed that implementing open access requires an overhaul of scheduling infrastructure, including decreasing appointment types, simplifying operational processes, using past appointment requests to predict future demand and developing contingency plans for times when demand exceeds prediction[225].

A study assessing the impact of overbooking on patient access and provider productivity found that patient access and provider productivity are significantly improved with overbooking, but that overbooking causes increases in both patient wait times and provider overtime. The study performed a series of simulation experiments and determined that overbooking provides greater utility when clinics serve larger numbers of patients, no-show rates are higher, and service variability is lower.[226]

There are studies that have applied text mining techniques like co-location/co-occurrence identification and topic modeling for various healthcare applications. These include: redundancies in EHR corpora[227], translation of medical summaries[228], semantic classification of consumer health content[229], and automatically labeling topics in drug-drug reactions[230].

3.1.3.2 Monitoring Productivity

Measuring provider productivity in ambulatory care has long since been an important tool used by healthcare managers to identify inefficiencies and control rising costs since at least since the 1980s[231]. Later the relative value units (RVUs) was developed by the Center for Medicare and Medicaid Services as a means to value clinical care since the early nineties[232]. RVUs are a measure of value used in the US Medicare reimbursement formula for physician services and is based on physician work, practice expense, and professional malpractice insurance expense[233]. Different from the practice expense RVU and the malpractice RVU, the work RVU (wRVU) is a constant value across specialties and geographic locations, providing a good standard for the value of work performed. It takes into account the physician's expertise, the time and technical skill spent in performing the entire service (including the mental effort and judgment expended by the physician prior to, during and after the patient encounter terminates), including documentation of the service[234].

An article discussing the practical applications of relative value units (RVUs) for managing a medical practice, showed that RVUs empower practice administrators to objectively measure and quantify a medical practice's physician productivity and performance data versus traditional productivity measures such as office visits, net charges, or net collections, as used in many organizations[235]. It also allows common denominator analyses and per-unit comparisons for both clinical productivity and expense data benchmarking across clinics. Despite the apparent benefits of using RVUs for productivity assessments, there was no evidence in the literature or practice of its use among community health centers. Lack of a financial incentive has been a reason for this status quo, given that the Medicaid reimbursement model is one that is visit-based[236]. Based on evidence gathered working closely with health centers, another reason for this is a lack of easy access to normalized Electronic Health Record/Practice Management (EHR/EPM) data and the limited analytics capabilities at health centers.

3.1.4 Knowledge Gaps

There are knowledge gaps uncovered from the review of the literature above. First, it is well-known that there is a need to develop patient-centered scheduling system that most commonly involve some form of open access scheduling. What is not known is how to best design open access scheduling interventions for CHCs and in particular, how advanced data analytics can help.

Second, previous studies have shown that implementing open access scheduling can overcome many of the challenges with traditional scheduling methods and that no-show analysis with simulation modeling has been used in the past to guide such interventions. However, prior studies show that patient demographic and appointment related variables that can predict appointment adherence are highly dependent on the care settings, such as primary care (adult, pediatrics, behavioral health), specialty care, hospitals, and academic health centers. CHCs as a care setting for no-show modeling has not been studied as much considering their unique characteristics in term of patient demographics and their revenue model. The significance of the time dimension (a key variable identified in the IOM report) as a predictor variable has not been analyzed. Mining of large scale clinic appointment data from a large number of clinics could potentially identify patterns of no shows and cancellations, guide consideration of modified open access timing [such as time of the day and day of the week, month of the year, season] and provide insight into patient behavior, preferences and practices.

Third, in terms of statistical/machine learning algorithms most no-show models are based on logistic regression based, with the exception of the airline industry, which has applied association rules based no-show modeling. Visual models like decision trees when used in combination of logistic regression provide unique insights into the appointment adherence in being easily interpretable by lay users. Additionally, text mining techniques, including n-gram analysis can be applied to the unstructured/free text corpus within the appointment scheduling systems and uncover hitherto unavailable new knowledge.

Fourth, assessing scheduling system requires a consideration of its impact on clinic efficiency and in particular provider productivity. In particular, CHCs are known to assess provider productivity in terms of billable visits/hour. However, studies have shown the limitation of these methods in representing true provider work and benchmarking across sites and between CHCs.

To ensure the validity of the knowledge discovered from databases (KDD), the process of data preparation, selection, cleaning, incorporation of appropriate prior knowledge, and proper interpretation of the results of mining is required[237]. In this specific research project, KDD process is used to guide patient-centered scheduling in community health centers, followed by operationalization of the findings.

3.2 Approach

The description of the data (source, extraction transformation and loading process, dataset creation) and the analytic methods applied are described in the sections ahead.

The CHCs in this study use variants of traditional scheduling methods, some with few hours a day allocated for same-day appointments (aka modified open access). The CHCs accommodated patients with acute care needs through either one or more of the following ways: asking patients to use open access hours or extended clinic hours (during the weekday or on Saturday); double booking appointments; or routing them to one of their urgent care clinics.

3.2.1 Obtaining Clinic and Patient Level Data

A multidimensional clinic data warehouse (CDW) has been developed based on the requirements gathered from the key stakeholders in the seven participating CHC systems. The data used in this project is devoid of any protected health information (PHI) under the HIPAA privacy rule and was obtained under institutional IRB approval and a business associate agreement with the participating health centers. The patient level data served as the most important component of the data warehouse. The data elements selected were those that are digitally captured as part of routine clinical practice captured using the clinic EHR and practice management (PM) systems.

Details on design and development of that data warehouse has been documented in our previous publication[238]. Briefly, there were three primary data sources loaded into the data warehouse: 1) Patient-level data from the EHR/PM systems and included: patient demographics, clinic scheduling/appointment data, date/time stamps of key events for patient-flow, insurance information, provider information, medical problem list, encounter diagnoses, immunizations, medications, laboratory data, care plan, procedures, and health care referral information; 2) CHC operational data including phone reports, clinic type, certification status, locations, size, patient mix, payer mix, service mix, staff utilization, scheduling system, enrollment system, quality measures and other dimensions of care; and 3) CHC financial data, including revenue, expense, margins, cash flow, and payroll. The clinic operational and financial data was obtained from the audited financial statements of the clinics, IRS Form 990, which is the annual tax return for certain federally tax-exempt organizations, and HRSA's Uniform Data Systems (UDS) filings. The data from the EHR and scheduling systems were pulled using one of three methods depending on the reporting functionalities of the EHR used by the CHC and our needs. This was supplemented with a structured questionnaire completed through an interview with the clinical manager, followed with workflow observations and opportunistic interviews as appropriate, in the clinic. The latter served to triangulate on the information inferred from the quantitative data.

3.2.2 Data-Set

The data analytics for this study was performed on three of the seven CHCs, which were selected for implementing expanded open access scheduling. We use the following pseudonyms CHC-S, CHC-O and CHC-P for the three clinics that constituted the intervention group, while the rest served as the control group for the larger project. The three health systems selected were particularly interested in sharing their data to participate in an intervention to improve clinic scheduling.

3.2.3 Data Analytics

Machine learning, statistical analysis and data visualization techniques were used to analyze the clinic appointment dataset. The first step of the data analysis was data cleaning and validation. In this case, the dataset was cleansed of any duplicates. As expected with database table normalization, there were several dictionary/reference tables containing information on appointment types, patient profile, delivery of care, and providers. Two categories of analyses, in particular, were considered to hold the greatest relevance – appointment adherence and provider productivity.

3.2.3.1 Appointment Adherence Analysis

Appointment adherence is defined in terms of missed appointment rates, be it from no-shows or cancellations. The dataset was mined and analyzed to assess the patterns of appointments associated with an extensive set of variables. These variables included: appointment/visit type, time dimension (time of day, day or week, month of year, season), lead-time from appointment, patient characteristics (such as age, income level, and gender), insurance type, provider characteristics (age, gender, specialty, longevity with clinic), and facility types (primary care, urgent care) and location (urban, rural, mobile clinics).

There are two major considerations for data analyses: a) One can optimize the open access hours based on the time dimension (i.e., identify the times most suitable for open access or walk-in appointments); and b) Incorporating patient profile data will help identify profiles of patients who are likely to miss appointments.

3.2.3.2 Clinic Efficiency Analysis

Efficiency (η) is often measured as the ratio of useful work output to total work input, or $\eta = \frac{W_{out}}{W_{in}}$. A closely related term is production efficiency or productivity in ambulatory healthcare operations, which is the work units/scheduled [bookable] clinical hour, where the work output is usually the number of patient seen, but can also be measured as billed revenue. Work input is usually a time interval defined in hours, days or weeks. Visit volume can quite simply be the total number of patient visits, but an alternative measure is what is known as the work relative value units (wRVUs) earned[239]. The wRVU is a constant value across specialties and geographic locations, providing a CPT (current procedural terminology) code-based medical standard for the value of work performed by a provider. The wRVU accounts for time, technical skill, mental effort and judgment, and stress to provide a service, while being blind to the payer mix or amount of revenue generated.

With regards to the provider productivity, most CHCs measure productivity based on a calculation of patient visits/hour and then incentivize providers when they exceed pre-set benchmarks (set at 2.5 visits/hour for Nurse Practitioners and 2.75 visits/hour for MD providers). While this method is simple, both the literature[240] and the clinic leadership recognize the problems with this approach, since the level of work required varies amongst the patient visits. The providers scheduling template does not reflect the reality of the hours the provider actually worked, since there is non-clinical time (admin, lunch, and meetings) as well as missed appointments, both of which directly impact provider productivity. Additionally, when an appointment slot is double booked, the appointment duration is to be counted only once even as the visit count will be two (if both scheduled patients show up).

In this study, we have automated the process to extract data from the EHR for the computation of provider productivity. We also compared and assessed the differences between wRVUs and traditional measures (such as visits/hour for primary care providers) at a large CHC.

3.2.3.3 Analytic Methods

To perform the analytics on the intervention CHCs in this study, we used RapidMiner 7 for machine learning (decision tree modeling)[241], R for statistical analysis/text mining (chi-square test, logistic regression, n-gram analysis)[242], and Tableau 10.0 for data visualization[243].

A decision tree builds classification or regression models in the form of a tree structure with decision nodes and leaf nodes. In a decision tree, each leaf node represents a value of the label attribute given the values of the input attributes represented by the path from the root to the leaf. The advantage of decision trees is that they are powerful classification methods which can also be understood easily.

The C4.5 algorithm[244] is an algorithm to generate a decision tree for classification. At each node of the tree, C4.5 chooses the attribute of the data that most effectively splits its set of samples into subsets enriched in one class or the other. The splitting criterion is the normalized information gain. The attribute with the highest gain in normalized information is chosen to make the decision. This process will be recursive with the C4.5 algorithm on the smaller sub-lists.

Chi-squared test is a statistical test applied to sets of categorical data to evaluate how likely it is that any observed difference between the sets arose by chance. A limitation of this test is that it does not give us much information about the strength of the relationship. Logistic regression is used to model dichotomous outcome variables (in this case show or missed status for the previous scheduled appointments). The log odds of the outcome are modeled as a linear combination of the predictor variables. A required package to perform this analyses using R was *nnet*, which is used for feed-forward neural networks and multinomial log-linear models. The logistic regression coefficients give the change in the odds of the outcome for a one-unit increase in the continuous predictor variable. If the predictor variable is categorical, the coefficient gives the change in the odds of the outcome vs. a predefined reference.

N-Gram Analysis: Use of n-grams is common in the field of computational linguistics and probability. N-gram is a contiguous sequence of 'n' items (typically words) from a given sequence of text. When n = 1, 2, 3, it referred to as a unigram, bigram or trigram respectively. We applied uni-, bi- and tri-gram analysis of the appointment reasons, which is a free text field in the EHR. We have developed an automatic process to extract the common patient visit reasons. The steps are highlighted in Figure 12.

Figure 12: Steps for n-gram analysis

The main packages within R required to perform this analysis include: tm (for text mining), ggplot2 (advanced graphing package), and wordcloud (for generating word/tag clouds).

3.3 Results

3.3.1 Results Overview

In this section, we will first summarize the clinic and patient characteristics of the three participating health centers in this study. Then, the we will present the results from the analyses of the patient appointment scheduling datasets and apply the analytical results to guide open access scheduling. The specific analyses include predictive modeling of appointment adherence based on the decision tree algorithm; chi-square and logistic regression to quantify the impact of the predictor variables; visual representation of the results using a stacked bar graph; text mining of the unstructured/free text field describing the reason for the visit when the appointment type is same-day; and lastly, productivity analysis.

3.3.2 Clinic Data Characteristics

First, the CHC-level information was gathered from the operational data stores and in particular the 2015 UDS filings. In addition, a clinic questionnaire for each participating CHC was used to understand the appointment scheduling procedures. The combined information was analyzed and the key characteristics for the year 2015 are summarized in Table 2.

CHC Characteristic	**CHC-S**	**CHC-O**	**CHC-P**
# of Patients [for all sites of a CHC]	9,502	20,610	4,989
CHC Care Setting/Location	Urban	Semi-Urban	Rural
Services Offered [Medical-M, Dental-D, Behavioral Health-BH]	M	M, D, BH	M, BH
# of Visits [incl. Physician, NP, Nurse]	54,344	63,108	13,110
# of Primary Care Sites [excl. Mobile Clinic]	1	3	2
# of Providers [MDs and NPs]; FTEs	1 MD; 22 NP	4.5 MD; 11.5 NP	No MDs; 8 NP
Provider Tenure [Avg. in mo., for MD/NPs]	16 mo; 11 mo	55 mo; 32 mo	64 mo
Ratio of Number of Visits to Providers	9,502/23 = 413	20,610/16 = 1288	4,989/8 = 624
Scheduling Method [Pre-booked and Open Access (OA)]	AM: Pre-booked PM: Mostly OA	1-2 PM & 6-7 PM are Open Access	1 hr/provider each day is OA
Appt. Adherence Rates [No Show-NS; Cancellations-C, %]	22% NS; 13% C	18% NS; 14% C	20% NS; 13% C

Table 2: CHC-level Characteristics for Patient Visits for 2015

The table shows the diversity of CHCs in terms of: the patient populations (5000 to 21,000 patients/yr.), care settings, medical services offered, the types and tenure of the providers, ratios of patient visits to providers [413 to 1288], and appointment adherence rates. One of the clinics had a high provider turnover, a problem brought out during the interviews with the CHC leadership. However, the relatively small number of physicians in these clinics emphasize the importance of provider stability in determining access to care. Another important finding is that while modified open access makes it easier to accommodate acute care visits, the open access hours are rather limited (1-2 hours/day). Mornings are typically pre-booked ahead of time, while select hours in the afternoons are reserved for same day/open access hours. This limited number of open access hours are not enough to mitigate the missed (no-show and cancellation) appointment rates (over 30% at all three of the participating CHCs). After reviewing the practice characteristics, it was clear that both the number of open access hours offered each day and the timing of the offerings could be optimized to maximize provider utilization at all three of the intervention CHCs.

One of the clinics (CHC-P) is unique in that it is nurse practitioner driven organization. It has two different sites forty miles apart in a largely rural setting and uses the traditional method of appointment scheduling. Mining the CHC-P appointment data brought forth an intriguing difference in their no-show and cancellation rates for the two clinical sites. The clinic headquarter site [Site-D] had a 57% higher no-show rate than the other site [Site-M] (22% vs 14%) and a 56% higher cancellation rate (14% vs 9%). We sought to understand the underlying reasons for these differences through numerous meetings with the site managers and further analysis of their EHR data, including patient demographics and visit characteristics. Site-M handles half as many patient-visits as Site-D [5500 vs. 11,000; has one-third as many providers [5 vs. 15]; and a relatively large Hispanic patient population [46% vs. 13%]. Analysis of the appointment types showed that while about 70% of the appointments are for chronic, acute and physicals, Site-M has 3 times as many pre-natal visits [6% vs. 2%]. Interviews with the site managers highlighted the dynamism and diligence of the lead NP at Site-M. Corroborating evidence from our interviews, Site-M's patient satisfaction surveys included positive comments about the lead NP's attention and diligence as well as the efficient operation of the clinic.

With this information on the clinic operational characteristic, we now describe the results from the data analytics, starting with the decision tree modeling.

3.3.3 Decision Tree Modeling

We represented the visit adherence in terms of appointment and patient characteristics using a decision tree model. The target class was binary: Completed (i.e. patient showed up for the appointment) or Missed (i.e. patient either no-showed or cancelled the appointment). The data corresponded to CHC-S (the one having the highest no-show rates) and the attributes included: *Appointment Duration, Insurance, Visit Type, Patient Web Portal Enabled Status, Lead Time, New/Established Patient, Patient Age Group, Ethnicity, Race, Language and Gender.* Data for 'lead time' was categorized into discrete ranges. The decision tree generated from the dataset is shown in Figure 13.

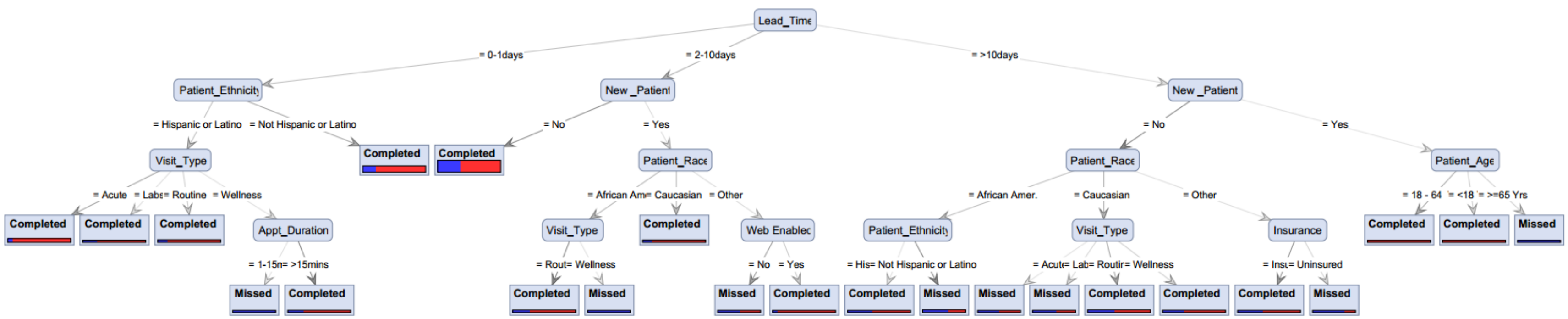

Figure 13: Decision Tree Model to predict Appointment Adherence

At the root of the decision tree is the 'lead time', which is defined as the time difference between the appointment creation date and the actual appointment date. Being at the root of the decision tree, implies that the lead time is the most significant predictor variable determining appointment adherence. This finding supports the notion of a scheduling system based on making more same-day appointment slots available (since a smaller lead time will lead to higher appointment adherence) as well as avoiding scheduling appointments too far in the future. Other attributes that held significance (i.e. provided the greatest information gain) included: new patient vs. established patient, patient ethnicity, race, age group, and visit type.

Furthermore, visually tracing through the decision tree allowed for examining various scenarios. As an example, an acute visit type for a patient with Hispanic ethnicity whose appointment is scheduled the same day, is highly likely to show up for the visit. In another scenario, a wellness visit for an African American patient who is new to the clinic and whose appointment is scheduled more than ten days before the actual appointment date is highly likely to miss his/her appointment. Thus, the decision tree model helps demonstrate the importance of different variables and their linkage to others, in determining the outcome of interest (adherence). A decision tree can also be converted into a set of rules since it creates if-then constructs to separate the data into increasingly homogenous sub-sets. Referring to the tree above,

IF (Patient > 65 years old)
and (Patient Type = New),
and (Appt. Lead Time > 10 days)
Then
there is high probability (> 90%) that patient will miss his/her appointment

This visual representation with simple *if-then* constructs makes the decision tree technique an extremely useful tool to guide open access scheduling. For instance, patients with high likelihood to miss their appointments are good candidates for double booking or more intensive appointment reminder systems.

3.3.4 Statistical Analysis

Following decision tree modeling, statistical approaches were applied to quantify the predictors for appointment adherence in terms of odd ratios from the logistic regression and also assess the strength of association between each variable and the appointment adherence using chi-square test of independence. The outcome variable of interest is 'show status', which is the binary values of 0 or 1. The *show status* of 0 implies the patient missed the appointment (no-shows or cancellations), while a status of 1 implied the patient showed up for the appointment.

The statistical analysis specifically focused on the predictor variables in the time dimension, given their importance for designing the optimal appointment schedule and the fact that knowledge of these would allow for deciding the hours allocated for open access. The specific variables in this dimension included: ApptLeadTime (same-day and not same-day), ApptStartHour (8:00am-06:00pm), ApptDay (Mon-Sat), ApptMonth (Jan-Dec), and ApptSeason (Winter and non-Winter).

The null hypothesis (H_0) was that the 'ShowStatus' is independent of these variables in the time dimension. The Pearson's Chi-Square test of independence revealed that the 'show status' is dependent on each of the time dimensions (p value was < 0.05) for CHC-S and CHC-O. For CHC-N, the show status was not significantly dependent on the season (p value was > 0.05), and the dependence on ApptHour and ApptMonth was not significant at the 99% confidence interval (p value was > 0.01). The results are summarized in Table 3.

Pearson's Chi-Square Test of Independence [Format: χ^2; df; p-value] for *'show status'* prediction			
	CHC-S	**CHC-O**	**CHC-N**
Sample Size	17,449	49,118	6,708
LeadTime	848.71; 1; < 2.2e-16	1376.1; 1; < 2.2e-16	399.95; 1; < 2.2e-16
Season	918.22; 1; < 2.2e-16	10.762; 1; 0.001036	0.45043; 1; 0.5021
ApptHr	230.71; 10; < 2.2e-16	165.65; 12; < 2.2e-16	20.388; 11; 0.04029
ApptDay	36.982; 4; 1.816e-07	22.751; 6; 0.0008843	19.406; 5; 0.001614
ApptMonth	2736.4; 11; < 2.2e-16	90.019; 11; 1.653e-14	20.181; 11; 0.04292

Table 3: Pearson's Chi-Square Test of Independence for prediction of 'show status'

For the logistic regression analysis, the log odd of the outcome was modeled as a linear combination of the predictor variables (i.e. the five time dimensions).

ShowStatus ~ ApptLeadTime + ApptHour + ApptDay + ApptMonth + ApptSeason

The change in the odds of the outcome is always with reference to a value. The reference values for the different variables are 8:00 am for ApptHr, Fri for ApptDay, April for ApptMonth, non-Winter for ApptSeason, and Non-Same Day appointment for the ApptLeadTime. We present some of the key odds ratios from the logistic regression analysis for each of the CHC in Table 4.

Logistic Regression: Key Odds Ratios				
Variable	**Reference**	**CHC-S**	**CHC-O**	**CHC-N**
ApptLeadTime	Non-Same Day	3.77	3.15	16.61
ApptSeason	Non-Winter	0.95	0.91	0.91

ApptHour	8:00 AM	0.92	1.21	0.98
11:00 AM				
2:00 PM		1.08	1.16	0.95
3:00 PM		0.89	1.12	1.16
4:00 PM		0.82	1.28	1.02
ApptMonth	April			
Jan		1.33	1.06	0.94
May		0.93	0.92	0.75
Aug		0.90	1.44	0.77
Sep		0.97	1.26	0.74
Oct		1.02	1.26	0.82
Dec		1.003	1.10	0.80
ApptDay	Fri			
Mon		0.83	1.07	1.20
Thu		0.98	1.14	1.19

Table 4: Key Odds Ratios of the time-based variables for prediction of 'ShowStatus'.

The odds ratio for same-day vs. non same-day, ranged from 3.15 to 16.61, showed that the ApptLeadTime was a strong determinant of appointment adherence. This implies that the likelihood of a patient showing up for their appointment scheduled in the same day is anywhere from 3 to 17 times higher than that for appointments not scheduled the same day. This represents a significant impact and support the progressive implementation of more same-day appointments to improve appointment adherence. The odds ratio for winter vs. rest of the year was 0.9 (i.e. 10% lower likelihood of appointment adherence in winter months) for CHC-O and CHC-N.

The results also showed that even for the same predictor variable, the effect may be different for different clinical settings due to a unique attribute such as the characteristics of the patient population or the organizational culture of the clinic. For example, one important odds ratio was between 4:00 pm relative to 8:00 am being equal to 1.28 for CHC-O (i.e. a 28% higher likelihood for appointment adherence at 4:00 pm relative to 8:00 am), but this was opposite for CHC-S where the odds ratio was 0.82 (i.e. showing appointment adherence was 18% lower for 4:00 when compared with 8:00 am).

3.3.5 Visual Plots

The analytical results were visualized through an interactive dashboard, developed using Tableau - visualization software. To visualize the appointment adherence for a provider at different times of the week, we started with an illustration of how a provider schedule spreads out over a typical week. Figure 14 illustrates appointment adherence for a random week (Monday through Friday) at a given CHC, organized by the time of the day along the y-axis (from 9:00 am to 5:00 pm). Appointment adherence was grouped into three categories, namely: Completed, Cancelled, and No-Show. The bubble size represents the appointment duration (legend on top right for 15 to 60 minutes). For example, looking at Friday between 4:00-5:00 pm, there were two completed visits: one for 30-minutes and the other for a 15minute duration. There is one 30-minute appointment cancellation which had been scheduled for 4:00 PM, and one 30-minute appointment scheduled at 4:30 pm where the patient didn't show or call to cancel (no show).

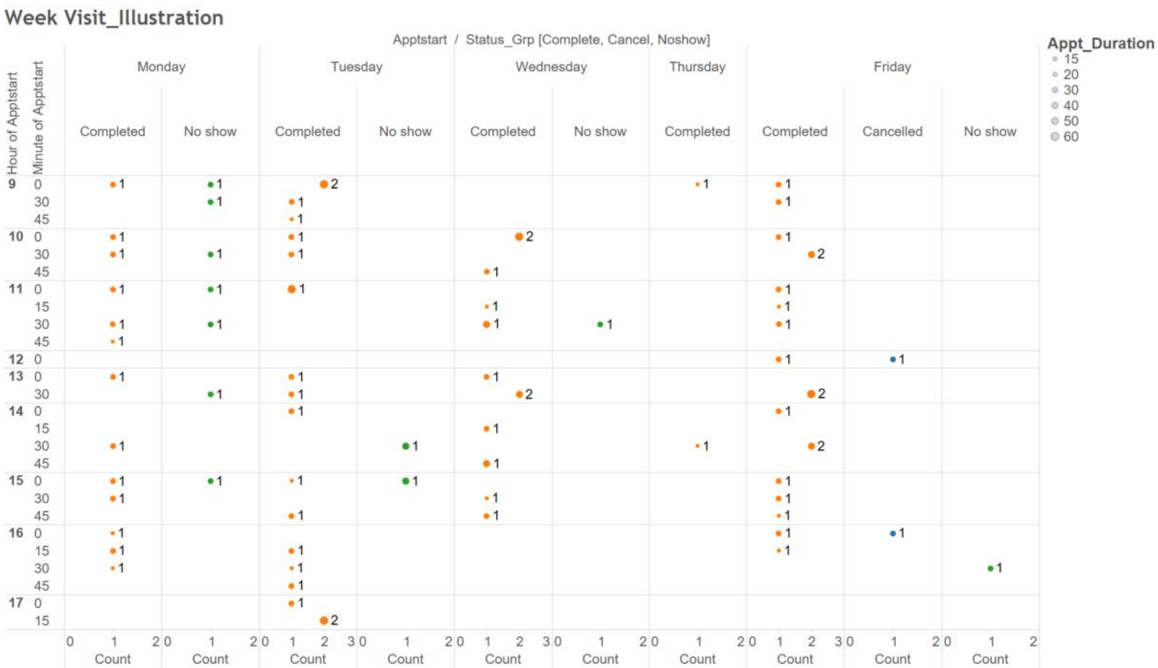

Figure 14: Illustration of Patient Visit Adherence for a randomly selected week at a clinic

Figure 15 provides another type of visualization in the form of stacked line and bar graphs of the appointment adherence (complete, cancelled, or no show) based on the cumulative data from January to December 2015. The data showed that fewer appointments are scheduled before 9 am, or at 12 pm, or after 4 pm. However, no show and cancellation rates are high at these time slots. In particular, there are more no-shows in the morning times [7:00 – 8:00 am] and relative high no shows around noon. For the rest of the times of the day, the no show rate is around 15%. These results are in line with the results from the logistic regression analysis.

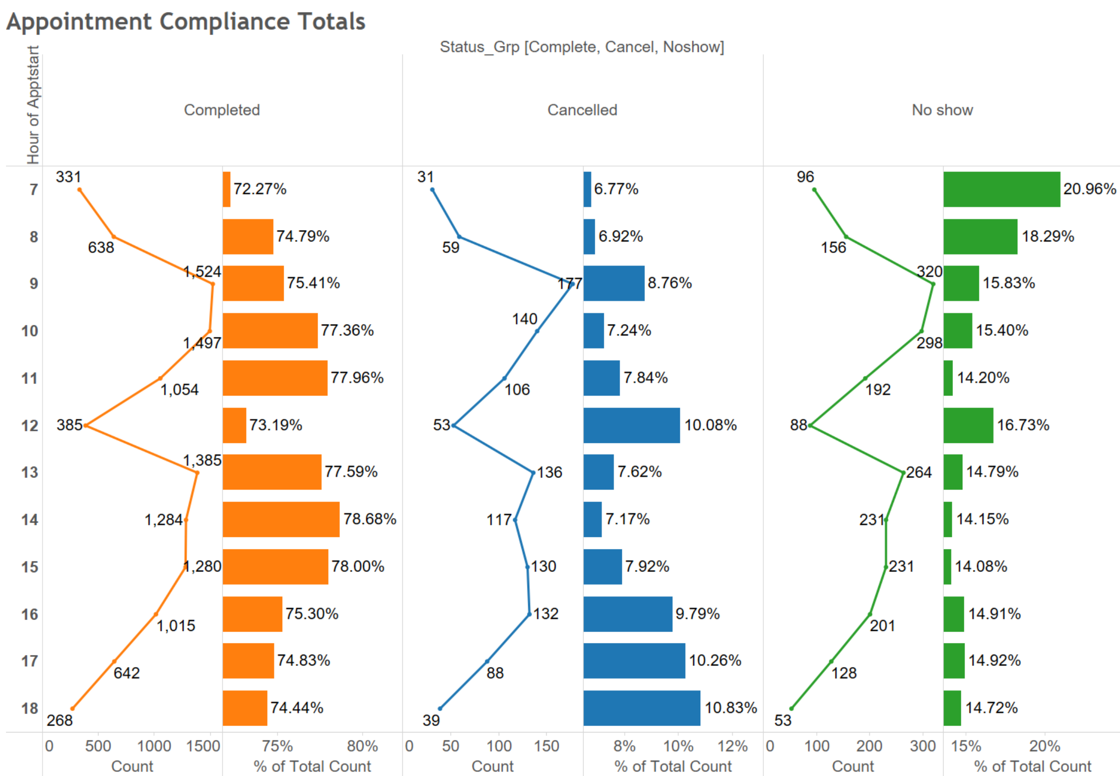

Figure 15: Appointment Adherence by the time of the day for an entire year

Figure 16 compares the visit adherence patterns for the three CHCs we evaluated. The average rates for no-show vary from 15 to 23%, cancellations rates from 8 to 17% and successful visit rates from 69 to 77%. There are distinct visit adherence patterns within the CHC hours of operation. For example, the no-show rates are consistently highest in the morning across each of the CHCs; the peak no-show rates vary from 7:00 am to 9:00 am among the CHCs. The lowest no-show rates occur in the early afternoon (1:00 – 2:00 pm) across all three CHCs. The pattern for cancellations rates are also fairly consistent, in that most of them occur late in the afternoon after 3:00 pm. Notably, appointments with high completion rates are in the early afternoon (1:00 – 3:00 pm) and in the evenings at 6:00 pm at all the three CHCs. In two out of the three CHCs, the 7:00 am time slot also has high visit success rates.

Figure 16: Appointment Adherence by the time of the date across the 3 CHCs in the study

After identifying these patterns, we could recommend to all three clinics that they schedule same-day/open access hours in the late morning [10:00 am – noon] because of the high no-show rates during those hours and it would still allow adequate time in the morning for patients to call in and schedule a same-day appointment. Offering more same-day appointment slots during this historically high no-show time period should improve access for patients while reducing the no-show rate for the clinic.

3.3.6 Operationalizing the Findings

The findings from of our statistical analysis was applied to the clinical operations of CHC-S, which had the highest no-show rates relative to the other intervention clinics. They had recently achieved PCMH recognition and were keen to optimize their same-day appointment scheduling, in particular, having a more sophisticated approach to selecting open access hours. This CHC had been using a modified open access scheduling system, where mornings were scheduled in advance and afternoons were being left open for walk-in appointments (although this was not strictly enforced).

We examined the appointment adherence pattern for CHC-S to guide the selection of open access hours. Prior analysis had provided guidance on the relative importance of various predictors for appointment adherence (in particular the lead time) and the patterns seen across a day. Specifically, based on the high missed appointments rates in the morning, we recommended that CHC-S keep the early morning hours as scheduled, but have open access hours in the late morning from 10:00 am to noon (to allow time for patients to call in for same-day appointments). For afternoons, we recommended CHC-S have scheduled appointments until 2:00 pm and then open access hours after that (to combat the higher cancelled appointment rates occurring during that time period).

CHC-S implemented the schedule change in June 2016 based on the recommended changes. Their overall missed appointment rates (no shows and cancellations) dropped from an average of 37% (during March-May) to 31% (during Jun-Sep) - a 16% reduction (Figure 17).

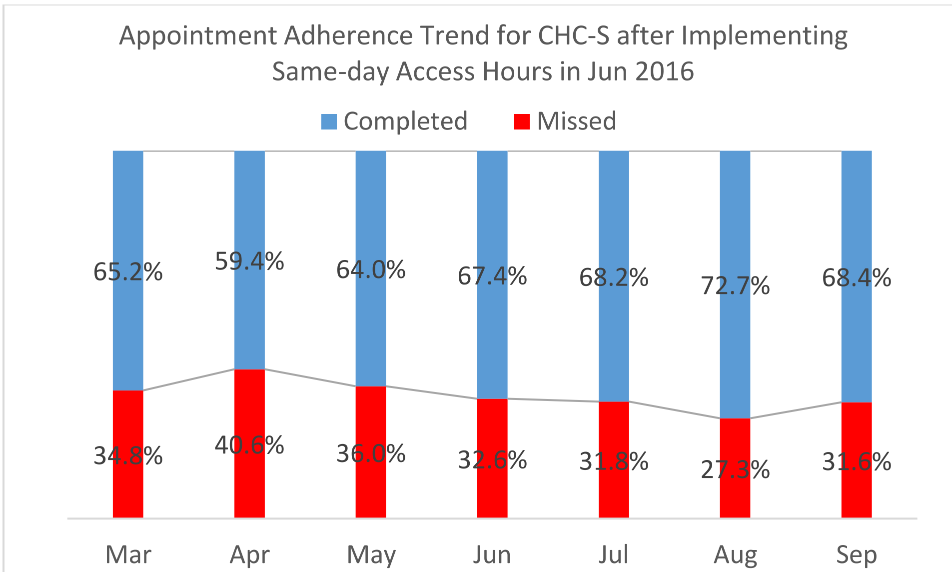

Figure 17: Appointment Adherence Trend for CHC-S

CHC-S was thus able to see an immediate benefit of modifying the open access hours based on knowledge extracted from their scheduling data.

3.3.7 N-Gram Analysis of Unstructured Text

PCMH standards for measuring access to care require that health centers report a certain percentage of appointments as reserved for same day access. To meet this requirement, CHC-S found it easier to create 'same-day' as an appointment type as a means to meet the PCMH reporting requirement. However, unlike other appointment types – such as wellness, hospital follow-up, or chronic care management – an appointment type labelled 'same-day' says nothing about the actual reason for appointment unless one were to analyze the free text entered in the 'reason for appointment.' Not only is that information useful to understand the trend in appointment purpose, it is critical information used to reduce the number of appointment types, which in turn helps with implementing an open access scheduling system.

We performed text mining of appointment reason (an unstructured text field) for CHC-S to understand the reasons for those appointments categorized under ‘same-day’ – they represent about 20% of the appointments for this CHC.

We found that the tri-gram analysis gave the most meaningful outcomes in the context of extracting the appointment reason. This is because a uni-gram analysis showed words, like cough, sore, lower, pain, pressure, which lack complete meaning. Bi-gram was better, but still had phrases that were not easy to interpret, such as ‘lower back’, ‘right knee’, and ‘high blood’. The tag cloud for the tri-gram analysis is shown in Figure 18.

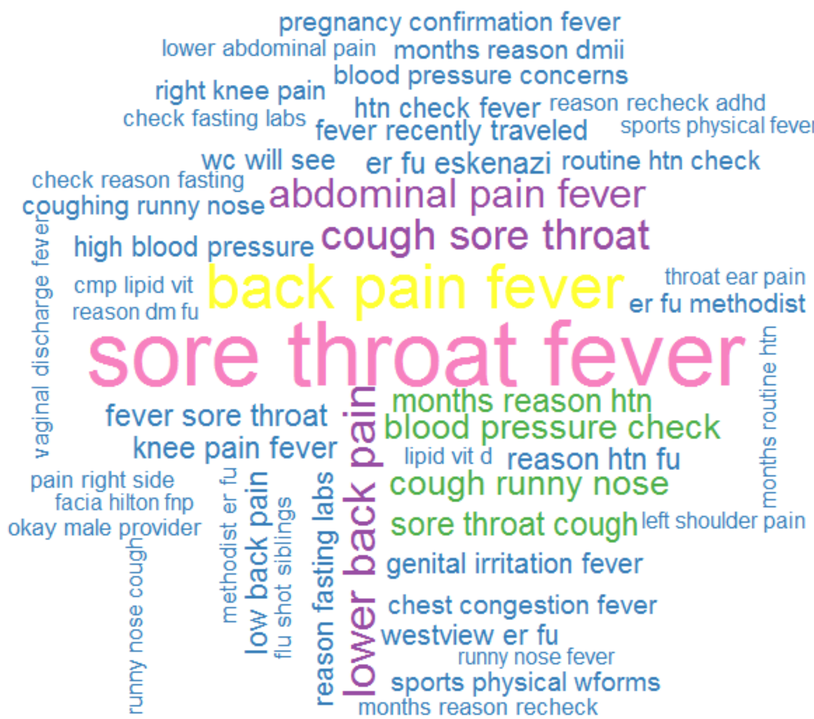

Figure 18: Tri-gram tag cloud of the appointment reason for same-day appointments

We can see that the most common reason for visits include sore throat, cough, fever and hypertension, all of which are non-urgent or chronic health conditions. This shows that same-day visits were not being used exclusively for acute conditions as one might expect. We also recommended CHC-S (as well as to CHC-O and CHC-N) to avoid using same-day as an appointment type (to simplify and minimize the number of appointment types for easier scheduling). Rather, we suggested using a report generated from their EHR data that identified same day appointments by comparing the ‘appointment creation date’ and the ‘appointment scheduled date.’

3.3.8 Productivity Analysis

Finally, we analyzed the appointment data for 12 months (July, 2015 through June, 2016] representing about 37, 400 appointments at the main primary care site of CHC-O – the largest of the three CHCs in this study. There were 13 providers who worked for each of those months, of which five were medical doctors (MDs) and the remaining eight were nurse practitioners (NPs). Table 5 shows the measured productivity values for each of the primary care providers at CHC-O with embedded data bars to visually represent the values.

Provider	Visits/hr	wRVUs/hr
P1, MD	2.86	1.85
P2, MD	2.79	1.77
P3, NP	2.77	1.98
P4, NP	2.70	1.84
P5, NP	2.70	2.33
P6, NP	2.61	2.14
P7, NP	2.58	2.19
P8, MD	2.40	2.02
P9, NP	2.32	1.87
P10, MD	2.25	2.33
P11, MD	2.07	1.81
P12, NP	2.04	1.39
P13, NP	1.66	2.08

Table 5: Provider productivity at a large clinic in terms of Visits and wRVUs per hour

Referring to the table, the highest productivity over the 12-month period was for P1 (2.86 visits/hour), while the lowest was for P13 (1.66 visits/hour). wRVUs was highest for P5 (2.33 wRVUs/hour) and lowest for P12 (1.39 wRVUs/hr). Interestingly though, while P1 had the highest visits/hour, the productivity in terms of wRVU was in the lowest quartile (1.85 wRVUs/hr). In contrast, P13 had the lowest productivity in visits/hour but their wRVU ranked in the top 40% (2.08 wRVUs/hr). These results show that the visits/hours can mask the complexity of work that a provider performs during a single visit and thus is not a representative metric to assess providers work output at the chosen CHC.

3.4 Discussion

3.4.1 Summary of Key Findings

The chapter sought to understand how clinic-specific analytics can guide appointment access. Firstly, predictive analytics was applied to optimize appointment scheduling with a focus on the time dimension of appointments and appointment adherence. Secondly, scheduling data was analyzed to accurately assess provider productivity and specifically the suitability of RVUs per hour in community health centers. The results for both of these sub-research questions are summarized ahead.

Exploratory and predictive modeling of data extracted from clinic EHR systems can provide important insights into key operational factors that govern the implementation of patient-centered scheduling. The analysis included a wide range of variables related to patient characteristics, visit types and time dimension that were evaluated to assess impact on appointment adherence. A decision tree algorithm was first applied to predict appointment adherence. Interestingly, the lead time for the appointment was the most significant predictor of appointment adherence. Although this variable has not been as widely studied, it is consistent with an article assessing the effectiveness of reminder systems to reduce the no-show rates[245]. Predictive analysis based on the decision tree model also provided insights on multiple demographic (age, race, ethnicity, insurance status), clinical (new patient, visit type) and operational (visit duration, web-enabled) variables that are associated with appointment adherence.

Logistic regression was then applied to quantify the impact of lead time and other variables in the time dimension (such as time of the day, day of the week, month of the year, and season) on appointment adherence. This showed that patients scheduled for same day appointments were at least three times more likely to show for their appointment when compared to patients who had scheduled non-same day appointments. This among other factors (such as visit type, ethnicity, race, new vs. established patient, and appointment duration) were possible predictors of appointment adherence. There were differences in the statistical measures (odds ratios and p-values) among the three participating CHCs, likely due to the differences in the operational and patient characteristics of the clinics. Knowledge discovered from the decision tree and logistic regression regarding the patient appointment adherence can directly guide effective schedule utilization through appropriate double booking (for example, double booking two patient appointments that are likely to be a no-show) and appointment reminder systems (targeting chronic no-show patients with more intensive reminder systems).

Visual analytics was used to assess the time dimension of the appointment and its association with appointment adherence. The visual plots reaffirmed that appointment adherence is significantly related to the time of day that the appointment is scheduled. These distinct patterns inform selection of booked and open access hours. Across each of the three CHC organizations, we consistently saw high no-show or cancellation rates in the morning from 8:00 to 9:00 am and in the late afternoon after 3:00 pm. Early afternoon and evenings have the lowest no-show rates across all of the CHCs. Based on these findings, clinics using a mix of booked and open access hours should consider having fewer booked appointments in the early morning, more booked appointments in the early afternoon and offer open access hours during late morning and late afternoon.

Graphical representation of these results in the form of decision trees and visualizations could be made available to schedulers, or incorporated in electronic scheduling algorithms and tools to maximize provider utilization. As a use case, this pattern analysis was applied to one of the CHCs that was keen to optimize the selection of their open access hours. The scheduling changes implemented as a result of the analysis led to a 16% reduction in missed appointment rates at this CHC. In this same CHC we used text mining, specifically trigram analysis, of the unstructured text in the appointment reason fields to efficiently discover the reason for patient visits (mostly, non-acute or chronic health conditions) under same-day appointment type. The tri-gram analysis was shown to be the most preferred route in discovering these reasons for appointment and is likely generalizable when dealing with short phrases of clinical terms.

Lastly, the productivity analysis for one of the largest CHCs in the study showed that wRVUs are much more accurate representation of provider work relative to visits/hour, which is the current standard in CHCs, and should be used for assessing productivity with any scheduling system (open access or otherwise). This productivity data analysis highlighted the importance of using wRVUs within the community health center enterprise and specifically showed that wRVUs/hour and visits/hour are not directly correlated (thereby disproving the hypothesis). The lack of data normalization and analytic capabilities at the health centers was one reason (in addition to the current visit volume based Medicaid reimbursement model) why this wasn't being applied despite its obvious importance for practice management.

3.4.2 Strengths and Limitations of the Study

We designed and implemented a CHC data warehouse merging data from multiple CHCs across Indiana. This is a rich repository of data from all seven participating CHCs for appointments since January 2010 to the present. Out of the seven CHCs participating in the larger project three were chosen for the analysis completed in this study as part of the intervention group. The CHCs selected were such that they served geographically and ethnically diverse patient populations and represented different modalities of appointment scheduling. Furthermore, CHCs included in the intervention group were chosen based on their interest in improving access, so may have represented those with the most to gain, or those most committed to change. However, comparison of characteristics of the clinics chosen for inclusion and those that were not, showed that they shared many similar clinic operational characteristics.

Predictive modeling provides likelihood (probability) information, which may be difficult to interpret at times. Predicting the future using relations derived from historical data is inherently challenging when system involves people (i.e. patients and providers). There is also the possibility of not considering variables that may be critical to the predicted outcome due to unavailability of data. Additionally, modifying appointment hours requires a flexibility in how the providers may be scheduled, which may pose challenges. These results should be considered as potentially important to providers, but reconfirmed within a given clinic system before changes in scheduling are made.

Finally, assessing provider utilization through productivity measures should take into account the 'value of the work' through such measures as wRVUs, as opposed to a simply counting the billable visit volume. wRVUs/hour give a more accurate representation of the provider work and can unmask the reasons why a provider may have high productivity numbers assessed using visits/hour.

We believe the knowledge discovered from mining the EHR/PM data will be relevant to many CHCs in the state and across the country. This is due to the relative homogeneity of the patient demographic served by CHC (specifically those with socio-economic challenges), and similar clinic operational and financial characteristics. The study findings however may not be generalizable to other non-CHC ambulatory health centers.

3.4.3 Implications in the Context of the Totality of Evidence

The study has many implications in the area of knowledge discovery from health center EHR and practice management databases, specifically when applied to designing patient-centered scheduling system. The current study applies advanced data analytics on a multidimensional CHC data warehouse to guide the design of unique scheduling interventions to improve access to care in three CHCs. Prior research has shown great variability in factors (e.g. patient demographics and appointment variables) that affect patient appointment adherence and potential interventions due to the difference in care settings. Prior studies at a primary care center demonstrated that no-show behavior was independently correlated with the patient's age and race, the presence of any physician-identified psychosocial problems, and the percent of non-cancelled appointments that were kept during the prior 12 months[246]. Notably though, in an academic practice site no-show analysis found that the lead time for an available appointment was not a significant factor in appointment keeping behavior and that implementation of the concepts of advanced access scheduling did not directly result in a reduction in no-show rates[247]. Instead, provider-patient continuity, wait time to an appointment, and other factors may be a more important measures. A pediatric clinic dealt with the problem of no-shows through overbooking[248]. However, a study assessing the impact of overbooking on patient access and provider productivity found that patient access and provider productivity are significantly improved with overbooking, even as overbooking caused increases in both patient wait times and provider overtime[249].

This study provides further evidence of predictive value of patient demographics and appointment variable through the application of multiple complementary analytic methods: machine learning, statistical analysis and visualization techniques, which are easy to interpret and apply in clinic practice. Delivering the PCMH promise of improved access to healthcare requires that the health center provide for same-day appointments and monitor its effectiveness. This study adds to the evidence on implementing operational changes that support the open access scheduling paradigm.

Assessing the impact of open access scheduling interventions on production efficiency is also critical and the study shows the need to consider provider time, skill, training and intensity required to provide care as opposed to the simplistic billable visit count for internal fairness as well as regional benchmarking. This is particularly relevant as the health care system moves towards rewarding value more than volume as noted in a perspective article[250] that calls for a reconfigured RVU system that accounts for physicians' services in a variety of delivery and payment contexts.

3.4.4 Future Research Directions

Further studies are needed to confirm these findings in other populations and determine how best to incorporate such factors in developing tools to optimize patient-centered scheduling. Open access scheduling methods may be optimized by appointment analysis and ongoing monitoring to derive the benefits associated with such an approach. Given the importance of effectively utilizing scarce health care resources and appointments, there is a need for further research into models, strategies and technological and organizational change solutions.

3.5 Conclusion

The study has applied advanced exploratory and predictive analytics on EHR data to inform patient-centered scheduling and has uncovered factors that govern appointment adherence based on a multi-CHC EHR and practice management data warehouse built for the purpose of improving efficiency and access to care. Appointment adherence was significantly correlated with the time dimension (lead time, time of day, day of the week, month of the year, season) and there were distinct patterns identified in the data from participating clinics. The process applied in this study can be used to guide the selection of open access hours and optimize the provider schedule utilization. Open access scheduling methods require ongoing monitoring, and analysis for matching of supply (providers, staffing, open slots) and demand (appointment requests), to derive the benefits associated with such an approach. The study highlights the need for greater 'in-CHC' data analytic capabilities as well as additional research aimed at re-designing the care delivery process at CHCs for improving access and efficiency in an increasingly value-driven healthcare system.

CHAPTER 4: A USER-CENTERED DASHBOARD FOR CARE DELIVERY IN COMMUNITY CLINICS: DESIGN, DEVELOPMENT AND EVALUATION

4.1 Introduction

'You cannot manage what you do not measure' is an old adage that applies to all industries, and health care is no exception. While health care has lagged behind other industries in the use of modern analytics, the Health Information Technology for Economic and Clinical Health (HITECH) Act has been recognized as the single most potent force to change this through promotion of 'meaningful use' of electronic health record data[251]. Additionally, the Affordable Care Act (ACA) has established comprehensive healthcare insurance reforms to increase access to healthcare, improve quality, lower costs, and provide comprehensive care through patient-centered medical homes (PCMH)[252, 253].

Within the context of primary care, Community Health Centers (CHCs) provide health care to low income uninsured/underinsured individuals. CHCs play a crucial role in implementing the provisions within the ACA, by providing access to high-quality patient-centered care - that which is responsive to individual patient preferences, needs and values[254]. The ACA also ties increased funding to demonstrated quality and efficiency improvements[255], which require CHCs to leverage their clinical and practice management data to drive improvements in their care delivery process[256]. Delivery of patient-centered care has much to do with the design and implementation of an effective appointment scheduling system, in particular the provision of open access scheduling. An optimized appointment schedule is an essential mechanism for the delivery of patient-centered care because it affords the flexibility needed to meet patient preference for appointment times, especially for urgent health care needs[257,258]. In order to evaluate the impact of efforts to optimize appointment scheduling or other facets of healthcare delivery, health systems need suitably designed business intelligence (BI) tools[259].

Performance dashboards, built on a data integration and BI infrastructure, are multilayered applications that can enable organizations to measure, monitor, and manage business performance more effectively[260]. Interactive self-serve cloud-based performance dashboards represent the 'new face' of BI that link strategy and performance across all types of users within an organization, not just senior management, business analyst and power users[261].

With the increasing prevalence of scorecards, dashboards and other visualization tools for decision makers to draw insights from their data, the issue of visual information design has assumed more importance than ever[262]. A key design element for effective data representation is understanding the information needs of the end-users: What metrics do the users need? What context does each metric require (target, variance, trend, breakdown by regions)? What is the visual representation that best communicates the metric (table, bar chart, pie graph, line chart, scatter plot or some entirely unique visual)? This underscores the need for visual tools to be appropriately designed and implemented for community health centers to support effective decision making.

4.1.1 Related Work

The premise of a learning healthcare system (LHS) is based on data capture, analysis, interpretation and then delivery of tailored messages to individual decision makers in the health care practice[263]. Despite the importance of tailoring messages for decision makers, there is poor understanding of the nature of customization for individual roles (providers, management and support staff) in health centers. In particular, alignment of information systems with organizational objectives and strategies is a key management objective in health care organizations[264]. Studies have shown that the vast majority of healthcare organizations experience conflicts varying from minor disagreements to fully disruptive behavior among nurses, physicians, physician assistants and administrators[265]. There often are inherent structural tensions or conflict between physicians, mid-level providers (nurse practitioner/physician assistant), clinical support staff (nurses, patient care assistants), and administrative staff[266]. Inadequate resolution of structural tensions and conflict stifles organizational innovation and growth[267]. By understanding the unique information needs of different stakeholders, Information systems can be designed to enable staff buy-in to support the core missions of the health center and help facilitate data-based vs. emotion-based conflict resolution to ease tensions[268].

A recent systematic review of evidence supporting the use of clinical and quality dashboards in health care environments found that there is considerable heterogeneity in the design of dashboards and the users targeted[269]. The review called for detailed research studies to establish guidelines for dashboard design and provide evidence that these tools make an impact. Several standardized measures have been developed to capture patient reported outcomes (such as symptom status, physical function, mental health, social function, and wellbeing) with a focus on improving the quality of care from the patient's perspective[270]. However, in order to leverage these metrics and operationalize evidence-based practice management, clinic managers need tools that facilitate the generation of knowledge and insight from data[271].

Despite substantial evidence showing the benefits of deploying EHRs and health information systems, healthcare providers often report only modest improvements in their ability to make better decisions based on using more comprehensive clinical information. This is largely due to a lack of appropriate visual analytics that can address information overload challenges[272]. Information overload is the problem that arises when individuals try to analyze a number of variables that surpass the limits of human cognition[273]. In fact, studies have shown that access to comprehensive and large-scale clinical data resources have resulted in analytical processes becoming more difficult[274]. The challenges noted in the study stem from (a) clinicians trying to combine data from different disparate systems and make sense of the patient's condition within the context of the patient's medical history, (b) administrators trying to make decisions grounded in data, (c) researchers trying to understand differences in population outcomes, and (d) patients trying to make use of their own medical data. The study also noted inadequate levels of stakeholder engagement in implementation of integrated data analytic systems as one of the barriers to benefiting from such systems. Visual analytics (the science of analytical reasoning facilitated by advanced interactive visual interfaces) has been noted as an emerging discipline with significant promise in addressing many of the information overload challenges and healthcare data-driven decision making inherent in multi-dimensional data analytics and presentation[275].

The Veteran Health Administration's (VHA's) foray into large scale data analytics has shown that managers and clinical end users of data are critical participants in the decision making process about which reports, analyses and associated data elements need to be prioritized for standardization and validation at the national level[276]. The VHA also recognized the need to include improved BI tools that can simplify sophisticated analyses, thereby minimizing the need for extensive end-user training.

There are several examples of dashboards that are built upon databases of community-based health indicators. Most of these are maintained by government agencies such as the Center for Diseases Prevention and Control (CDC) and several non-profit organizations. While these dashboards are designed for a general audience (often including the lay public), there are dashboards specifically designed for healthcare settings to guide decision making. User-centered design (UCD) principles (i.e. designing for and involving users in the design of information systems[277]) have been applied in the design of dashboards for clinicians and patients that address specific health care situations such as: reflecting on quality indicators on cancer care[278], accessing data for diabetes care[279], quality of life reports for patients with prostate cancer[280], longitudinal data for post-kidney pediatric transplant patients[281], psychotherapy progress tracking[282], and integrating patient-reported outcomes into spine surgical care[283].

Despite the existence of many health related dashboards, there remains a need for additional research to more rigorously test their effectiveness. Some of the key topics for future research include metrics selection, relationships among metrics, and the ultimate question of whether dashboards provide sufficient benefits to justify the resource allocation required for adoption[284]. Validated survey instruments to measure the effectiveness of dashboards to support decision-making in healthcare have been developed[285]. One such instrument measures satisfaction with a web-based decision support system[286] and another examines the effect of dashboards on practitioner performance and patient outcomes[287]. There have also been studies examining the impact of clinical dashboards on workflow, care quality and patient safety in an intensive care unit environment[288], as well as in primary care environments and community health programs[289]. The latter demonstrated the use of dashboards that assess the short-term effects of health programs on performance indicators of interest to program managers – including awareness of ongoing programs, program enrollment, and cost per client.

In summary, the paradigm of a learning health system calls for leveraging the data infrastructure for tailoring analytics and messaging for end users with a goal to align internal stakeholders and minimize structural conflicts. This requires an understanding the different roles and their performance monitoring measures. What is also important is to benefit from designing user-centered dashboards containing visualizations that reduce the problem of information overload. There is also a need to use robust methods to assess the usability and impact of the dashboard on decision making. From the above review of the literature, we identify the specific knowledge gaps that warrant this study.

4.1.2 Knowledge Gap

Review of the related research in the space of data analytics and presentation in the healthcare space has provided the foundation for our research questions. First, we understand that within the paradigm of a learning healthcare system (LHS), any decision support system needs to be tailored to the end user. Second, we know from multiple studies that there are inherent structural tensions or conflict among clinic staff – including providers (physicians, nurse practitioners), clinical support staff (nurses, medical assistants), administrative staff (front desk, appointment schedulers, enrollment specialists) and clinic management (Chief Medical, Executive, and Operating Officers – CMO, CEO, COO), and other ancillary support staff members (such as technology support). By understanding the unique performance monitoring needs of all these distinct roles within a CHC healthcare delivery environment, a tailored information dashboard designed to support the monitoring of role-specific goals and priorities for a shared core mission may improve communication and mitigate internal staff conflict.

Research on dashboards has thus far been focused on either health indicators (public facing) or patient reported outcomes (patient or clinician facing). In contrast, management facing dashboards have largely been serviced by commercial vendors who have focused more on financial and/or quality measures with little emphasis on the health care operational characteristics that underlie and affect these measures. Operational measures have assumed greater importance in the era of patient-centered care delivery environment, with particular emphasis on operational efficiency and improved patient access to healthcare.

Reporting packages that are part of EHR systems, are often inadequate in meeting requirements of individual CHCs[290]. They are geared towards meeting the reporting requirements at a federal level in accordance with meaningful use incentives and grant requirements[291]. These reporting solutions are important to CHCs, but fall short of providing operationally meaningful analytics, let alone understanding role-specific needs within CHCs, user-centered design of the interface or visual analytics with real-time interactivity. Similar issues were uncovered as VHA made their foray into advanced analytics[292]. Beyond the challenge of dealing with information overload and data governance, VHA faced organizational challenges in not having a sound understanding of which reports and analyses are most critical to care delivery and which data elements need to be prioritized for standardization and validation at the national level. They have explicitly called for more research on understanding user needs, build evidence about the efficacy of providing dashboards to clinic staff, and for validated guidelines for the design of clinic dashboards. The VHA experience also highlighted the need for implementing self-serve BI tools that can enable automated sophisticated analyses without the user having to know how to write complicated executable query code for ad hoc analyses and reports. Dashboards that incorporate visual analytics and usability principles during the design and assessment have a potential to address some of these challenges. Lastly, post-dashboard implementation examination of the effectiveness of dashboard as a decision support tool has not been adequately studied in the era of modern visual analytics.

In summary, there is a gap in knowledge when it comes to understanding what are the various performance measures of interest to various roles within a community health center. This is at the foundation of a LHS platform that tailors analytics to stakeholders. Furthermore, existing CHC information solutions from EHR and BI vendors are more focused on providing mandatory reporting solutions as opposed to delivering dashboard visualizations that are designed, developed and implemented using user-centered principles. There is also a need for appropriate methods to assess the usability and efficacy of dashboards on decision making. In this research study, we apply usability principles and agile development methods to engage with stakeholders at community health centers to design, implement and assess visual analytics based dashboards to support operational performance monitoring and decision making.

4.1.3 Research Question

The overarching research question that addresses existing gaps in the literature and governs this study is: How can clinic-specific data be harnessed to promote reflection and decision making among varying roles in community healthcare delivery settings? Specifically: 1) What role-specific measures are effective for performance monitoring in health centers? and 2) What are the design and assessment needs for interactive dashboards of user-defined performance metrics for CHCs?

The work presented in this paper identifies the role-specific measures for performance monitoring in community health centers and delivers these measures via interactive user-center designed dashboard.

4.2 Methods

4.2.1 Background

The CHC dashboard design work is a part of a larger 3-year research project funded by the Patient Centered Outcomes Research Institute (PCORI). The larger project aims to improve patient access to healthcare at CHCs in three steps: 1) identifying common barriers and successful patient-centered strategies to accessing healthcare; 2) applying Delphi expert panel methods along with workflow and simulation modeling to identify feasible interventions to reduce patient barriers to getting needed healthcare and develop patient-centered appointment scheduling strategies; and 3) determining if implementing patient-centered interventions and scheduling systems can improve access to care, relative to existing approaches. As part of this larger project, a multidimensional data warehouse was developed using the EHR and various operational and financial data stores to support simulation modeling and development of performance monitoring process[293]. There are seven CHC organizations representing over 40 different clinical sites spread across the state of Indiana that are participating in the study. The clinics selected were such that they represent urban, semi-urban and rural settings and provide care to diverse patient population in terms of race, ethnicity, insurance types, and income levels.

4.2.2 Overall Approach

An iterative, user-centered design (UCD) approach was employed to create a dashboard that is tailored to the unique needs of CHC staff including administrators, providers and support staff (clinical and non-clinical). The design approach included a user needs assessment phase to determine the ideal collection of measures to support role-specific decision-making. User preferences for navigation and data organization were then evaluated using a high fidelity prototype in a second phase of our user-centered design approach. Together, these two phases informed the development of the final interactive dashboard that was deployed in the field at our study CHC site. Post-deployment, the usability of the dashboard was assessed and user perceptions about the value of the tool were collected. Following the needs assessment phase, there were periodic group meetings, where the lead author demonstrated the dashboard to the clinic staff. This facilitated rapid iterative modifications to the dashboard design.

The study approach shares similarity to typical phases of the systems development life-cycle from requirements gathering to deployment but applied the principles of the agile application software development[294]. Figure 19 shows the specific phases of the dashboard deployment, with the associated rapid iterations and associated tasks within each phase of this study.

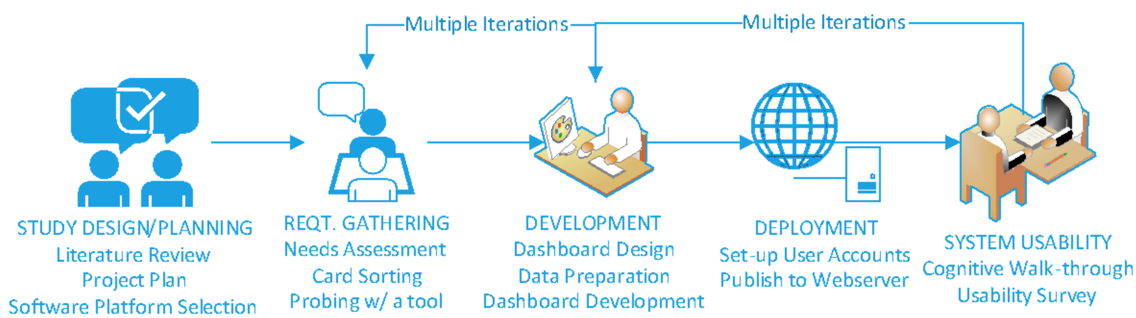

Figure 19: Dashboard Development Phases and the Associated Usability Evaluation

The specific details on each phase of the dashboard study from initial planning through usability assessment are described next.

4.2.3 Study Design

A high level project plan outlined the road-map of the dashboard implementation. This included, the selection of the participating health center, the timeline, the software platform to be used, and the overall approach to be followed. A baseline list of measures was developed based on prior work done by the researchers on re-designing health systems to improve access to care as well as performance measures discussed in the health services literature. Efficient face to face communication between the researchers and the clinic staff was built in throughout the development lifecycle. In addition, we incorporated short feedback loops and process adaptability as expected of an agile development methodology.

4.2.3.1 Participants

The input on design requirements of a data analytics platform was obtained from the seven CHCs participating in the larger study via clinic questionnaires, workflow observations and key informant interviews – all of which provided information on the challenges and needs of analyzing EHR data and generating reports. However, a formal sequential user-centered design process (interviews to uncover role-specific measures, card sorting exercise, cognitive task analysis and a SUS survey) was applied to only one of these CHCs (namely, CHC-O). There were multiple reasons for taking this approach. First, CHC-O was among the larger semi-urban multi-site CHC and had characteristics that were fairly representative of other CHCs and thus allowed us to developed a dashboard that was more generalizable to our other study sites. Specifically, CHC-O had four physical primary care sites that served both an urban and rural population, had a mix of MD and NP providers, and offered the wide range of services (medical, dental, behavioral health) and sites (including school based and mobile clinics). The second and equally important reason was that CHC-O was most responsive and eager about the potential benefits of an enhanced analytics platform. Additionally, CHC-O had also switched from an older EHR (client-server architecture based) to a modern cloud based EHR in mid-2015 thus offering a deeper perspective about differences in EHR reporting functionality when compared with our other partner CHCs. The characteristics of CHC-O are outlined in Table 6.

CHC Characteristic	**CHC-O**
# of Patients [Across all sites]	20,610
# of Visits [incl. Physician, NP, Nurse Visits]	63,108
Care Setting	Semi-Urban
Services Offered [Medical-M, Dental-D, Behavioral Health-BH]	M, D, BH
# of Primary Care Sites [excl. Mobile Clinics]	4
# of Providers [MDs and Nurse Practitioners], FTEs	4.5 MD; 11.5 NP

Table 6: Characteristics of the Participating CHC in the Dashboard Study

4.2.3.2 *Selection of Staff for Interviews*

The clinic leadership at CHC-O was asked to identify the various roles within the CHC and the appropriate personnel who should be interviewed to develop the list of performance measures that would feed into a dashboard. The different roles identified at CHC-O included: Administrators [CEO, COO, CMO]; Admin Support [Quality Assurance/Improvement - QA/QI], Operations Manager, Clinic Supervisor, IT Staff]; Providers [MD, Nurse Practitioner (NP)]; Provider Support [Medical Assistant (MA), Registered Nurse (RN), Scheduler, and Patient Care Rep (PCR)]. At least 1-2 key individuals representing each role were selected at CHC-O. Seven staff were interviewed in the pre-dashboard user-centered design phase, while eight staff were involved in the usability testing phase post-dashboard implementation. Six interviewees were common across both groups and in both cases there was adequate representation of the different roles within the CHC.

4.2.3.3 *Software Platform Selection*

Various potential software systems were evaluated, including open source applications: Elastic's Kibana[295] and D3.js[296] [both of which are based on JavaScript] and commercial data visualization software: Tableau[297], and Power BI[298]. The governing criteria for selection were: richness of visualization features; data security, web-based access for the end-user, ease of programmability [for sustainability], and one that could better leverage the Indiana University (IU) research infrastructure. The software that best met these needs was Tableau – a popular interactive data visualization software focused on business intelligence. The product queries databases and flat files to generate a number of graph types that can be combined into dashboards and shared over a network or the web[299].

4.2.3.4 *Requirements Gathering*

Several methods from the field of human computer interaction were applied in a lightweight manner during requirements gathering phase. First, the specific information needs of potential dashboard users were evaluated through contextual interviews. This was followed with an application of the card-sorting technique to define preferences for performance measures and information architecture design. Finally, a high fidelity prototype of the dashboard was tested with users and feedback was implemented iteratively in the design. About one hour was allocated for each individual session with the following break-down of time: 30 minutes for Contextual Interview; 20 minutes for the Card Sorting activity; and 10 minutes for the Prototype Dashboard evaluation.

4.2.3.5 Contextual Interviews for User Needs

Role-specific contextual interviews were carried out at the clinic site. Contextual interview is a critical method in gathering design requirements and involves naturally probing into users work in their own space (as opposed to being in a controlled lab)[300]. The purpose was to gain an understanding of the individual's decision making requirements and how it is currently accomplished, independent of the technological capabilities. The types of questions asked were along the following lines: Describe your role in the clinic? What decisions are you responsible for? Do you currently use data to make these decisions? If so, where do you get data? Is there any additional data/info that would help you with your decisions? How do you know you are doing a good job? How do you assess the impact of changes you implement? Participants were also asked to show the researcher, if possible, where they currently get information to make decisions in their work and assess their work performance.

Participants were probed with questions and encouraged to think beyond what is currently available to identify new performance measures. The interviewee was encouraged to think of any and all types of measures they felt would be useful in their work, without being concerned about the practicality of accessing such measures (in particular data availability). This allowed for an extensive list of measures to be gathered for this study. This process brought out new measures of interest, which were noted down on index cards to be used in combination with the predefined deck of cards (derived from apriori knowledge of the researchers) as part of the card sorting session. This allowed for data gained from contextual interviews to be quickly integrated into the iterative design of the dashboard. Field notes were also collected during the interviews and were reviewed by the lead author after the all the sessions were complete for important themes that crossed over different clinic roles.

4.2.3.6 Card Sorting for Information Architecture

The Card Sorting method[301] is an established technique in the area of user experience design and is used to develop the information architecture (the structure and organization of the information) for the dashboard. This technique is used to gain insight on how users mentally categorize and prioritize information[302]. Information architecture is an important determinant for the usability of system design[303], given that an effective information system enables the user to find what they need efficiently and accurately. Different users have different views and priorities regarding the information contained in the system.

Individual card sorting sessions were chosen over group sessions in order to focus on understanding the individual role specific measures, which can be very difficult to discern in a group setting. And since this was a small face to face study, we used index cards (the measures from the baseline list were individually typed up, while the measures emerging from contextual interviews were hand written). Each card contained the name of the metric, a description, potential filters for slicing/dicing the data, and the data refresh frequency. Table 7 shows the format of one of the measures that was presented to the user in the card sorting exercise.

Name of Metric	Description	Filters [Slice/Dice Options]	Frequency
Appointment Adherence	% of no-shows and cancellations	By Site; Provider; Appt Type; Appt DateTime;	Biweekly

Table 7: Example of a dashboard measure presented during the card sorting technique

Cards containing the list of measures from apriori knowledge and new ones uncovered in the preceding contextual interview step were presented to the user. To begin with there were 25 cards and that expanded to over 45 by the end of the needs assessment phase. The end-user was asked to sort each card into categories that made sense to them. They were also asked to assign a priority level - high, low, intermediate, to each measure.

All the measures that were uncovered from each interviewee were then organized into a master list. The data model was reviewed to assess if each particular measure could be implemented into the dashboard. Those measures that were supported by the existing dataset were implemented, while those that require development of new informatics tools to gather the supporting data were noted as such, but not implemented in the dashboard.

4.2.3.7 Prototype for Interaction Design

A high fidelity dashboard prototype, using sample EHR data from CHC-O, was built using Tableau. The dashboard prototype included the common visualizations such as bar graphs, line graphs, bubble charts, and a geo-map. Figure 20 shows the dashboard prototype used to obtain feedback on the design of the interactivity.

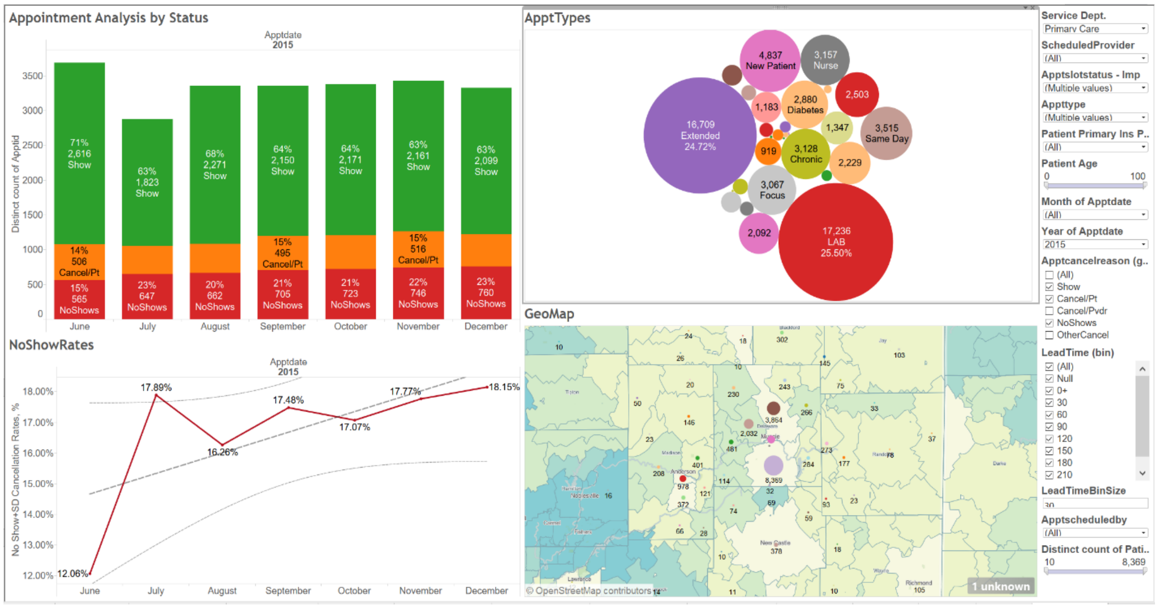

Figure 20: Prototype Dashboard used to obtain feedback on interaction design

Providing a high fidelity prototype (one that shares foundational similarities to the end product), allowed more reliable assessment of user preferences as well as uncovered new preferences for data visualization. After a demonstration of the prototype dashboard, participants were probed on the various interaction features including: preferred user interface elements (including input controls, navigational and informational components); preferred visual design (typography, images and colors); and preferred mode of delivery of the dashboard (static reports vs. interactive visualizations delivered from a web portal). These sessions were audio recorded and field notes were gathered. Audio recordings were transcribed to text and the first author reviewed the transcription and notes looking for common preferences across participants, which were incorporated into the dashboard design where feasible.

4.2.3.8 Development of the Dashboard

Tableau software (version 10) was installed on a Windows Server that housed the data warehouse containing EHR, operational and financial data from all CHCs participating in this study. All the visualizations were developed based on the analytic needs identified as being of value to the end user. De-identified data from the clinic health IT systems was pulled and loaded into the data warehouse. Each dashboard visualization was linked to one or more database tables. All table joins and other data processing tasks were performing within the Tableau interface to allow for rapid automated updates dynamically every time the source database was updated with the newer data set.

4.2.3.9 Deployment, Training and Iteration on the Design

To allow for secure web based access of the dashboard visualizations hosted on the IU Tableau server, IU affiliate accounts (designed for external collaborators) were created for the CHC staff. Once staff at the clinic were assigned user names and agreed to the terms of use, access to their clinic-specific Tableau dashboards was provided, along with the ability to interact with this new visual interface to their data.

Basic training was provided using a customized train the trainer model – in this case training was offered to a super user of the system identified within CHC-O. The training was carried out over two sessions. In the first session, the super user was provided high level training following which, the super user provided detailed feedback on each visualization to make sure it was truly represented their clinic operations. Following the completion of the first training session, changes to the dashboard were implemented and then in a second session the super user was provided with a demo of the revised dashboard. The super user then moved forward with training the other participating staff members at the clinic during an internal meeting.

The dashboards were published three months before the usability study was performed. During this interim period, the dashboard was refined iteratively based on feedback received during biweekly team meetings, called as AIM meetings. AIM stood for 'access is meaningful'. These were meetings with key members of the CHC staff and also served as a source of information on the impact of the dashboard on decision making. The AIM meetings included the members of the research team (including the lead author) and a AIM committee (n = 6) from the CHC. The latter included a committee chair (the 'champion') and five other staff members representing different functional roles at the CHC, all of whom were also part of the dashboard design and assessment group (n = 8).

4.2.3.10 System Usability Assessment

4.2.3.10.1 Usability Overview

Usability testing refers to the evaluation of information systems that involves testing of subjects as they perform tasks using the system. There are multiple methods to assess usability of a system. Survey questionnaires are often used as a primary method of assessing the impact of a system on work tasks, given their advantages (insignificant cost, ease of administering via the web, and quick analysis). However, relying on surveys alone provides limited information on the actual process of using an information system in performing complex tasks. In addition, the pre-determined questions in a survey instrument may not identify new or emergent issues that may not have been considered in the design of the survey. There are also issues surround poor recall when asking users to rate a system following gap in time from last use. The above deficiencies can be overcome when using survey questionnaires as one type of data collected in a multi-method evaluation of information systems[304].

For this study, two methods were employed to assess the usability of the dashboard: Pluralistic Cognitive Walkthrough (a task-based usability inspection method) and Usability Survey (a usability test method). Each of these is now described.

4.2.3.10.2 Pluralistic Cognitive Walk-through

Cognitive walk-through (CW) is a task-oriented usability inspection method which is used to identify usability issues in interactive systems and emphasizes simulation of cognitive issues, such as learnability, and user actions needed to perform those tasks[305]. A pluralistic cognitive walkthrough is a modification to CW in that it involves conducting review meetings where the end-user and the developer step through specific scenarios together and discuss usability issues that may arise[306]. An onsite pluralistic CW was performed with each of the dashboard eight users at CHC-O to inspect the usability of the dashboard. Each participating user was provided with a laptop and a mouse to perform the specified tasks.

The aim of the pluralistic cognitive walkthrough was to allow participants to interact with the dashboard while assessing both the usefulness and ease of use. Three specific tasks were developed and these related to: appointment adherence, productivity, and patient demographics. These represented common scenarios of user interaction with the dashboard. Each of these tasks had specific sub-tasks. Field notes were used to capture the user actions and system responses and corresponding usability issues. Each participant was encouraged to 'think aloud' while executing the required task so that the audio recording could be used as a reliable method to capture the session activities. Screen capture software was used to record the user interaction with the dashboard, which complimented the field notes taken by the researcher. The field notes informed the usability of the system in terms of user actions, system responses and corresponding usability issues (if any). The audio was later transcribed and the author read through the transcription to check for any missed information.

4.2.3.10.3 Usability Survey

The System Usability Scale (SUS)[307] is a well validated tool with references in over 1200 articles and publications[308]. The SUS is a Likert scale survey instrument which includes 10 questions to which users respond by ranking 1 to 5 based on how much they agree or disagree with the statement they are reading. 1 means they strongly disagree and 5 means they strongly agree.

It may be noted here that besides SUS, there are other options to judge the usability of a dashboard, such as SUMI (Software Usability Measurement Inventory), SUPR-Q (Standardized User Experience Percentile Rank Questionnaire), and QUIS (Questionnaire for User Interaction Satisfaction), but the unique advantage of SUS is that it is comparatively quick, easy and inexpensive, whilst still being a reliable way of gauging usability, particularly when used in early stages of a software tool such as this dashboard on small sample sizes.

The original SUS was slightly modified to include two sections. The first section contained the standard ten SUS questions on a five point Likert scale. The second section obtained background information on the participant: their role in CHC, length of service, comfort level with computer and the internet, willingness to spend time to learn to use the dashboard effectively, and perception on the dashboard allowing them to their job better. The survey was delivered online at the end of the usability assessment phase and the users completed it on the spot.

4.3 Results

Results are shared starting with the user requirement gathering and analysis phase of the dashboard development process. This phase involved contextual interviews, card sorting and interaction design evaluation using a prototype dashboard. We then describe the dashboard design, development and finally, the usability evaluation.

4.3.1 Results from the Needs Assessment Phase

We had originally identified 25 different measures that were found to be relevant for ongoing performance monitoring of a community health center. These measures were identified through a review of the literature, key informant interviews with staff at each of the seven CHCs on the topic of improving efficiency and access to care, a clinic questionnaire, and workflow observations.

The contextual interviews explored staff roles and their performance monitoring needs at CHC-O. The latter added 20 additional measures, bringing the total to 45 different measures. Using the card sorting technique, each participant selected measures that were important for their role, grouped them into categories, and then prioritized based on the order of importance.

While there was some amount of overlap in the measures selected, there were unique measures (‘dominant themes’) tied to specific roles in the CHC. For example, the dominant theme that related to the CEO of the clinic was a measure to assess how effective their internal communications and directives were. Specifically, the CEO wanted to increase collaboration and engagement with staff so they could take ownership of the organization values and priorities. While this was the dominant theme, there was also an interest in operational measures such as monitoring patient flow and throughput within the clinic (particularly when there are deviations from the norms), and appointment adherence (such as no-show and cancellations rates). The interest of the operations manager of the CHC were driven by a desire to improve provider schedule management and tracking supply/demand and productivity. The CMO was concerned about closure of all open issues (such as review lab results, medication refills, visit sign offs) and wanted to track provider productivity and adherence to scheduling templates. An MD provider was concerned about care team implementation and wanted to track appointment adherence, patient throughput and quality scores. The Nurse Practitioner's dominant concern related to reducing the patient check-in time. The Lead Medical Assistant (MA) of the clinic had great interest in knowing about all of the things they do right when supporting their providers as part of the care team. Her concern was that too often the feedback mechanism was used as a sounding board for negativity, while reinforcement of positive aspects was lacking. In terms of measures, the MA wanted to track schedule coverage, open patient requests and appointment adherence. The scheduler was concerned about not having enough appointment slots available to meet patient demand, while wanting to track the incoming phone wait times and appointment demand. Finally, the patient care representative (PCR) at the front office was interested in being able to see patient registration information in a single view rather than moving through multiple screens. She also wanted to be able to assess door to door time as well as receive feedback from providers on how their scheduling performance.

Table 8 shows this information grouped by the 8 different roles within the clinic, the top 3 measures of interest of them and the dominant theme that emerged from interaction during the interviews.

Role/Specific Title	**Top 3 Key Operational Measures**	**Dominant Theme**
CHC Leadership/CEO	Patient flow Status; Patient Throughput; Appointment Adherence	Communication Effectiveness
Operations Manager	Supply and Demand; Productivity; Continuity of Care Assessment	Schedule Management
CMO	Productivity; Appointment Lead Times; Patient Satisfaction Scores	Closure of Open Issues [Labs, Medication, visit sign-offs]
Provider (MD)	Appointment Adherence; Door to Door Time; QA/QI Scores	Evaluation of Care Teams

Provider (NP)	Appointment Adherence; Door to Door Time; Schedule Coverage	Reducing time from Arrival to Exam
Lead MA	Schedule Coverage; Open Patient Requests; Appointment Adherence	Feedback from Providers and Patients [including positive]
Scheduler	Incoming Phone Wait Times; Appt Demand; Cont. of Care Assessment	Non-Availability of Appointment Slots
Front Desk/PCR	Door to Door Time; Provider Feedback; Open Patient Requests	Quick view of the Patient Information in the EMR

Table 8: Role-specific measures of interest uncovered during the contextual interviews

4.3.1.1 Prototype Dashboard Evaluation

A sample dashboard containing four different types of visualizations was shown to the users and they were asked to use the dashboard and provide feedback in terms of: organization, interactivity, colors, graph types, filters for data slicing, access and delivery mechanism. The users were thus far only used to seeing tables and line charts in static spreadsheets generated internally or reports on financial performance. All users found the ability to interact with the dashboard by running scenarios using the various filters associated with each visualization to be beneficial. For example, the prototype dashboard allowed the users to see the graphic on no-show rates and see how those numbers varied for patients having different insurance types (such as Medicaid, private/commercial, uninsured/self-pay,) and visit types (such as lab visit, wellness, chronic). There was an interest in layering, i.e. the ability to clicking into the visualizations and see the underlying data. With regard to the mode of delivery of the dashboard, most users expressed interested in periodic email updates (as is typically the norm with traditional reporting solutions). However, when the users were made aware of the possibility of an interactive dashboard always accessible via a web portal, all users preferred the same (but still wanted to receive notifications via email to serve as a trigger to click and view the online dashboard).

4.3.2 Dashboard Design and Development

4.3.2.1 Overview

The visualizations were developed in Tableau Desktop version and then deployed on the Tableau Server hosted by the university. Tableau allowed development of visualizations by connecting to various data sources, including Microsoft SQL Server Database, which was used for the clinic data warehouse in this research project.

Visualizations were designed in such a fashion that each encompasses more than one measure. For example, a single visualization sheet, called ‘Appointment Analysis’ addressed 2 measures, namely appointment adherence (i.e. no show and cancellation rates) and the visit volume (equivalent to the demand for clinic services). Similarly, the scheduled hours available for appointment (equivalent to supply or capacity of providers), productivity (defined in terms of visits/hour), and the number of new patients, are 3 different measures available as a single visualization sheet, called 'Productivity'.

4.3.2.2 Core Measures

Each CHC had different data generation capabilities, which was governed by their EHR/Practice Management (PM) system being used, and the level of data extraction rights available. However, there are a core set of dashboard visualizations that was made available to each of the CHC organizations participating in this study. These 15 core visualizations are listed in Table 9.

Core Measures for Project	**Filter Criteria/Selections**	**Source**
Appointment Volume	Appt Status, Provider, Site, Date, Pt. Profile	EHR/PM
Patient Booking to Appt. Time	By Appt Status, Provider, Site, Time Period	EHR/PM
CPT Visits [CPT Code]	Provider, Site, Date	EHR/PM
Appointment Types	Provider, Site, Date	EHR/PM
Diagnostic Visits [ICD Code]	Provider, Site, Date	EHR/PM
Patient Throughput	Workflow Stage, Provider, Site, Date	EHR/PM
Productivity [Visits/hour]	Provider, Site, New Patients, Date	EHR/PM
Schedule Coverage/Utilization	Provider, Site, Date	EHR/PM
Patient Geo-Map	Superimposed on US Census Statistics	EHR/PM
Revenue, Expense	By Source, Year	IRS 990
Federal Quality Measures	Measure type, CHC, Year	UDS[309]
Patient Demographics	By Race, Gender, Age, Ethnicity, Language	UDS
Patient Characteristics	Insurance, Poverty Level	UDS
Clinic Services Offered	By Service Type	UDS
Patient Satisfaction Score	CHC, Pre and Post Intervention	Survey

Table 9: Core set of dashboard measures available for each CHC

Note that these 15 core visualizations represent 25 different measures. In addition, there were 20 additional measures which were CHC-specific. These depended on the EHR capabilities or the use of specific information systems (such as sophisticated phone systems). These measures include: *Charge Entry Lag, Enabling Services Coded, Phone Abandonment Rates, Auto registry calls, Phone Requests, Incoming Phone Wait times, Internal Phone Wait times, Continuity of care assessment, New Employee Retention, Patient Payments, Open Patient Request, Medical Records availability, Errors in Orders, Provider Feedback, Patient Feedback, Transfers to Urgent Care and ER Visits, Panel Sizes, Provider Complaints, Provider Orders, and Email communication types and volume.*

4.3.2.3 Visualizations

Developing the dashboard visualizations required transformation of the data using the data visualization software. Some of the common transformation activities included: creating calculated fields, embedding multiple if/else conditions, data binning, grouping picklists, creating aliases for source database fieldnames, and changing data types. There was a particular challenge related to productivity calculations. None of the four different EHRs used by the seven CHCs in the larger study had the capability to provide accurate productivity calculations, per the administration's needs. This made productivity calculations (required for provider incentive calculations and monitoring by the board of directors) a laborious process every month. The primary challenges were in computing the number of scheduled appointment duration/hours (including shows, no-shows, cancellations), while at the same not double counting the duration, if an appointment is double booked. To accomplish this, a custom de-duplicated database table was created to compute the appointment hours and then an 'outer join' was performed with the appointments table.

Screenshots of a select few data visualizations showing the variety of graphical displays and data filtering options are presented. Figure 21 shows example of landing page once the user logs into the web portal using their affiliate account. The landing page in this case shows all the visualizations pertaining to the statewide comparative data on community health centers, such as size of each health center by patient volume, patient demographics (age, race, ethnicity, insurance types), services offered, clinical quality measures, cost of care delivery, and site locations.

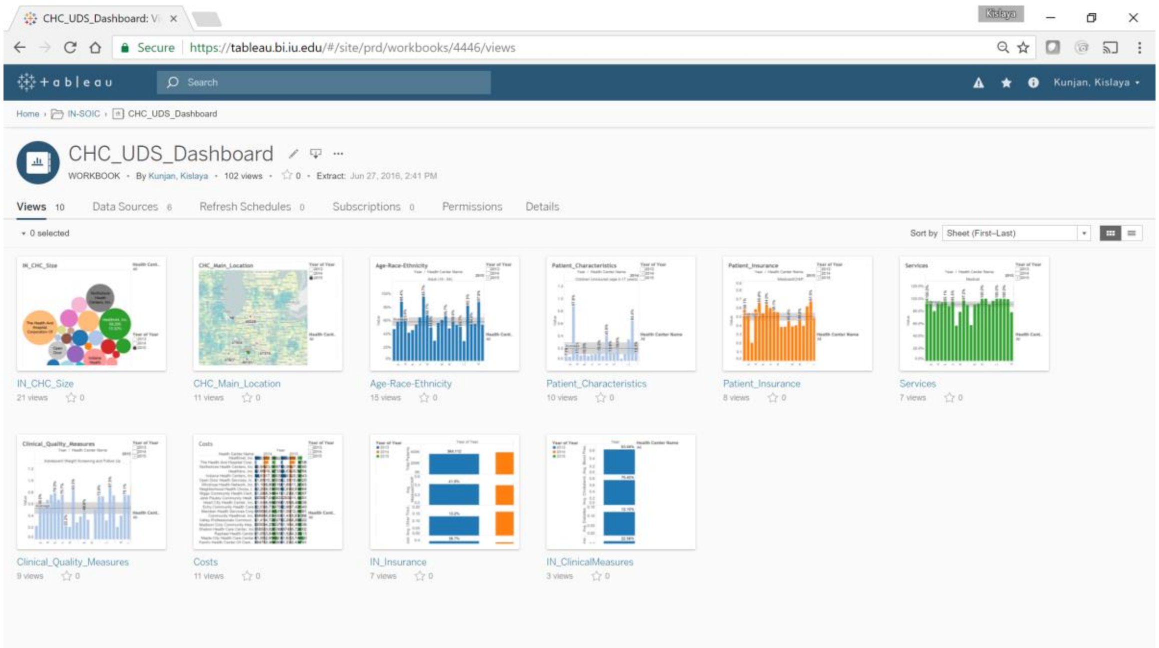

Figure 21: Web portal for secure access to the dashboard

Figure 22 shows the geo-map of the incoming patient population (user definable threshold of > 10 unique patients per zip code) for CHC-O, superimposed on the per-capita income by the zip code. This helped the administration to visualize the 'blank spaces' in the map where the clinic would focus its outreach efforts.

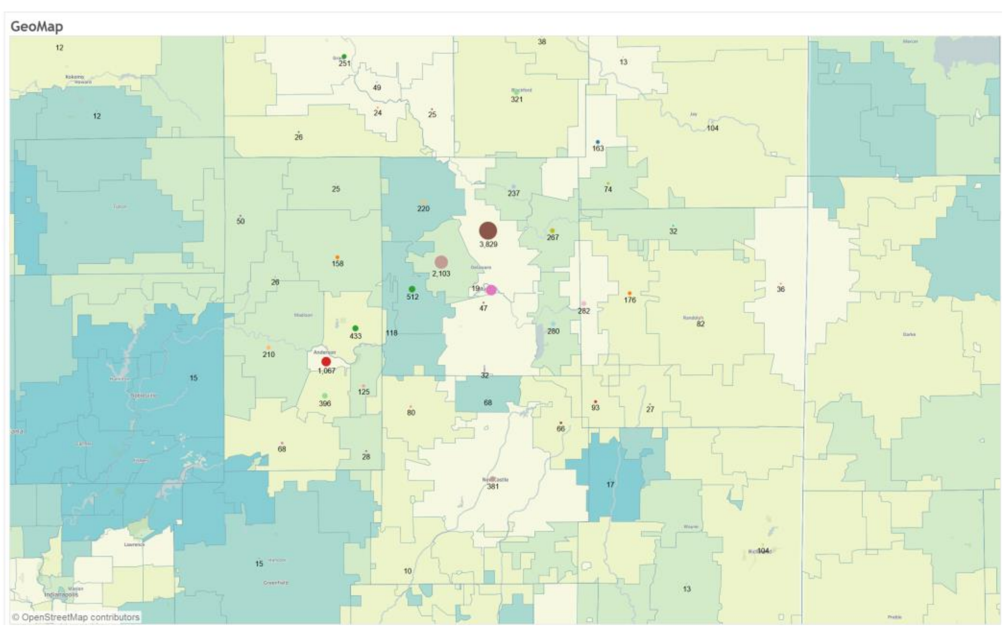

Figure 22: Geo-map of the patients at CHC-O layered on per capital income by zip code

Figure 23 shows the trend of ICD 10 codes (above an editable threshold count) for the primary patient diagnoses. The clinic administrators were able to spot a clear downward trend in the volume of patients with hypertension and diabetes starting in Jan of 2016, triggering a root cause analysis of this problem.

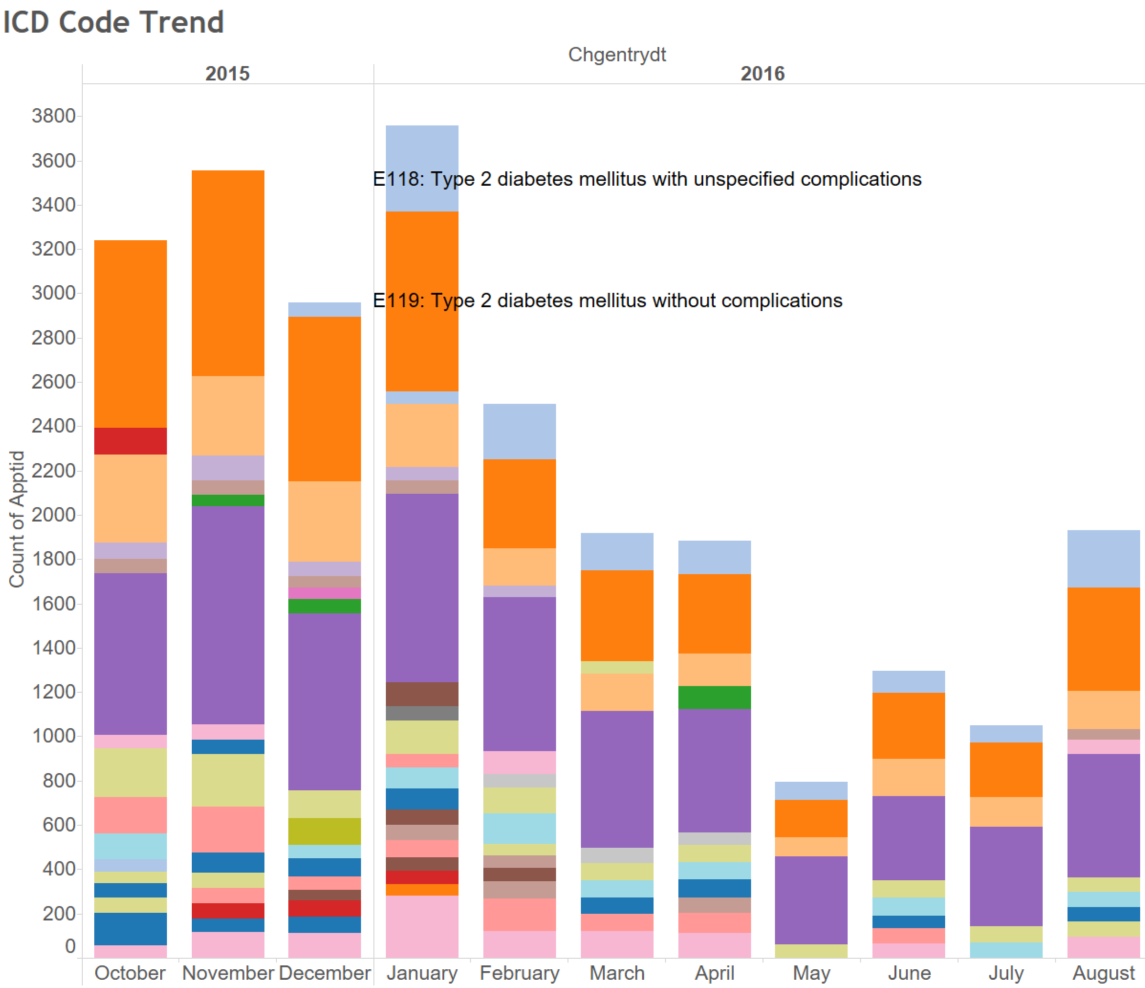

Figure 23: ICD 10 codes for the patient diagnoses

4.3.2.4 Design Iterations

We used an agile iterative design process for implementing the dashboard. Biweekly meetings with the clinic team insured that existing visualizations could be modified in rapid design cycles such that the needs of the end user were being integrated throughout the development process.

The needs assessment phase informed the measures of relevance as well as the interaction design, but the types of graphical displays/visualization for each measure was assessed post implementation of the dashboard. In some cases, the chosen visualization met the needs of the users, while in others the staff had specific requirements. As an example, the staff found the appointment adherence visualization in the form of a stacked bar to be appropriate, since it showed the numbers and percentages for the appointment status (show, no-show, cancellations by patient, cancellations by provider, reschedules) by the month. However, there were specific preferences in the use of the color scheme in the stacked bars. They specifically wanted red for no-shows to display their non-desirability, while green to be used for completed visits.

The staff did not want to see the productivity report graphically, preferring instead to see it in the form of a table so that the numbers could be shared with the individual providers. The CHC management were open about sharing productivity data of all providers to promote a healthy competitive spirit. They also wanted to see a service line level summary for presentation to the CHC board, and even had a preference in the ordering of the service lines (primary care, followed by family practice, OB, pediatrics, behavioral health, family planning, mobile clinics and urgent care). This ranking was based on the relative importance (patient visit volume and revenue) of the departments for the health center practice.

On the primary diagnoses by ICD 10 code visualization, within primary care the most frequent code was I10 – Essential (primary) Hypertension, followed by E11.9 – Type 2 Diabetes without complications. However, the CMO noted that there was variability in which particular ICD code was used by the provider. Taking diabetes as an example, there was interest in also being able to group the diagnoses at higher level. So E11.x series would be one group related to Type 2 diabetes irrespective of the complication (noted by the 'x'). After implementing that change, the reported incidence of Type 2 diabetes diagnosis was 3 times higher than any form of hypertension.

On the patient geo-map the clinic leadership wanted to see more granularity on the patient residential location (down to the street level as opposed to just the zip code level) in order to optimize their marketing and outreach efforts.

4.3.2.5 Impact on Decision Making

We sought to gain an insight into the how the dashboard was facilitating decision making at the CHC to better manage their operations. This assessment was done during biweekly staff meetings, wherein at least 30 minutes was allocated for reflection work with the dashboard. These discussions were also transcribed. A selection of use cases demonstrating the impact of the dashboard on decision making is described. Select quotes from the transcripts are included, where applicable, to offer context.

4.3.2.6 Enabling a culture of data-driven decision making

We found that the dashboard appears to be enabling a culture of data-driven decision making for CHC management and staff. The CEO of CHC-O informed the researcher team that the entire administrative staff is "*excited about this dashboard and what the data is telling them."* The CEO also stated that they now need to work on "*addressing the operational challenges being brought out by the data visualizations*." The Indiana Primary Healthcare Care Association (IPHCA) has been briefed on this 7-CHC data warehouse and dashboard platform. IPHCA believes that this work will directly support their plan to implement a state-wide community health information network to support greater data-driven operational and clinical quality improvements for all 27+ CHCs in the state.

4.3.2.7 Guiding the Implementation of Open Access Scheduling

Implementing open access scheduling, wherein a majority of the appointment slots are left open for same day walk-in appointment, is an important component of a patient-centered medical home (PCMH)[310] – a care delivery model which is patient-centered (responsive to individual patient preferences, needs and values) and coordinated through a primary care physician while being committed to quality and safety. The dashboard supports the implementation of such a scheduling system by showing the benefit in terms of a significantly reduced no-show rates for same-day appointments (< 8% no show rate for same day appointment vs. > 25% for appointments not scheduled the same day). The CMO stated that, "*we have always felt that scheduling appointments too far out will have a high no-show rate associated with them, but now your data [visualization] is proving that."* The CMO was also able to identify the variability among provider/scheduler compliance with implementing open access hours. After seeing the lead time distribution for providers not following the open access appointment template, and the associated 2-3x higher now show rates, the CMO remarked that *"our schedule is being sabotaged by our schedulers, with some tacit approval from some providers."* This led to convening of a special 'lunch-n-learn' session with all providers to emphasize the importance of open access scheduling in not only reducing no-show rates, but also in streamlining clinic operations and improving patient satisfaction.

The lead scheduler at CHC-O, asserted that the dashboard "*will be very helpful in doing my job better*." In particular, she commented that knowing the open access hours and the associated call volume will help her with '*planning call center staffing needs."* When looking at the data visualization, the staff also noticed an unusually high no-show rate during the period of

January, 2016 through February 2016. Upon identifying this spike in no-shows, they were able to reflect and understand this increase knowing that their automated appointment reminder system was temporarily dysfunctional during that period which was the reason identified for this spike.

4.3.2.8 *Reporting of Productivity*

CHCs typically measure productivity based on a calculation of patient visits/hour. Like most of CHCs in our study, CHC-O performs this computation manually using spreadsheets. Not only is this a time consuming process, it was found to be prone to errors. The researchers were able to automate the calculation and presentation of the provider productivity data based on an extract of their EHR data. We worked with the clinic operations manager to validate the data (particularly identifying what all visits counts as billable) and appropriately accounting for the patient contact hours. The CMO stated that their EHR vendor had not been able to create a productivity report that met their needs and applauded the work done by the researchers in "*providing an automated accurate productivity report*." The IT manager went far as suggesting that *the "the productivity report design shown in the dashboard should be copyrighted!"*

4.3.2.9 *Patient Flow Tracking*

Tracking the patient flow within a CHC from arrival through departure is critical for streamlining CHC operations. Some EHRs are better than others in the tracking capabilities. The EHR used by CHC-O was among the better EHRs that provide good tracking functionality, but staff were still unable to effectively extract meaning from the data. The research team developed a patient flow tracking visualization, which took the form of an interactive storyline we called *'5-stage patient throughput.'* The storyline combined multiple workbooks on patient flow in CHC-O in the form frequency distribution of time duration, including the intermediate wait times. Data could be filtered by the responsible provider/clinic staff. The clinic manager noted that every time she performed process audits or tried to identify bottlenecks, she had to manually follow patients to track the time intervals from arrival through departure. The clinic manager commented that "*but now with your dashboard I can analyze all of this, without having to go through a manual process*." The Lead MA of CHC-O, commented that *"I knew that Dr. [---] took a long time completing his patient encounter, and now your chart is confirming that."* She decided to share this data with the concerned provider.

4.3.3 System Usability Study Results

The usability of the dashboard was assessed using two methods: pluralistic cognitive walk-through and a usability survey. The results from each of these methods are presented next.

4.3.3.1 *Results from the Cognitive Walk-through*

There were three tasks and associated sub-tasks to perform the cognitive walkthrough. Prior to performing these tasks, the user signed into the web portal containing the dashboard for CHC-O. There was minimal guidance offered for completion of the various tasks. There were eight users, of which five were interacting with the dashboard for the first time, while three out of the eight users had used the dashboard multiple times in the month prior to this session. Except for one user (who had joined CHC-O a month prior), all else were familiar with the project and had seen the author perform the demo during earlier team meetings. First time users were intentionally asked to participate in the task analysis to assess the learnability and level of training warranted. Field notes, screen captures, and video/audio recordings of each cognitive-walk through session were reviewed and the findings (expected user actions, system response, and problems experienced by the users) are presented below, organized by each task category and sub-task (framed as a question). For each sub-task, we state the expected user interaction with the dashboard, the system (dashboard) response, and usability problems, if applicable.

Task 1 relates to appointment analysis and the associated visualization (Figure 24) illustrates the user interaction with the dashboard. On the right are the various filter selections to run various interactive scenarios.

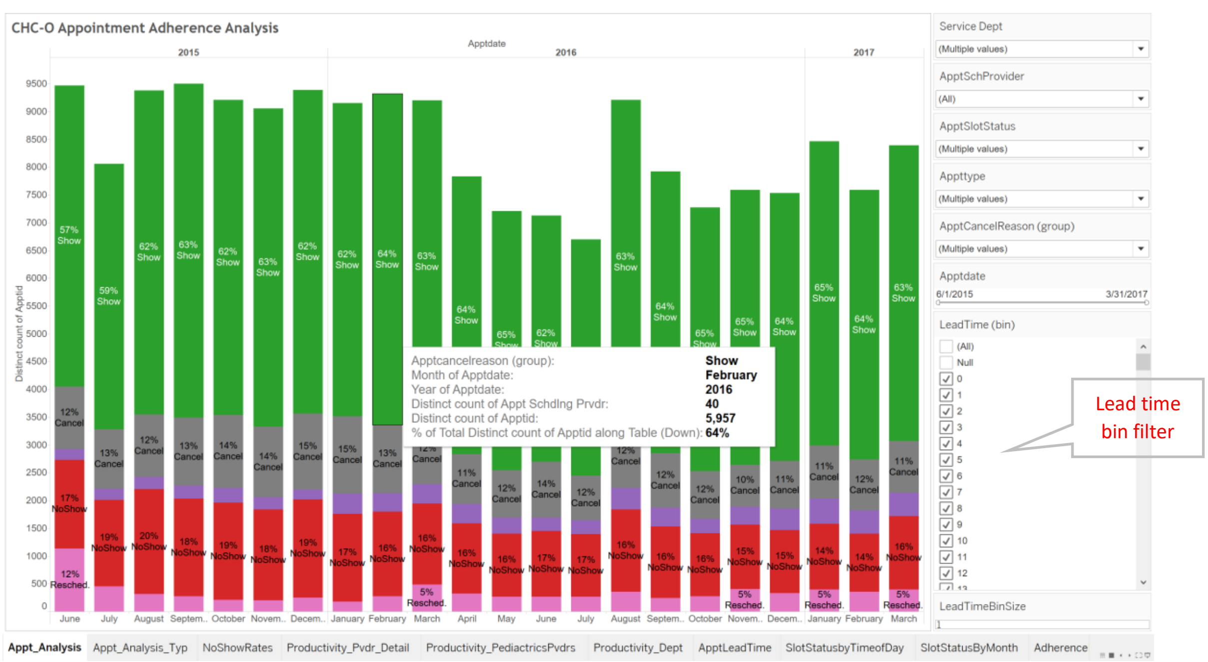

Figure 24: Visualization on Appointment Analysis

Task 1: User analyzed appointment adherence among primary care patients

- Sub-task 1: What was the no-show and cancellation rate during June 2016 within primary care?
    - Expected User Interaction: The visualization tab, labeled as 'Appointment Analysis' showed the appointment adherence (complete, no-show, cancellations, reschedules) for each month of the year since January, 2015 in the form a stacked bar graph. The user would notice the percentage of no-shows and cancellations for the month of June, 2016 referenced in the horizontal axis.
    - Dashboard Response: The percentage numbers displayed on the stacked bar for each month. Additional information (such as number of patients) shows up upon hovering or clicking over the particular bar for the month of June. No-show was 20% and cancellations rates was 14% (cancellations were summed up by provider and patient)
    - Problems: The context sensitive display of the text box upon hovering over the no-show and cancellation zone on the bar graph was not obvious and bothersome to a (first time) user (refer again to Figure 24 to see the context sensitive text box that shows up upon hovering over stacked bar). The others did not express any concern with that and saw its purpose.
- Sub-task 2: What was the impact of scheduling appointments with a wait time of ≥ 30 days on the no-show and cancellation rates?
    - Expected User Interaction: The same visualization tab, labeled 'Appointment Analysis' showed the appointment adherence (Figure 26). There was a filter labeled 'lead time', which had a multi-select picklist with bin size increments of 30 days. The user had to select all and then unselect the first option for the lead time that corresponds to 0-29 days.
    - Dashboard Response: The visualization updated automatically to show the revised no-show and cancellation rates. No-show rate went up to 28% and cancellations rates went up to 24% when the lead time bin sizes selected were ≥ 30 days.
    - Problem: The method of de-selecting all options in a multi-select filter was not immediately apparent (the user is required to check 'All' and then uncheck the 'All' box to deselect all options). Secondly, the bin size categories selection was not intuitive. So, if there were 5 bins for the appointment lead time with a bin

size of 30 days then it was not readily clear what range of days, does 0, 30, 60, 90, 120 and so on meant (the lead time bin filters are displayed on Figure 24 on the bottom right). When asked to select days greater than equal to 30 days, the users were unclear if they should include or leave out 30. Bin sizes are forward looking, i.e. 0 means 0-29 days, 30 means 30-59 days, and so on, with each bin label designating the lower limit of the range of numbers that is assigned to the bin (with the lower limit being inclusive).

Task 2: Find out the provider productivity in primary care

- Sub-task 1: Which provider(s) had met their productivity targets [>= 2.5 visits/hour] for June, 2016?
    - Expected User Action: The user had to go to the visualization tab, labeled as ‘Provider Productivity’ which showed a table with the various measures related to productivity (number of billable visits, scheduled patient contact hours, visits/hour, number of no-shows, number of cancellations, and number of new patients) for individual providers grouped by department (primary care, pediatrics, OB/GYN, behavioral health, family planning, urgent care). The user had to go to the pane for the primary care department and find those providers whose productivity exceeded 2.5 visits/hour.
    - Dashboard Response: There was only 1 out of the 15 primary care providers who exceeded 2.5 visits/hour in the month of June.
    - Problem: None
- Sub-task 2: What was the complete visit count and productivity of the primary care department?
    - Expected User Interaction: Within the same ‘Provider Productivity’ tab, the user had to look at the grand total of visits for the primary care department for that particular month. The underlying calculation for productivity was total visits divided by total scheduled patient contact hours.
    - Dashboard Response: The department productivity was displayed in the summary row.
    - Problem: None

Task 3: Analyze the patient demographics and disease conditions

- Sub-task 1: Which zip code brought in the second largest number of patients to CHC-O?

- Expected User Interaction: The relevant visualization tab was 'Patient Map,' which was a geo-map of the patients based on their residential zip code. The count of patients from each zip code was displayed as a bubble (whose size based on the number) along with the textual information on the patient count and zip code. This was superimposed on top of a per capita income map by the zip code.
- Dashboard Response: The zip code with the second largest number of patients was displayed on the map, which was from an adjoining zip code to where the primary care site was located.
- Problem: None

- Sub-task 2: What was the most common primary diagnosis for urgent care visits in January, 2015?
  - Expected User Interaction: The associated visualization tab was labeled as 'ICD Codes,' which was a stacked bar graph of the count of appointments by the primary patient diagnoses and month on the horizontal axis. The user had to use the department filter to only show the count of the diagnoses for urgent care.
  - Dashboard Response: For the month of January, the most common primary ICD 10 diagnoses in the urgent care department was 'acute sinusitis' with 'acute pharyngitis' being a close second.
  - Problem: None

All users were able to perform the assigned tasks without much guidance. This is not as surprising considering that, for one, the users were all comfortable with computers and the web (their EHR is cloud-based) and secondly, the data visualization software interface is known to have an easy user interface.

4.3.3.2 Common Usability Challenges

Beyond the task specific usability issues, there were some common usability challenges encountered, particularly by the first time untrained users (five out of the eight participants). For one, the context sensitive text box upon hovering across the graphical sub-objects came across as a surprise to many participants (and annoyance to one). However, the participants later realized the value of the large hovering text in that they would otherwise be required to visually match the graph elements to a legend. Secondly, the participants struggled to easily navigate the various tabs corresponding different visualizations on the web portal. There was an inconspicuous scroll bar on the top that the first time users struggled noticing. This made navigating the tabs easier once they found this scroll bar. Three out of the eight users, were actually returning to the home page of the dashboard and then opening up the particular visualization, instead of clicking on the adjacent navigations tabs. Thirdly, two users tended to overuse the filters (for example, selecting a single month from the filter, instead of reading the values straight off the graph showing data for all months). Most of these initial issues were resolved with continued use of the dashboard for the remaining tasks and suggestions made by the researcher.

4.3.3.3 SUS Survey Results

Soon after completion of the cognitive walkthrough, the users were asked to fill out the system usability scale (SUS) survey. As discussed earlier, the survey had 2 sections. The first section contained the standard ten SUS questions on a five point Likert scale. The second section obtained background information on the participant (their role in CHC-O, length of service, comfort level with computer and the internet, willingness to spend time to learn to use the dashboard effectively, and perception on the dashboard allowing them to their job better). For the latter two questions, seven out of the eight answered a score of five (strongly agree).

For the SUS survey used for the dashboard, the average score turned out to be 81.25 (range: 50 to 100 and standard deviation of 17.1). All SUS and supplementary questions with the individual scores by the role and the average scores are shown in Table 10.

SUS Questions (↓) and Role/Job Title (→)	**Scheduler**	**Grants & Comms. Mgr**	**CMO**	**Lead MA**	**Operations Mgr**	**Clinic Supervisor**	**CEO**	**Compliance & IS Mgr**	**Avg.**
I think that I would like to use this system frequently	5	5	4	5	5	5	1	5	**4.4**
I found the system unnecessarily complex	1	1	1	2	2	1	3	1	**1.5**
I thought the system was easy to use	5	5	4	3	4	5	3	5	**4.3**
I think that I would need the support of a technical person to be able to use the system	2	1	2	3	2	1	5	1	**2.1**
I found the various functions in this system were well integrated	5	5	4	4	4	5	4	5	**4.5**
I thought there was way too much inconsistency in this system	1	1	1	3	2	1	1	1	**1.4**
I would imagine that most people would learn to use this system very quickly	5	3	3	2	4	5	2	4	**3.5**
I found the system very cumbersome to use	1	1	1	2	3	1	3	1	**1.6**
I felt very confident using the system	4	4	4	3	4	5	4	5	**4.1**
I needed to learn a lot of things before I could get going with this system	2	1	1	3	2	1	2	1	**1.6**
Raw SUS Score	**37**	**37**	**33**	**24**	**30**	**40**	**20**	**39**	**32.5**
Computed SUS Score	**92.5**	**92.5**	**82.5**	**60**	**75**	**100**	**50**	**97.5**	**81.3**
<u>Supplementary Questions</u>									
How many years have you worked here?	13	12	4	5	6	0.2	0.2	18	**7.3**
Please rate your ability to use computers and the internet	4	5	4	4	4	4	4	5	**4.3**
I am willing to spend time to learn to use the dashboard effectively	5	5	4	5	5	5	5	5	**4.9**
I believe the dashboard will allow me to do my job better	5	5	4	5	5	5	5	5	**4.9**

Table 10: Results of the SUS survey and the supplementary information

There are guidelines to interpret the SUS score. The average SUS score is 68 from 500 studies that have used this test over the years. A recent research[311] has shown that two questions in the test also evaluate learnability while the remaining eight assess usability. Studies that score 80.3 or higher are given an A grade and interpreted as 'People love your site/application and will recommend it to their friends/colleagues'.

4.4 Discussion

4.4.1 Summary of Key Findings

The current study on the design and implementation of a CHC and role-specific dashboard is part of a larger research project aimed at improving efficiency and access to care in seven CHCs across the state of Indiana. It leverages multi-EHR clinic data warehouse and CHC-specific workflow models developed as part of the larger study. In this particular study, key informant interviews of CHC staff was performed to uncover the information needs and gaps in meeting those needs to support clinic operations. Then a user-centered design (UCD) methodology was applied which included open card sorting and interactive assessment using a high fidelity prototype. Following completion of the needs assessment phase, dashboards were designed and deployed on a secure web portal. Agile and iterative design methods were applied to refine the dashboard to efficiently meet end user needs. Finally, a study to assess the usability and effectiveness of the dashboard for supporting decision making was performed. Each aspect of the study, namely, needs assessment, dashboard implementation, and finally usability testing is further described.

Needs Assessment: We started off with a baseline set of 25 dashboard measures based on a review of the literature and our prior experience working with health centers to improve access to care. As part of the needs assessment phase, the number of measures increased to 45. Importantly, the study discovered unique and hitherto unknown measures based on the specific role of the users in the CHC, covering a wide range of roles from the CEO to the patient care representative at the front desk.

Dashboard Implementation: The visualizations were all interactive and combined multiple measures into one. There were measures for which existing information systems were inadequate, such as surveys on staff satisfaction and team work. The study is informing the development of appropriate informatics tools to capture that information in an automated and time efficient manner to support future implementation in a CHC dashboard. The dashboard clearly elicited significant enthusiasm by all staff within the CHC. Reports and insights, previously not easily available were now made possible through the dashboard (the productivity analysis and patient throughput visualization embedded in the dashboard being prime examples).

Usability Testing: The task analysis brought forth some interactivity and navigation related issues for first time untrained users, but resolved with continued use. The usefulness of the dashboard as tool for operational decision making was ascertained through frequent meetings between the researchers and the clinic staff. In general, the dashboard acts to support a culture of data-driven vs. emotion-driven decision making. Specific areas of impact included, guiding the design of open access scheduling systems, automating the assessment of provider productivity, improving patient-throughput, and supporting patient outreach efforts. Ranking of the system, per SUS scores, shows the participants found the dashboard to be both easy to learn and usable.

4.4.2 Limitations of the Study

The study has certain limitations. The first relates to the generalizability of the findings from the dashboard design. The study is based on knowledge gained from working with seven CHC organizations with over 40 clinic sites in the state of Indiana. However, a great majority of the work to specifically inform the dashboard design development was done with a single clinic, CHC-O, one of the large CHCs in the state. All six of the other CHCs received a similar dashboard as CHC-O, but it there was no design work or testing completed at those sites. CHC-O was selected because it was representative of other CHCs in terms of its characteristics: mid-size with rural and urban sites, serving a racially diverse patient population, and having mixed providers (NPs and MDs). In comparison with other clinics, CHC-O already employed modern information and communication technologies (cloud based EHR/Practice Management System, advanced phone systems) and management was particularly open and willing to share its data and work with it to form insights into clinic operations. Nevertheless, each CHC is unique in their own way that extends beyond the type of the patient population they serve. There are subtle differences related to organization culture and leadership, and the data infrastructure. These warrant additional user-centered design work at other CHCs.

Secondly, the sample size of the usability study was small (n = 8). While this informs the design of the dashboard, it limits the statistical significance of the findings. A larger study involving all seven CHCs and their key staff members [n ~ 50] will shed more light on the findings and allow for high powered statistical analyses.

There is also the issue of the data latency involved with the data warehouse linked to the dashboard. There are practical limitations on how frequently the source data from the EHR could be queried, extracted and loaded into the data warehouse. While the current frequency (fortnightly), was adequate for many of the measures being tracked, there are some measures (such as ones that track the real-time patient flow through the clinic) which would benefit from a live connection to the EHR database. However, this would require creating secure web services based interface with the EHR vendors.

The study represents an early phase implementation of the dashboard. Ongoing assessment of the dashboard is needed to evaluate its continued utilization, through means such as user log files in the server. The inclusion of key operational measures such as appointment adherence and productivity in the dashboard will encourage its continued use, while visualizations such as patient geo-maps will help guide CHC outreach efforts.

Finally, there is the issue of sustainability of the dashboard beyond the larger project end date. Currently the data warehouse and dashboard is hosted on an infrastructure hosted by the research team, which necessitates the users log in to a separate web portal. The ongoing plans to set up a health center controlled network by state primary care association may allow for continued provision of data analytics as a service to the enrolled CHC organizations.

4.4.3 Study Implications

We describe the implications of this study in terms of the following areas: a) supporting a learning health care system; b) uncovering clinic role-specific measures; and c) implementing a user-centered dashboard

Supporting a LHS: The widespread adoption of EHRs continues to make increasing volume of health care information available in computable form as well. Secured and trusted use of these data is the cornerstone for health care transformation[312]. The learning health care system framework when applied to health center redesign is based upon leveraging routinely captured EHR/PM data to continuously improve their clinic operations and quality measures. One of the key components of that framework is the delivery of tailored messages for stakeholders to support effective decision making. This study leverages a multi-CHC heterogeneous data warehouse developed and validated by the researchers to support tailored operational decision making.

Clinic Role-Specific Measures: Structural conflicts among internal stakeholders leads to poor alignment of goals in an organization and healthcare is no exception. While this is a multifaceted problem with many reasons, studies have shown that this conflict originate from differences in expectations and desires regarding data, interests, structure, relationships, and values[313]. An understanding of the different roles within a health center and their performance monitoring needs can help alleviate some of these issues. Research has not yet shown how information systems like dashboards may be designed based on an understanding of role-specific decision-making within ambulatory health centers. Typical roles within CHCs that benefit from a dashboard include: clinic administrators (such as CEO, COO, CMO), support staff for administrators (such as clinic site managers, quality assurance, analysts, front office staff), providers (MD and Nurse Practitioners) and support staff for providers (such as nurses, medical assistants, schedulers, and patient care representatives). The study has tried to uncover the performance monitoring needs and dominant themes of personnel in the front office to the CEO at a large semi-urban health center.

User-Centered Dashboard: Management information systems have been studied since the 1980s[314]. Tools for rapid data analytics and decisions support within clinic settings are increasingly becoming important even as the analytical processes have become more difficult. Inadequate levels of stakeholder engagement in implementation of integrated data analytic systems is one of the barriers to benefiting from such systems[315]. Visual analytics has been noted as an emerging discipline with significant promise in addressing many of the information overload challenges and healthcare data-driven decision making inherent in multi-dimensional data analytics and presentation[316]. One of the key contributions of the study beyond the understanding of the clinic role-specific performance measures of interest is how to engage with the stakeholders in implementing data visualizations to support decision making.

The study prioritized a core set of measures that were implemented in the dashboard for which data was available within different EHR data models, and other operational data stores like phone systems. However, there also were several performance measures that emerged from the study for which data does not currently exist or is not easily extractable. Another key contribution of the study is that it informs the development and implementation of tools that can enable the capture of such data as part of routine clinic operations. These can then be part of an enhanced performance dashboard of the future. Even though this dashboard was designed to be clinic-facing, the study had a patient advocate whose perspective enriched the design of the dashboard visualizations. The methods used in this study also can inform important work on developing a patient-facing dashboard of clinic measures.

4.5 Conclusion

The learning health system paradigm calls for a need to understand clinic role-specific measures of interest and then deliver those measures in the form of interactive user-centered visualizations. Examining underlying information systems as a means to support rapid decision making is essential in the resource constrained environments prevalent in community health centers where meaningful information can drive important change to improve access to care. A user-centered agile system development methodology was employed to design and develop a dashboard that addressed the direct needs of the end-users at community health centers. The dashboard was delivered over the web and linked at the back-end to a data warehouse that is refreshed periodically. The usefulness and usability of the dashboard was assessed using multiple methods of usability evaluation, including, cognitive walk-through and a usability survey. The study found the dashboard to be easily learnable and highly usable. Larger studies on multiple CHC organizations in the state and beyond can improve the generalizability of the findings and guide development of health informatics tools.

CHAPTER 5: CLOSNG REMARKS

5.1 Overview

Community health centers (CHCs) play a pivotal role in primary healthcare delivery to the underserved and have been at the forefront in the adoption of health information technologies. An in-depth understanding of the information needs of these organizations in the post HITECH era is vital to designing an information architecture that is appropriate for CHCs as they seek to improve efficiency and access to care. It may be pointed out here that information needs arise when people or organizations recognize gaps in their state of knowledge and wish to resolve that gap by applying the needed information[317]. CHCs often lack the resources and data infrastructure to benefit from a modern analytics platform that can support clinical, operational and financial decision making across the practice. Delivering patient-centered care has assumed vital importance at CHCs in light of recent federal legislation, and one of the key components of that is improving patient access to appointments. In particular, greater open access scheduling (wherein a portion of the schedule is left open for same-day appointments – a use case for this research) can potentially overcome the challenges of inefficiencies and poor access associated with traditional scheduling protocols, but also require careful optimization of the provider schedule supported with an overhaul of the data infrastructure [318]. There is inadequate understanding of the analytic needs and in particular patient and appointment-related variables that can predict the adherence of patient visits in community health center (CHC) settings. While no-show modeling has been performed in a variety of settings, there is great variation in the significant predictors of appointment adherence depending on the care settings[319]. Notably, a prior study in an academic medical center did not find lead time for an appointment to be a significant predictor of no-show behavior and that implementation of the concepts of advanced access scheduling did not directly result in a reduction in no-show rates[320]. There is also a lack of research on how EHR and practice management data can be presented to help CHC staff align with the core mission, goals, and priorities of the health center [321]. Little is known about user-specific needs among the various CHC staff roles and how dashboards can be designed that address those needs in an effective way [322].

The overall hypothesis driving this research has been that a user-centered data analytics platform would support implementation of interventions aimed at improving efficiency and access to care within the framework of a learning health system (LHS). Within the LHS framework, the CHC stakeholders must find the analytics platform to have content validity (the extent to which the measures represent all facets of the given construct) and support organizational leadership and staff buy-in. With that in mind, this research has sought to understand the information needs within CHCs, build a multidimensional data architecture and analytics platform to support that need, and lastly incorporate role-specific analytics into a user-centered interactive dashboard to serve as a novel decision support tool geared towards improving efficiency and access to care. Figure 25 shows the LHS framework showing the three RQs and as well as the outcomes (text in blue) from the application of the framework.

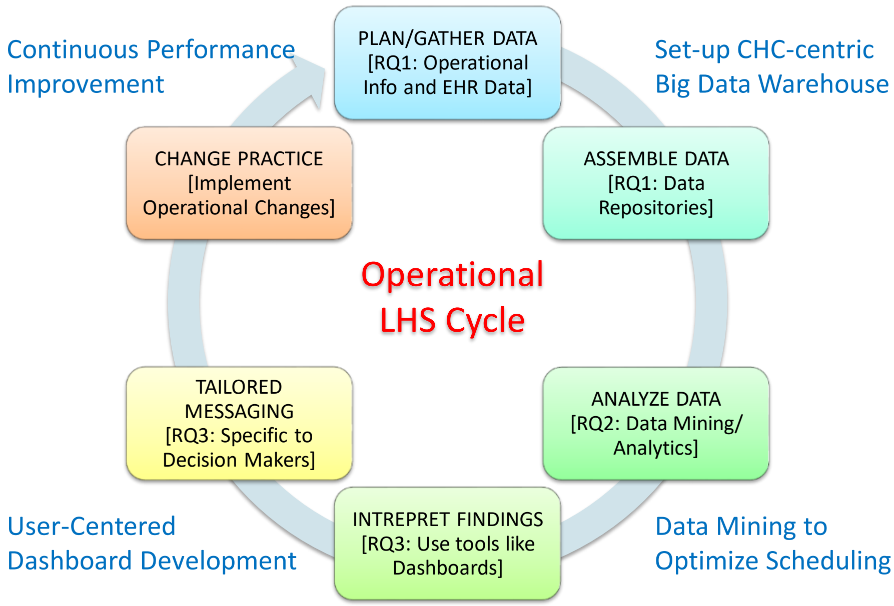

Figure 25: Application of the LHS framework Research Summaries

There were three research questions (RQs) that formed the basis of this research. We re-state each RQ governing the individual chapters (2, 3 and 4) and summarize how the results from each of these chapters addressed gaps in the literature as well as identified areas for future research in health informatics.

5.1.1 The Information Architecture

The first research question was referred to as the 'information architecture'. RQ1 had two parts to it: 1) What is the information needs to support operations at CHCs? and, 2) What is the optimal data architecture to support those needs?

Chapter 2 of this document addresses RQ1. We assessed the information needs of seven CHC systems across the state of Indiana, using a detailed information questionnaire, workflow observations (that leveraged the rapid ethnography method) and semi-structured interviews. We found that CHCs faced numerous resource challenges in leveraging electronic health records

(EHRs) to support improvements in clinical, operational and financial outcomes across the practice. Based on the unique constraints of CHCs and their distinct work and information flow, we designed and implemented a multi-CHC data warehouse architecture. The clinic data warehouse (CDW) comprised multi-EHR databases that integrated patient-level EHRs along with operational, financial and quality data stores. The patient level data serves as a backbone of the data warehouse and has been extracted from each CHC's EHR system with data dating back to 2010. The various categories of the EHR data included: patient profiles, appointment/encounter details, diagnoses, labs, immunizations, observations, vitals, labs, medication, and billing. Additionally, clinic operational and financial data was obtained from the CHCs through their audited financial statements, tax filings and annual UDS (Uniform Data Systems) submissions to the US Bureau of Primary Health Care.

This chapter detailed the rationale for the research project, the information needs, the data architecture employed, the content of the data warehouse, and the challenges experienced and strategies used in the development of this repository that may help other researchers, managers and leaders in health informatics. The resulting multidimensional data warehouse is highly practical. It provides a foundation for a community health information network (CHIN) complimented with wide-ranging healthcare data analytics (RQ2) to support the community health center research enterprise.

5.1.2 The Data Analytics

The second research question was referred to as the 'data analytics', with the primary research question being: How can clinic-specific analytics guide open access scheduling. There were two sub-research questions, each with an associated hypothesis. The first sub-RQ being, how can predictive analytics be used to optimize appointment access and in particular, what is the association between the time dimension of appointments and patient adherence to their appointments. The second sub-RQ being, how data analytics can be used to better evaluate provider productivity. We analyzed large scale appointment data on a multi-CHC data warehouse using a combination of decision tree, logistic regression, visual analytics, and n-gram modeling to examine the patterns of patient volume, missed appointments and optimal capacity utilization with open access scheduling interventions. We investigated three uniquely different CHCs that cared for a diverse patient population and operate in rural, urban and semi-urban areas, respectively. Analytical results showed that lead time (the duration between the time when the appointment was made and the scheduled appointment time) was notably the most significant predictor of appointment adherence. Patients scheduled the same-day for their appointment were at least three times more likely to show up for their appointment when compared to patients with non-same day scheduled appointments. The knowledge discovered from the analytics was applied to the optimization of open access hours, which subsequently resulted in more than 15% improvement in appointment adherence (no-shows and cancellations) at a health center that piloted this intervention. Our analysis also revealed that conventional measurement of provider productivity, using visits/hour, was not a true representation of provider work. Instead, the work relative value units (wRVUs) – a standard for the value of physician work performed accounting for the physician's expertise, the time and technical skill spent in performing the entire service, is a better productivity measurement. We found that the wRVUs per hour was not correlated with the simplistic patient visits per unit of provider time. There was no evidence in the literature or practice of its use among community health centers, likely stemming from a simple visit volume-based Medicaid reimbursement model[323]. This is despite the known benefits of using RVUs to objectively measure and quantify a medical practice's physician productivity and performance data versus traditional productivity measures (such as office visits, net charges, or net collections) as used in many organizations[324]. This research made use of CHC practice management data to highlight the importance of using wRVUs for equitable assessment and benchmarking of provider work vis-à-vis existing approaches.

Open access scheduling methods require ongoing monitoring, and analysis for matching of supply (providers, staffing, open slots) and demand (appointment requests), to derive the benefits associated with such an approach. The study highlighted the need for greater in-CHC visual data analytic capabilities (RQ3) tied with re-designing the care delivery process at CHCs for improving access and efficiency in an increasingly value-driven healthcare system.

5.1.3 The Dashboard

The third RQ was referred to as the 'dashboard', with the primary question being: How can clinic-specific data be harnessed to promote reflection and decision making among varying roles in community healthcare delivery settings? Specifically: 1) What role-specific measures are effective for performance monitoring in health centers? and 2) What are the design and assessment needs for interactive dashboards of user-defined performance metrics for CHCs? The work presented in this chapter identified the role-specific measures for performance monitoring in community health centers and delivered these measures via interactive user-center designed dashboard. The study leveraged a multidimensional data warehouse developed as part of a collaborative of CHCs across the state of Indiana. We applied a user-centered design process to uncover the analytic needs of relevance to different role-based users within a large semi-urban multi-site CHC.

Over 45 different measures were revealed along with a categorization of the dominant themes specific to different roles within the organization. A core-set of these measures were then implemented in a web-based dashboard prototype developed using iterative agile principles. The dashboard contained interactive visualizations grounded in data from electronic health record (EHR) and practice management (PM) systems along with operational data stores for assessing the care delivery process. The usability of the dashboard prototype was assessed through a pluralistic cognitive task walk-through and a modified System Usability Scale (SUS) survey. The dashboard was evaluated as having high learnability, ease of use and applicability for community health center practice management. But more importantly, it helped the CHC staff align with their core mission, goals and priorities. The emergent role-specific measures and lessons learned from the user-centered design process made a contribution to implementation science and design research, as well as had practical applications in supporting health center operations.

5.2 Scientific Contribution and Future Research

The contribution to the discipline of health informatics stems from provisioning a rich data repository of longitudinal healthcare and administrative data with an associated clinic-specific analytics and a user-centered dashboard to sustain informatics research and provide educational opportunities. The work falls within the framework of a learning health care system from a health center operational perspective. The key contributions to the field and the associated areas for future research are described in the sections ahead.

5.2.1 Guiding a Health Centered Controlled Network

The research has uncovered information needs and operational challenges of CHCs, considering the resource constraints they face. A solution common across all these health centers was enabling economies of scale for health informatics support. A notable mechanism to achieve that are state-wide Health Center Controlled Networks (HCCNs) which until 2016, Indiana has not benefited from. HCCNs are supported by the Health Resources and Services Administration (HRSA – an agency within the US Department of Health and Human Services). They are a group of safety net providers working together to improve access to care, enhance quality of care and achieve cost efficiencies through the redesign of practices to integrate services, optimize patient outcomes, or negotiate managed care contracts on behalf of the participating members[325]. The

HCCNs are based on a foundation of shared health IT infrastructure and assist with: Adoption and implementation of certified EHR technology; Meeting meaningful use requirements under the CMS EHR Incentive Programs; and Improve clinical and operational quality, reduce health disparities, improve population health through health information technology, and achieve patient centered medical home recognition.

The data warehousing and analytics platform has demonstrated the feasibility of a multi-site Community Health Information Network for practice-based research. The work is directly benefiting the setting up a state-wide HCCN that will sustain research, education and operational improvements in CHCs across the state. This will impact the lives of about half a million patients each year in this state as clinics leverage data to deliver more patient-centered care.

Future research in this area will leverage the HCCN for real-time data warehousing and analytics (as opposed to the current batch analytics which has a data latency period). Additional research on web services will support more efficient interoperability. There are many population health management solutions that focus on reporting of clinical quality measures derived from EHRs, but additional research is needed on standardizing practice management measures to support improvements in efficiency, effectiveness and value – one of the core objectives of HCCNs.

5.2.2 Modeling and Analytics

Research using machine learning algorithms and simulation modeling have already been made possible through the clinic data warehouse (CDW) based on different EHR data models (NextGen, AthenaHealth, eClinicalWorks, and GE Centricity). There are innumerable research opportunities possible from the CDW and the associated analytics. The latter are broadly classified as being, descriptive, predictive, and prescriptive analytics. A few research examples under, 'descriptive analytics' are: development of enriched care-flow/process models based on EHR time-stamps; ranking clinics based on adherence to evidence based best practices; and peer comparisons at the state and national level. Under 'predictive analytics' the goal is to go beyond building machine-learning models of patient no-show behavior and look at other aspects for population health management, such as risk stratification modeling and treatment adherence analytics; perform agent-based modeling and simulation of clinic operations to improve efficiency; and to perform longitudinal network analysis of diseases. Assessing the impact of the affordable care act (ACA) on access to care could be of particular and timely relevance.

Under ‘prescriptive analytics,’ findings from the predictive analytics can be incorporated into standalone software applications or embedded within patient management systems to provide personalized decision support. The data can be further enriched by including external variables, such as data from community and geo-spatial information systems. There are opportunities for research and development of new technologies for patient engagement and management in the community, as well as integration of disparate care providers and care sites (inpatient, outpatient, community, “minute” clinics, prisons, other) as part of a care management team. The community clinics stand to benefit from such analytics since they lack the infrastructure that is usually available to academic medical centers. The study also shows that there is great opportunity to foster community informatics research and development using this data to improve care delivery, patient engagement and efficiency.

5.2.3 Informatics Tools for Health Centers

The dashboard visualizations have enabled rapid discovery of insights that hitherto required days of data cleaning and analysis. The research has created a methodology for developing ‘rapid’ user-centered visual analytics platform using multiple EHR data models. But more importantly, the research has identified key role-based operational performance monitoring metrics. The study prioritized a core set of measures that were implemented in the dashboard for which data was available within different EHR data models, and other operational data stores like phone systems. However, there also were several performance measures that emerged from the study for which data does not currently exist or is not easily extractable. As such, another key contribution of the study is that it informs the development and implementation of tools that can enable the capture of such data as part of routine clinic operations. These can then be part of an enhanced performance dashboard of the future. Even though this dashboard was designed to be clinic-facing, the study had a patient advocate whose perspective enriched the design of the dashboard visualizations. The methods used in this study can also inform work on developing a patient-facing dashboard of clinic and operational measures.

5.2.4 Training and Education

The longitudinal (2010-16) and de-identified repository of EHR and Practice Management data has been used as a basis for multiple undergraduate and graduate level courses in the Department of BioHealth Informatics at the Indiana University School of Informatics and Computing (Indianapolis, IN). One of the CHCs formally granted permission to use all of their de-identified data in the CDW for educational purposes. This rich data set is proving to be of great value, considering how challenging it has been for the department to access a real life dataset for teaching.

The courses that have already made use this dataset for teaching include: B535 - Clinical Information Systems, B513 - The Design, Implementation, and Evaluation of Electronic Health Record Systems; and M200 - Database Design for Health Information Management. For the latter 2 courses, the EHR/PM data was imported into OpenMRS[326] (an open source medical record system platform led by the Regenstrief Institute[327]) to facilitate real-world experience for students.

The author was also invited for multiple guest lectures to showcase parts of his research. These include presentations on 'Application of Mixed Methods Research' as part of the curriculum for B575 (Informatics Research Design), and on Clinical and Business Intelligence for B641 (Business of Health Informatics). Finally, the research work and resulting analytics platform (CDW and the dashboard) are the subject of multiple grant applications and collaboration opportunities.

5.3 Limitations

There is a broad representation of CHCs of Indiana in this research, particularly in terms of information needs assessment and the development of the CDW. The latter was based on working with seven CHC organizations with over 40 clinic sites. However, specific aspects of the research, including mining of appointment scheduling data and user-centered dashboard development was done with a smaller sub-set of three CHCs. In particular, a great majority of the work to inform the dashboard design development was done with one CHC (CHC-O), one of the largest CHCs in the state. The six other CHCs received a similar dashboard as CHC-O, although the differences in the EHR data models limited the types of visualizations incorporated into their dashboard. CHC-O was selected as a basis, because it was representative of other CHCs in terms of its characteristics: mid-size with rural and urban sites, serving a racially diverse patient population, and having mixed providers (NPs and MDs). A similar level of work is warranted with other clinics across Indiana and beyond to increase generalizability, especially considering the uniqueness of all CHCs in such attributes as organizational culture and resource availability.

Studies are also needed to confirm these findings in other populations and to determine how to incorporate such factors in developing tools to optimize patient-centered scheduling. Open access scheduling methods may be optimized by appointment analysis and ongoing monitoring to derive the benefits associated with such an approach. Given the importance of effectively utilizing scarce health care resources and appointments, there is a need for further research into models, strategies and technological and organizational change solutions.

5.4 Conclusion

The research work presented in this dissertation is a result of a partnership with seven CHCs across the state of Indiana that offer primary care services through about 40 clinic sites caring for over 150,000 patients annually. Using the framework of an operational LHS, the research has considered the unique work and information flow needs at these CHCs and identified knowledge gaps in the literature to develop a novel multidimensional clinic data warehouse of CHCs to support data analytics and dashboards for improving efficiency and access to care.

The study has developed a detailed understanding of the CHC information needs using multiple complimentary methods: questionnaires, key informant interviews, and workflow observations. The information needs defined through this work called for the design of a multidimensional clinic data warehouse - one that integrates patient, operational, financial, and quality data. A centralized data architecture was chosen that was relatively easy to implement and sustain, yet rich in scope and content, and most importantly relevant to the participating clinics. The goal was to enable free standing CHCs to participate in this data warehouse, in the same way that large hospitals participate in HIEs and benefit from large scale healthcare data analytics to improve overall clinic operational performance and achieve better scores on the ever evolving clinical quality and operational measures. Developing strategies for interacting with technology developers and entrepreneurs, in order to test the feasibility of new tools and prototypes should speed innovation and dissemination of effective IT. Ultimately, expanding on the model of a CDW should help CHCs better meet the needs of their patients in continually evaluating, learning from and redesigning care.

The study has applied advanced exploratory and predictive analytics on EHR data to inform patient-centered scheduling and has uncovered the factors that govern appointment adherence based on multi-CHC EHR and practice management data warehouse purpose-built for improving efficiency and access to care. Appointment adherence is significantly correlated with the time dimension and there are distinct patterns that clinics can use to guide the selection of open access hours and optimize the provider schedule utilization. Operationalizing this at a CHC led to over a 15% reduction in no-show and cancellation rates post intervention. We uncovered new data analytic needs and developed a novel low-cost model to give clinics access to actionable data. It also highlighted the need for additional research aimed at re-designing the care delivery process at CHCs for improving access and efficiency in an increasingly value-driven healthcare system.

The learning health system paradigm also calls for a need to understand clinic role-specific measures of interest and then deliver those measures in the form of interactive user-centered visualizations that efficiently tap into underlying information systems in resource constrained environments prevalent in CHCs. A user-centered agile system development methodology was employed to design and develop a dashboard that addressed the direct needs of the end-users at community health centers. The study discovered at least 45 unique measures tied to specific role of the users in the CHC, from the CEO to the patient care representative at the front desk many of which were hitherto unknown. The dashboard was delivered over the web and linked at the back-end to a data warehouse that was refreshed periodically. The usefulness and usability of the dashboard was assessed using multiple methods of usability evaluation, including, cognitive walk-through and a usability survey. The study found the dashboard to be easily learnable and highly usable. Larger studies on multiple CHC organizations in the state and beyond can improve the generalizability of the findings and guide development of health informatics tools.

To be successful in the evolving era of healthcare and information technology, health centers that provide care for the underserved populations, must not only skillfully adopt EHR systems, but also use data in new ways to manage patient care and their clinic operations to realize improvements. This research has carefully studied the unique workflow and information needs of CHCs, with the goal of supporting thoughtful planning and effective decision-making among all roles in CHCs – providers, staff, and administrators. The development of the data warehouse and the analytics platform in this work, however technologically involved, is only a means to an end. The real challenge that this research has tried to understand is how to make the information technology fit into the operational goals and technical capabilities of the concerning healthcare organization. Without this link, there can be no success in a technology project.

Secondly, the exploratory and predictive analytics of this research work is of value to the health center administrators and clinicians, but operationalizing those required two things: one is to be able to visualize the impact and the second is to ensure that the visualizations are made role specific. Each stakeholder has their own set of priorities and concerns, which may also lead to internal conflicts. While there are organizational values and culture that govern health center operations, each job function in a health center has certain uniqueness, and provisioning role-centered tools is a novel aspect addressed by this research.

Referring back to the learning health systems framework, this research has made a contribution in understanding how to help CHCs participate in the virtuous cycle of data-driven continuous learning and improvement. It has created models for design and development of a data warehouse driven analytic and visualization tools that are sustainable in resource constrained health systems and tailored to their unique needs. With the recent creation of an HCCN in Indiana under the auspices of the state primary care association and the partnership with the school, this research work is going to be expanded. The research is poised to enable data-driven systems redesign, research, education and training, technology development and testing, and informed decision making by the participating stakeholders.

REFERENCES

[1] Adams, J., Mounib, E. L., Pai, A., Stuart, N., Thomas, R., & Tomaszewicz, P. (2006). Healthcare 2015: Win-win or lose-lose? IBM Institute for Business Value.

[2] Davis, K., Schoen, C., & Stremikis, K. (2014). Mirror, mirror on the wall: How the performance of the US health re system compares internationally: 2014 update. Commonwealth Fund.

[3] Osborn, R., Moulds, D., Schneider, E. C., Doty, M. M., Squires, D., & Sarnak, D. O. (2015). Primary care physicians in ten countries report challenges caring for patients with complex health needs. Health Affairs, 34(12), 2104-2112.

[4] US Department of Health and Human Services. (2013). HITECH Act enforcement interim final rule. US Department of.

[5] Jha, A. K. (2010). Meaningful use of electronic health records: the road ahead. Jama, 304(15), 1709-1710.

[6] Adler-Milstein, J., DesRoches, C. M., & Jha, A. K. (2011). Health information exchange among US hospitals. The American journal of managed care, 17(11), 761-768.

[7] Protection, P., & Act, A. C. (2010). Public Law 111-148. Title IV, x4207, USC HR, 3590, 2010.

[8] Kocher, R., Emanuel, E. J., & DeParle, N. A. M. (2010). The Affordable Care Act and the future of clinical medicine: the opportunities and challenges. Annals of internal medicine, 153(8), 536-539.

[9] Friedman, C. P., Wong, A. K., & Blumenthal, D. (2010). Achieving a nationwide learning health system. Science translational medicine, 2(57), 57cm29-57cm29.

[10] The Institute for Healthcare Improvement: http://www.ihi.org/; Accessed Sep 3, 2016

[11] Bodenheimer, T., & Sinsky, C. (2014). From triple to quadruple aim: care of the patient requires care of the provider. The Annals of Family Medicine, 12(6), 573-576.

[12] Sheikh, A., Sood, H. S., & Bates, D. W. (2015). Leveraging health information technology to achieve the "triple aim" of healthcare reform. Journal of the American Medical Informatics Association, 22(4), 849-856.

[13] Pruinelli, L., Delaney, C. W., Garcia, A., Caspers, B., & Westra, B. L. (2016). Nursing management minimum data set: cost-effective tool to demonstrate the value of nurse staffing in the big data science era. Nursing Economics, 34(2), 66.

[14] Shi, L., & Stevens, G. D. The role of community health centers in delivering primary care to the underserved: experiences of the uninsured and Medicaid insured. The Journal of ambulatory care management 2007;30(2):159-170

[15] HRSA Health Center Data: http://bphc.hrsa.gov/datareporting/index.html; Accessed Sep 2, 2016

[16] Shin, P., Sharac, J., Barber, Z., Rosenbaum, S. J., & Paradise, J. (2015). Community health centers: a 2013 profile and prospects as ACA implementation proceeds.

[17] Miller, S. C., Frogner, B. K., Saganic, L. M., Cole, A. M., & Rosenblatt, R. (2016). Affordable Care

Act Impact on Community Health Center Staffing and Enrollment: A Cross-Sectional Study. The Journal of Ambulatory Care Management, 39(4), 299-307.

[18] Amico, P. R., Chilingerian, J. A., & Hasselt, M. (2014). Community health center efficiency: the role of grant revenues in health center efficiency. Health services research, 49(2), 666-682.

[19] Hawkins, D., & Groves, D. (2011). The future role of community health centers in a changing health care landscape. The Journal of ambulatory care management, 34(1), 90-99.

[20] Shields, A. E., Shin, P., Leu, M. G., Levy, D. E., Betancourt, R. M., Hawkins, D., & Proser, M. (2007). Adoption of health information technology in community health centers: results of a national survey. Health Affairs, 26(5), 1373-1383.

[21] Jones, E. B., & Furukawa, M. F. (2014). Adoption and use of electronic health records among federally qualified health centers grew substantially during 2010–12. Health Affairs, 33(7), 1254-1261.

[22] Hsiao C-J, Hing E. Use and characteristics of electronic health record systems among office-based physician practices: United States, 2001–2013. NCHS data brief, no 143. Hyattsville, MD: National Center for Health Statistics. 2014.

[23] Patient Centered Medical Home: https://www.pcmh.ahrq.gov/ ; Accessed Sep 2, 2016

[24] Carrier, E., Gourevitch, M. N., & Shah, N. R. (2009). Medical homes: challenges in translating theory into practice. Medical care, 47(7), 714.

[25] Cashman, S. B., Savageau, J. A., Lemay, C. A., & Ferguson, W. (2004). Patient health status and appointment keeping in an urban community health center. Journal of Health Care for the Poor and Underserved, 15(3), 474-488.

[26] Ciechanowski, P. S., Katon, W. J., Russo, J. E., & Walker, E. A. (2001). The patient-provider relationship: attachment theory and adherence to treatment in diabetes. American Journal of Psychiatry, 158(1), 29-35.

[27] LaGanga, L. R., & Lawrence, S. R. (2007). Clinic overbooking to improve patient access and increase provider productivity. Decision Sciences, 38(2), 251-276.

[28] Mahon, M., & Fox, B. (2014). US health system ranks last among eleven countries on measures of access, equity, quality, efficiency and healthy lives. The Common Wealth Fund.

[29] Office of National Coordinator's Health IT Site: https://www.healthit.gov/; Accessed Sep 2, 2016

[30] CMS Accountable Care Organizations: http://www.cms.gov/aco ; Accessed Aug 28, 2014

[31] Davis, K., Abrams, M., & Stremikis, K. (2011). How the Affordable Care Act will strengthen the nation's primary care foundation? Journal of general internal medicine, 26(10), 1201-1203.

[32] PCORI Supporting CHCs: http://www.pcori.org/blog/how-pcori-supports-community-health-centers/; Accessed Aug 28, 2015

[33] NIH Big Data to Knowledge: http://commonfund.nih.gov/bd2k/; Accessed Aug 28, 2016

[34] National Patient-Centered Research Network: http://pcornet.org/ ; Accessed Aug 28, 2014

[35] National Healthcare Quality & Disparities Reports. June 2015. Agency for Healthcare Research and Quality, Rockville, MD. http://www.ahrq.gov/research/findings/nhqrdr/index.html; Accessed Aug 28, 2016

[36] CDC Fast Facts on Physician Office Visits: http://www.cdc.gov/nchs/fastats/physician-visits.htm Accessed on Nov 10, 2015

[37] Phillips, R. L., & Bazemore, A. W. (2010). Primary care and why it matters for US health system reform. Health Affairs, 29(5), 806-810.

[38] AHRQ Publication No. 09-0029-EF; http://healthit.ahrq.gov/sites/default/files/docs/page/09-0029-EF_cdm_1.pdf; Accessed Aug 23, 2015

[39] Ranallo, P. A., Kilbourne, A. M., Whatley, A. S., & Pincus, H. A. (2016). Behavioral Health Information Technology: From Chaos To Clarity. Health Affairs, 35(6), 1106-1113.

[40] Sukumar, S. R., & Ferrell, R. K. 'Big Data' collaboration: Exploring, recording and sharing enterprise knowledge. Information Services and Use 2013;33(3), 257-270.

[41] McGinnis, M., Olsen, L., & Goodby, A. W. Clinical Data as the Basic Staple of Health Learning: Creating and Protecting a Public Good: Workshop Summary. National Academies Press. Chicago 2010

[42] Ledbetter, Craig S., and Matthew W. Morgan. "Toward best practice: leveraging the electronic patient record as a clinical data warehouse." Journal of Healthcare Information Management 2001;15.2:119-132.

[43] Sukumar, S. R., & Ferrell, R. K. (2013). 'Big Data'collaboration: Exploring, recording and sharing enterprise knowledge. Information Services & Use, 33(3-4), 257-270.

[44] Millman, M. (Ed.). (1993). Access to health care in America. National Academies Press.

[45] Chapter 9. Access to Health Care. Content last reviewed October 2014. Agency for Healthcare Research and Quality, MD; http://www.ahrq.gov/research/findings/nhqrdr/nhqr11/chap9.html

[46] IHI Improvements Measures: http://www.ihi.org/resources/Pages/Measures/; Accessed Sep 5, 2016

[47] Open Access Scheduling: http://www.ahrq.gov/cahps/quality-improvement/ ; Accessed Sep 5, 2016

[48] Cayirli, T., Veral, E., & Rosen, H. (2006). Designing appointment scheduling systems for ambulatory care services. Health care management science, 9(1), 47-58.

[49] Ulmer, T., & Troxler, C. (2006). The economic cost of missed appointments and the open access system. Community Health Scholars.

[50] Johnson, S. E., & Newton, W. P. (2002). Resource-based relative value units: a primer for academic family physicians. Family Medicine-Kansas City, 34(3), 172-176.

[51] Bates, D. W., & Bitton, A. (2010). The future of health information technology in the patient-

centered medical home. Health Affairs, 29(4), 614-621.

[52] Blumenthal, D., & Tavenner, M. (2010). The “meaningful use” regulation for electronic health records. New England Journal of Medicine, 363(6), 501-504.

[53] Johnson CM, Johnston D, Crowley PK, et al. (2011, August). EHR Usability Toolkit: A Background Report on Usability and Electronic Health Records (Prepared by Westat under Contract No. HHSA 290-2009-00023I). AHRQ Publication No. 11-0084-EF. Rockville, MD: Agency for Healthcare Research and Quality.

[54] Center for Care Innovations, Safety-Net Analytics Program: http://www.careinnovations.org/uploads/SNAP_RFA.pdf; Accessed Aug 28, 2016

[55] Jones, E. B., & Furukawa, M. F. (2014). Adoption and use of electronic health records among federally qualified health centers grew substantially during 2010–12. Health Affairs, 33(7), 1254-1261.

[56] Jones, E., & Wittie, M. (2015). Accelerated adoption of advanced health information technology in Beacon community health centers. The Journal of the American Board of Family Medicine, 28(5), 565-575.

[57] Sukumar, S. R., & Ferrell, R. K. (2013). ‘Big Data’ collaboration: Exploring, recording and sharing enterprise knowledge. Information Services & Use, 33(3-4), 257-270.

[58] Kellermann, A. L., & Jones, S. S. (2013). What it will take to achieve the as-yet-unfulfilled promises of health information technology. Health Affairs, 32(1), 63-68.

[59] Sheikh, A., Sood, H. S., & Bates, D. W. (2015). Leveraging health information technology to achieve the “triple aim” of healthcare reform. Journal of the American Medical Informatics Association, 22(4), 849-856.

[60] Payne, T. H., Corley, S., Cullen, T. A., Gandhi, T. K., Harrington, L., Kuperman, G. J., ... & Tierney, W. M. (2015). Report of the AMIA EHR 2020 task force on the status and future direction of EHRs. Journal of the American Medical Informatics Association, ocv066.

[61] McAlearney, A. S., Robbins, J., Hirsch, A., Jorina, M., & Harrop, J. P. (2010). Perceived efficiency impacts following electronic health record implementation: an exploratory study of an urban community health center network. International journal of medical informatics, 79(12), 807-816.

[62] Jones, E., & Wittie, M. (2015). Accelerated adoption of advanced health information technology in Beacon community health centers. The Journal of the American Board of Family Medicine, 28(5), 565-575.

[63] Zeng, X., Forrestal, E. J., Cellucci, L. W., Kennedy, M. H., & Smith, D. (2013). Using Electronic Health Records and Data Warehouse Collaboratively in Community Health Centers. Journal of Cases on Information Technology (JCIT), 15(4), 45-62.

[64] Stolba, N., & Tjoa, A. M. (2006). The relevance of data warehousing and data mining in the field of evidence-based medicine to support healthcare decision making. International Journal of Computer Systems Science and Engineering, 3(3), 143-148.

[65] Forjuoh, S. N., Averitt, W. M., Cauthen, D. B., Couchman, G. R., Symm, B., & Mitchell, M. (2001). Open-access appointment scheduling in family practice: comparison of a demand prediction grid with actual appointments. The Journal of the American Board of Family Practice, 14(4), 259-265.

[66] Gupta, D., & Denton, B. (2008). Appointment scheduling in health care: Challenges and opportunities. IIE transactions, 40(9), 800-819.

[67] Daggy, J., Lawley, M., Willis, D., Thayer, D., Suelzer, C., DeLaurentis, P. C., ... & Sands, [67] LaGanga, L. R., & Lawrence, S. R. (2007). Clinic overbooking to improve patient access and Underserved, 15(3), 474-488.

[68] Cashman, S. B., Savageau, J. A., Lemay, C. A., & Ferguson, W. (2004). Patient health status and appointment keeping in an urban community health center. Journal of Health Care for the Poor and Underserved, 15(3), 474-488.

[69] LaGanga, L. R., & Lawrence, S. R. (2007). Clinic overbooking to improve patient access and increase provider productivity. Decision Sciences, 38(2), 251-276.

[70] Daggy, J., Lawley, M., Willis, D., Thayer, D., Suelzer, C., DeLaurentis, P. C., ... & Sands, L. (2010). Using no-show modeling to improve clinic performance. Health Informatics Journal, 16(4), 246-259.

[71] Goldman, L., Freidin, R., Cook, E. F., Eigner, J., & Grich, P. (1982). A multivariate approach to the prediction of no-show behavior in a primary care center. Archives of Internal Medicine, 142(3), 563-567.

[72] Bennett, K. J., & Baxley, E. G. (2009). The effect of a carve-out advanced access scheduling system on no-show rates. Family medicine, 41(1), 51.

[73] O'Hare, C. D., & Corlett, J. (2004). The outcomes of open-access scheduling. Fam Pract Manag, 11(2), 35-8.

[74] Garrow, L. A., & Koppelman, F. S. (2004). Predicting air travelers' no-show and standby behavior using passenger and directional itinerary information. Journal of Air Transport Management, 10(6), 401-411.

[75] Lawrence, R. D., Hong, S. J., & Cherrier, J. (2003, August). Passenger-based predictive modeling of airline no-show rates. In Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 397-406). ACM.

[76] Cohen, R., Elhadad, M., & Elhadad, N. (2013). Redundancy in electronic health record corpora: analysis, impact on text mining performance and mitigation strategies. BMC bioinformatics, 14(1), 10.

[77] Zeng, Q. T., & Crowell, J. (2006). Semantic classification of consumer health content. MEDNET Retrieved May, 19.[77] Jiang, Y. (2011). A computational semantics

[78] Glass, K. P., & Anderson, J. R. (2001). Relative value units and productivity: Part 2 of 4. The Journal of medical practice management: MPM, 17(6), 285-290

[79] Vazquez, M., Krallinger, M., Leitner, F., & Valencia, A. (2011). Text mining for drugs and chemical

compounds: methods, tools and applications. Molecular Informatics, 30(6‐7), 506-519.

[80] Garrett, J. J. (2010). Elements of user experience, the: user-centered design for the web and beyond. Pearson Education.

[81] Feigenberg, L. (1975). Care and understanding of the dying: A patient-centered approach. OMEGA-Journal of Death and Dying, 6(2), 81-94.

[82] Shaller, D. (2007). Patient-centered care: What does it take? (pp. 1-26). Commonwealth Fund. Chicago

[83] Miriovsky, B. J., Shulman, L. N., & Abernethy, A. P. (2012). Importance of health information technology, electronic health records, and continuously aggregating data to comparative effectiveness research and learning health care. Journal of Clinical Oncology, 30(34), 4243-4248

[84] Bush, M., Lederer, A. L., Li, X., Palmisano, J., & Rao, S. (2009). The alignment of information systems with organizational objectives and strategies in health care. International journal of medical informatics, 78(7), 446-456.

[85] American Association of Critical-Care Nurses. (2004). Zero tolerance for abuse position statement. Aliso Viejo, Calif: American Association of Critical-Care Nurses.

[86] Stecker, M., Epstein, N., & Stecker, M. (2013). Analysis of inter-provider conflicts among healthcare providers. Surgical neurology international, 4, 375.

[87] Wickramasinghe, N. (2000). IS/IT as a tool to achieve goal alignment. International journal of healthcare technology and management, 2(1-4), 163-180.

[88] Dowding, D., Randell, R., Gardner, P., Fitzpatrick, G., Dykes, P., Favela, J., ... & Currie, L. (2015). Dashboards for improving patient care: review of the literature. International journal of medical informatics, 84(2), 87-100.

[89] Nelson, E. C., Eftimovska, E., Lind, C., Hager, A., Wasson, J. H., & Lindblad, S. (2015). Patient reported outcome measures in practice. BMJ, 350, g7818.

[90] Mick, J. (2011). Data-driven decision making: a nursing research and evidence-based practice dashboard. Journal of Nursing Administration, 41(10), 391-393.

[91] Wyatt, J. (2004). Scorecards, dashboards, and KPIs keys to integrated performance measurement: to effectively measure performance, providers need more than just a balanced scorecard; they also must be able to access performance data from a central data warehouse. Healthcare financial management, 58(2), 76-81.

[92] Pfeffer, J., & Sutton, R. I. (2006). Evidence-based management. Harvard business review, 84(1)

[93] Rohrer, J. C. (2010). Dashboard impact evaluation for primary care and community health programs. Journal of primary care & community health, 1(1), 25-28.

[94] Cook, K. A., & Thomas, J. J. (2005). Illuminating the path: The research and development agenda for visual analytics (No. PNNL-SA-45230). Pacific Northwest National Laboratory (PNNL), Richland, WA (US).

[95] Caban, J. J., & Gotz, D. (2015). Visual analytics in healthcare–opportunities and research challenges. Journal of the American Medical Informatics Association, 22(2), 260-262.

[96] California Safety Net Foundation' Analytics Program: http://www.chcf.org/rfps/2014/rfa-safety-net-analytics ; Accessed Aug 28, 2015

[97] Institute of Medicine http://www. iom.nationalacademies.org/; Accessed Sep 4, 2016

[98] Foley, T., & Fairmichael, F. (2015). The Potential of Learning Healthcare Systems.

[99] Friedman, C. P., Wong, A. K., & Blumenthal, D. (2010). Achieving a nationwide learning health system. Science translational medicine, 2(57), 57cm29-57cm29.

[100] The Learning Healthcare Project: http://www.learninghealthcareproject.org/ ; Accessed Sep 4, 2016

[101] Mandl, K. D., & Lee, T. H. (2002). Integrating medical informatics and health services research. Journal of the American Medical Informatics Association, 9(2), 127-132.

[102] Stolba, N., & Tjoa, A. M. (2006). The relevance of data warehousing and data mining in the field of evidence-based medicine to support healthcare decision making. International Journal of Computer Systems Science and Engineering, 3(3), 143-148.

[103] ACA: www.HealthCare.gov/glossary; Accessed Sep 3, 2016

[104] Acute Care: www.HealthCare.gov/glossary; Accessed Sep 3, 2016

[105] Big Data: https://en.wikipedia.org/wiki/Big_data; Accessed Sep 3, 2016

[106] Chronic Disease Management: www.HealthCare.gov/glossary; Accessed Sep 3, 2016

[107] FQHC: www.HealthCare.gov/glossary ; Accessed Sep 3, 2016

[108] Healthcare Analytics: https://en.wikipedia.org/wiki/Health_care_analytics; Accessed Sep 3, 2016

[109] Piatetsky-Shapiro, G. (1996). Advances in knowledge discovery and data mining (Vol. 21). U. M. Fayyad, P. Smyth, & R. Uthurusamy (Eds.). Menlo Park: AAAI press.

[110] Data Warehouse: https://en.wikipedia.org/wiki/Data_warehouse; Accessed Sep 3, 2016

[111] Decison Making: https://en.wikipedia.org/wiki/Decision-making; Accessed Sep 3, 2016

[112] EHRs: https://www.healthit.gov/providers-professionals; Accessed Sep 3, 2016

[113] Federal Poverty Level: www.HealthCare.gov/glossary; Accessed Sep 3, 2016

[114] Health IT: https://www.healthit.gov/; Accessed Sep 3, 2016

[115] Information Architecture: https://en.wikipedia.org/wiki/Information_architecture; Accessed Sep 3, 2016

[116] Data Management Association: https://www.dama.org/; Accessed Feb 12, 2017

[117] Belkin, N. J., & Vickery, A. (1985). Interaction in information systems: A review of research from

document retrieval to knowledge-based systems (No. 025.04 BEL. CIMMYT.)

[118] Nicholas, D. (2003). Assessing information needs: tools, techniques and concepts for the internet age. Routledge.

[119] Medicaid: www.HealthCare.gov/glossary; Accessed Sep 3, 2016

[120] Business or Organizational Informatics: https://en.wikipedia.org/wiki/Business_informatics; Accessed Sep 3, 2016

[121] Patient-Centered Outcomes Research: www.HealthCare.gov/glossary; Accessed Sep 3, 2016

[122] Preventive Services: www.HealthCare.gov/glossary; Accessed Sep 3, 2016

[123] Primary Care: www.HealthCare.gov/glossary ; Accessed Sep 3, 2016

[124] Primary Care Providers: www.HealthCare.gov/glossary; Accessed Sep 3, 2016

[125] Usability: https://en.wikipedia.org/wiki/Usability; Accessed Sep 3, 2016

[126] User-Centered Design: www.usability.gov; Accessed Sep 3, 2016

[127] User Experience [UX]: www.usability.gov; Accessed Sep 3, 2016

[128] Workflow (under Health IT Tools) : https://healthit.ahrq.gov/; Accessed Sep 3, 2016

[129] Starfield, B., Shi, L., & Macinko, J. (2005). Contribution of primary care to health systems and health. Milbank quarterly, 83(3), 457-502.

[130] Shi, L., & Stevens, G. D. The role of community health centers in delivering primary care to the underserved: experiences of the uninsured and Medicaid insured. The Journal of ambulatory care management 2007;30(2):159-170

[131] Shin, P., Sharac, J., Barber, Z., Rosenbaum, S. J., & Paradise, J. (2015). Community health centers: a 2013 profile and prospects as ACA implementation proceeds.

[132] HRSA Health Center Data: http://bphc.hrsa.gov/datareporting/index.html; Accessed Sep 2, 2016

[133] Buntin, M. B., Burke, M. F., Hoaglin, M. C., & Blumenthal, D. The benefits of health information technology: a review of the recent literature shows predominantly positive results. Health affairs 2011;30(3):464-471.

[134] National Healthcare Quality & Disparities Reports. June 2015. Agency for Healthcare Research and Quality, Rockville, MD. http://www.ahrq.gov/research/findings/nhqrdr/index.html

[135] CDC Fast Facts on Physician Office Visits: http://www.cdc.gov/nchs/fastats/physician-visits.htm Accessed on Nov 10, 2015

[136] AHRQ Publication No. 09-0029-EF; http://healthit.ahrq.gov/sites/default/files/docs/page/09-0029-EF_cdm_1.pdf ; Accessed Aug 23, 2014

[137] Ranallo, P. A., Kilbourne, A. M., Whatley, A. S., & Pincus, H. A. (2016). Behavioral Health

Information Technology: From Chaos To Clarity. Health Affairs, 35(6), 1106-1113.

[138] Sukumar, S. R., & Ferrell, R. K. 'Big Data'collaboration: Exploring, recording and sharing enterprise knowledge. Information Services and Use 2013;33(3), 257-270.

[139] McGinnis, M., Olsen, L., & Goodby, A. W. Clinical Data as the Basic Staple of Health Learning: Creating and Protecting a Public Good: Workshop Summary. National Academies Press. Chicago 2010

[140] Ledbetter, Craig S., and Matthew W. Morgan. "Toward best practice: leveraging the electronic patient record as a clinical data warehouse." Journal of Healthcare Information Management 2001;15.2:119-132.

[141] Sukumar, S. R., & Ferrell, R. K. (2013). 'Big Data' collaboration: Exploring, recording and sharing enterprise knowledge. Information Services & Use, 33(3-4), 257-270.

[142] Sukumar, S. R., & Ferrell, R. K. (2013). 'Big Data'collaboration: Exploring, recording and sharing enterprise knowledge. Information Services & Use, 33(3-4), 257-270.

[143] Shin, P., Alvarez, C., Sharac, J., Rosenbaum, S. J., Vleet, A. V., Paradise, J., & Garfield, R. (2013). A profile of community health center patients: Implications for policy.

[144] Shields, A. E., Shin, P., Leu, M. G., Levy, D. E., Betancourt, R. M., Hawkins, D., & Proser, M. (2007). Adoption of health information technology in community health centers: results of a national survey. Health Affairs, 26(5), 1373-1383.

[145] Jones, E. B., & Furukawa, M. F. (2014). Adoption and use of electronic health records among federally qualified health centers grew substantially during 2010–12. Health Affairs, 33(7), 1254-1261.

[146] Blumenthal, D., & Tavenner, M. (2010). The "meaningful use" regulation for electronic health records. New England Journal of Medicine, 363(6), 501-504.

[147] Johnson CM, Johnston D, Crowley PK, et al. (2011, August). EHR Usability Toolkit: A Background Report on Usability and Electronic Health Records (Prepared by Westat under Contract No. HHSA 290-2009-00023I). AHRQ Publication No. 11-0084-EF. Rockville, MD: Agency for Healthcare Research and Quality.

[148] Center for Care Innovations, Safety-Net Analytics Program: http://www.careinnovations.org/programs-grants/grants/safety-net-analytics-program; Accessed Aug 28, 2014

[149] Jones, E., & Wittie, M. (2015). Accelerated adoption of advanced health information technology in Beacon community health centers. The Journal of the American Board of Family Medicine, 28(5), 565-575.

[150] Sukumar, S. R., & Ferrell, R. K. (2013). 'Big Data' collaboration: Exploring, recording and sharing enterprise knowledge. Information Services & Use, 33(3-4), 257-270.

[151] Bates, D. W., & Bitton, A. (2010). The future of health information technology in the patient-centered medical home. Health Affairs, 29(4), 614-621.

[152] Buntin, M. B., Burke, M. F., Hoaglin, M. C., & Blumenthal, D. The benefits of health information technology: a review of the recent literature shows predominantly positive results. Health affairs 2011;30(3):464-471.

[153] Jones, E. B., & Furukawa, M. F. (2014). Adoption and use of electronic health records among federally qualified health centers grew substantially during 2010–12. Health Affairs, 33(7), 1254-1261.

[154] Sukumar, S. R., & Ferrell, R. K. 'Big Data' collaboration: Exploring, recording and sharing enterprise knowledge. Information Services and Use 2013;33(3), 257-270.

[155] Rosenblatt, R. A., Andrilla, C. H. A., Curtin, T., & Hart, L. G. (2006). Shortages of medical personnel at community health centers: implications for planned expansion. Jama, 295(9), 1042-1049.

[156] Agénor, P. R. (2008). Health and infrastructure in a model of endogenous growth. Journal of Macroeconomics, 30(4), 1407-1422.

[157] Dorsch, J. L. (2000). Information needs of rural health professionals: a review of the literature. Bulletin of the Medical Library Association, 88(4), 346.

[158] Dorsch, J. L. (2000). Information needs of rural health professionals: a review of the literature. Bulletin of the Medical Library Association, 88(4), 346.

[159] Indiana Health Information Exchange: http://www.ihie.org/; Accessed Feb 30, 2015

[160] Parnes, B., Fernald, D., Coombs, L., DeAlleaume, L., Brandt, E., Webster, B., ... & West, D. (2011). Improving the management of skin and soft tissue infections in primary care: a report from State Networks of Colorado Ambulatory Practices and Partners (SNOCAP-USA) and the Distributed Ambulatory Research in Therapeutics Network (DARTNet). The Journal of the American Board of Family Medicine, 24(5), 534-542.

[161] Zeng, X., Forrestal, E. J., Cellucci, L. W., Kennedy, M. H., & Smith, D. (2013). Using Electronic Health Records and Data Warehouse Collaboratively in Community Health Centers. Journal of Cases on Information Technology (JCIT), 15(4), 45-62.

[162] Zeng, X., Forrestal, E. J., Cellucci, L. W., Kennedy, M. H., & Smith, D. (2013). Using Electronic Health Records and Data Warehouse Collaboratively in Community Health Centers. Journal of Cases on Information Technology (JCIT), 15(4), 45-62.

[163] Nicholas, D. (2003). Assessing information needs: tools, techniques and concepts for the internet age. Routledge.

[164] Information Architecture: https://en.wikipedia.org/wiki/Information_architecture; Accessed Aug 28, 2016

[165] Data Management Association: https://www.dama.org/; Accessed Feb 12, 2017

[166] Patient Centered Outcome Research Institute (PCORI) http://www.pcori.org/, Accessed Nov 26, 2016

[167] Bradley EH, Curry LA, Ramanadhan S, Rowe L, Nembhard IM, Krumholz HM. Research in action:

using positive deviance to improve quality of health care. Implementation science : IS. 2009;4:25.

[168] Bradley, E. H., Curry, L. A., Ramanadhan, S., Rowe, L., Nembhard, I. M., & Krumholz, H. M. (2009). Research in action: using positive deviance to improve quality of health care. Implementation Science, 4(1), 25.

[169] Seidman, W., & McCauley, M. (2008). Positive deviants rule!. Cutter IT Journal, 21(7), 16.

[170] AHRQ National Resource Center on HIT Tools/Surveys for Workflow: https://healthit.ahrq.gov/health-it-tools-and-resources/workflow-assessment-health-it-toolkit/all-workflow-tools/survey; Accessed Feb 12, 2017

[171] AHRQ Workflow Assessment Health IT Toolkit: http://healthit.ahrq.gov/health-it-tools-and-resources/workflow-assessment-health-it-toolkit; Accessed Oct 23, 2014

[172] Millen DR. Rapid ethnography: time deepening strategies for HCI field research. . The 3rd conference on Designing interactive systems: processes, practices, methods, and techniques 2000:280-286.

[173] Denzin, NK. (1978). Sociological Methods. New York: McGraw-Hill.

[174] Patton, MQ. (1999). "Enhancing the quality and credibility of qualitative analysis." HSR: Health Services Research. 34 (5) Part II. pp. 1189-1208.

[175] Wright, M. D., Flanagan, M. E., Kunjan, K., Doebbeling, B. N., & Toscos, T. (2016, May). Missing links: challenges in engaging the underserved with health information and communication technology. In Proceedings of the 10th EAI International Conference on Pervasive Computing Technologies for Healthcare (pp. 122-129). ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering).

[176] Inmon, William H. Building the data warehouse. John Wiley & Sons, 2005.

[177] Sahama, Tony R., and Peter R. Croll. "A data warehouse architecture for clinical data warehousing." Proceedings of the fifth Australasian symposium on ACSW frontiers-Volume 68. Australian Computer Society, Inc., 2007.

[178] Verma, R., & Harper, J. (2001). Life cycle of a data warehousing project in healthcare. Journal of healthcare information management, 15(2), 107-118.

[179] HL7: http://www.hl7.org ; Accessed Aug 1, 2016

[180] PCORnet: http://www.pcornet.org/ ; Accessed Aug 1, 2016

[181] FDA Sentinel Initiative: http://mini-sentinel.org/ ; Accessed Aug 1, 2016

[182] US Department of Health and Human Services. "Guidance regarding methods for de-identification of protected health information in accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule." 2013

[183] Estabrooks, P. A., Boyle, M., Emmons, K. M., Glasgow, R. E., Hesse, B. W., Kaplan, R. M., ... & Taylor, M. V. (2012). Harmonized patient-reported data elements in the electronic health record: supporting meaningful use by primary care action on health behaviors and key psychosocial

factors. Journal of the American Medical Informatics Association, 19(4), 575-582.

[184] Jones, E. B., & Furukawa, M. F. (2014). Adoption and use of electronic health records among federally qualified health centers grew substantially during 2010–12. Health Affairs, 33(7), 1254-1261.

[185] Sukumar, S. R., & Ferrell, R. K. ‘Big Data’ collaboration: Exploring, recording and sharing enterprise knowledge. Information Services and Use 2013;33(3), 257-270.

[186] Zeng, X., Forrestal, E. J., Cellucci, L. W., Kennedy, M. H., & Smith, D. (2013). Using Electronic Health Records and Data Warehouse Collaboratively in Community Health Centers. Journal of Cases on Information Technology (JCIT), 15(4), 45-62.

[187] Miriovsky, B. J., Shulman, L. N., & Abernethy, A. P. (2012). Importance of health information technology, electronic health records, and continuously aggregating data to comparative effectiveness research and learning health care. Journal of Clinical Oncology, 30(34), 4243-4248.

[188] McAlearney, A. S., Robbins, J., Hirsch, A., Jorina, M., & Harrop, J. P. (2010). Perceived efficiency impacts following electronic health record implementation: an exploratory study of an urban community health center network. International journal of medical informatics, 79(12), 807-816.

[189] Rosenblatt, R. A., Andrilla, C. H. A., Curtin, T., & Hart, L. G. (2006). Shortages of medical personnel at community health centers: implications for planned expansion. Jama, 295(9), 1042-1049.

[190] Agénor, P. R. (2008). Health and infrastructure in a model of endogenous growth. Journal of Macroeconomics, 30(4), 1407-1422.

[191] Health Centers and the Data Warehouse: http://www.rchnfoundation.org/wp-content/uploads/2013/02/Data-Warehouse-and-Health-Centers_12_2_08.pdf ; Accessed Sep 18, 2016

[192] Bodenheimer, T., & Sinsky, C. (2014). From triple to quadruple aim: care of the patient requires care of the provider. The Annals of Family Medicine, 12(6), 573-576.

[193] Bukhari, S. I. A. (2013). Real Time Data Warehouse. arXiv preprint arXiv:1310.5254.

[194] Institute of Medicine (US). Committee on Quality of Health Care in America, 2001. Crossing the quality chasm: a new health system for the 21st century. National Academy Press.

[195] Institute of Medicine (US). Committee on Optimizing Scheduling in Health care, 2015. Transforming Health Care Scheduling and Access: Getting to Now. National Academy Press

[196] Derose, K. P., & Varda, D. M. (2009). Social capital and health care access: a systematic review. Medical Care Research and Review.

[197] Litaker, D., Koroukian, S. M., & Love, T. E. (2005). Context and health care access: looking beyond the individual. Medical care, 43(6), 531-540.

[198] Health care under the Affordable Care Act http://www.hhs.gov/health care/ ; Accessed Jan 24, 2016

[199] Health Insurance Marketplace: https://www.health care.gov/ ; Accessed May 25, 2016

[200] Hawkins, D., & Groves, D. (2011). The future role of community health centers in a changing health care landscape. The Journal of ambulatory care management, 34(1), 90-99.

[201] Gulliford, M., Figueroa-Munoz, J., Morgan, M., Hughes, D., Gibson, B., Beech, R., & Hudson, M. (2002). What does' access to health care' mean?. Journal of health services research & policy, 7(3), 186-188.

[202] IHI Improvements Measures: http://www.ihi.org/resources/Pages/Measures/ ; Accessed Sep 5, 2016

[203] Phan, K. and Brown, S.R., 2009. Decreased continuity in a residency clinic: a consequence of open access scheduling. Fam Med, 41(1), pp.46-50.

[204] Murray, M. and Berwick, D.M., 2003. Advanced access: reducing waiting and delays in primary care. Jama, 289(8), pp.1035-1040.

[205] Forjuoh, S. N., Averitt, W. M., Cauthen, D. B., Couchman, G. R., Symm, B., & Mitchell, M. (2001). Open-access appointment scheduling in family practice: comparison of a demand prediction grid with actual appointments. The Journal of the American Board of Family Practice, 14(4), 259-265.

[206] Ulmer, T., & Troxler, C. (2006). The economic cost of missed appointments and the open access system. Community Health Scholars.

[207] Murray, M., Bodenheimer, T., Rittenhouse, D. and Grumbach, K., 2003. Improving timely access to primary care: case studies of the advanced access model. Jama, 289(8), pp.1042-1046.

[208] Robinson, L.W. and Chen, R.R., 2010. A comparison of traditional and open-access policies for appointment scheduling. Manufacturing & Service Operations Management, 12(2), pp.330-346.

[209] Parente, D. H., Pinto, M. B., & Barber, J. C. (2005). A pre-post comparison of service operational efficiency and patient satisfaction under open access scheduling. Health care management review, 30(3), 220-228.

[210] Murray, M.M. and Tantau, C., 2000. Same-day appointments: exploding the access paradigm. Family practice management, 7(8), pp.45-45.

[211] Patient-Centered Medical Home: https://pcmh.ahrq.gov/ ; Accessed Jan 24, 2016

[212] Carrier, E., Gourevitch, M. N., & Shah, N. R. (2009). Medical homes: challenges in translating theory into practice. Medical care, 47(7), 714.

[213] Patient Centered Outcome Research Institute (PCORI) http://www.pcori.org/, Accessed Nov 26, 2016

[214] Stolba, N., & Tjoa, A. M. (2006). The relevance of data warehousing and data mining in the field of evidence-based medicine to support healthcare decision making. International Journal of Computer Systems Science and Engineering, 3(3), 143-148.

[215] Cashman, S. B., Savageau, J. A., Lemay, C. A., & Ferguson, W. (2004). Patient health status and appointment keeping in an urban community health center. Journal of Health Care for the Poor

and Underserved, 15(3), 474-488.

[216] LaGanga, L. R., & Lawrence, S. R. (2007). Clinic overbooking to improve patient access and increase provider productivity. Decision Sciences, 38(2), 251-276.

[217] Forjuoh, S. N., Averitt, W. M., Cauthen, D. B., Couchman, G. R., Symm, B., & Mitchell, M. (2001). Open-access appointment scheduling in family practice: comparison of a demand prediction grid with actual appointments. The Journal of the American Board of Family Practice, 14(4), 259-265.

[218] Gupta, D. and Denton, B., 2008. Appointment scheduling in health care: Challenges and opportunities. IIE transactions, 40(9), pp.800-819.

[219] Jacobson, S.H., Hall, S.N. and Swisher, J.R., 2006. Discrete-event simulation of health care systems. In Patient flow: Reducing delay in health care delivery (pp. 211-252). Springer US.

[220] Cayirli, T. and Veral, E., 2003. OUTPATIENT SCHEDULING IN HEALTH CARE: A REVIEW OF LITERATURE*. Production and Operations Management, 12(4), p.519.

[221] Huang, Y., & Hanauer, D. A. (2014). Patient no-show predictive model development using multiple data sources for an effective overbooking approach. Applied clinical informatics, 5(3), 836-860.

[222] Daggy, J., Lawley, M., Willis, D., Thayer, D., Suelzer, C., DeLaurentis, P. C., ... & Sands, L. (2010). Using no-show modeling to improve clinic performance. Health Informatics Journal, 16(4), 246-259.

[223] Goldman, L., Freidin, R., Cook, E. F., Eigner, J., & Grich, P. (1982). A multivariate approach to the prediction of no-show behavior in a primary care center. Archives of Internal Medicine, 142(3), 563-567.

[224] Bennett, K. J., & Baxley, E. G. (2009). The effect of a carve-out advanced access scheduling system on no-show rates. Family medicine, 41(1), 51.

[225] O'Hare, C. D., & Corlett, J. (2004). The outcomes of open-access scheduling. Fam Pract Manag, 11(2), 35-8.

[226] LaGanga, L. R., & Lawrence, S. R. (2007). Clinic overbooking to improve patient access and increase provider productivity. Decision Sciences, 38(2), 251-276.

[227] Cohen, R., Elhadad, M., & Elhadad, N. (2013). Redundancy in electronic health record corpora: analysis, impact on text mining performance and mitigation strategies. BMC bioinformatics, 14(1), 1.

[228] Chen, H. B., Huang, H. H., Tan, C. T., Tjiu, J., & Chen, H. H. (2012, January). A statistical medical summary translation system. In Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium (pp. 101-110). ACM.

[229] Zeng, Q. T., & Crowell, J. (2006). Semantic classification of consumer health content. MEDNET Retrieved May, 19.

[230] Jiang, Y. (2011). A computational semantics system for detecting drug reactions and patient

outcomes in personal health messages.

[231] Deguchi, Janet J., Thomas S. Inui, and Diane P. Martin. "Measuring provider productivity in ambulatory care." The Journal of ambulatory care management 7.2 (1984): 29-38.

[232] Center for Medicare and Medicaid Services: https://www.cms.gov/apps/physician-fee-schedule/ Accessed Feb 3, 2017

[233] Gallagher, P. E., Smith, S. L., & American Medical Association. (2000). Medicare RBRVS: The Physicians' Guide, 2000. Amer Medical Assn.

[234] AAPC (American Academy of Professional Coders) Work RVU Calculator https://www.aapc.com/practice-management/rvu-calculator.aspx; Accessed Feb 5, 2017

[235] Glass, K. P., & Anderson, J. R. (2001). Relative value units and productivity: Part 2 of 4. The Journal of medical practice management: MPM, 17(6), 285-290.

[236] Koppen, C. (2001). Understanding the Medicaid Prospective Payment System for Federally Qualified Health Centers (FQHCs). National Association of Community Health Centers, 13.

[237] Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. AI magazine, 17(3), 37.

[238] Kunjan, K., Toscos, T., Turkcan, A., & Doebbeling, B. N. (2015). A Multidimensional Data Warehouse for Community Health Centers. In AMIA Annual Symposium Proceedings (Vol. 2015, p. 1976). American Medical Informatics Association.

[239] Hariri, S., Bozic, K. J., Lavernia, C., Prestipino, A., & Rubash, H. E. (2007). Medicare physician reimbursement: past, present, and future. J Bone Joint Surg Am, 89(11), 2536-2546.

[240] Hariri, S., Bozic, K. J., Lavernia, C., Prestipino, A., & Rubash, H. E. (2007). Medicare physician reimbursement: past, present, and future. J Bone Joint Surg Am, 89(11), 2536-2546. Chicago

[241] RapidMiner Data Science Platform: https://rapidminer.com/ ; Accessed Nov 26, 2017

[242] The R-Project for Statistical Computing: https://www.r-project.org/; Accessed Nov 26, 2016

[243] Tableau Software for Business Intelligence and Analytics http://www.tableau.com/; Accessed Nov 26, 2016

[244] Quinlan, J. R. (2014). C4. 5: programs for machine learning. Elsevier.

[245] Parikh, A., Gupta, K., Wilson, A. C., Fields, K., Cosgrove, N. M., & Kostis, J. B. (2010). The effectiveness of outpatient appointment reminder systems in reducing no-show rates. The American journal of medicine, 123(6), 542-548. Chicago

[246] Goldman, L., Freidin, R., Cook, E. F., Eigner, J., & Grich, P. (1982). A multivariate approach to the prediction of no-show behavior in a primary care center. Archives of Internal Medicine, 142(3), 563-567.

[247] Bennett, K. J., & Baxley, E. G. (2009). The effect of a carve-out advanced access scheduling system on no-show rates. Family medicine, 41(1), 51.

[248] Huang, Y., & Hanauer, D. A. (2014). Patient no-show predictive model development using multiple data sources for an effective overbooking approach. Applied clinical informatics, 5(3), 836-860.

[249] LaGanga, L. R., & Lawrence, S. R. (2007). Clinic overbooking to improve patient access and increase provider productivity. Decision Sciences, 38(2), 251-276.

[250] Stecker, E. C., & Schroeder, S. A. (2013). Adding value to relative-value units. New England Journal of Medicine, 369(23), 2176-2179.

[251] Jha, A. K. (2010). Meaningful use of electronic health records: the road ahead. Jama, 304(15), 1709-1710.

[252] Rittenhouse, D. R., Shortell, S. M., & Fisher, E. S. (2009). Primary care and accountable care—two essential elements of delivery-system reform. New England Journal of Medicine, 361(24), 2301-2303.

[253] Davis, K., Abrams, M., & Stremikis, K. (2011). How the Affordable Care Act will strengthen the nation's primary care foundation. Journal of general internal medicine, 26(10), 1201-1203.

[254] Hawkins, D., & Groves, D. (2011). The future role of community health centers in a changing health care landscape. The Journal of ambulatory care management, 34(1), 90-99.

[255] Davis, K., Abrams, M., & Stremikis, K. (2011). How the Affordable Care Act will strengthen the nation's primary care foundation. Journal of general internal medicine, 26(10), 1201-1203.

[256] Finkelstein, J., Barr, M. S., Kothari, P. P., Nace, D. K., & Quinn, M. (2011). Patient-centered medical home cyberinfrastructure: current and future landscape. American journal of preventive medicine, 40(5), S225-S233.

[257] Cayirli, T., Veral, E., & Rosen, H. (2006). Designing appointment scheduling systems for ambulatory care services. Health care management science, 9(1), 47-58.

[258] Tantau, C. (2009). Accessing Patient-Centered Care Using the Advanced Access Model. The Journal of ambulatory care management, 32(1), 32-43. Chicago

[259] Zhang, X., Gallagher, K. P., & Goh, S. (2011). Bi Application: Dashboards for Healthcare. In AMCIS.

[260] Eckerson, W. W. (2006). Performance Dashboards: Measuring. Monitoring, and Managing. Chicago

[261] Eckerson, W. W. (2010). Performance dashboards: measuring, monitoring, and managing your business. John Wiley & Sons. Chicago

[262] Brath, R., & Peters, M. (2004). Dashboard design: Why design is important. DM Direct.

[263] Miriovsky, B. J., Shulman, L. N., & Abernethy, A. P. (2012). Importance of health information technology, electronic health records, and continuously aggregating data to comparative effectiveness research and learning health care. Journal of Clinical Oncology, 30(34), 4243-4248.

[264] Bush, M., Lederer, A. L., Li, X., Palmisano, J., & Rao, S. (2009). The alignment of information

systems with organizational objectives and strategies in health care. International journal of medical informatics, 78(7), 446-456.

[265] American Association of Critical-Care Nurses. (2004). Zero tolerance for abuse position statement. Aliso Viejo, Calif: American Association of Critical-Care Nurses.

[266] Stecker, M., Epstein, N., & Stecker, M. (2013). Analysis of inter-provider conflicts among healthcare providers. Surgical neurology international, 4, 375.

[267] Boateng, I. A. (2014). Conflict resolution in organizations – an analysis. European Journal of Business and Innovation Research, 2(6), 1-8.

[268] Wickramasinghe, N. (2000). IS/IT as a tool to achieve goal alignment. International journal of healthcare technology and management, 2(1-4), 163-180.

[269] Dowding, D., Randell, R., Gardner, P., Fitzpatrick, G., Dykes, P., Favela, J., ... & Currie, L. (2015). Dashboards for improving patient care: review of the literature. International journal of medical informatics, 84(2), 87-100.

[270] Nelson, E. C., Eftimovska, E., Lind, C., Hager, A., Wasson, J. H., & Lindblad, S. (2015). Patient reported outcome measures in practice. BMJ, 350, g7818.

[271] Pfeffer, J., & Sutton, R. I. (2006). Evidence-based management. Harvard business review, 84(1), 62.

[272] Caban, J. J., & Gotz, D. (2015). Visual analytics in healthcare–opportunities and research challenges. Journal of the American Medical Informatics Association, 22(2), 260-262.

[273] Halford, G. S., Baker, R., McCredden, J. E., & Bain, J. D. (2005). How many variables can humans process?. Psychological science, 16(1), 70-76.

[274] Amarasingham, R., Patzer, R. E., Huesch, M., Nguyen, N. Q., & Xie, B. (2014). Implementing electronic health care predictive analytics: considerations and challenges. Health Affairs, 33(7), 1148-1154.

[275] Cook, K. A., & Thomas, J. J. (2005). Illuminating the path: The research and development agenda for visual analytics (No. PNNL-SA-45230). Pacific Northwest National Laboratory (PNNL), Richland, WA (US).

[276] Fihn, S. D., Francis, J., Clancy, C., Nielson, C., Nelson, K., Rumsfeld, J., ... & Graham, G. L. (2014). Insights from advanced analytics at the Veterans Health Administration. Health affairs, 33(7), 1203-1211.

[277] Abras, C., Maloney-Krichmar, D., & Preece, J. (2004). User-centered design. Bainbridge, W. Encyclopedia of Human-Computer Interaction. Thousand Oaks: Sage Publications, 37(4), 445-456.

[278] Izard, J., Hartzler, A., Avery, D. I., Shih, C., Dalkin, B. L., & Gore, J. L. (2014). User-centered design of quality of life reports for clinical care of patients with prostate cancer. Surgery, 155(5), 789-796.

[279] Koopman, R. J., Kochendorfer, K. M., Moore, J. L., Mehr, D. R., Wakefield, D. S., Yadamsuren, B., ... & Belden, J. L. (2011). A diabetes dashboard and physician efficiency and accuracy in

accessing data needed for high-quality diabetes care. The Annals of Family Medicine, 9(5), 398-405.

[280] Izard, J., Hartzler, A., Avery, D. I., Shih, C., Dalkin, B. L., & Gore, J. L. (2014). User-centered design of quality of life reports for clinical care of patients with prostate cancer. Surgery, 155(5), 789-796.

[281] Ziehm, A., Cheedalla, A., Geist, N., Mcinnis, A., & Valdez, R. (2015, April). Improving efficiency of electronic health records: longitudinal dashboard design for post-kidney transplant pediatric patients. In Systems and Information Engineering Design Symposium (SIEDS), 2015 (pp. 218-223). IEEE.

[282] Koerner, K., Dailey, D., Lipp, M., Connor, H., & Sharma, R. (2012, October). Data visualization for psychotherapy progress tracking. In Proceedings of the 30th ACM international conference on Design of communication (pp. 213-218). ACM.

[283] Hartzler, A. L., Chaudhuri, S., Fey, B. C., Flum, D. R., & Lavallee, D. (2015). Integrating Patient-Reported Outcomes into Spine Surgical Care through Visual Dashboards: Lessons Learned from Human-Centered Design. eGEMs (Generating Evidence & Methods to improve patient outcomes), 3(2), 2.

[284] Pauwels, K., Ambler, T., Clark, B. H., LaPointe, P., Reibstein, D., Skiera, B., ... & Wiesel, T. (2009). Dashboards as a service: why, what, how, and what research is needed?. Journal of Service Research.

[285] Elwyn, G., Edwards, A., Mowle, S., Wensing, M., Wilkinson, C., Kinnersley, P., & Grol, R. (2001). Measuring the involvement of patients in shared decision-making: a systematic review of instruments. Patient education and counseling, 43(1), 5-22.

[286] Bharati, P., & Chaudhury, A. (2004). An empirical investigation of decision-making satisfaction in web-based decision support systems. Decision support systems, 37(2), 187-197.

[287] Garg, A. X., Adhikari, N. K., McDonald, H., Rosas-Arellano, M. P., Devereaux, P. J., Beyene, J., ... & Haynes, R. B. (2005). Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review. Jama, 293(10), 1223-1238.

[288] Egan, M. (2006). Clinical dashboards: impact on workflow, care quality, and patient safety. Critical care nursing quarterly, 29(4), 354-361.

[289] Rohrer, J. C. (2010). Dashboard impact evaluation for primary care and community health programs. Journal of primary care & community health, 1(1), 25-28. Chicago

[290] Miller, R. H., & West, C. E. (2007). The value of electronic health records in community health centers: policy implications. Health Affairs, 26(1), 206-214.

[291] Blumenthal, D. (2011). Implementation of the federal health information technology initiative. New England Journal of Medicine, 365(25), 2426-2431.

[292] Fihn, S. D., Francis, J., Clancy, C., Nielson, C., Nelson, K., Rumsfeld, J., ... & Graham, G. L. (2014). Insights from advanced analytics at the Veterans Health Administration. Health affairs, 33(7),

1203-1211.

[293] Kunjan, K., Toscos, T., Turkcan, A., & Doebbeling, B. N. (2015). A Multidimensional Data Warehouse for Community Health Centers. In AMIA Annual Symposium Proceedings (Vol. 2015, p. 1976). American Medical Informatics Association.

[294] Beck, K., Beedle, M., Van Bennekum, A., Cockburn, A., Cunningham, W., Fowler, M., ... & Kern, J. (2001). Manifesto for agile software development.

[295] Elastic Kibana: https://www.elastic.co/products/kibana; Accessed Jul 28, 2016

[296] Data-Driven Documents: https://d3js.org/; Accessed Jul 18, 2016

[297] Tableau Software: http://www.tableau.com/ ; Accessed Jul 28, 2016

[298] Microsoft Power BI Tools: https://powerbi.microsoft.com/en-us/; Accessed Jul 28, 2016

[299] Murray, D. G. (2013). Tableau Your Data!: Fast and Easy Visual Analysis with Tableau Software. John Wiley & Sons.

[300] Holtzblatt, K., & Beyer, H. (2009). Contextual design. Human-Computer Interaction: Development Process, 55.

[301] Spencer, D. (2009). Card sorting: Designing usable categories. Rosenfeld Media.

[302] Ahlstrom, V., & Allendoerfer, K. (2004). Information organization for a portal using a card-sorting technique.

[303] Rosenfeld, L., & Morville, P. (2002). Information architecture for the world wide web. " O'Reilly Media, Inc.".

[304] Kushniruk, A. W., & Patel, V. L. (2004). Cognitive and usability engineering methods for the evaluation of clinical information systems. Journal of biomedical informatics, 37(1), 56-76.

[305] Holzinger, A. (2005). Usability engineering methods for software developers. Communications of the ACM, 48(1), 71-74.

[306] Wharton, C., Rieman, J., Lewis, C., & Polson, P. (1994, June). The cognitive walkthrough method: A practitioner's guide. In Usability inspection methods (pp. 105-140). John Wiley & Sons, Inc..

[307] Brooke, J. (2013). SUS: a retrospective. Journal of usability studies, 8(2), 29-40.

[308] Brooke, J. (1996). SUS-A quick and dirty usability scale. Usability evaluation in industry, 189(194), 4-7.

[309] Uniform Data Systems: http://bphc.hrsa.gov/datareporting/index.html ; Accessed Sep 8 2016

[310] Patient Centered Medical Home: https://www.pcmh.ahrq.gov/ ; Accessed Sep 2, 2016

[311] Lewis, J. R., & Sauro, J. (2009, July). The factor structure of the system usability scale. In International Conference on Human Centered Design (pp. 94-103). Springer Berlin Heidelberg.

[312] Friedman, C. P., Wong, A. K., & Blumenthal, D. (2010). Achieving a nationwide learning health

system. Science translational medicine, 2(57), 57cm29-57cm29.

[313] Moore, J. B., & Kordick, M. F. (2006). Sources of conflict between families and health care professionals. Journal of Pediatric Oncology Nursing, 23(2), 82-91.

[314] Robey, D. (1979). User attitudes and management information system use. Academy of Management Journal, 22(3), 527-538.

[315] Amarasingham, R., Patzer, R. E., Huesch, M., Nguyen, N. Q., & Xie, B. (2014). Implementing electronic health care predictive analytics: considerations and challenges. Health Affairs, 33(7), 1148-1154.

[316] Cook, K. A., & Thomas, J. J. (2005). Illuminating the path: The research and development agenda for visual analytics (No. PNNL-SA-45230). Pacific Northwest National Laboratory (PNNL), Richland, WA (US).

[317] Nicholas, D. (2003). Assessing information needs: tools, techniques and concepts for the internet age. Routledge.

[318] O'Hare, C. D., & Corlett, J. (2004). The outcomes of open-access scheduling. Fam Pract Manag, 11(2), 35-8.

[319] Goldman, L., Freidin, R., Cook, E. F., Eigner, J., & Grich, P. (1982). A multivariate approach to the prediction of no-show behavior in a primary care center. Archives of Internal Medicine, 142(3), 563-567.

[320] Bennett, K. J., & Baxley, E. G. (2009). The effect of a carve-out advanced access scheduling system on no-show rates. Family medicine, 41(1), 51.

[321] Bush, M., Lederer, A. L., Li, X., Palmisano, J., & Rao, S. (2009). The alignment of information systems with organizational objectives and strategies in health care. International journal of medical informatics, 78(7), 446-456.

[322] Wickramasinghe, N. (2000). IS/IT as a tool to achieve goal alignment. International journal of healthcare technology and management, 2(1-4), 163-180.

[323] Koppen, C. (2001). Understanding the Medicaid Prospective Payment System for Federally Qualified Health Centers (FQHCs). National Association of Community Health Centers, 13.

[324] Glass, K. P., & Anderson, J. R. (2001). Relative value units and productivity: Part 2 of 4. The Journal of medical practice management: MPM, 17(6), 285-290.

[325] HRSA's Health Center Controlled Networks: http://www.hrsa.gov/healthit/networkguide/ Accessed Dec 4, 2016

[326] Open Medical Record System: http://openmrs.org/; Accessed Dec 4, 2016

[327] The Regenstrief Institute: http://www.regenstrief.org/ Access Dec 4, 2016

CURRICULUM VITAE

Kislaya Kunjan

ACADEMICS

	INDIANA UNIVERSITY	**Indianapolis, IN**
2011-2017	**Ph.D. (Health Informatics)**	**GPA 3.95**
	Dissertation: A Big Data Augmented Analytics Platforms to Operationalize Efficiencies in Community Clinics	
	CORNELL UNIVERSITY	**Ithaca, NY**
2004-05	**MBA (Marketing & Finance)**	
	PURDUE UNIVERSITY	**West Lafayette, IN**
2001-03	**M.S. (Mechanical Engineering)**	
	Thesis: Design and development of a Mid Infrared Glucose Monitoring System	
	NATIONAL INSTITUTE OF TECHNOLOGY	**Jalandhar, India**
1994-98	**B.Tech. (Mechanical Engineering)**	

PROFESSIONAL EXPERIENCE

Indiana Primary Healthcare Association, Inc. — **Indianapolis, IN**

2016-Present — **Health IT Director**

- Leading all health IT initiatives at the Indiana Quality Improvement Network - a first state-wide health center controlled network for community health centers
- Implementing initiatives in the area of health information exchange, data aggregation and analytics to support population health management
- Developing and managing relationships with health center staff, vendors, and other stakeholders

Cascade Metrix, LLC. — **Indianapolis, IN**

2006-Present — **Co-founder and Chief Technology Officer**

- Leading the design, development, clinical validation and commercialization of the world’s first automated blood analyte monitoring system for intensive care units

- Managing strategic collaborations with medical technology players, academic institutions, FDA/cGMP compliant product development sub-contractors
- Hands-on development and validation of blood sampling systems and biosensors based on infrared spectroscopy and electrochemistry

Indiana University School of Informatics and Computing Indianapolis, IN

2011-2016 **Research Assistant**

- Collaborative Workflow Modeling, Data Warehousing and Advanced Analytics

Marion County Indianapolis, IN

2009-2016 **Consultant (Finance & IT)**

- Budget manager to the office of the Marion county coroner
- Project Lead, developer and system administrator of Salesforce CRM Solution for case management
- Led a major 6-sigma process improvement project that reduced cycle time for case closure by 50%, receiving special recognition from the Mayor

Indiana University School of Informatics and Computing Indianapolis, IN

2015-2016 **Adjunct Faculty and Guest Lecturer**

- Course Instructor in Health Information Management [BS in Informatics Program]
    - Data Organization and Presentation in the Healthcare Environment [Spring 2016]
- Invited Guest Lectures [Topics]
    - Clinical and Business Intelligence [Fall 2015 and Fall 2016]
    - Mixed Methods Research in Informatics [Fall and Spring 2016]
    - Advanced Research Methods in BioHealth Informatics [Spring 2014]

Johnson & Johnson/LifeScan India

2005-2005 **Product Manager**

Critically evaluated the consumer and hospital glucose monitoring market through countrywide field visits and made strategic recommendations to senior management

Vista Biosciences, LLC (A Purdue Research Park Co.) **West Lafayette, IN**

2003-2004 **Lead Engineer**

- Helped raise $250,000 from the Purdue Trask Pre-Seed Venture Fund
- Led engineering feasibility studies and clinical testing of an infrared medical diagnostic system in collaboration with a leading device development firm

Purdue University (School of Mechanical Engineering) **West Lafayette, IN**

2001-2003 **Research Assistant/Teaching Assistant (RA/TA)**

Completed an interdisciplinary project involving the design, build and testing of a portable infrared spectroscopy based system for glucose sensing and pathogen detection; Also served as a TA for undergraduate courses in engineering

Saint-Gobain (India) Ltd **India**

1998-2001 **Sales Engineer (Industrial Abrasives, Non Wovens and Lubricants)**

Exemplary sales performer in a $2M sales territory, recording >25% growth through a network of dealers and key direct accounts across a wide spectrum of manufacturing industries

TECHNOLOGY SKILLS

Extremely proficient in the entire MS Office Suite, including Access, Project, Visio, SharePoint

Software Development: Web [HTML/CSS, XML, PHP/MySQL, JavaScript]; CMS [Joomla, Drupal, Moodle], Databases (MS SQL Server, MySQL Workbench, Crystal Reports); Salesforce Apex

Application Software: R, SAS (Statistics); RapidMiner (Machine Learning); Tableau (Visualization); Simul8 (Simulation); nVivo (Qualitative Analysis); Salesforce (CRM); Maple (Engineering Analysis); SolidWorks (CAD); ANSYS (FEA); NI LabVIEW (Data Acquisition)

PEER REVIEWED JOURNAL ARTICLES [PUBLISHED]

- Wright, M. D., Flanagan, M. E., **Kunjan, K.**, Doebbeling, B. N., & Toscos, T. (2016, May). **Missing links: challenges in engaging the underserved with health information and communication technology.** In Proceedings of the 10th EAI International Conference on Pervasive Computing Technologies for Healthcare (pp. 122-129). ICST (Institute for

Computer Sciences, Social-Informatics and Telecommunications Engineering).

- **Kunjan, K.,** Toscos, T., Turkcan, A., & Doebbeling, B. N. (2015). **A Multidimensional Data Warehouse for Community Health Centers**. In AMIA Annual Symposium Proceedings (Vol. 2015, p. 1976). American Medical Informatics Association.
- **Kunjan, K.**, & Lloyd, F. P. (2008). **Automated blood sampling and glucose sensing in critical care settings.** Journal of diabetes science and technology, 2(2), 194-200.

PEER REVIEWED JOURNAL ARTICLES [MANUSCRIPTS]

- **Kunjan, K.**, Toscos, T., Doebbeling, B.N., et. al. **Information Needs at Community Health Centers: Challenges and Solutions**. Journal of Medical Informatics Association [**In Preparation**]
- **Kunjan, K.**, Wu, H., Toscos, T., Doebbeling, B.N. **Large Scale Appointment Data Mining: Informing Patient-Centered Scheduling**. Journal of Health Services Research [**In Preparation**]
- **Kunjan, K.**, Doebbeling, B.N. Toscos, T., et. al. **A User-Centered Dashboard for Community Clinics: Design, Development and Evaluation**. Interacting with Computers **[In Preparation]**
- Toscos, T., Wright, M, Flanagan, M, **Kunjan, K,** Miller, A; Doebbeling, B (2016) **Tailored, theory-based strategies for engaging low income populations with a personal health record; Submitted to EAI Endorsed Transactions, 2016**
- Toscos, T., Carpenter, M., Flanagan, M, **Kunjan, K**; Doebbeling, B. **Identifying Successful Practices to Overcome Access to Care Challenges in Community Health Centers: A "Positive Deviance"** Approach; **Submitted to Medical Care**

PANELS, SHORT PAPERS, ABSTRACTS, POSTERS

- Doebbeling, B.N., Toscos, T., **Kunjan, K**, Wu, H., **Improving Access to Care and Efficiency: Patient-Centered Methods and Strategies**; Accepted Panel for iHealth 2017 Clinical Informatics Conference, Philadelphia, PA, May 2-4, 2017
- **Kunjan, K** (2016). **A Big Data Augmented Heterogeneous Warehouse and Analytics Platform for Community Health Centers**; Doctoral Consortium at the 10th EAI International Conference on Pervasive Computing Technologies for Healthcare 2016, Cancun, Mexico
- **Kunjan, K**, Lloyd, F., Giddings, D. (2015) **Pre-Clinical Evaluation of an Automated Blood Glucose Monitoring System for Critical Care Settings**. Abstract and Poster: International

Hospital Diabetes Meeting, San Francisco, Feb 2015.

- Mohammadi, I., Turkcan, A., **Kunjan, K.**, Toscos, T., Doebbeling, B.N., Miller, A. (2015). **Assessing and Simulating Scheduling Processes in Community Health Centers.** In AMIA Annual Symposium Proceedings.
- Mehrabi, S., Mohammadi, I., **Kunjan, K.,** & Kharrazi, H. (2013). **Effects of data transformation methods on classification of patients diagnosed with myocardial infarction.** In MedInfo (p.1203).
- **Kunjan, K**, and Lloyd, F.; **An AutoSampler for Blood Analysis**; Abstract: International Hospital Diabetes Meeting Abstract, San Diego, Oct 2010
- **Kunjan, K** and Lloyd, F.; **A Continuous Blood Glucose Monitoring System Using Mid IR Sensing Technology for Critical Care Settings**; Abstract and Poster Presentation: Diabetes Technology Meeting, Atlanta, Nov 2006

PATENTS

- **Kunjan, K** and Lloyd, F; **Automated Blood Sampler and Analyzer;** US Pat. 8,384,844; Dec 2008
- **Kunjan, K** and Lloyd, F.; **Continuous Whole Blood Glucose Monitor;** US Pat. 11982565; Nov 2006
- Gore, J, Kumar S, **Kunjan, K**, et. al.; **Method and apparatus for measuring a substance in a biological sample**; US Pat. 10692996; 2003

GRANTS

- **National Science Foundation Travel Grant** (**$2000)** to Pervasive Health Conference, Cancun, Mexico, May 16-18, 2016
- **National Institute of Justice** (NIJ) Paul Coverdell Forensic Science Improvement Grant **($300,000)** for the Marion County Coroner's Office, Indianapolis, IN; FY 2009-10 [$125,000] and FY 2011-12 [$175,000]
- **Purdue University Technology Innovation Award ($250,000)** for commercialization of university research in the area of mid infrared sensing for biological applications; Gore, J and Kunjan, K; 2002-04

REVIEWER OF SCIENTIFIC ARTICLES

- American Medical Informatics Association [AMIA] 2017 Annual Symposium Nov, 2017 Washington DC

- American Medical Informatics Association [AMIA] 2015 Annual Symposium Nov, 2015 San Francisco
- American Medical Informatics Association [AMIA] 2014 Annual Symposium Nov, 2014 Washington, DC
- The Journal of the American Medical Association [JAMA], Feb, 2014

SPECIAL RECOGNITION

- Student Spotlight Video: Featured in the school website and marketing campaign as a doctoral candidate in the Indiana University School of Informatics in Computing, Indianapolis, IN [2015]
- Featured in video highlighting my research work that improved efficiency and access to care at community health centers across the state of Indiana.
- Received recognition by Mayor Ballard of Indianapolis, IN as a consultant for successfully implementing a lean six sigma project, reducing the cycle time by half for forensic investigations [2009]
- Photo Featured in Nanotechnology/MEMS in a book 'Engineering and Technology' by Michael Hacker, et. al., [ISBN 13: 9781418073893] [2010]
- Featured in the Journal and Courier (www.jconline.com) Top 8 in the country in the Purdue University Burton D Morgan Business Plan Competition. Won $2500 in prize money [2009]
- Photo featured in Purdue Research Park Innovation News for a company developing a continuous glucose sensor for diabetes monitoring [2004]
- $10,000 in prize money [Fourth place] as the Masters student (sole) presenter in the Purdue University Burton D Morgan Business Plan Competition [2002]