# SIMULATION AND MODELING FOR IMPROVING ACCESS TO CARE FOR UNDERSERVED POPULATIONS

Iman Mohammadi

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

December 2018

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

Doctoral Committee

Josette F. Jones, RN, PhD, Chair

Edward Berbari, PhD

October 29, 2018

Xiaowen Liu, PhD

Huanmei Wu, PhD

## DEDICATION

Dedicated to my parents and my sister.

ACKNOWLEDGEMENTS

I would like to start by expressing my deepest gratitude to my advisor, Dr. Huanmei Wu, for her brilliant guidance, caring, patience, and providing me with support for conducting this research. I am sincerely grateful to my thesis committee chair Dr. Josette Jones for her tremendous help during my PhD studies and thesis work. She has been supporting me both morally and scientifically since the days before I was even admitted to the PhD program. I would like to thank my thesis committee member, Dr. Xiaowen Liu for graciously accepting to be part of my thesis committee. I want to specifically acknowledge my gratitude to Dr. Edward Berbari for his guidance since I started my graduate endeavor in engineering space. I want to send my special thanks to Dr. Brad Doebbeling, Dr. Tammy Toscos, Dr. Ayten Turkcan, and Ms. Amy Miller for their guidance and contributions during my PhD study. Finally, I thank Patient-Centered Outcomes Research Institute (PCORI) for supporting this work with PCORI award (IH-12-11-5488).

Iman Mohammadi

SIMULATION AND MODELING FOR IMPROVING ACCESS TO CARE FOR UNDERSERVED POPULATIONS

This research, through partnership with seven Community Health Centers (CHCs) in Indiana, constructed effective outpatient appointment scheduling systems by determining care needs of CHC patients, designing an infrastructure for meaningful use of patient health records and clinic operational data, and developing prediction and simulation models for improving access to care for underserved populations. The aims of this study are 1) redesigning appointment scheduling templates based on patient characteristics, diagnoses, and clinic capacities in underserved populations; 2) utilizing predictive modeling to improve understanding the complexity of appointment adherence in underserved populations; and 3) developing simulation models with complex data to guide operational decision-making in community health centers. This research addresses its aims by applying a multi-method approach from different disciplines, such as statistics, industrial engineering, computer science, health informatics, and social sciences. First, a novel method was developed to use Electronic Health Record (EHR) data for better understanding appointment needs of the target populations based on their characteristics and reasons for seeking health, which helped simplify, improve, and redesign current appointment type and duration models. Second, comprehensive and informative predictive models were developed to better understand appointment non-adherence in community health centers. Logistic Regression, Naïve Bayes Classifier, and Artificial Neural Network found factors contributing to patient no-show. Predictors of appointment non-adherence might be used by outpatient clinics to design interventions reducing overall clinic no-show rates. Third, a simulation model was developed to assess and simulate scheduling systems in CHCs, and necessary steps to extract information for simulation modeling of scheduling systems in CHCs are described. Agent-Based Models were built in AnyLogic to test different scenarios of scheduling methods, and to identify how these scenarios could impact clinic access performance. This research potentially improves well-being of and care quality and timeliness for uninsured, underinsured, and underserved patients, and it helps clinics predict appointment no-shows and ensures scheduling systems are capable of properly meeting the populations' care needs.

Josette F. Jones, RN, PhD, Chair

TABLE OF CONTENTS

# CHAPTER ONE: INTRODUCTION & BACKGROUND

## 1.1 Opening Remarks

Underserved populations are individuals suffering from economical disadvantages, being in racial and ethnic minorities, and being under- or un-insured. Community Health Centers (CHCs) provide primary and mental care for underserved populations, and are designed as safety net for these populations. This research, through partnership with seven CHCs in Indiana, sought to determine care needs of CHC patients, develop an infrastructure for meaningful use of patient health data, and develop predictive modeling and simulation tools for improving access to care for our target populations.

Electronic Health Records (EHRs) are "secure, real-time, point-of-care, patient centric information resource for clinicians[1]." EHRs cover a wide range of different information, consisting of both unstructured narrative text as well as structured data. Many different health care professionals and facilities adopt and contribute to an individual medical record that is an in-house EHR. For example, Community Health Centers, federally qualified health centers (FQHCs), FQHC public clinics, private nonprofit clinics, rural health clinics, free clinics, and others serving diverse composition of patients have been adopting EHR systems. EHR data can be leveraged to develop tools for decision-making, research, and health policy purposes[2].

Analysis of population-scale clinical data can support discovering appointment needs, barriers to accessing healthcare, strategies to reduce access barriers, and operational strategies for CHCs. Preliminary research and literature review completed as part of this research find that: 1) there is an association between the characteristics of patients, diagnoses, and appointment type, and there is a need to redesign appointments based on those characteristics[3]; 2) patient and provider characteristics and visit features help predict appointment adherence[4]; and 3) clinical workflow analysis and EHR data analysis can be used to build the simulation models to identify effective scheduling processes and test alternate strategies to improve timely access to care[5]. The goal of this research is to identify care needs of underserved populations and to redesign healthcare systems to better meet these needs. Specifically, the objective of this project is to improve access to care for underserved population by using patient and clinic operational data. This project provides an opportunity to expose important correlation between target populations socio-economic statuses and access to care and health conditions.

## 1.2 Underserved Populations

This study focuses on improving access to care for poor, underserved, uninsured, or underinsured populations. Poverty and poor health outcomes are highly linked. According to World Health Organization, the determinants of health, i.e. factors affecting individual's health, may include income and social status, health services, education, physical environment, genetics, and gender[6]. Those determinants of health can significantly impact on patients' quality of life. Table 1.1 shows the U.S. Census Bureau demographics comparison between Indiana and the nation in 2016[7]. Indiana had higher percentage of children compared to the rest of the country. The rate of high school or bachelor's degree graduates are lower compared to the rest of the country. Indiana had higher percentages of poverty, uninsured, or disables comparing to the average in the United States. This data articulates importance of focus on underserved populations in Indiana.

**Table 1.1.** Demographic distribution comparison between the US and Indiana[7].

| People | Indiana | US |
|---|---|---|
| Population | | |
| Population estimates, July 1, 2016 | 6,633,053 | 323,127,513 |
| **Age and Sex** | | |
| Persons under 5 years, percent, July 1, 2015 | 6.3 | 6.2 |
| Persons under 18 years, percent, July 1, 2015 | 23.9 | 22.9 |
| Persons 65 years and over, percent, July 1, 2015 | 14.6 | 14.9 |
| Female persons, percent, July 1, 2015 | 50.7 | 50.8 |
| **Race and Hispanic Origin** | | |
| White alone, percent, July 1, 2015 | 85.8 | 77.1 |
| Black or African American alone, percent, July 1, 2015 | 9.6 | 13.3 |
| American Indian and Alaska Native alone, percent, July 1, 2015 | 0.4 | 1.2 |
| Asian alone, percent, July 1, 2015 | 2.1 | 5.6 |
| Native Hawaiian and Other Pacific Islander alone, percent, July 1, 2015 | 0.1 | 0.2 |
| Two or More Races, percent, July 1, 2015 | 1.9 | 2.6 |
| Hispanic or Latino, percent, July 1, 2015 | 6.7 | 17.6 |
| **Families and Living Arrangements** | | |
| Persons per household, 2011-2015 | 2.55 | 2.64 |
| Living in same house 1 year ago, percent of persons age 1 year+, 2011-2015 | 85.0 | 85.1 |
| Language other than English, percent of persons age 5 years+, 2011-2015 | 8.3 | 21.0 |
| **Education** | | |
| High school graduate or higher, percent of persons age 25 years+, 2011-2015 | 87.8 | 86.7 |
| Bachelor's degree or higher, percent of persons age 25 years+, 2011-2015 | 24.1 | 29.8 |
| **Health** | | |
| With a disability, under age 65 years, percent, 2011-2015 | 9.7 | 8.6 |
| Persons without health insurance, under age 65 years, percent | 11.2 | 10.5 |
| **Economy** | | |
| In civilian labor force, total, percent of population age 16 years+, 2011-2015 | 63.9 | 63.3 |
| **Income and Poverty** | | |
| Median household income (in 2015 dollars), 2011-2015 | 49,255 | 53,889 |
| Per capita income in past 12 months (in 2015 dollars), 2011-2015 | 25,346 | 28,930 |
| Persons in poverty, percent | 14.5 | 13.5 |

## 1.3 Community Health Centers

Community Health Centers (CHCs) are safety-net clinics providing primary care for underserved and under/uninsured populations. CHCs provide primary care services for acute and chronic diseases, injuries, and preventive services. CHCs frequently provide services, such as dental and vision care, counseling services, behavioral health, and substance abuse services[8]. CHCs may also offer other social or public health services, such as health education, translation and transportation. Majority of CHC patients are typically female and children[9]. About half of the patients in CHCs are from ethnic minorities. Currently, there are 48 Indiana Community Health Centers serving high-need communities, such as medically underserved areas, medically underserved populations, and health professional shortage areas. They provide services, such as comprehensive primary care and supportive services including translation and transportation. They are committed to be available to all residents of their service areas regardless of income status.

### 1.4 Access to Health Care

Access to health care is defined as having "the timely use of personal health services to achieve the best health outcomes[11]." The National Healthcare Quality Report[12] discusses that good access can be achieved in three steps: (1) getting into the health care system; (2) having access to sites providing needed health care services; and (3) finding providers that can meet the individual patient's medical needs. One of the key points in advancement and development of a healthy society is to improve access to care. Access to health care is not only important for improving the patient experience, but also health outcomes[13].

The Affordable Care Act[14] has provided a foundation to expand the number of insured Americans; hence, increased demand for healthcare systems. If accessibility and availability of primary care providers could not meet the high demand, it may lead to hospital emergency rooms becoming the default provider of primary care, and that is costly to the healthcare system[15]. Many studies have reported that limited access to care is associated with bad health outcomes and poor population health outcomes[16]. Improving access to care is critical for establishing health equity, increasing quality, and improving health outcomes. High missed appointment rates have been identified as one of the most significant barriers to access to care for underserved populations[5,17]. In semi-structured interviews conducted at our partner CHCs, clinic staff and providers agreed that a high missed appointment rate is a major problem[18].

Availability, accommodation, and affordability of care are the major areas of access. Availability barriers are the shortage of primary care providers or patients not knowing how to get care. Timeliness of appointments, forgetting about having appointments, and transportation are some examples of accommodation barriers. Not being insured and not being able to pay for medical expenses are some of affordability barriers[19]. Improving timeliness of getting care and eliminating barriers to health care access are critical for promoting optimal health and wellness for poor and uninsured individuals[20]. As it is shown in table 1.1, Indiana has a high population of underserved. According to National Healthcare Quality and Disparities Report 2014, Indiana is among the states with the lowest quality (shown in figure 1.1) of and access to health care[12].

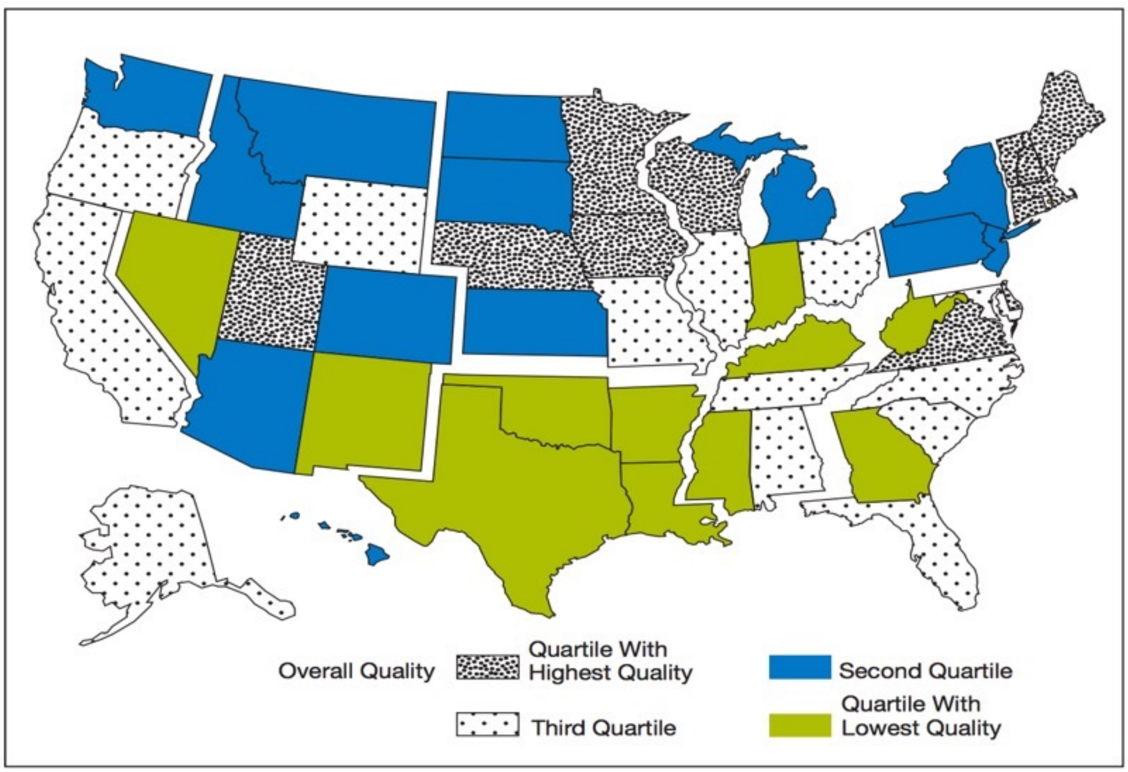



**Figure 1.1.** Overall healthcare quality in the US (source: National Healthcare Quality and Disparities Report[12]).

## 1.5 Electronic Health Records

Electronic health records (EHR) can assist in providing vital information and data to better understand the patient and the utilization of CHCs. EHR is “a secure, real-time, point-of-care, patient centric information resource for clinicians[1,21].” EHR in CHCs is still limited with a need to a proper EHR data collection, aggregation, and analysis[22]. An example of EHR critical patient-related data include medical history, medication and allergies, immunization status, laboratory test results, vital signs, demographic, and billing data[21]. Comprehensive data analysis of EHR supports assessments of community health patients by offering real-time (up-to-date) data linked with their social determinant of health[23]. For example, researchers used linked EHRs and found a trend in diabetes, hypertension, and hypercholesterolemia across community hospitals population in the large metropolitan area of Indianapolis[23]. To document these health care needs and risks in uninsured population, health conditions were analyzed from the National Health and Nutrition Survey 2007-2010 for about 1042 uninsured adults in comparison with 421 low-income adults with Medicaid insurance[24]. The uninsured samples were characterized with a low income of no more than 138% of the federal poverty level. The results indicate that uninsured adults are likely to be obese, “to report a physical, mental, or emotional limitation”, and to have several chronic conditions[24].

Most previous studies addressed the importance of health information technology (HIT) for better care quality in general, but a few publications studied use of EHR data in underserved, rural, and other safety-net populations. Improving quality of care for minority and low-income population is one of the focuses of the Commonwealth Fund[25]. High performance care models, patient-centered care models, and improvement of policy making are the main objectives of the Commonwealth Fund. In primary care, they found that CHCs have been able to deliver comparable care to other care delivery systems. In a literature review on impacts of EHRs on quality of care in underserved population, they found evidence showing EHRs can improve quality of care[26]. The literature review suggests that the effectiveness of EHR data-driven tools should be further studied. In our study, we propose methodologies to make use of EHR data, and to test the effectiveness of such EHR data-driven tools.

### 1.6 Problem Statement and Proposed Approaches

In our multisite and multiyear Patient-Centered Outcomes Research Institute awarded study entitled "Improving Healthcare Systems for Access to Care and Efficiency by Underserved Patients", we aimed to 1) identify and prioritize patient-centered strategies to reduce accommodation barriers to care; 2) develop a computer simulation model of community health centers to model and test the feasibility and impact of the proposed strategies; and 3) coach clinics to implement multicomponent strategies to enhance access and continuity of care for all patients. In this study, we partnered with 7 CHCs across Indiana and found that effective scheduling addresses accommodation barriers, and improving CHC appointment scheduling is key to improve access to care for underserved populations. In this PhD dissertation work, we seek to apply simulation and modeling techniques to improve existing CHC appointment scheduling that would consequently improve access to care for underserved populations. This research does not only potentially improve the well-being of underserved populations, it also provides a basis for the development of decision-making tools for healthcare systems redesign. This research reaches its goals by applying a multi-method approach from different disciplines, such as statistics, industrial engineering, computer science, health informatics, and social sciences. The specific aims of this study are illustrated in the following.

**Specific aim 1) Redesigning appointment scheduling based on patient characteristics, complaints, and diagnoses in underserved populations.** A key to a more efficient scheduling systems is to ensure appointments are designed to meet patient's needs and to design and simplify appointment scheduling less prone to error. In this specific aim, we used EHR data to investigate the relationship between diagnoses, patient characteristics, and reasons for visit to help providers redesign healthcare systems that can meet the needs of underserved patients.

**Specific aim 2) Utilizing predictive modeling to improve understanding the complexity of appointment adherence in underserved populations.** One key measure for improving access to care is reducing the number of appointment no-shows. In this aim, we applied predictive modeling and machine learning approaches to identifying factors impacting appointment adherence in underserved populations.

**Specific aim 3) Developing simulation models with complex data to guide decision-making in community health centers.** Effective scheduling processes can reduce clinic no-show rates and patient waiting time while improving continuity of care and overall clinic performance. In this aim, computer simulation models were developed to assess and simulate the scheduling processes in CHCs, and provide a decision-making tool for clinic managers to analyze the impact of alternate scheduling modalities.

This work tackled the issue of access to care for underserved populations from three different angles (shown in figure 1.2). First, a novel method, a combination of natural language processing and unsupervised machine learning, was developed to use EHR data for better understanding of appointment needs of the target populations based on their characteristics, which helps simplify, improve, and redesign current appointment type and time models. Second, comprehensive and informative predictive models were developed to better understand appointment non-adherence in community health centers. Third, a guide tool described how to extract necessary information and how to develop of simulation modeling of appointment scheduling systems in community health centers.

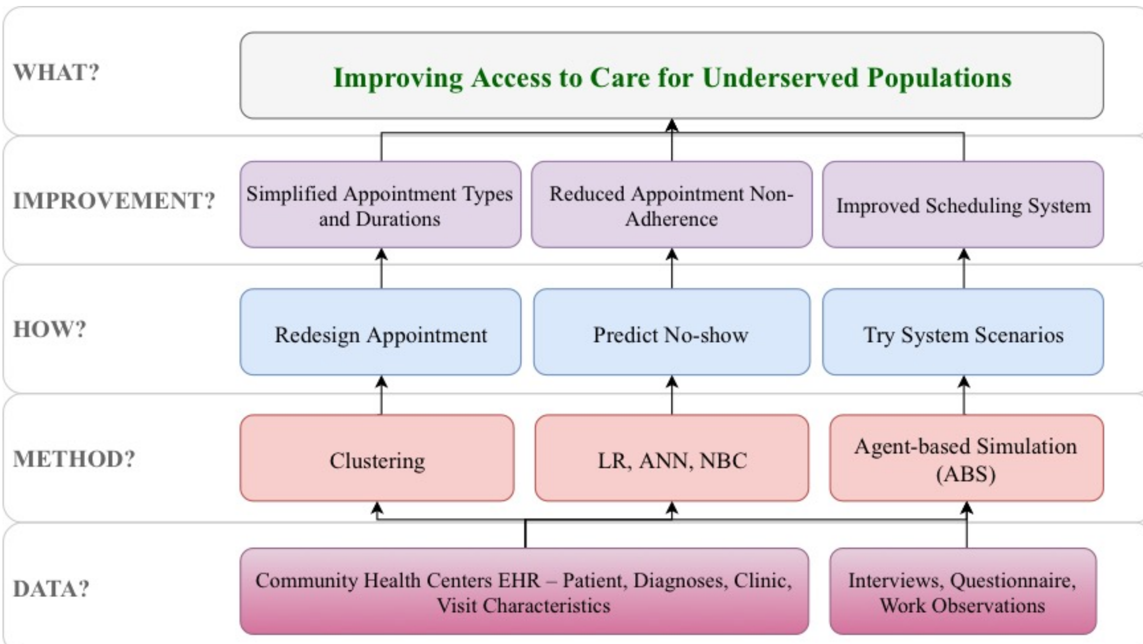



**Figure 1.2.** Research strategy. Abbreviations: EHR, Electronic Health Records; LR, Logistic Regression; ANN, Artificial Neural Network; NBC, Naïve Bayes Classifier.

In response to the aims of this research, we answered the following groups of research questions:

***RQ1:*** What are the care needs in underserved population? What are the common diagnoses or disease trends in the target population? How is the socio-economic status associated with diagnoses? What are the potential analytics approaches that can address the population needs? What is the most effective way to redesign appointments in community health centers to meet health care needs of underserved populations?

***RQ2:*** What are the important factors impacting appointment adherence in community health centers? What are the approaches to improve understanding and predicting of no-show behavior of underserved populations? Will machine learning improve accuracy of no-show prediction models compared to statistical models?

***RQ3:*** What are the necessary steps to develop a simulation toolkit for improving scheduling systems in community health centers? What are the approaches to collect data for simulation modeling? How do simulation models of a community health clinic become more realistic?

This research and its findings are described in five chapters. In chapter one, we describe an introduction to the problem and our target populations. In chapter two, we describe the clustering techniques used to redesign appointment types based on patients' reasons to seek health. It also describes results and discussions around appointment scheduling redesign characterized as aim one. Chapter three describes developments and results of three appointment no-show predictive modeling methodologies that are Logistic Regression, Artificial Neural Network, and Naïve Bayes Classifier. We also discuss the impacts of clinic characteristics, localities, and patient social determinant of health on appointment adherence behavior. Chapter four describes application and development of agent based modeling in healthcare system improvement, and we thoroughly explain the step by step requirements of clinic scheduling and patient flow simulation modeling development. Chapter four also includes an example of a real world problem that was undertaken in our research accompanied by the intervention designed for improved scheduling and access. In chapter five, we discuss potential real world application of this study, and it also includes clinical and scientific significance, and contributions of this study to the field of Health Informatics.

# CHAPTER TWO: APPOINTMENT SCHEDULING REDESIGN

Meeting the health care needs for underserved populations is crucial[3]. Appointment scheduling that is efficient, patient centered, and effective can satisfy health care needs of those populations. In this chapter, we aimed to redesign appointment scheduling in community health centers by identifying care needs of clinic patients and simplifying scheduling templates. In this chapter, we discuss application of “word embedding”, a natural language processing approach, to extract important information from patient records. We used extracted information to cluster patients into groups based on similarity of their reasons for seeking health. We then used patient clusters to redesign appointment types and durations that are simpler and more efficient without adding additional burden on clinics. CHCs managers and other stakeholders might use the findings of this study to restructure their health care systems. The approaches in this chapter can also be a roadmap for developing automated appointment scheduling tools for ambulatory care settings.

## 2.1 Appointment Scheduling

Appointment scheduling in health care is different from scheduling in other industries as patient medical, physiologic, and mental state is dynamic that would cause uncertainty in patient flow[27]. In acute health care systems, triaging is applied to evaluate acuity and meet demands; however, in non-acute settings, triage scheduling is not the most effective way of scheduling. Open access (same day access) models have also been proposed for scheduling appointments in outpatient setting[28]. Outpatient settings should consider factors, such as number of services, number of providers, patient arrival process, number of appointments, service times, and provider punctualities to design their scheduling systems[29]. In this chapter, we focus on redesigning appointment scheduling to meet the needs of patients. Three aspects of the new appointment scheduling include relationship between appointment scheduling and diagnoses, and patient and visit characteristics; redesigning appointment types and durations to help improve access; and focusing on appointment scheduling in underserved settings.

## 2.2 Previous Works in Appointment Redesign

Many appointment-scheduling methods have been developed to address issues such as demand uncertainty, urgent care, and no-shows. These studies were mostly aimed at improving access to care and clinic service quality that would potentially reduce costs, patient waiting time, and provider idle and over time[30]. They redesigned appointment types and time slots using only the clinical constraints rather than the traditional way that required an objective function. Many studies focused on open access scheduling which allows patients to see a provider on the same day of requesting an appointment[31]. Most of these studies focused on implementation of open access scheduling and determining the impact of this scheduling system on patient and clinical outcomes. Common conclusion of these studies is that simplifying appointment types is an important principle to implement open access scheduling[32-34]. Other studies suggest simpler appointment types reduce complexity in scheduling leading to less error and better access to care[35].

One study focused on how to redesign scheduling based on patient characteristics[36]. They redesigned a scheduling scheme in a hospital based on patient characteristics to improve utilization of medical resources. They used patient’s characteristics to determine the length of required procedure time. They used a decision tree technique for reclassification of patient groups. They found that the proposed scheduling scheme reduced patient wait time, increased the radiographer utilization, and reduced overall cost. Another study focused on impact of standardizing appointment time slots on quality of care and access to care[37,38]. They concluded that enforcing standardization on appointment durations could have negative impact on access and productivity measures.

Most studies on designing appointment scheduling focused on implementation on open access scheduling or the assessment of new scheduling systems. One study proposed a scheduling template by taking clinic goals to considerations rather than designing appointment scheduling based on care needs and characteristics of patients. In studies focusing on improving access and implementation of open access scheduling, decreasing number of appointment types or simplifying appoint types in scheduling systems was recognized as a key step towards successful implementation. However, in these studies, they did not discuss the most appropriate ways to simplify appointment types. They offered appointment types such as new, established, acute, and postoperative as decreased number of appointments. Concerns about management of established patients still remain unanswered. We, through this introductory literature review, found that previous work in redesigning appointment scheduling did not propose patient-centered appointment windows for optimizing scheduling systems. Few studies focused on improving appointment scheduling based on patient characteristics, but the proposals were to accommodate health care settings like emergency department, radiology department, and inpatient settings rather than helping community health centers or outpatient settings providing care for underserved populations. In this chapter, we utilize real world encounter data in community health centers to identify appointment needs of underserved populations. We discuss how we might leverage patients encounter data including diagnoses and reasons for seeking care to construct patient-centered appointment scheduling.

### 2.3 Care Needs of Underserved Populations

Meeting the health care needs for underserved populations is crucial. We used EHR data to investigate the relationship between diagnoses and patient characteristics to help providers redesign healthcare systems that can meet the needs of underserved patients. Our analyses help healthcare providers and stakeholders identify strategies to improve access to care and health outcomes for underserved populations. Here, we present results of an analysis of EHR data that highlights the relationship between diagnoses, identified by International Classification of Disease (ICD-9) codes, and patient characteristics.

#### 2.3.1 Consensus Clustering for Patient Diagnoses

In our multiyear and multisite study[3], we have collected EHR data from seven CHCs across the state of Indiana. The semi-structured EHR data were extracted using SQL queries for each of the EHR systems utilized by CHCs in our study. We ran SQL queries on the clinics databases, and EHR data was received from CHCs in “csv” format. We followed the Health Insurance Portability and Accountability Act (HIPAA) rules of data de-identification. We use Microsoft SQL database to organize, manage, and query the data. Our database includes five years of EHR data with more than 3M encounters data from 700k patients. In this project, we used a dataset from a CHC in Indianapolis including 10 clinics spread out across the city. The dataset included appointment data (>700k appointments) for more than 85,000 patients from January 1, 2014 to April 30, 2016. We removed appointments scheduled for dental care, vision care, and telephone encounters (e.g. medication refill).

Data fields were pseudo patient ID, encounter date, age, ethnicity, race, sex, zip code, marital status, primary insurance type, income, tobacco use, and ICD-9 codes. We grouped ICD codes based on the "2015 ICD-9-CM Diagnosis Codes" chapters (http://www.icd9data.com/). Labels for each ICD code chapter can be seen in figure 2.1. ICD codes co-occurred in patients, and patients had between 2 to 4 ICD codes on average (Median=2, Mean=2.7). In order to find the associations between chapters in the population, we used consensus clustering to visualize the association using a heat map (shown in figure 2.1). We created three clusters of diagnoses shown in figure 2.1. If 75% or more of ICD codes for a given patient were from a specific cluster, we considered the patient to be in the cluster. We tested the association between patient demographics and clusters to understand what groups of patients were more likely to have a certain cluster of diagnoses. We used SAS 9.4 to perform ANOVA test for continuous variables and Chi-Square test for categorical variables.

### 2.3.2 Association of Diagnosis and Access to Care

The final dataset included diagnoses of 76,454 patients. Figure 2.1 is the heat map showing the association between diagnosis chapters and the three clusters. Cluster1 included 3,666 patients where at least 75% of the ICD codes belong to one of the chapters in the cluster. Cluster2 included 596 patients. Cluster 3 included 1,912 patients.


**Figure 2.1.** Eighteen groups of ICD codes (right). Heat map showing association between chapters (left).

Age, ethnicity, gender, and marital status were found to be statistically significantly associated with clusters (Figure 2.2). Hispanic or Latino patients are more likely to be in cluster 2, and non-Hispanic or non-Latino patients are more likely to be in cluster 3 ($p=.0002$). Female patients are more likely to be in cluster 2 while male patients are more likely to be in cluster 3 ($p<.0001$). Single or separated patients are more likely to be in cluster 3 while married patients are more likely to be in cluster 2 ($p<.0001$). Most widowed patients fall in cluster 1 ($p<.0001$). Patients in cluster 2 are older than patients in cluster 1 and cluster 3 ($p<.0001$). Income, smoking status and race are not statistically significantly associated with any of the diagnosis clusters.

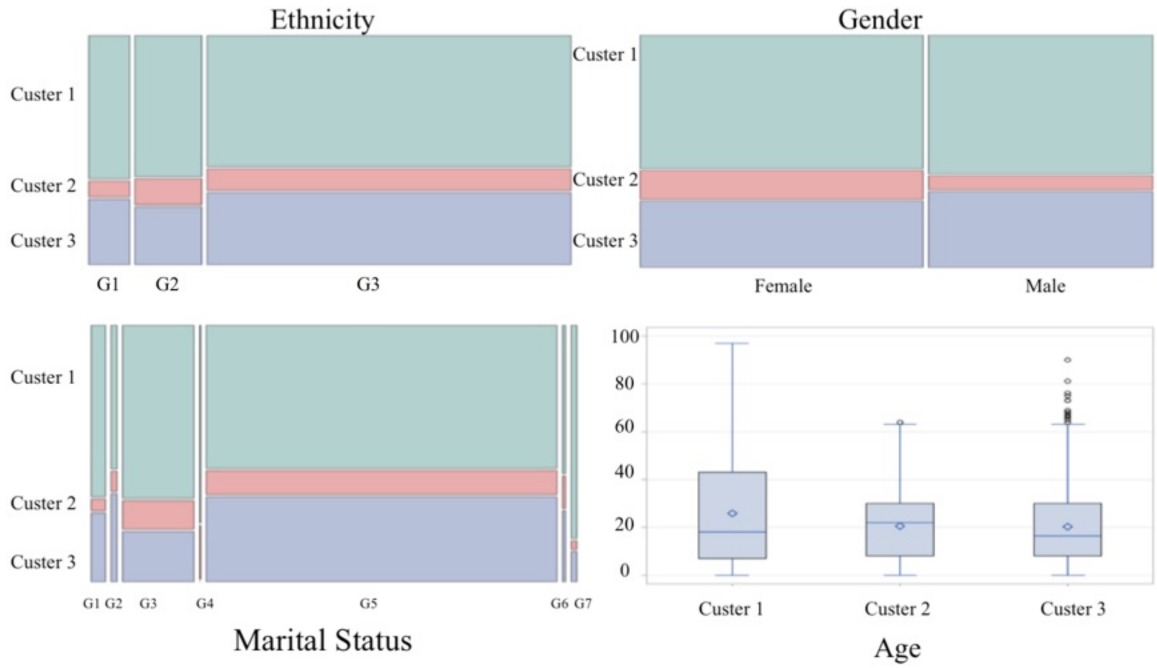



**Figure 2.2.** Distribution of statistically significant patient characteristics in 3 clusters. Ethnicity groups: G1, Unspecified; G2, Hispanic or Latino; G3, Not Hispanic or Latino. Marital Status groups: G1, Divorced; G2, Legally separated; G3, Married; G4, Partner; G5, Single; G6, Unknown; G7, Widowed.

This exploratory work to identify the health care needs of underserved patients revealed factors such as age, ethnicity, gender, and marital status in underserved populations are related to specific group patient types and diagnoses. This information may help clinics define how to cluster appointments for patients to improve access to care. It may also help CHCs prioritize utilization of resources. For example, results from our work could help clinics design culturally appropriate intervention aimed at Hispanic patients targeting diagnoses in cluster 2. While income was not found to be a significant factor to cluster diagnoses, we believe this could be due to the fact that more than 80% of the patients in the population have income less than 100% federal poverty level, therefore patients have similar income level. These insights provide a tool and approach to address the objectives of patient-centered redesign of appointment scheduling.

## 2.4 Word Embedding

Natural Language Processing (NLP) has been widely used to enable computers understand free text and use the information derived from free texts[39]. NLP includes wide range of computational techniques used by machines to human-like language processing[40]. Word embedding is one of feature learning techniques in NLP that word, phrases, or sentences are mapped to number of vector of numbers[41]. Word embedding is able to derive semantic relationships between words using deep learning algorithms[42]. Many studies in areas, such as sentiment analysis, information retrieval, and information extractions have applied word embedding[42]. It also has been utilized in healthcare arena, such medical synonym extraction[43], drug-drug interaction[44], and medical abbreviation disambiguation[45]. There are some open source word embeddings, such as Google News and GloVe. In a recent study of NLP, word embedding was found successful to capture semantics of medical terms and similarity between them[42]. The findings of their study were comparable to human judgments. The source of free text data in our project is patient compliant for seeking health, also known as reason for appointment. This field is entered by schedulers into the partner CHC EHR systems. The objective of this study is to utilize word embedding to extract information from reasons for appointments, and then aggregate the similar reasons into single concepts. Those concepts are used to create new appointment types and durations.

There are various word embedding models that map words to vector of real numbers that generally can be categorized to two methods of matrix factorizations and shallow window-based models. Matrix factorization methods capture the statistical information about the corpus. Approaches, such as latent semantic analysis (LSA)[46], capturing the term document frequencies or Hyperspace Analogue to Language (HAL)[47] capturing the term-term frequency are of some of matrix factorization methods. The problems with these methods are that most frequent words contribute a disproportionate amount to the similarity measure for instance co-occurrence with words such as “the” or “a” has large effect on the similarity measure despite lack of semantic relatedness. An entropy-based or correlation-based normalization can address this issue as a preprocessing. Shallow Window-Based Methods [48] is another approach to learn the word representation. The skip-gram and continuous bag-of-words (CBOW) models are two of the widely used word2vec approaches that use neural network structures in learning word representations. In the CBOW, the objective is to predict a word given its context while in skip-gram models the objective is to predict a word’s context given the word itself. These methods suffer from taking advantage of the vast amount of repetition in the data as they scan the corpus on small window size across the corpus.

### 2.4.1 Methods

Data: We collected the EHR data from an urban community health center included patient, visit, and provider characteristics. The chief complaint or schedulers’ notes was the main data point to extract information. Scheduler note is the reason patient seeking to come to the clinic. For example, when patient calls the clinic and ask for an appointment, the scheduler enters the patient explanation into the EHR system. We used schedulers’ notes to cluster patients based on the similarity of reasons seeking health. Schedulers’ notes are free text fields with many abbreviations; therefore, any attempt to extract information should include dealing with abbreviations.

Text mining: MedTagger was developed by Mayo Clinic and contains a suite of programs indexing based on dictionaries[49]. We used MedTagger[50] dictionary list to expand the abbreviation to their full forms, for example "DM" is transformed to "diabetes mellitus". We then used Global Vectors for Word Representation (GloVe)[51] method to represent each text column with their real-valued vectors. In this work, we used the GloVe model developed to answer the shortcoming of the earlier models. GloVe captures the benefits of count data while simultaneously capturing the meaningful linear substructures prevalent in recent log-bilinear prediction-based methods like word2vec. We used the GloVe pre-trained vectors on 6 billion token corpus of 2014 Wikipedia and constructed a 50 dimensional vector for every word in the text that appeared in the pre-trained model, and a normalized vector for words without representation in the pre trained vector model. We then averaged all the vectors for the words in the sentence to calculate the final representation of each sentence. For each patient encounter, there exists a chief complaint that is a free text. We ran the word2vec algorithm on each encounter complaint. Each encounter was converted to a row with 50 columns representing the 50 dimensional vector that is derived from the complaint. Figure 2.3 shows analysis engines used for redesigning appointment scheduling templates.

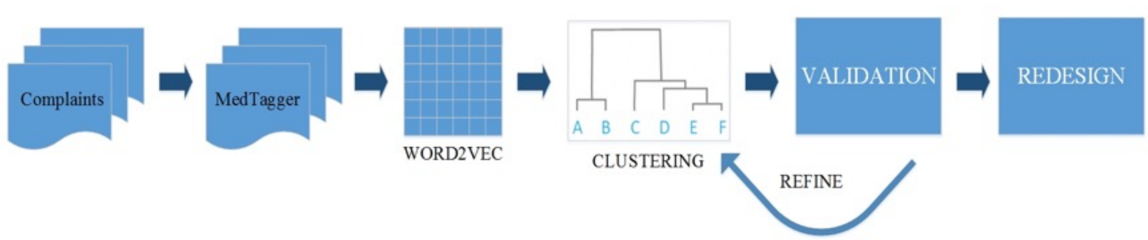



**Figure 2.3.** Analysis engines used for redesigning appointment scheduling templates

Clustering: The data was then fed into an Agglomerative Clustering algorithm. Agglomerative clustering is a bottom-up hierarchical clustering approach by merging pair of clusters, while the clusters generated in earlier step might be nested within the ones generated later. This approach does not necessarily neglect the small clusters; hence, it is useful for the discovery of the smaller groups. It also allows for utilizing the different distance matrices based on the nature of the given problem[52]. We used “AgglomerativeClustering” from “sklearn” library in Python to develop our clustering algorithm. To find the optimal number of clusters, we started with 2 clusters and stepwise increased the number of clusters to 20. In each run, we compared the results of clustering by analyzing the profile each cluster. Attributes such as age, gender, and provider specialty were used to objectively validate appropriate number of clusters. We also subjectively evaluated the clusters by reading 100 notes per cluster on average to see whether clustered notes are aligned with human judgment. We found optimal number of clusters is between 10 to 12. We chose 10 as our final number clusters for this study.

Appointment type and duration redesign: In this step, we assigned new appointment durations to the new appointment types, i.e. the 10 clusters. we assumed that the clinic capacity and demand do not change to investigate how standalone simplification of appointment types and durations could potentially impact access measures, such as number of patients seen per year and patient satisfaction defined as patient time spent in-clinic. Patient time spent in clinic is the difference between patient arrival and departure times and includes sum of waiting time to see the provider, time with the provider, and time spent for check out and payments. Proposed appointment durations were calculated based on the capacity that clinic must accommodate patients on a daily basis. The sum of provider hours allocated to see patients per day was defined as daily clinic capacity. For example, if the clinic had two providers on a given day, who each allocates 4 hours to see patients, the total capacity of the clinic on that day is 8 hours (i.e. 480 minutes), that is 240 minutes per provider. We used the distribution of current appointment durations per cluster to determine the most effective appointment durations for each cluster. For each cluster, we assigned the percentiles of current appointment durations (from minimum to maximum by 1 percentile) to proposed durations. We then used the capacity and demand of clinic to calculate performance measures, such as number of patients seen per year, provider time with patient, and predicted patient time spent in clinic. We fit the current daily demand to the current daily capacity using the proposed types and durations. We found the most effective duration for each cluster by maximizing number of patients per year and provider time with patient while minimizing overall patient time spent in-clinic.

Access measure in the proposed appointment system: The difference between current durations and proposed durations was calculated as time available to see more patients. We then calculated number of additional patients can be seen in the proposed system by dividing the time available to see more patients by new appointment durations per cluster. To normalize this measure, we calculated the number of additional patients that the clinic can see in the proposed scheduling system by year.

Time spent in clinic: The time patients spent in clinic, that includes in-clinic waiting time plus time spent seeing the provider, was calculated as the difference between arrive time and departure time recorded in the EHR data. We used current appointment durations, arrival time (AM vs PM), gender, provider specialty, number of provider available in the day of appointment, day of week, and patient age as independent variables and in-clinic time as dependent variable to develop multivariate linear regression model. We used the prediction model to predict time spent in clinic using the proposed appointment durations. Paired wise t-test calculated significance level of difference between current and proposed systems.

### 2.4.2 Results

Clustering: We collected 17,722 encounters of an urban community health clinic in 2014. The dataset included deidentified patient ID, day and time of encounter, patients' arrival and departure times, age, gender, provider ID and specialty, appointment type (102 types recoded in the EHR), and chief complaint (or schedulers' notes). The dataset included complaints from 7,061 patients in 2014. Following data processing, NLP implementation, and clustering, appointment types were grouped into 10 clusters using patients' complaints in the current scheduling systems (shown in table 2.1).

Tables 2.1, 2.3, and 2.4 were used to determine the most accurate number of clusters. Our proposed scheduling system has 10 types of appointments (noted as clusters). Table 2.1 shows examples of free texts that were aggregated into one concept. Cluster 1 seems to be appointments that are assigned to patients with complex issues. Cluster 2 represents acute women problems or patients with behavioral health complaints. Cluster 3, that is the largest in terms of number of complaints, is for acute care encounters that need to be scheduled as soon as possible. Clusters 6 and 7 are assigned to patient with chronic pain problems and other chronic problems. Clusters 8 and 9 are dominantly for pregnant and other woman problems. Cluster 10 are for well and other child care patients.

**Table 2.1.** Cluster profiles and examples of reasons grouped into clusters.

| Clusters | Number of appointment types in current scheduling | Number of chief complaints | Examples |
|---|---|---|---|
| **1** | 6 | 203 | • knot on left breast is more tender and now hot touch not hot today has tried ibuprofen and tylenol f<br>• NS-f/u adhd & med mom needs refill on Eucerin Cream mom is unfamiliar which dose of Concerta Perry i<br>• SVB, no epidural but used IV pain Rx " I breastfed for two weeks and then stopped. My nipples were b<br>• Infertility eval per CT per pt has been trying to get preg x 3 yrs with no success has one child<br>• colpo R/S due to + Trich in pap test and pt did not come in to R/O via urine Needs Urine Testing for |
| **2** | 27 | 859 | • trouble sleeping<br>• med f/u-anxiety<br>• hx of BV, pap, hx of, urine concentrated, but not now, burning on urination<br>• poss UTI/ pt is about 7 weeks PG<br>• Bipolar, Anxiety med f/u, |
| **3** | 37 | 4646 | • lightheaded, vomiting, intermittent umbilical pain,<br>• fever, cough with green mucus, increased seizures/Comm ins<br>• sinus congestion, vag discharge with odor, hx of BV<br>• pe dizziness @ times low abd. cramping @ times<br>• cough, congestion, runny nose, tired, decreased appetite |
| **4** | 27 | 494 | • irregular bleeding pelvic pain-P/S 8 x 2 mons. pain with sex and nausea , trying to get preg.<br>• rob headaches x 3 wks- P/S 5 @ times vaginal odor with no other syms. x 3 wks<br>• STD check pt states "menstrual cramps last week like I was going to start"<br>• birth control consult, here with involved mother has tried depo last IM 12/2013, reports not happy w<br>• asthma check mother concern speech not clear -history of father having speech problems child |
| **5** | 37 | 2619 | • DEPO Needs current start depo order, last one expired 4/29/14<br>• newborn/jaundice,vfc NB records given to JM-per Kennecia<br>• fu multiple ED visits for abdominal pain, N&V - MCARE<br>• not sleeping x3-4 months COMMERCIAL INSURANCE first<br>• Postpartum, del 6/19; never had PP visit, wants depo |
| **6** | 30 | 2078 | • low back pain, pain in legs Has been taking wife's medications for pain<br>• left hand pain; chronic back pain/need new referral for pain management<br>• f/u cough, had LTBI appt last month, still cough and night sweats, sore throat<br>• Somali Interpreter - left shoulder pain and numbness, numbness in left foot<br>• stomach and chest pain (pt wanted to wait until this day for appt) |
| **7** | 34 | 1589 | • wants to get off work due to side effects of medication/"other problems"<br>• meds/gallstones-upper mid-abd.pain since Sat. -P/S 10 @times needs meds for bipolar<br>• Wt Management f/u (last visit 1/2014) - only came one time Has gained weight since January<br>• WCC REFILL MIRALAX-PHARMACY VERIFIED (chammans) COMMERCIAL<br>• F/U hospitalization for asthma, forms to be completed Needs asthma action plan |
| **8** | 24 | 317 | • New OB HX @10:15 nob packet given and instructed on verbal consent for uds and hiv declines mfm refe<br>• ROB 37 wks wants cx checked, increase in contractions and increase in pressure<br>• Pt to have biopsy review MRI Mammogram results- need addendum (on desk) review ortho referral, corr<br>• NS NB wt check mom and dad took pt to ER at st francis 5 days ago for abnormal breathing,pt was chec<br>• Repeat CS on 4-1-14, little boy baby "I breastfeed for the first few days." Had PPTL Smiling and tal |
| **9** | 31 | 728 | • vaginal discharge/ possible yeast *Unable to reach,please have pt call<br>• pregnancy symptoms, no period x10wks/ neg upt 01/14/14 trying to concieve x 5 years nausea, irritabl<br>• f/u labs/pelvic pain pelvic pain x 2 week c/o clear vaginal discharge +odor- itch<br>• Yearly Birth Control consult pt gets depo every 9 wks per C. Bonsack for BTB<br>• Right breast lump, grape sized, 5 o'clock, red around lump, hot to touch, painful |
| **10** | 16 | 511 | • 9 month wcc cough, cold sx's "twitching" episodes<br>• Well child 12 mos Commercial Insurance Vaccines UTD (CHIRPS Printed) chammans<br>• WCC CBC AND LEAD check on audiology appt.<br>• 14 month wcc (almost 15 months) & f/u rash<br>• wcc cough, cold sx's "twitching" episodes |

Table 2.2 illustrates the distribution of current appointment durations per each cluster. Appointment durations typically ranges between 10 to 60 minutes. Cluster 1 has

the highest durations, and this is aligned with the visit reasons shown in table 2.1, because it is given to complex patients. Cluster 10 has the lowest durations as it is given to well child care.

**Table 2.2.** Distribution of current appointment durations within each cluster.

| Appointment cluster | Average appointment duration in current scheduling system (minutes) | Standard Deviation | Percent of Appointments | Min | Median | 75th Percentile | Max |
|---|---|---|---|---|---|---|---|
| **1** | 20.0 | 7.7 | 1 | 15 | 15 | 30 | 30 |
| **2** | 17.6 | 5.8 | 6 | 10 | 15 | 20 | 60 |
| **3** | 17.5 | 6.3 | 35 | 10 | 15 | 15 | 60 |
| **4** | 17.6 | 5.4 | 3 | 10 | 15 | 20 | 60 |
| **5** | 18.1 | 7.0 | 20 | 10 | 15 | 15 | 60 |
| **6** | 16.8 | 4.8 | 13 | 10 | 15 | 15 | 60 |
| **7** | 17.8 | 6.2 | 10 | 10 | 15 | 20 | 60 |
| **8** | 19.9 | 7.3 | 2 | 10 | 15 | 30 | 60 |
| **9** | 18.9 | 7.0 | 5 | 10 | 15 | 20 | 60 |
| **10** | 17.1 | 4.2 | 5 | 15 | 15 | 20 | 60 |

The clinic overall patient gender distribution was 63 to 37 for female to male. Table 2.3 breaks down age and gender profiles of each cluster. Clusters 1 represents young patients from both genders. Cluster 10 shows 95% of patients are younger than 13 years old, and it represent children population. Cluster 8 and 9 are dominantly female patients. Cluster 3, that was concluded to be acute care base on table 2.1, represents all ages and genders. The gender and age profile of each cluster seems to be reasonable based on review of examples of reasons for visits.

**Table 2.3.** Distributions of patient age and gender within each cluster.

| Appointment cluster | Age | | | | | | | Gender | |
|---|---|---|---|---|---|---|---|---|---|
| | Mean | SD | Min | 5th percentile | Median | 95th percentile | Max | Female | Male |
| **1** | 17 | 11 | 2 | 2 | 18 | 33 | 33 | 66.67 | 33.33 |
| **2** | 27 | 19 | 0 | 1 | 26 | 61 | 89 | 75.1 | 24.9 |
| **3** | 24 | 21 | 0 | 0 | 18 | 64 | 89 | 65.71 | 34.29 |
| **4** | 22 | 18 | 0 | 0 | 21 | 59 | 96 | 76.36 | 23.64 |
| **5** | 21 | 19 | 0 | 0 | 19 | 59 | 96 | 70.77 | 29.23 |
| **6** | 22 | 20 | 0 | 0 | 16 | 62 | 88 | 66.16 | 33.84 |
| **7** | 22 | 19 | 0 | 0 | 17 | 62 | 90 | 65.11 | 34.89 |
| **8** | 28 | 16 | 0 | 0 | 25 | 61 | 90 | 83.81 | 16.19 |
| **9** | 29 | 19 | 0 | 0 | 27 | 65 | 86 | 80.89 | 19.11 |
| **10** | 3 | 7 | 0 | 0 | 1 | 13 | 76 | 49.86 | 50.14 |

Table 2.4 shows percentages of appointments within a cluster that were scheduled with various provider specialties. Cluster 2 is a mix of behavioral health and all other specialties. Cluster 3 (acute care) patients were scheduled with all types of specialties. Cluster 10 patient are dominantly scheduled with pediatricians.

**Table 2.4.** Percentages of appointments scheduled with each specialty within cluster.

| Appointment cluster | Behavioral Health | Family Practice | Internal Medicine | Nurse Practitioner | Woman | Pediatrics |
|---|---|---|---|---|---|---|
| **1** | 0 | 0 | 17 | 0 | 67 | 17 |
| **2** | 10 | 29 | 15 | 9 | 23 | 14 |
| **3** | 3 | 22 | 25 | 11 | 10 | 29 |
| **4** | 0 | 17 | 12 | 12 | 36 | 24 |
| **5** | 2 | 18 | 18 | 11 | 26 | 26 |
| **6** | 0 | 26 | 25 | 16 | 6 | 27 |
| **7** | 2 | 20 | 20 | 11 | 18 | 30 |
| **8** | 0 | 15 | 9 | 14 | 52 | 10 |
| **9** | 1 | 20 | 26 | 10 | 28 | 15 |
| **10** | 0 | 5 | 26 | 8 | 0 | 61 |

<u>Duration redesign:</u> Figure 2.4 shows scheduling performance measures per several potential durations for new appointment types. Performance measures are percentage of reduction in average patient in-clinic time, ratio of patients seen in new practice compared to current practice, and ratio of provider time spent with patient compared to their capacity. For example, if we consider the value of 65$^{th}$ percentile of all durations within a cluster to the new appointment duration for that cluster, we would see 11% decrease in patient time in the clinic, 35% increase in number of patients accommodated, and 30% decrease in provider time with patients. The results in figure 2.4 include iterations from 65$^{th}$ to 80$^{th}$ percentiles. We did not see changes outside this range, so they are not included in the figure. We chose 75$^{th}$ percentile duration of each cluster as the new proposed appointment durations, because it can reduce average patient in-clinic time by 10%, increase overall number of patients to see by 9%, insignificantly change provider time spent with patient.

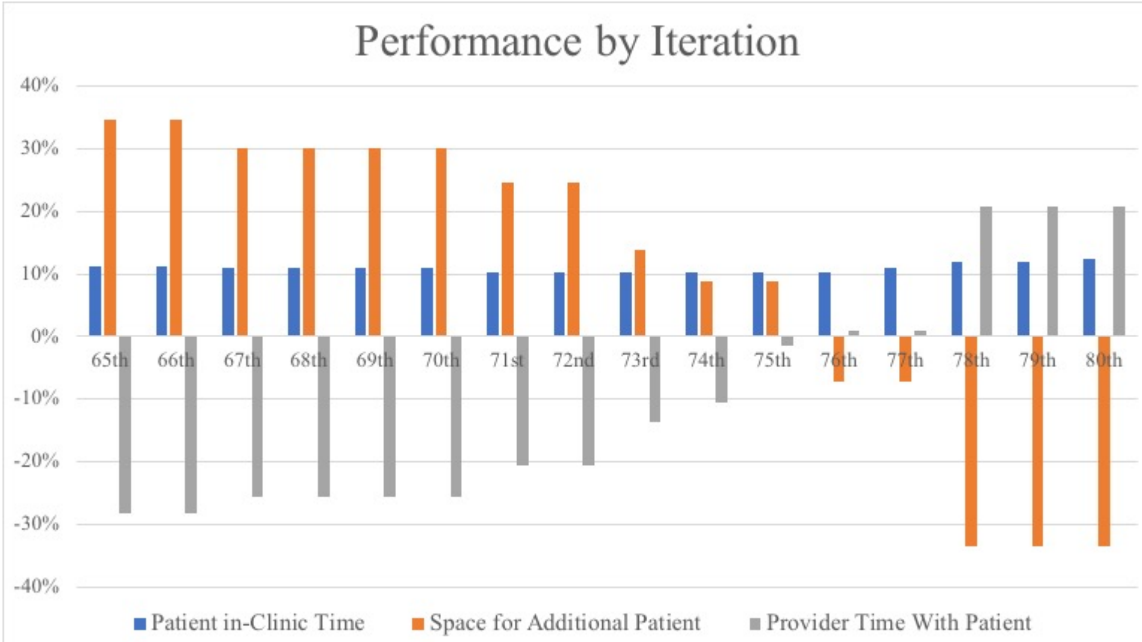



**Figure 2.4.** Performance measures by iterations. $N^{th}$ iteration means assigning the $N^{th}$ percentile of appointment durations within a cluster to the cluster.

Table 2.5 shows comparisons of the current scheduling system and the proposed scheduling system. Average appointment duration in current scheduling system is the average of current durations by cluster. They are higher for clusters 1, 2, 4, 7, 8, 9, and 10, and lower for clusters 3, 5, and 6 compared to the average of current durations (p-value<0.0001).

**Table 2.5.** Comparison of current versus proposed appointment scheduling templates.

| **Appointment cluster** | **Average appointment duration in *current* scheduling system (minutes)** | **Average appointment duration in *proposed* scheduling system (minutes)** | **Average time spent in clinic in *current* scheduling system (minutes)** | **Average predicted time spent in clinic in *proposed* scheduling system (minutes)** | **Number of additional/less patients clinic can see in the *proposed* scheduling system (patient/year)** |
|---|---|---|---|---|---|
| **1** | 20.0 | 30 | 50.8 | 67.1 | -4 |
| **2** | 17.6 | 20 | 63.0 | 61.5 | -21 |
| **3** | 17.5 | 15 | 66.0 | 57.7 | 649 |
| **4** | 17.6 | 20 | 64.3 | 61.0 | -58 |
| **5** | 18.1 | 15 | 68.0 | 57.4 | 368 |
| **6** | 16.8 | 15 | 64.7 | 57.6 | 115 |
| **7** | 17.8 | 20 | 66.9 | 60.9 | -87 |
| **8** | 19.9 | 30 | 70.3 | 68.3 | -61 |
| **9** | 18.9 | 20 | 67.5 | 61.7 | -138 |
| **10** | 17.1 | 20 | 65.4 | 58.4 | -47 |
| **Total** | | | | | **716 (10%)** |

The time patients spent in clinic per visits is calculated based on the EHR patient's arrival and departure times. Predicted time spent in clinic was calculated using a linear regression model trained using the current scheduling. Table 2.5 shows the proposed scheduling system could open space to see overall 716 additional patients per year, which is about 10 percent more patients. Figure 2.5 shows distributions of patient time spent in clinic per visit. Our results suggest that the new scheduling systems and appointment duration could decrease patient in-clinic time by 3.7 minutes on average (p-value<0.0001).

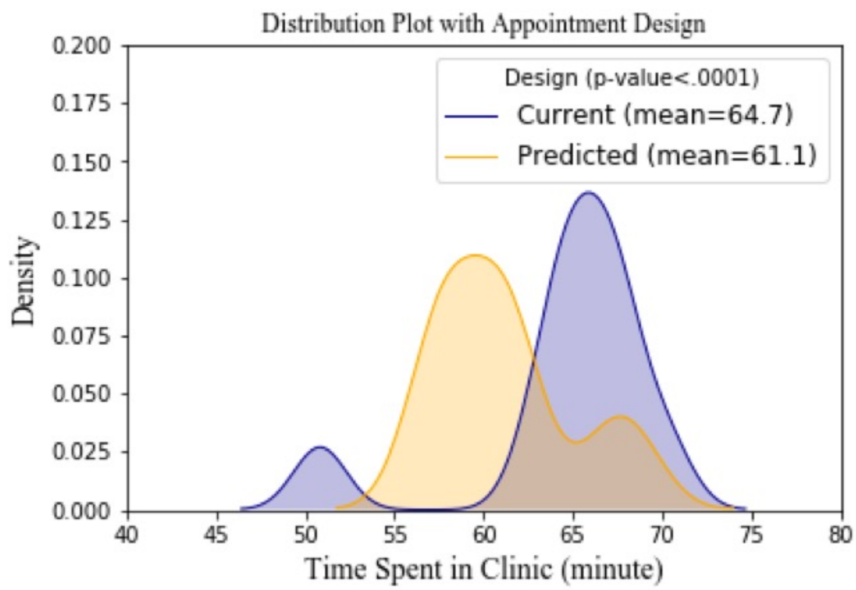



**Figure 2.5.** Comparison of distributions of patient time spent in clinic between the current and proposed appointment types and durations.

### 2.4.3 Discussion

We studied possibility of using patients' reasons for seeking health along with patient, visit, and provider characteristics to design new appointment types and durations for community health centers. Our study has three major findings. First, word embedding, that is an NLP approach, can be used to extract information from schedulers notes for improving scheduling systems. Second, unsupervised machine learning approach can be applied to simplify appointment scheduling in community health centers. Third, patient centered appointment scheduling can be achieved by simplifying and redesigning appointment types and durations that could improve access measures, such as increasing availability of time and patient satisfaction.

In this work, we expanded utilization of word embedding trained models by applying it on scheduler notes in primary care settings. We found word embedding trained on EHR scheduler notes, MedTagger, and GloVe can capture semantics of medical terms, and the results are aligned with human judgment (shown in table 2.1).

The Institute of Medicine defines health care quality as "the degree to which health care services for individuals and populations increase the likelihood of desired health outcomes and are consistent with current professional knowledge[53]." One of the domains of health care quality is efficiency. Our study found simplification of scheduling based on patient, provider, and clinic characteristics could improve efficiency. In this work, we designed a methodology to simplify appointment types and times because complex schedule templates could lead to mismatching patient problem to incorrect answer[54]. Simplifying appointment types and times is one of the requirements of transitioning from traditional access models to advance access models[54]. The approaches in this study could simplify appointment scheduling to match daily supply and demand. We found simplifying scheduling templates could improve overall clinic performance, such as improving provider productivity, decreasing patient in-clinic waiting time, and improving clinic accommodations. Our methodology is significant because improved overall performance could be achieved without additional supply, more resources, or extended hours.

Patient centeredness is another domain of health care quality that is achieved by meeting patient needs and preferences. In this study, we designed an infrastructure for patient specific resource allocation. Patients with different reasons for seeking health, age, and gender have different resource requirement[55]. Our proposed appointment scheduling template clusters patients into classes based on reasons for seeking health.

Timeliness and patient satisfaction are other aspects of a good health care delivery system[56]. Our study found that simplified scheduling can reduce in-clinic time that could consequently lead to improved timeliness and satisfaction.

Our study had some limitations. First, our patient encounter data lacked clinical information such as diagnoses, procedures, lab results, and clinicians' notes. In any future work these features can also be used to design stronger patient specific resource distribution. Another limitation of this study was that our dataset did not include information about in-clinic patient journey, such as step by step activities and timestamps from the moment that patient checks in to departure of patients, and information about daily number of staffs, medical assistants, and nurses. Those factors could be predictors of in-clinic waiting time.

Future work in this area might focus on three objectives. First, refinements of expansion of abbreviations by utilizing more comprehensive dictionaries that would be able to deal with less commonly used abbreviations. Second, other unsupervised clustering methods such as deep learning or reinforcement learning might be able to extract more relations between notes; that would lead to more precise clusters. Third, researchers might use the findings of this study to either implement the algorithms in current EHR interface or design a new interface for a decision support system. Future research in this area could be evaluating the effectiveness of the proposed algorithms in a real world clinical practice.

## 2.5 Potential Medical Applications

In this chapter, we found that health care needs of patients vary by factors such as age, ethnicity, gender, and marital status in underserved populations. Clinics might use the findings of this study to prioritize the most feasible and impactful patient-centered approaches to reduce barriers of accommodation. One of the steps of moving from traditional appointment scheduling to optimized open access scheduling is to simplify appointment types and times. CHCs or other primary care settings can leverage methodologies and findings of this chapter to achieve optimized open access scheduling. Previous studies did not discuss the most appropriate ways to simplify appointment types. They offered appointment types such as new, established, acute, and postoperative as decreased number of appointments[34]. A key advantage of methodology presented in this chapter is that the simplification of appointment template is not only helping clinics achieve advanced open access scheduling system, but they are also patient centered and patient specific. Proposed appointment scheduling templates is designed based on reasons patients seeking health.

Another potential medical application of this study is to utilize the unsupervised machine learning approach presented in this chapter to design automated appointment scheduling tools for healthcare settings. These tools can be in form of online appointment scheduling or automated phone call scheduling. These potential tools ask patients why they need appointment and the system find the most appropriate appointment type and time for the patient. Methodologies presented in this chapter can be applied on both scheduler and clinician notes to find care needs and gaps for patients, and design interventions to close the gaps.

## 2.6 Conclusion

A key to a more efficient scheduling systems is to ensure appointments are designed to meet patient's needs, and to design and simplify appointment scheduling less prone to error. In this chapter, we presented approaches for redesigning appointment scheduling based on patient characteristics, complaints, and diagnoses in underserved populations. We used EHR data to investigate the relationship between diagnoses, patient characteristics, and reasons for visit to help providers redesign healthcare systems that can meet the needs of underserved patients. We found care needs in underserved population by applying consensus clustering on patient records to reveal the common diagnoses or disease trend in the target population. We found the socio-economic status of patients is associated with diagnoses in underserved populations. We applied word embedding and unsupervised machine learning methods to design most effective appointments in community health centers. We found simplifying appointment types and times can help healthcare systems achieve improved access and patient satisfaction without adding additional resources.

# CHAPTER THREE: APPOINTMENT NO-SHOW MODELING

In this chapter, using predictive modeling techniques, we developed and compared appointment no-show prediction models to better understand appointment adherence in underserved populations. We collected Electronic Health Records (EHR) data and appointment data including patient, provider, and clinical visit characteristics over a three-year period. All patient data came from an urban system of community health centers (CHCs) with 10 facilities. We sought to identify critical variables through Logistic Regression, Artificial Neural Network, and Naïve Bayes Classifier models to predict missed appointments. We used 10-fold cross-validation to assess the models' ability to identify patients missing their appointments. Following data preprocessing and cleaning, the final dataset included 73,811 unique appointments with 12,392 missed appointments. Predictors of missed appointments versus attended appointments included lead-time (time between scheduling and the appointment), patient prior missed appointments, cell phone ownership, tobacco use, and the number of days since last appointment. Models had a relatively high area under the curve for all three models (e.g., 0.86 for Naïve Bayes Classifier). Patient appointment adherence varies across clinics within a healthcare system. Data analytics results demonstrate the value of existing clinical and operational data to address important operational and management issues. EHR data including patient and scheduling information predicted the missed appointments of underserved populations in urban CHCs. Our application of predictive modeling techniques helped prioritize the design and implementation of interventions that may improve efficiency in community health centers for more timely access to care. CHCs would benefit from investing in the technical resources needed to make these data readily available as a means to inform important operational and policy questions.

## 3.1 No-show Phenomenon

Given financial challenges of delivering quality health care in the United States, finding ways to improve performance is critical in the plight to provide greater access to care. Optimizing scheduling systems has been identified as one system level approach to address access needs. For example, reducing the number of missed appointments is crucial as when appointment slots go unused, and it effectively reduces access to others in need of an appointment[57]. In addition to underutilizing providers' time, missed appointments impact waits and delays for others, increase health care costs, and increase possibility for adverse health outcomes[58,59]. Research has shown that lowering missed appointment rates can improve clinical efficiency and utilization, reduce waste, improve provider satisfaction, and lead to better health outcomes for patients[60,61]. Missed appointment rates range from 10% to 50% across healthcare settings in the world with an average rate of 27% in North America[59]. Patients with higher missed appointment rates are significantly more likely to have incomplete preventive cancer screening, worse chronic disease control and increased rates of acute care utilization[62]. In previous studies, missed appointments have been due to logistical issues, lack of understanding of the scheduling system, patients not feeling respected by healthcare providers or the health system, affordability, timeliness, patients forgetting appointment, and patient severity of illness[59,63].

Researchers studied impact of no-show on health outcomes. They found patients with higher non-adherence were significantly more likely to have incomplete preventive cancer screening, worse chronic disease control, and increased rates of acute care utilization[62]. In another study, missed clinic visits was found as a significant factor contributing to in viral load outcomes among black and injection drug use HIV patients[64]. In underserved populations, poverty and health outcomes are highly linked[65]. According to World Health Organization (WHO), health determinants include income and social status, education, physical environment, health services, genetics, and gender[6]. Approximately 80% of U.S. patients in the underserved populations have income less than 100% of the federal poverty level[66]. Medicaid covers most patients in CHCs, once eligibility is established, and an application is filed. The racial mix of patients in Indiana includes about 60% white, 30% black, 10% other races, and approximately 20% Hispanic or Latino ethnicity. Primary healthcare services in Indiana are provided at rates below the national average. Indiana has a lower proportion of citizens with a bachelor's degree (24 versus 30% in the U.S), yet higher rates of disability, the absence of health insurance, and poverty, compared to the rest of the country[67].

## 3.2 Previous Studies on No-show Modeling

Studies on no-show modeling have been conducted in different health care setting providing primary or specialty care. Outpatient clinics, Veterans Health Administration (VHA), cancer care clinics, colonoscopy clinic, pediatrics clinic, endocrinology clinic, otolaryngology clinic, HIV clinic, mental care clinics and other specialty clinics are the healthcare settings in which appointment non-adherence is important and has been studied[59]. One study aimed to identify predictors of a missed appointment and develop a model to predict no-show in an underserved outpatient internal medicine clinic from January 2008 to June 2011[68].

Different datasets, variables, and data volumes have been proposed for understanding of the complexity appointment adherence in different healthcare settings. Medium scale studies included no-show studies including 6,700 patients and 7,988 patients and variables that most health care clinics collect included appointment characteristics, demographics, provider characteristics, and insurance and billing information[69,70]. Other researchers studied a medium scale no-show modeling including 7,291 patients, but only aiming at a single component (time) to improve predictability[38]. The study of no-show modeling in undeserved setting collected 2.5 years of patient data excluding children patients (< 18 years of age) and contained about 11,546 patients[68].

Most studies developed regression models to predict appointment non-adherence[70,71]. Researchers used a simple heuristic and found that giving same-day appointments to likely shows and future-day appointments to likely no-shows can lead to a substantial increase in performance[69]. One study performed ANOVA to determine patterns of no-show variation by patient age and gender in the United States Veterans Health Administration (VHA)[72]. In a different study, they aimed to determine the variables that contribute to repeated patient no-show. The outcome variable was 3 or more no-shows versus less than 3 no-shows[73]. Study of no-show modeling in CHCs aimed to identify predictors of a missed appointment and develop a model to predict no-show in an underserved outpatient internal medicine clinic from January 2008 to June 2011[68]. They found that percent previously missed appointments (no-show or cancellation within 24 hours), wait time from booking to appointment, season, day of the week, provider type, and patient age, sex, and language proficiency were all associated with a missed appointment.

### 3.3 Predictive Modeling

In order to understand the complexity of appointment adherence in different healthcare settings, different datasets, variables, and data volumes have been studied. Medium scale studies (ranging from 6,000 to 8,000 patients) focused on a few patient characteristics or a single (for example time) component[38,69,71]. For example, a large-scale no-show modeling of a VA outpatient clinic included 555,183 patients, which scheduled 25,050,479 appointments; however, the study only considered a few variables such as the patient gender, the date of the appointment, and new vs. established patients[72]. Most studies developed regression models to predict appointment non-adherence[70,71]. Most similar to the present study, one study identified predictors of missed clinic appointments among an underserved population[68]. These results revealed predictors for a missed appointment included percent previously missed appointments (no-show or cancellation within 24 hours), wait time from scheduling to appointment, season, day of the week, provider type, and patient age, sex, and language proficiency. In other studies of predictive modeling in healthcare arena using EHR data, other predictive modeling techniques such as Naïve Bayes Classifier[74] and Neural Network[75] were used to predict hospital readmissions. These methodologies can be used in our no-show modeling study as they have similarly predicted a binary outcome using EHR data. In this chapter, we applied and built on these techniques to predict appointment no-show in CHCs. Although number of studies has been done on no-show modeling and predicting appointment no-shows, not much research work has focused on fully understanding no-show behavior of underserved populations. One study on no-show modeling for a clinic in underserved area was done using a small dataset from a single clinic serving only adults with a very high no-show rate of 45%.

Logistic regression has been the most frequent used methodology for no-show modeling. But, previous studies have not proposed a proper feature selection for no-show modeling. New variables, more specific patient related variables, or interaction between features would be helpful for developing more accurate models and useful for improving understanding of appointment non-adherence.

Here, we tested missed appointment prediction models by analyzing EHR and scheduling data. We aimed to fully utilize predictive modeling to improve understanding of the complexity of appointment adherence in underserved populations. Information about patients, providers, appointments and time were used to predict patients' adherence to appointments. The main contributions of this chapter include 1) build on previous no-show modeling in community health centers by expanding the focus on various outpatient specialties and underserved population specific predictors; 2) compare different predictive modeling methodologies, that are Logistic Regression, Naïve Bayes Classifier and Artificial Neural Networks, and 3) investigate the impact of clinic characteristics on predictors of the no-show.

## 3.4 Materials and Methods

### 3.4.1 Participants

Data for this project was collected from a large urban multisite community health center, involving 10 locations in Indianapolis, most of which are considered federally qualified health centers (FQHC). This CHC has provided care for more than 100,000 patients during 2014 to 2016. Healthcare services provided by this CHC include but not limited to primary care, pediatrics, family practice, internal medicine, OB/GYN, dental care, vision care, behavioral health services and preventive care. The goal of this no-show modeling was to focus on primary care, so data on dental and vision care visits was not considered. All study methods were approved by our institutional review board.

### 3.4.2 Data collection and sample size

We extracted and de-identified semi-structured data from over 17 tables in the CHC's database from 2010 to 2016 to address the study aim. EHR data including clinic (i.e., operational and financial data) and patient (i.e., patient demographics and clinical characteristics) information were included and linked at the patient level. The data was stored in a secure Microsoft SQL Server with limited access. For this study, we created a dataset of patients' encounters from January 1, 2014 to April 30, 2016. The dataset included 599,636 appointments by 76,453 unique patients (Table 3.1).

### 3.4.3 Data preprocessing

Appointment compliance field was the dependent variable in this analysis, which included the categories of checkout (i.e., complete) appointment, no-show, cancelled, rescheduled, and others. A no-show appointment is defined as a patient who did not keep the prescheduled appointment and did not cancel the appointment at least 24 hours ahead of the appointment time. We performed the following data filtering steps:

- Filtering appointment categories: To create the binary outcome variable in this study, we only included no-show and checkout appointments in the final analysis, and observations having other appointment compliance such as rescheduled, cancelled, etc. were censored from the dataset.
- Ensuring appointment independences: To ensure observations are independent from each other, we only included the last appointment of each patient in the final analysis.

- Handling missing information: Some unstructured free text fields such as schedulers' notes were used to complete any missing values in fields, such as appointment type, patient age, or gender. All other observations with missing information were removed from the dataset.

The final dataset included 73,811 observations of unique individuals and whether they showed for their last appointment during the study period.

### 3.4.4 Variables Preparation

Data fields included visit characteristics (facility/clinic type, date of visit, date contacted the clinic for scheduling the visit, time of visit, visit duration and visit type), patient characteristics (patient pseudo-ID, age, race, ethnicity, gender, marital status, cell phone ownership, email availability, whether using patient portal, employment status, tobacco use, income, needing translator and primary insurance), provider characteristics (whether seeing the patients' primary care practitioner (PCP) or not, specialty and medical license) and appointment compliance ("no-show" or "check out").

In addition to the existing variables in the EHR, we created the following variables to consider in our no-show modeling:

a. Lead-time: which is the time difference (in days) between the date of visit and date the patient had contacted the clinic to arrange an appointment.
b. Prior no-show rate: which is the number of no-shows for a given patient prior to the last appointment, divided by the patient's total number of appointments prior to the last appointment. We used this to test the effect of patient no-show behavior on appointment adherence.

c. Days since the last appointment: which is the difference between the date of the last visit and the date of appointment before the last visit.

**Table 3.1.** Distribution of patient characteristics versus appointment adherence. Note: T-TEST for continuous variables and Chi-Square for categorical.

| Variables | | Appointment Adherence | | |
|---|---|---|---|---|
| | | Attended N=61,419 | Missed N=12,392 | p-value |
| **Patient Characteristics** | | | | |
| | | Categorical (Percentages) | | |
| **New Patient** | Yes | 2.1 | 2.4 | 0.0455 |
| **Translator Needed** | Yes | 15.2 | 8 | <.0001 |
| **Ethnicity** | Hispanic or Latino | 19.6 | 11.9 | <.0001 |
| | Not Hispanic or Latino | 75 | 80.2 | |
| | Unspecified | 5.4 | 7.9 | |
| **Race** | American Indian or Alaska Native | 0.1 | 0.1 | <.0001 |
| | Asian | 4.2 | 2 | |
| | Black | 30.3 | 37.7 | |
| | Multiple Races | 3.9 | 3.7 | |
| | Native Hawaiian and Other Pacific Islander | 1.1 | 0.7 | |
| | White | 60.4 | 55.7 | |
| **Gender** | Female | 61.4 | 64.8 | <.0001 |
| **Marital Status** | Divorced | 3.3 | 3.1 | <.0001 |
| | Legally Separated | 1.3 | 1.7 | |
| | Married | 12.8 | 9.5 | |
| | Partner | 0.4 | 0.3 | |
| | Single | 80.8 | 83.4 | |
| | Widowed | 1.2 | 0.8 | |
| **Cell phone Ownership** | No | 18.2 | 26.4 | <.0001 |
| **Email Availability** | No | 70.6 | 74.5 | <.0001 |
| **Using patient portal** | No | 78.2 | 83.5 | <.0001 |
| **Employment Status** | Employed full-time | 13 | 10.8 | <.0001 |
| | Employed part-time | 5.1 | 5.5 | |
| | Not employed | 79.6 | 82.4 | |
| | Retired | 1.5 | 0.4 | |
| | Self-employed | 0.5 | 0.3 | |
| **Insurance** | Commercial | 14.8 | 8.4 | <.0001 |
| | Marketplace | 0.6 | 0.3 | |
| | Medicaid | 66.8 | 69 | |
| | Medicare | 5.6 | 3.6 | |
| | Self-Pay | 12.2 | 18.7 | |
| **Tobacco Use** | Current every day smoker | 22.8 | 35.5 | <.0001 |
| | Current some day smoker | 2.8 | 3.4 | |
| | Former Smoker | 13 | 12 | |
| | Never Smoker | 61.3 | 49.1 | |
| | | Continuous (Mean (SD)) | | |
| **Age** | Mean (SD) | 21.1 (19.4) | 21.4 (16.9) | 0.1393 |
| **Annual Income** | Mean (SD) | $2748(8421) | $2046(7109) | <.0001 |
| **Prior No-show Rate** | Mean (SD) | 0.11 (0.2) | 0.2 (0.3) | <.0001 |

### 3.4.5 Statistical Analyses

We hypothesized that patient and provider characteristics and visit features were all predictors of appointment no-show in CHCs. We tested variables individually for relationships with the appointment adherence using a Chi-Square test for categorical variables and t-test for continuous variables. We chose variables with a p-value less than 0.2 to enter into the model development step. Tables 3.1, 3.2, and 3.3 list variables that were included in the modeling. The dataset included 73,811 observations, 83% arrived and 17% no-show.

### 3.4.6 Prediction Model Development

We randomly split the dataset into two samples: 70% for the training (or derivation) set, and 30% for the test (or validation) set. This train and test set selection was repeated 10 times to overcome selection bias. We used the training subsets to develop the no-show prediction model using three methodologies:

a) Logistic Regression: We used logistic regression in SAS 9.4 to develop the prediction model with a stepwise selection and significance level of $\alpha=0.01$. All the variables, shown in tables 3.1, 3.2, and 3.3 and their interactions, were included in the model development.

b) Artificial Neural Network: The large number of features and observations in this study led us to utilize more complex machine learning algorithms such as Artificial Neural Networks. Neural Networks consist of multiple linear regression models are advantageous when there is a large number of features (variables) with complex relations among them[76]. Categorical variables were transformed to numeric variables. For example, if a patient is a "New Patient", the numeric variable of New Patients would be created with a value of 1. Continuous, binary, and numeric variables were used as inputs for a Neural Network and one binary variable (No-show = 1 or 0) was used as output. Matlab software was used to develop the Neural Network in this project having three layers of the input layer, hidden layer including 25 nodes and output layer. The training data subsets were used to train the network by minimizing the mean-square error (MSE) between the desired output and the actual output of the network. The value of the output node determined the classification using a range (between 0 and 1) of cutoff thresholds. Here, we used absolute value of weights for input layer nodes to identify and rank the most important variables contributing to no-show prediction.

c) Naïve Bayes Classifier: The majority of predictors in our datasets were categorical; hence, we applied a Naïve Bayes Classifier that is appropriate to categorical data[77]. This classifier computes a conditional probability of each category in each variable given the outcome. Then, Bayes rules are applied to calculate the probability of the outcome given different categories of variables in the data. We applied the Naïve Bayes classifier algorithm implemented in "scikit-learn" in Python over the randomly selected train and test datasets. The smoothing value of 0.1 provided the best performance for naïve Bayes classifier.

### 3.4.7 Model Validation

Models were assessed by calculating the area under the curve for the receiver operating characteristic (AUC-ROC) curve. Test dataset was used to validate models' ability to discriminate between patients who no-showed versus those who attended. 10-fold cross-validation was used to validate the three models, and average AUCs, sensitivities to predicting no-show and overall model accuracy were the key indicators of model validation.

## 3.5 Results

Statistical analyses: The final dataset included 73,811 observations with 12,392 missed appointments. Comparative analyses of patient characteristics revealed that black, non-Hispanic or non-Latino, female, single, not employed, Medicaid, self-pay, or smoker patients had a higher chance of missed appointments (p<0.0001; see Table 3.1). The average annual income is lower, and the average prior missed appointment rate is higher in patients who no-showed in their last appointment (p<0.0001). Patients without a cell phone, email or patient portal had a higher chance of a missed appointment (p<0.0001).

**Table 3.2.** Distribution of provider characteristics versus appointment adherence. Note: T-TEST for continuous variables and Chi-Square for categorical.

| | | Appointment Adherence | | |
|---|---|---|---|---|
| **Variables** | | Attended<br>N=61,419 | Missed<br>N=12,392 | p-value |
| | | Categorical (Percentages) | | |
| **Provider Specialty** | Behavioral Health | 1.8 | 4.3 | <.0001 |
| | Certified Nurse Midwife | 9.5 | 12.7 | |
| | Family Medicine | 17.1 | 14.7 | |
| | Internal Medicine | 11.5 | 11.7 | |
| | Nurse Practitioner | 9.9 | 7.3 | |
| | OB - Gynecology | 4.3 | 5.9 | |
| | Pediatrics | 33.6 | 30.3 | |
| | Podiatry | 0.7 | 1.4 | |
| **Patient's PCP?** | No | 83.2 | 86.6 | <.0001 |

The comparative analysis of the provider characteristics showed that patients scheduled with behavioral health or OB-GYN providers or not scheduled with their primary care providers have higher missed appointment rates compared to other appointment types (p<0.0001), as demonstrated in Table 3.2.

The appointment duration, the time between appointment days, and the day appointment requested, the time (daytime, weekday or season) of an appointment, and the type of an appointment are statistically significantly different between checkout and missed appointment patients (p<0.0001), as shown in Table 3.3. Table 3.4 shows characteristics of 10 facilities within this CHC system. Clinics are different in terms of missed appointment rates and distributions of patient type, visit type, and provider type.

**Table 3.3.** Distribution of visit characteristics versus appointment adherence. Note: T-TEST for continuous variables and Chi-Square for categorical.

| Variables | | Appointment Adherence | | p-value |
|---|---|---|---|---|
| | | Attended N=61,419 | Missed N=12,392 | |
| **Appointment Duration** | 10 minutes | 0.8 | 0.1 | <.0001 |
| | 15 minutes | 68.3 | 60.3 | |
| | 20 minutes | 14.3 | 14.7 | |
| | 30 minutes | 15.6 | 22.1 | |
| | 45 minutes | 0.5 | 1.7 | |
| | 60 minutes | 0.5 | 1.1 | |
| **Lead Time** | Same day | 31.4 | 8.4 | <.0001 |
| | Next day | 9 | 7.1 | |
| | Within two weeks | 31.6 | 35.4 | |
| | Between 2 weeks and 1 month | 13 | 20.7 | |
| | More than 1 month | 15 | 28.5 | |
| **Days Since Last Appointment** | Within a week | 1.4 | 1.9 | <.0001 |
| | Between 1 and 2 weeks | 1.1 | 1.8 | |
| | Between 2 weeks and 1 month | 2.3 | 4.1 | |
| | Between 1 and 3 months | 5.6 | 9.3 | |
| | Between 3 and 6 months | 6.7 | 10.2 | |
| | Between 6 months and a year | 14.5 | 16.2 | |
| | More than a year | 53.7 | 39.6 | |
| | No prior appointment since 2014 | 14.8 | 16.9 | |
| **Appointment time** | AM | 43.8 | 44.5 | 0.1294 |
| **Season** | Fall | 18.1 | 19.8 | <.0001 |
| | Spring | 29.9 | 28.9 | |
| | Summer | 15.1 | 18.3 | |
| | Winter | 36.9 | 33 | |
| **Weekday** | Monday | 22.3 | 23.4 | <.0001 |
| | Tuesday | 21.9 | 22.1 | |
| | Wednesday | 20.1 | 19.1 | |
| | Thursday | 18.8 | 19.2 | |
| | Friday | 15.8 | 15.3 | |
| | Saturday | 1.1 | 1 | |
| **Visit Type** | Acute care | 27.7 | 12.1 | <.0001 |
| | Adult Routine/Follow up | 17 | 24.4 | |
| | Behavioral Health | 2 | 4.8 | |
| | Podiatry | 0.7 | 1.4 | |
| | Pediatric | 37.6 | 37.5 | |
| | Pregnant | 4.5 | 6.5 | |
| | Women | 10.5 | 13.4 | |

Predictive modeling: As shown in table 3.4, clinics had different population sizes, characteristics, and no-show rates. Therefore, we developed a separate logistic regression model for each clinic. Table S.1 shows the results from regression model development. These separate models corresponding to individual clinics yielded different predictors for missed appointments. Notably, lead-time, prior missed appointment rate, age, insurance type, tobacco use, days since the last appointment and cell phone ownership were consistent significant factors across clinics.

**Table 3.4.** Clinic characteristics. Note: T-TEST for continuous variables and Chi-Square for categorical.

| Facility | Total Patients | No-show Frequency | No-show Percent | Clinic Characteristics | Percentage/Mean Among all clinics |
|---|---|---|---|---|---|
| **Clinic1** | 10,633 | 2,248 | 21% | • Large number (23%) of patients needing translator<br>• Large number (20%) of Asian patients<br>• Highest mean lead-time (28.6 days) | • 14% - p-value<.0001<br>• 4% - p-value<.0001<br>• 17days-p-value<.0001 |
| **Clinic2** | 3,680 | 660 | 18% | • Higher percentage of new patients (10.3%)<br>• Dominantly pregnant and woman patients (98%)<br>• Dominantly CNM and OB-GYN providers (95%)<br>• Dominantly female patients (98%)<br>• Dominantly adult patients (95%)<br>• Patients with lower prior no-show rates (0.08) | • 2.1% - p-value<.0001<br>• 16% - p-value<.0001<br>• 15% - p-value<.0001<br>• 62% - p-value<.0001<br>• 43% - p-value<.0001<br>• 0.12 - p-value<.0001 |
| **Clinic3** | 3,206 | 392 | 12% | • Mostly scheduled with patients' PCPs (56%)<br>• Patients with lower prior no-show rates (0.08) | • 16% - p-value<.0001<br>• 0.12 - p-value<.0001 |
| **Clinic4** | 6,731 | 803 | 12% | • Majority Black (77%)<br>• Mostly same-day appointments (67%)<br>• Higher number of acute care appointments (46%) | • 32% - p-value<.0001<br>• 27% - p-value<.0001<br>• 25% - p-value<.0001 |
| **Clinic5** | 2,216 | 480 | 22% | • Highest no-show rate | • 17% - p-value<.0001 |
| **Clinic6** | 7,870 | 1,543 | 20% | • Mostly 20 minute appointments (79%)<br>• Dominantly children (97%)<br>• Majority Black (63%)<br>• Dominantly not employed (98%) | • 14% - p-value<.0001<br>• 55% - p-value<.0001<br>• 32% - p-value<.0001<br>• 80% - p-value<.0001 |
| **Clinic7** | 10,703 | 1,916 | 18% | • Large number (23%) of patients needing translator<br>• Higher number of Hispanic or Latino (34%) | • 14% - p-value<.0001<br>• 18% - p-value<.0001 |
| **Clinic8** | 12,016 | 1,659 | 14% | • Large number (22%) of patients needing translator<br>• Higher number of Hispanic or Latino (31%)<br>• Highest income level ($4553/year) | • 14% - p-value<.0001<br>• 18% - p-value<.0001<br>• $2665 - p-value<.0001 |
| **Clinic9** | 11,521 | 1,942 | 17% | • Dominantly White (85%) | • 60% - p-value<.0001 |
| **Clinic10** | 5,235 | 749 | 14% | • Patients with lower prior no-show rates (0.08) | • 0.12 - p-value<.0001 |

<u>Patient characteristics:</u> Table 3.4 demonstrates that Clinic 2 patients had lower prior missed appointment rates compared to other clinics. In all clinics except clinic 6, patients between 18 to 64 years old were 1.59 (99%CI [1.48 1.64]) and 3.65 (99%CI [2.92 4.57]) times more likely to no-show their next appointments compared to patients between 0 to 17 years old and 65 and older patients, respectively. Notably, clinic 6 is a pediatric clinic and patients are dominantly between 0 to 17 years old. Patients who needed a translator in their appointments, particularly in clinic 7 (with a high proportion of Hispanic or Latinos), were 0.5 times less likely to no-show in their next appointments (99%CI [0.4, 0.5]). In two clinics, the interaction between age and gender also influenced no-shows.

Insurance status was another significant predictor of missed appointments, such that insured patients were less likely to keep their appointments. In most clinics, patients insured by commercial, marketplace, Medicaid and Medicare plans were 0.37 (99%CI [0.33 0.41]), 0.33 (99%CI [0.21 0.51]), 0.67 (99%CI [0.63 0.72]) and 0.42 (99%CI [0.37 0.49]) times as likely to miss appointments, compared to their uninsured counterparts. Smoking daily increased the likelihood of missed appointments by 95%, compared to patients who never smoked (OR=1.95, 99%CI [1.8 2.12]). Patients using patient portal (web-enabled) were less likely to no-show in their appointments (OR=0.7, 99%CI [0.7, 0.8]). In clinic 5, patients without an email address recorded in the EHR system are 1.22 times more likely to no-show (99%CI [1.15 1.23]).

Scheduling characteristics: Lead-time was the most consistent significant factor across all the clinics. Longer lead-time provides greater opportunity for a missed appointment ($p<0.0001$). Appointments made more than one month in advance are 7.12 (99%CI [6.46 7.48]), 2.43 (99%CI [2.18 2.7]), 1.70 (99%CI [1.59 1.89]), and 1.2 (99%CI [1.11 1.29]) times more likely to become a no-show, compared to appointments made on same day, one day, two weeks and between two weeks and one month in advance respectively. Next day appointments were 2.93 times more likely to become a missed appointment than same day appointments (99%CI [2.59 3.32]). Patients with a history of missed appointments were 4.89 times more likely to miss their next appointments (99%CI [4.4 5.83]), in all clinics except clinic 2. Patients who had an appointment between 1 and 2 weeks prior to their last appointment were more likely to miss that last appointment compared to patients who had a prior appointment in the last 6 to 12 months (OR=1.5, 99%CI [1.2 1.8]), more than 12 months (OR=2.2, 99%CI [1. 8 2.7]), or patients who had no prior appointments (OR=1.4, 99%CI [1.1 1.7]). Patients without a cell phone number available in the records were 1.6 times more likely to no-show (99%CI [1.52 1.71]).

Clinic visit characteristics: In one-half of the clinics, type of visit predicted appointment adherence. Table 3.3 shows that acute visits had lower missed appointment rates than all other visit types, while behavioral health visits had the highest missed appointment rates. Seasonality of the appointments predicted missed appointments such that appointments occurring during spring or summer had higher missed appointment rates than winter appointments. Notably, patients scheduled with their own PCP were less likely to miss the appointment than the ones scheduled with other providers (OR=0.8, 99%CI [0.7, 0.8]). Appointment duration was also a significant factor (particularly in clinics 3 and 5). Longer durations such as 1 hour or 45 minutes were more likely to be no-show than shorter durations such as 15 or 20.

Machine learning approaches: The ranking of variables contributing to prediction of no-show in the Neural Network are shown in figure 3.1. The ranking is based on the weights of nodes in the input layer of Neural Network.

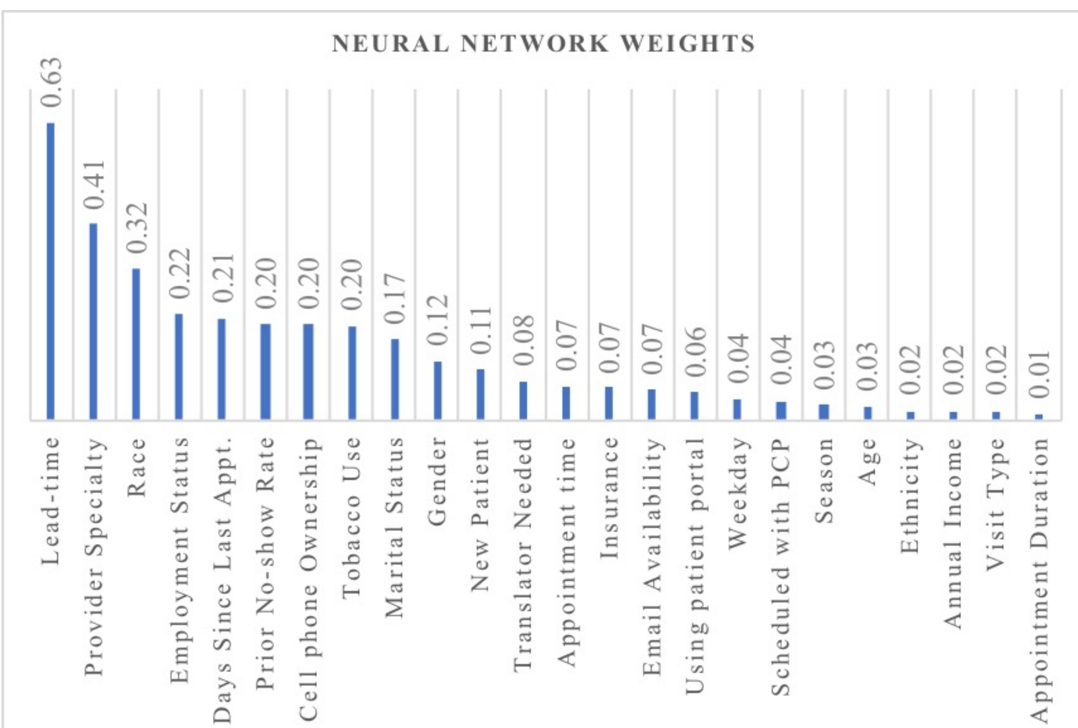



**Figure 3.1.** Neural Network weights showing the ranks of predictors of no-show.

The top ten predictors of the no-show in our Neural Network analyses included: lead-time, provider specialty, race, employment status, days since last appointment, prior no-show rate, cell phone ownership, tobacco use, marital status and gender.

There were multiple variables contributing to no-show (figure 3.2) using the Naïve Bayes classifier. Prior no-show rate, age group, visit type, lead-time, days since last appointment, duration, insurance, cell phone ownership, tobacco use, and ethnicity are the top ten factors predicting next appointment no-show. Those variables important in all three types of models included: lead-time, patient prior no-show behavior, cell phone ownership, tobacco use, and the number of days since the last appointment of patient. Logistic Regression and Naïve Bayes classifier have commonly identified visit type, age and insurance as top ten predictors.

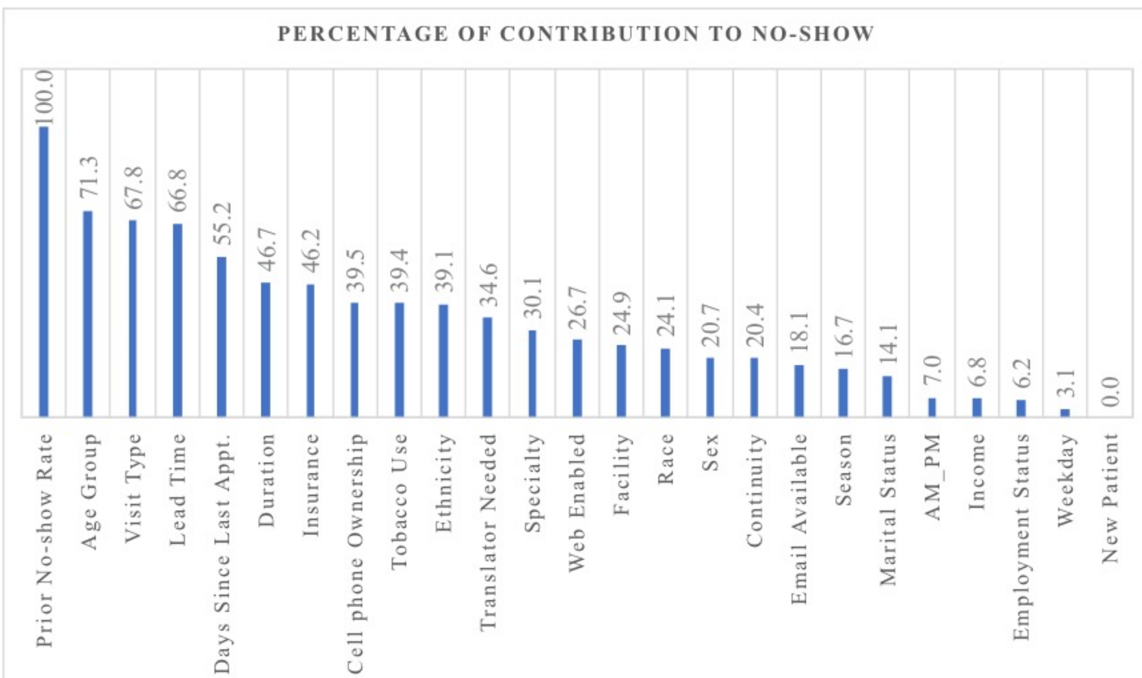



**Figure 3.2.** Naïve Bayes Classifier no-show predictor contributions and ranks.

<u>Model Validation:</u> Table 3.5 shows the validation results for three models. Overall accuracy in table 3.5 is the correct classification ratio for the model. The AUC for Logistic

Regression and the Naïve Bayes classifier are respectively 0.81 and 0.86, which are considered excellent for discriminating between two outcomes[59] . Neural network had low AUC of 0.66.

**Table 3.5.** Validation and comparison of prediction models. Abbreviations: AUC: Area under the curve for the receiver operating characteristic curve.

| Modeling Method | Train Set | | | | Test Set | | | |
|---|---|---|---|---|---|---|---|---|
| | AUC | Sensitivity | Positive (No-show) Predictive Value | Overall Accuracy | AUC | Sensitivity | Positive (No-show) Predictive Value | Overall Accuracy |
| Logistic Regression | 0.91 | 0.84 | 0.58 | 80% | 0.81 | 0.72 | 0.54 | 73% |
| Neural Network | 0.77 | 0.73 | 0.43 | 79% | 0.66 | 0.63 | 0.35 | 71% |
| Naïve Bayes Classifier | 0.96 | 0.82 | 0.67 | 92% | 0.86 | 0.73 | 0.59 | 82% |

## 3.6 Discussion

### 3.6.1 Findings

We studied missed appointments in 10 separate clinics within one urban community healthcare system. Our study shows that clinics have different population characteristics, specialties and patient demographics; thus, it is not surprising that appointment adherence varies across geographic sites. For example, specialty clinics such as pediatric or woman clinics have higher missed appointment rates than the ones providing acute or general primary care. Appointment lead-time, past missed appointments and age group of patients are the common important factors differentiating clinics' overall missed appointment rate. Our study suggests that any attempt to create a missed appointment prediction model or to design interventions for reducing missed appointment rates should be clinic/facility specific and tailored based on clinic, facility or department characteristics.

Our study has four major findings. First, patient, scheduling, and visit characteristics differ across missed and arrived appointments. These characteristics, such as patient, scheduling and visit characteristics, should be of interest to managers and policy makers, to better design interventions and policies to reduce missed appointments. Second, the consensus of the logistic regression, neural network, and naïve Bayes classification was that lead-time, patient prior missed appointments, cellphone ownership, tobacco use, and the number of days since the last appointment of a patient are the most significant predictors of missed appointments. Other factors were important in certain clinics, even after control for these factors. These findings should help managers in healthcare systems prioritize the design and implementation of interventions to reduce missed appointments. Third, patient appointment adherence had different determinants in different clinics or facilities within a single healthcare system. This finding makes sense in a large urban area, where neighborhood, population and clinic characteristics, as well as policies and procedures. It also underlines the importance of looking at data at the clinic level, because different clinics, even within the same system may have an important population and organizational differences. Fourth, according to the accuracy of the predictions, logistic regression and Bayes classifiers concluded similarly and perform better in missed appointment modeling than an artificial neural network classifier. This might be because of categorical nature of our data. Studies have reported that the discrimination ability of neural networks versus other statistical modeling techniques is data specific[78].

<u>Poverty, employment, and access to health information technology:</u> One key social determinant of health in populations is economic stability; this includes measures such as education, poverty and employment status[79]. We found that lower income and unemployment were associated with more missed medical appointments that would likely impair the health and/or health outcomes of patients. Studies found that socioeconomic characteristics have negative impact on health outcomes[80].

The role of poverty and employment are obviously complex and multifactorial across the United States. Our findings point to the need for social, financial and educational interventions to help indigent people prosper and communities thrive. Access to emerging technologies such as cell phones, the Internet and social media is another social and financial determinant. We found that patients without access to cell phone, email and a patient portal were more likely to miss their medical appointments. Therefore, lack of access to these technologies may impact health outcomes. Future research should examine if the provision of these consumer health technologies alone can enhance access to health for individuals in poverty or if our finding is more directly related to financial status alone. Our results show that patients without insurance for medical services are at risk of not adhering to their appointments and consequently their care plans. This factor is highly correlated with unemployment, which was very high (approximately 80%) in our study population.

Patient engagement, tobacco use, and promoting patient appointment adherence: In our study, smoking was one of the most significant factors related to missing medical appointments. We hypothesize that this variable as a health behavior, which may be highly related to other health practices, including adherence to scheduled clinic visits. It is beyond the scope of this study to determine whether this variable is a marker for adherence with recommendations or a confounder. Regardless, its importance underscores the importance of engagement of the underserved populations in their care and the role of individual health behaviors, attitudes, and practices.

Our study found that behavioral health patients were more likely to miss their next appointments than any other type of patients. Differences in adherence with appointments here could either be related to different systems for scheduling and reminding patients of appointments between medical and behavioral health systems, or related to intrinsic differences in practices, attitudes or adherence among behavioral health patients. Further investigation of this problem should focus on differences between the practices and policies for such patients, before efforts to make special accommodations for the population.

Scheduling redesign: Common reasons for missed appointments found in prior research include forgetting about the appointment, competing priorities, and demands (such as the need to work or inability to leave work), availability of transportation, or feeling better at the time of the appointment[81]. These reasons can be magnified if the lead-time (the most important predictor in our study) for appointments is elongated. Interventions such as increasing number of open access (same-day) hours and decreasing number of appointments made more than one month in advance should be considered to improve access to care in community health centers. Past missed appointment is an important predictor of future appointment adherence. Our findings are consistent with other research that operationalized passed missed appointments using clinicians' notes containing phrases like "no-show", "did not present", "failed to attend" and "missed appointment". These researchers found that patients that previously missed appointments were more likely to miss future appointments[82].

Methodologies proposed in this study have high performance of the prediction models. Logistic Regression and Naïve Bayes Classifier yielded prediction accuracy over 80 percent. Other machine learning algorithms, such as random forest, might or might not improve the accuracy, but interpretability of results remains unanswered. Therefore, a key advantage of our proposed methods is to have both high performance and interpretability that is significant in prediction modeling in health care where interpretation of results is crucial for designing interventions.

### 3.6.2 Limitations

One of the limitations of this study is that it includes only patients from one CHC system in Indianapolis. However, this CHC system involves multiple geographic sites and is very diverse from the patient characteristics perspective. Another limitation of this study is that the dataset used in this study did not have information on the clinical, physical and functional status of patients (e.g., diabetes, depression, congestive heart failure, etc.). These attributes can be significant predictors of the no-show. However, visit type variable in our dataset did relate to a patient's clinical characteristics. Findings of this study are drawn from FQHC clinics providing primary care to underserved populations. Whether these results are generalizable to other patient populations will need to be addressed in other studies. Another limitation of this study is that the dataset did not include information about new patients who no-showed in their first appointments; however, sufficient number of observations did not significantly impact the outcomes of this study.

### 3.6.3 Future Work

These results demonstrate the value of using existing clinical and operational data to address important operational issues. Further resources are needed in CHCs to make these data readily available and to inform important operational and policy questions. Another area of future work can be application of other machine learning algorithms, such as random forest, to potentially improve accuracy of predictions; however, interpretability of results is questionable[83]. Interpretability is important because one of the objectives of our no-show modeling was to design interventions based on the significance of predictors.

Future work might also focus on linking billing information and claims data with EHR to extract important information about patients and appointments. One example could be using evaluation and management codes to adequately identify provider type or provider time spent with patients.

## 3.7 Potential Medical Applications

Our study used large patient datasets with multiple potential explanatory variables to develop prediction models of various clinics within a healthcare system. We also used multiple methods to develop and compare the models. Access to health care can affect individuals' health status and quality of life. Missed appointments are one of the most important factors determining access to care. High levels of no-shows are not only an expensive waste of limited provider resources, but they can also lead to unmet health needs and delays in receiving appropriate care. Therefore, predicting and preventing missed appointments can potentially improve access to care[84]. The outcomes of this study could help clinics predict appointment no-shows that can potentially reduce no-show rates in CHCs. Researchers have reported lower no-show rates can improve clinical efficiency and utilization, reduce waste, improve provider satisfaction and lead to better health[85]. Redesigning and testing the alternate scheduling processes will help patients get appointments in a timelier manner. These better scheduling systems will improve access for acute patients, increase continuity of care for chronic patients and essentially positively impact health outcomes.

There are two possible real-world applications of this study. First, the methodologies and findings of this study can be used to redesign scheduling systems in CHCs to reduce the number of no-show appointments. Second, no-show predictions models can be implemented in EHR systems as decision support systems that would identify patients with a high risk of appointment no-show. Appointments with high risk of no-show may be double booked, or patients with high risk of no-show may be reminded more rigorously.

## 3.8 Conclusion

This project developed the statistical model and machine learning models that can be used to predict patients' chance of no-showing to their next medical appointment. Logistic regression, Neural Network, and Naïve Bayes Classifiers were utilized to develop and compare the no-show prediction models that resulted in finding lead-time, patient prior no-show behavior, cell phone ownership, tobacco use, and the number of days since the last appointment of a patient as significant predictors of appointment adherence. The application of these findings may be used to design new interventions to improve scheduling processes and other policies and practices for better and timelier access to care. We suggest that redesigned operations and policies, from scheduling practices to reminder systems and other technological tools to improve adherence can improve clinic revenues, utilization of resources and ultimately improve health outcomes.

# CHAPTER FOUR: SIMULATION OF SCHEDULING SYSTEMS

Effective scheduling processes can reduce clinic no-show rates and patient waiting time while improving continuity of care and overall clinic performance. Computer simulation models are used to assess the scheduling processes in CHCs, and provide a decision-making tool for clinic managers to analyze the impact of several scheduling scenarios. This chapter describes methodologies and approaches to develop simulation models using qualitative and quantitative data collected from CHCs. In this chapter, we illustrate the data requirements, methods used for data collection, data analyses needed to represent the clinical practice realistically, and methods used for validation of simulation results. Structured questionnaires and interviews were used to gather data about type of services offered in each clinic, number of providers, nurses and staff, access modalities, processes to make an appointment, scheduling methods and horizons, and insurance plan enrollment processes. Clinic managers, staff, quality assurance directors, schedulers, financial advisors, nurse managers, call center staff and front desk staff were the key questionnaire respondents. Workflow observations are necessary to fully understand the scheduling process in each CHC. The clinic staff working at the front desk (check in/check out), call center, scheduling and enrollment stations were observation targets for mapping the scheduling processes. Electronic Health Records (EHR) data was used to build patient population characteristics, and visit frequencies. Data requirements for the simulation model included provider characteristics, patient characteristics, appointment types, visit frequencies, and scheduling methods. No-show and cancellation probabilities and rates were calculated using statistical modeling or machine learning algorithms. Simulation models were validated by clinic teams and by comparing the simulation output results with the CHC actual output. Patient flows, patient demographics and characteristics, scheduling algorithms and rules, provider capacities and schedules, and overall clinic demands were used to build agent-based simulation models of scheduling systems.

## 4.1 Scheduling Systems

Predicting use of and improving effectiveness of health services are some of approaches to improve access to care[86]. In our project to improve access to care for underserved populations in Indiana, we found offering modified open access, reducing the cost of insurance, providing urgent, walk-in and after hour care, and better customer service as top solutions to overcome access barriers[18]. More effective scheduling addresses availability and accommodation barriers. Therefore, here, we present a guide to develop a computer simulation model to assess and simulate the scheduling processes in community health centers (CHCs), and provide a decision-making tool for clinic managers to analyze the impact of several scheduling systems.

### 4.1.1 Appointment Modalities

Community Health Centers use appointment scheduling to manage patient access to providers. Appointment scheduling system is a set of processes and rules determining how patients and providers get access to the resource within healthcare systems. Appointment systems are usually designed based on clinic demand, hour availability, available software, and patient and provider preferences[55]. Appointment scheduling ought to be efficient, satisfy patients, and provide timely access to health services[29,87]. There are three main scheduling methods in CHCs. First, traditional scheduling method with triage appointments where only acute appointments are considered as triage appointments. Second, open access scheduling, also known as same-day scheduling where all patients can get an appointment on the day they call, and follow-up appointments are scheduled when the time comes. Third, modified open access scheduling is a hybrid of open access and traditional scheduling. In modified open access scheduling, a portion of clinic hours is allocated for same-day appointments, also called "open access" hours, and the rest are allocated for pre-scheduled appointments. Some clinics might provide walk-in hours and urgent care and allow appointment overbooking in their scheduling systems. Advantages of open access scheduling are reducing delays in patient care, higher patient satisfaction, lower wait times and no-show rates[88,89]. Disadvantages of open access scheduling are reduction in health service utilization, no changes in clinical outcomes, not working well for chronic care patients, requiring overhauling scheduling systems, and potential negative impact on continuity of care[31,90].

### 4.1.2 Appointment Scheduling Optimization

Community Health Centers face several appointment scheduling challenges including provider shortage, limited provider availability, multiple patient visit types, appointment no-shows and cancellations. Patients in CHCs are concerned about long waiting times to get an appointment, not getting appointments at convenient times, and not getting appointments with preferred providers[5,91]. Effective scheduling processes can reduce clinic no-show rates and patient waiting time while improving continuity of care and overall clinic performance[5,91]. Assessing and testing new scheduling systems can help improve current practices, but repeatedly changing a real system could be disruptive, expensive and dangerous[92]. Simulation modeling is a tool that enables engineers to understand, assess, and test alternate systems designs. Simulation modeling allows for mistakes to be made in computers prior to any change or implementation[93]. This technology may also be used to facilitate decision making for primary care providers and clinic managers who often struggle with decisions about scheduling, such as when to offer open access hours, how to utilize staff resources (e.g. medical assistants, nurses, and mid-level providers). Here, we present the necessary steps for utilizing simulation modeling techniques to improve scheduling systems in community health centers. Through partnership with seven CHCs in Indiana, we aimed to find patient centered solutions and strategies to improve access to care. We developed simulation models of scheduling systems in our partner clinics to test and find best alternate scheduling system that could potentially improve access measure. This chapter describes step by step approach to develop, validate, and run healthcare workflow simulation models for making more informed decisions.

## 4.2 Simulation Modeling Background

Simulation is the process of designing a model of a real system (airports, factories, hospitals, primary care clinics, emergency departments, etc.)[94]. In a simulation analysis, operations of a facility, system, or process are imitated based on assumptions, estimations, or mathematical calculations. A simulation modeling study consists of inputs to build the model, "what if" scenarios to be tested, and outputs to be measured. Generally, inputs of simulation models of real systems are persons, tools, materials, resources, processes, workflows, demands, and capacities, and outputs are efficiency, safety, efficacy, quality, and optimization[95]. Models must be validated either numerically or by experts in order to represent the realities[93].

In healthcare arena, simulation modeling has been used to improve efficiency, workflow, patient flows, and operations in different settings. One simulation research studied reducing the waiting time of patients and increasing the utilization of resources of the Magnetic Resonance Imaging (MRI) in radiology department[96]. Researchers have used simulation modeling to study healthcare settings such as pediatric clinic, pediatric CICU, surgery unit, family health clinic, ophthalmic clinic, and oncology clinic [97]. One study simulated outpatient clinic scheduling of X-ray examination[98]. Another study focused on using simulation modeling for improving patient flow and scheduling in an primary care setting providing services such as walk ins, child care, and family health by interns and residents for underserved patients[99].

Researchers have applied different methodologies to simulate, understand, and optimize scheduling, flow, and operations of healthcare systems. Discrete event simulation (DES) has been the most common used methodology for studies aiming at reducing patient waiting time or length of stay[96,100-102]. Discrete event simulation has also been used for studies aiming at improving scheduling systems. Discrete event simulation was used to understand scheduling and patient flow in a family health center for underserved patients[99]. They used discrete event simulation to evaluate process flow, scheduling, and staffing in an oncology clinic[103]. Other simulation methodologies in healthcare have used multivariate linear regression approach to maximize operating room (OR) utilization[104], mathematical, and Monte Carlo simulation model to minimize patient waiting time and doctor idle time[86]. Other researchers used complex mathematical model to improve scheduling for spine surgeries[105].

Since primary care settings consist of active players (patients and providers), it seems using Agent Based Simulation (ABS) is more effective than DES[106]. Our pilot study[91] focused on using ABS to improve scheduling systems in CHCs. This small pilot study was limited to two weeks of data provided by two clinics that use traditional scheduling method. Research on agent based modeling for improving scheduling systems needs to be expanded, and in this chapter, we discuss approaches to develop ABS to model the patient flow and scheduling process.

Data collection is a crucial step to understand and simulate the behavior of a healthcare system. In simulation study of inpatient settings, data on arrival patterns, late arrivals, and service times were found important data points to collect[96]. To simulate patient flow inside outpatient clinics, information about provider full time equivalents (FTEs), appointment durations, patient trajectory, and time stamps were vital data points to collect[102]. In studies on surgery unit or operating room scheduling, variables such as procedure types, patient related variables impacting length of surgery and technical operative were important[104]. One study collected electronic scheduling and timekeeping system information on 3,245 open and endovascular vascular surgery operations to maximize operating room (OR) utilization[104]. Other researchers observed 234 patients encounter data (check in, check out times, and etc.) in an underserved setting between June and September 2013, in order to understand patient flow, scheduling, and operating procedures in the setting[99].

In studies focusing on improving inpatient setting processes and scheduling, performance measures such as patient waiting times, length of stay, staff utilization, and resource utilization were analyzed[96,101,104]. Performance measures for improving processes in outpatient setting were patient waiting times, length of stay, provider idleness, staff overtime, and patient throughput[98,100,102]. Performance measures for improving scheduling system in outpatient setting were wait time to get an appointment, provider idle time, waiting time for treatment, and length of stay[86,103,105]. No-show rate, the difference between discharge time and appointment time, resource utilization, and closing time were some measures simulate a family health clinic for underserved patients [99].

### 4.2.1 Basic Steps of a Simulation Study in Clinical Settings

Workflow study, data collection and analysis, and modeling are three major steps of a simulation study. Figure 4.1 shows details of each step. Simulation studies begin with workflow study by conducting interviews, creating questionnaires, and observing work and activities to fully understand clinic operations, patient flows, and scheduling processes. Then, patient level, operational, financial, and quality measure data (mostly found in EHR) are collected and analyzed to determine capacities, demands, population information, and appointment characteristics. Last step of simulation modeling is to validate assumptions, input data and model, and to prepare model for scenario analysis for testing alternate system strategies to improve performance of the system. This chapter thoroughly explains all steps and sub-steps of simulation modeling shown in figure 4.1.

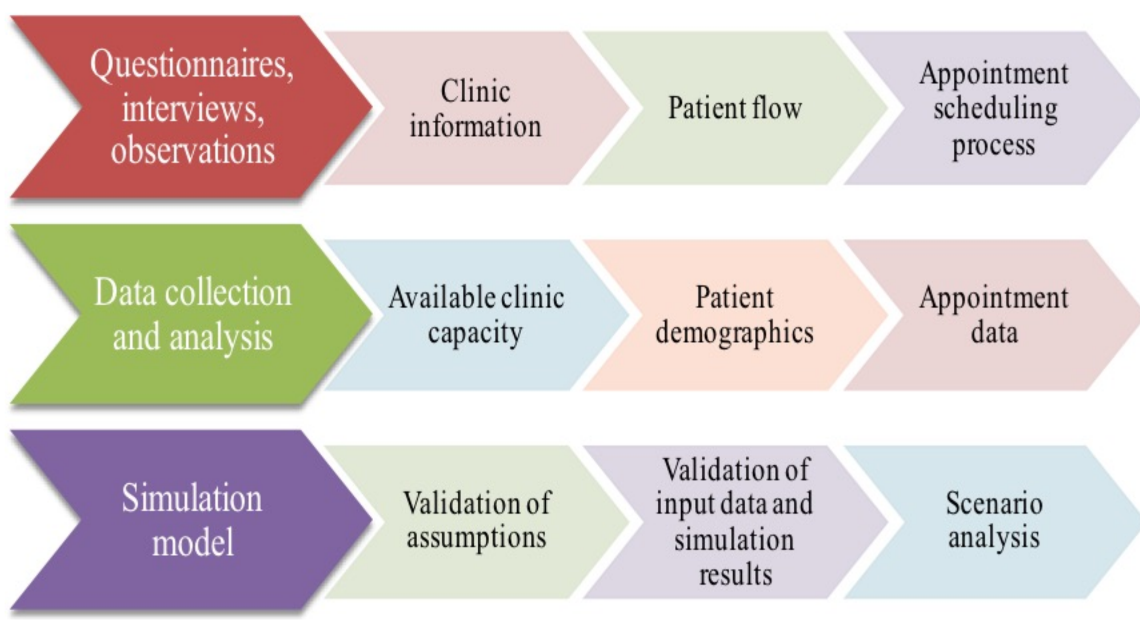



**Figure 4.1.** Basic steps of a simulation study in clinical setting.

### 4.3 Workflow Study

The purpose of a workflow study is to collect and analyze workflows and processes of real systems. A healthcare workflow study is performed by designing questionnaires and interviews and conducting work and activity observations to understand detailed system operations and information, patient flow, and appointment scheduling processes[107]. This not only allows for building scheduling algorithms and patient flows in the simulation models, but it also helps with developing scenarios to be tested before implementation in the real world. Activity diagrams are created from the perspective of various subjects (patient, physician, nurse, administrator, etc.), their goals, and tools, rules, and roles that mediate their interaction with the overall clinic setting. Workflow study identifies bottlenecks and problems in a real system[108], and defines objectives of simulation modeling. Outcomes and performance measures and alternative system strategies are mostly determined during workflow study; however, some of these items can change and be redefined during model or assumption validation steps (shown in figure 4.1).

#### 4.3.1 Work and Activity Observation

Work observations are necessary to fully understand the scheduling processes in healthcare systems. Activities, persons, and trajectories that need to be observed or recorded are determined based on the research questions and during the interview and questionnaire steps. In this study, the problem is improving access to care via scheduling systems, so targets of work observations would be persons who schedule patients or are in charge of moving patients inside the clinics. This includes clinic staff working at the front desk (check in/check out), call center, scheduling and enrollment stations and nurses or medical assistant in charge of taking care of patients and scheduling follow up appointments. In our study with 7 CHCs, a group of 5 to 6 persons observed schedulers, nurses and medical assistants, check in and check out staff, and enrollment and financial specialists. Observations for each clinic took about 6 hours on average. Clinic operational officers introduced observers to the participants in the beginning. Observers switched their stations each hour and took notes during observations. At the end of observations, notes were shared and combined. Table S.2 shows some example of what was recorded during observation meetings. Our observations show that the scheduling process changes according to different patient types (e.g. new vs. established). Also, acute, non-acute, follow up and enrollment (e.g. Medicaid and Medicare) visits are typically the four main appointment types.

### 4.3.2 Patient Flow in Scheduling System

Information derived from interviews, questionnaires, and work observations were used to create a representation of patient flow and develop the scheduling algorithm. We drew patients and process flowcharts in this step. Figure 4.2 shows a general patient flow in CHC scheduling system. Patients might be in one of the five main states that are 1) no appointment (i.e. needing an appointment); 2) scheduling appointment; 3) waiting for appointment day after scheduling; 4) visiting enrollment or financial specialists; and 5) visiting the medical provider. There are 2 paths for seeing a medical provider, call-in and walk-in. Clinics usually have walk-in hours, and walk-ins are only allowed during those hours. New patients might not be allowed to walk-in. Patients who schedule appointments to see providers might cancel their appointments or wait until the day of appointment and no-show. Some clinics consider appointments cancelled within 24 hours of the appointment as no-show. Enrollment appointments are appointments with enrollment or financial specialist in CHCs to enroll new patients to the clinic, get more information about the clinic, learn about clinic policies or apply for Medicaid, Medicare, or sliding fee scale.

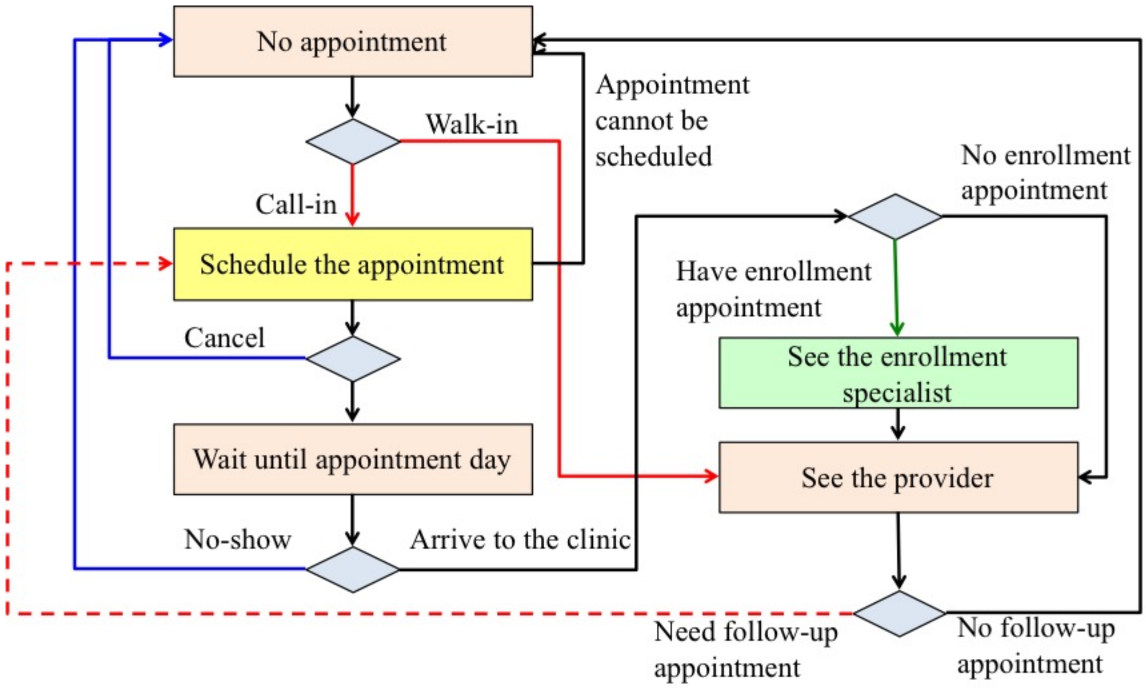



**Figure 4.2.** General patient flow in outpatient clinic scheduling system.

Usually, new patients or patients with expired Medicare/Medicaid need enrollment appointments. Some clinics have the exception to allow pregnant low income women to schedule and do the documents when they come for the first appointment. Enrollment appointments are usually scheduled on the same day or before the medical provider appointment. Follow up appointment and time are determined by medical providers.

### 4.3.3 Patient Flow During Clinic Visit

Figure S.1 shows general flow and potential activities of patients in during a clinic visit. The in clinic patient flow was drawn based on observations and interview results. In clinic patient journey starts with patients arriving at the clinic and going to check-in station. In the check in, the staff check patient’s identification, insurance card, new address or phone number, and medications. Following waiting in the lobby, patient might need to see an enrollment specialist. The next step starts with a nurse or medical assistant taking the patient to exam room and recording their vital signs and entering them into the EHR system. After that, patient is seen by the scheduled or rendering provider. Patients might need to go to laboratory for sample draws, or they might go to imaging room for radiology. Following this step, patient might need to be referred to specialty or hospital. CHCs usually have referral coordinator office that facilitate patient referrals, and help with finding and scheduling referral doctors, or transportation assistance. The last step is the check-out station that facilitates follow up scheduling, payments, and pharmacy referrals. Incorporating the time aspect of each activity into this patient flow completes the in-clinic work flow study. Work Observation Method By Activity Timing (WOMBAT) software is one the tools that might be used for recording amounts of time each staff or care provider spends with patients, it could be used to record waiting time and resource utilizations in different stages of care delivery in the clinics[109]. WOMBAT measures the sequence of tasks in order to assess changes in workflow. WOMBAT records all time related data, interruptions during care, and multitasking of personnel.

### 4.3.4 Appointment Scheduling Rules

Any clinic scheduling system includes a set of rules and constraints. These are business decisions usually made by operational managers. These rules include, but not limited to, scheduling horizons and priorities, provider restrictions, and overall appointment scheduling policies. We derived this information by interviews, questionnaires, and work observations.

Scheduling horizon: Scheduling horizon is a threshold for how far an appointment might be scheduled. Scheduling horizons are controlled by clinics operations for acute, non-acute, and follow-up appointments. In traditional appointment scheduling, appointments can be scheduled 30, 60, and 180 days in advance for acute, non-acute, and follow-up appointments respectively. While, in open access scheduling, planning horizon might be as low as 2 and 5 days for acute and non-acute appointments.

Scheduling algorithm and priorities: Appointments are divided to three major categories of acute, non-acute and follow-up in terms of priorities and acuity. Acute appointments are for patients who need immediate attention and care. Primary care clinics find appointments as early as possible for acute patients. Non-acute, non-urgent, or routine appointments are for patients with less rapid change or well/checkup patients. Clinics usually schedule these appointments with a reasonable waiting time, preferably with patient's primary care provider (PCP). Follow-up appointments are for patients with an initial action that needs to be followed with additional steps for more effective care. For example, patients might need a follow up appointment after a colonoscopy. Lead time and provider for follow up appointments are recommended by provider of first action. Figure 4.3 shows a general schema of algorithm for scheduling medical appointments in community health centers.

Provider restrictions: Doctor, nurse, and nurse practitioner are three main categories of provider medical license. Primary care clinics have several providers with different medical licenses. Providers are also categorized based on specialties including psychiatrist, certified nurse midwife, internal medicine, gynecologist, pediatrician, podiatrist, dentist, and optometrist. Some clinics form care team, that is doctors, nurses, NPs and medical assistants working together as a team, and assign patients to teams to improve performance. Patients are scheduled with providers or teams based on their care needs. Therefore, provider specialty is another condition shaping scheduling systems. Patients can also be clustered to complex and non-complex based on their health statuses. Complex patients are those with chronic conditions such as diabetes, hypertension, or comorbidity. Clinics usually assign these patients to physicians, and non-complex patients to physicians, nurse practitioners, or nurses.

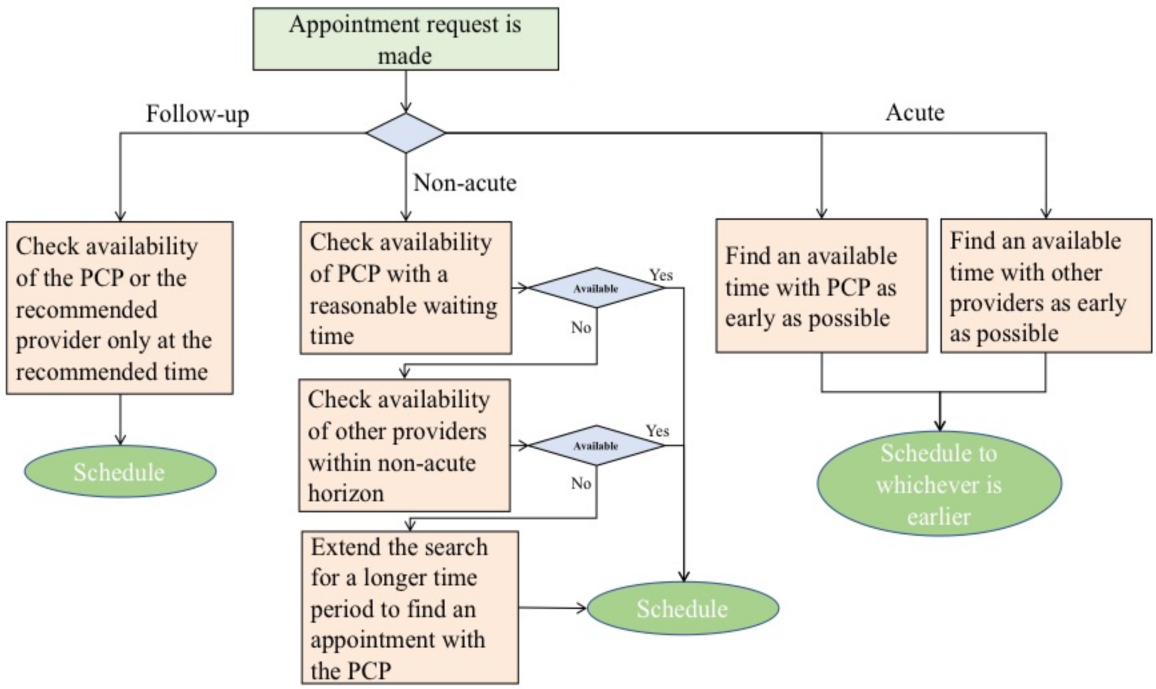



**Figure 4.3.** General schema of algorithm for scheduling medical appointments in community health centers. Abbreviations: PCP, primary care provider.

<u>Appointment scheduling policies:</u> Traditional scheduling method with triage appointments, open access scheduling, and modified open access scheduling are the three methods of appointment scheduling in outpatient clinics. Interviews and observations revealed details and policies for each of the scheduling methods. Modified open access scheduling is a hybrid of traditional and open access. Clinics might have different modified open access policies and hours. For example, some clinics only allow same day scheduling if the patient calls before certain time in the morning and the patients may not be able to

see their PCP. Some clinics might not allow scheduling appointments more than 4 weeks in advance, or they allow it only for patients who need transportation or translation services.

In this step, we also determined clinic's policies for walk in appointments. For example, we determined when is the walk-in hours and conditions that are and are not allowed during walk in hour, and these conditions could be acute illnesses such as sinus infection, UTI, flu, and diarrhea or minor procedures such as cuts, bruises, minor burns, removal of stiches, etc. Clinics have different overbooking policies if overbooking is allowed. For example, a clinic might not allow back to back double booking, or it might only allow double booking for certain providers on certain day of week. Clinics might not double book complex patients or patients with a disability.

Building the most realistic model of scheduling systems in community health centers required detailed workflows, rules, algorithms, restrictions, and policies plus providers willingness to change their schedules.

## 4.4 Data Collection and Analysis After Workflow Study

Data collection and analysis is a necessary step for calculating a system's capacity and demand. Clinics are different in terms of capacity, patient population, provider types, and types of interventions they are willing to test or implement. EHR and clinic management system data determine provider and clinic capacity, patient demographic, demand for acute, non-acute and follow-up appointments, and visits characteristics. Because the clinic performance measures are usually a yearly metric, one year of encounter data is sufficient for building clinic specific simulation model. Population characteristics, patient health status, treatment guidelines, and adherence behavior can be used to determine health care demand and needs.

### 4.4.1 Patient Distributions

Age, gender, race, ethnicity, marital status, education, language, income, insurance, distance to health care facility, reason of the visit, diagnoses and co-morbidities are found in EHR data. Patients with different age or gender have different care needs. Figure 4.4 shows number of patients seeking medical appointment based on age and gender in one year. Patient age ranges from zero to 90. As it is shown, number of female patients and male patients seeking health care is not different between ages of 0 to 15. Between ages of 16 to 64, women seek health more frequently than men do. As patients become older, number of patients is getting more similar for men and women. These distributions are used to create artificial patients reflecting reality of current practice in simulated clinic. Variables such as race, ethnicity, marital status, education, language, income, or distance to health care facility are collected because they might be predictors of appointment adherence.

In addition to EHR, this information might also be collected from Uniform Data System (UDS) Resources in the Bureau of Primary Health Care of Health Resources and Services Administration. Insurance data is another important information in the EHR systems. Insurance is not only important for determining needs for enrollment appointment, but it also might predict patient compliance (e.g. no-show, cancel or arrive) to the appointment. Clinical characteristics of patients and schedulers' notes (discussed in chapter two) could also be used to group patients in order to find care needs of patient population[3].

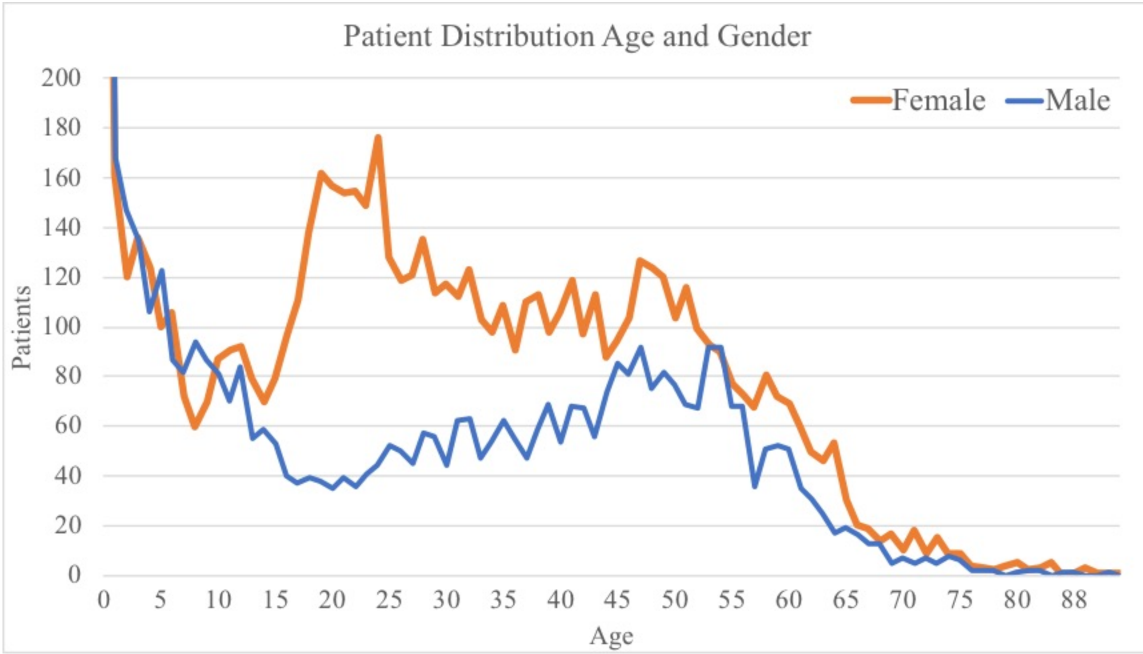



**Figure 4.4.** Number of patients seeking health in different Age and Gender groups.

### 4.4.2 Care Demands

The number of patients coming to a healthcare system or number of appointments requested is our definition of demand for healthcare systems. Patients enter into the scheduling systems by walking in, calling in, or scheduling follow up appointments (figure 4.2). If the clinics allow walk-ins, the assumption is that walk-ins are allowed only for established patients during the walk-in hours. EHR data determines the percentage of patients comes to the clinic as walk-in. Majority of patient demand is from call-ins. Each patient has an overall visit frequency (number of visits per year), which depends on gender and age. Each patient has a frequency (number of visits per year), which depends on gender, age, or chronic condition. The time between phone calls to request an appointment depends on these visit frequencies and the time of last visit. If the EHR data has information about follow up appointments such as recommender provider, recommended provider, or recommended time, the data can be used to calculate follow up appointment demand; else, follow-up appointment will be treated like call-ins. Children need to be seen at regular intervals for well-child visits. Adults need to be seen at regular intervals for preventive visits. Patients with chronic conditions must be seen at regular intervals (e.g. diabetes patients are seen 2-4 times per year). A certain percentage of patients with an acute appointment might need a follow-up appointment (after lab tests or treatment). The lead time of the next follow-up appointment is determined according to guidelines for well-visits and chronic conditions.

Visit frequencies. Treatment guidelines were used to estimate frequency of well child checks, well women exams, and preventive visits for chronic conditions. Visit distributions and frequencies were calculated using the one-year encounter data. Visit frequencies are age and gender specific and were calculated by dividing number of visits in an age group by number of patients in that age group in one year. Sick and well visit distributions were calculated for all age and gender groups. These distributions are clinic specific. Figure 4.5 shows density plot of visit frequency distributions for sick males, sick females and all visits in a clinic during 2014. The mean of female visit frequency (2.57 visits per year) is larger than the mean of male (2.11 visits per year) visit frequency (p-value=0.0003). These distributions as well as patient distributions were used to create patient agents in simulation modeling; therefore, the simulated patient population could represent the reality of care demands in CHCs. The findings shown in figure 4.5 do not necessarily imply that women are sicker; however, it magnifies the importance of reality in that particular clinic. These distributions are used to create random sick visits per year for each patient group based on the realty of current practice in simulated clinic.

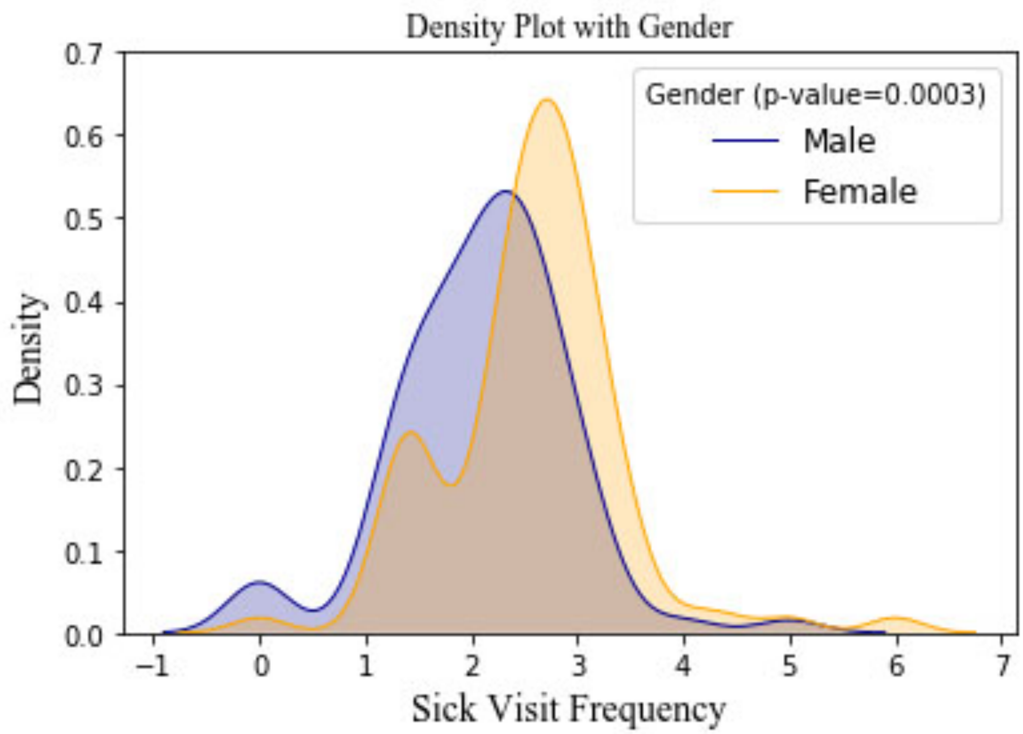



**Figure 4.5.** Distributions of sick visit types in different gender groups.

### 4.4.3 Clinic Supply and Capacities

Visit types in clinics: Visits were grouped based on patient characteristics and appointment types in EHR or clinic patient management systems. Figure 4.6 shows categories of visits from an urban CHC.  There are usually many visit types in EHR systems. These types could be grouped based on similarities and patient age and gender to fewer groups to simplify simulation modeling. Methodology described in chapter two could also be used for simplifying the appointment types for simulation modeling. The visit types could also be grouped based on patient type into the main categories of sick, adult, behavioral, pediatric, pregnant, and woman patients (see table S.3).

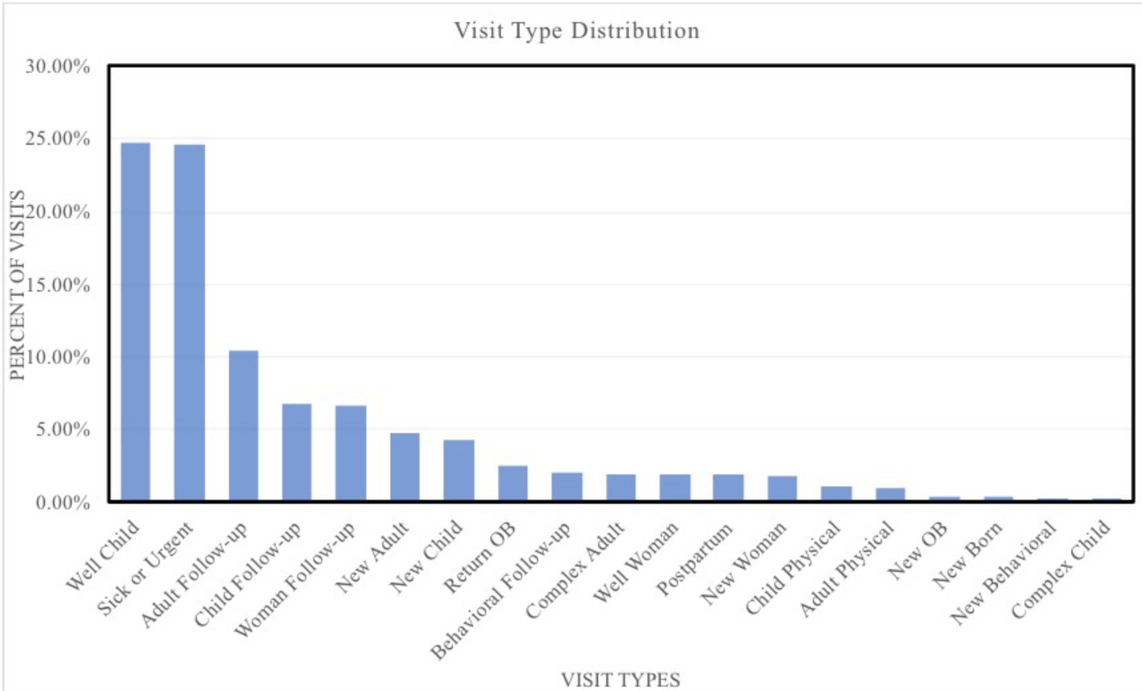



**Figure 4.6.** Visit type distribution in one year of clinic operation.

Provider specific appointment types and capacities: Providers within a healthcare system have distinct patient populations, specialties and capacities. Analysis of EHR encounter data provides this information. EHR data revealed provider of the visit and the appointment type for each encounter. Contingency table of provider versus appointment type revealed provider capacities for each appointment type. Table 4.1 shows provider capacities, patient group and appointment types in a CHC during 2014. It shows number of patients that a given provider sees in each category of visit type per year. This table (4.1) was used to create provider agents, their specialties, patient groups, and capacities in the simulation model.

**Table 4.1.** Providers capacities, number of unique patients seen per year per visit type.

| Provider ID | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Total |
|---|---|---|---|---|---|---|---|---|
| Patient Group | Woman, Pregnant | Behavioral | Acute, Adult | All | Woman, Pregnant | Pediatric | All | |
| Acute/Sick | 2 | 0 | 15 | 301 | 3 | 312 | 323 | 956 |
| Complex Adult | 0 | 0 | 3 | 148 | 0 | 2 | 156 | 309 |
| Complex Child | 0 | 0 | 1 | 9 | 0 | 17 | 2 | 29 |
| Adult Follow-up | 8 | 0 | 31 | 210 | 1 | 26 | 146 | 422 |
| Behavioral Follow | 0 | 111 | 0 | 0 | 0 | 11 | 0 | 122 |
| Child Follow-up | 0 | 0 | 8 | 104 | 0 | 110 | 118 | 340 |
| Woman Follow-up | 90 | 0 | 0 | 16 | 161 | 3 | 3 | 273 |
| New Adult | 0 | 0 | 23 | 91 | 0 | 92 | 24 | 230 |
| New Behavioral | 0 | 14 | 0 | 0 | 0 | 0 | 0 | 14 |
| New Child | 0 | 0 | 3 | 166 | 0 | 122 | 30 | 321 |
| New OB | 0 | 0 | 0 | 0 | 33 | 0 | 0 | 33 |
| New Woman | 15 | 0 | 0 | 0 | 60 | 0 | 0 | 75 |
| Adult Physical | 0 | 0 | 0 | 12 | 0 | 1 | 7 | 20 |
| Child Physical | 0 | 0 | 0 | 1 | 0 | 4 | 6 | 11 |
| Postpartum | 5 | 0 | 0 | 0 | 84 | 0 | 0 | 89 |
| Return OB | 23 | 0 | 0 | 0 | 89 | 0 | 1 | 113 |
| Well Child | 0 | 0 | 24 | 463 | 0 | 253 | 485 | 1225 |
| Well Woman | 9 | 0 | 1 | 6 | 49 | 3 | 1 | 69 |
| Total | 152 | 125 | 109 | 1530 | 480 | 965 | 1306 | 4667 |

The length of appointments varies by appointment type and provider. Appointment type and provider specific durations were also obtained from analysis of EHR data. Table 4.2 shows that length of appointments varies by appointment type and provider. For example, appointment duration for “New Adult” patients could be 30, 45, or 60 minutes.

**Table 4.2.** Providers visit duration (minute) distribution that is per provider per visit type.

| **Provider ID** | **Acute/Sick** | **Complex Adult** | **Complex Child** | **Adult Follow-up** | **Behavioral Follow up** | **Child Follow-up** | **Woman Follow-up** | **New Adult** | **New Behavioral** | **New Child** | **New OB** | **New Woman** | **Adult Physical** | **Child Physical** | **Postpartum** | **Return OB** | **Well Child** | **Well Woman** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **1** | 15 | | | 15 | | | 15 | | | | | 30 | | | 15 | 15 | | 15 |
| **2** | | | | | 30 | | | | 60 | | | | | | | | | |
| **3** | 30 | 30 | 30 | 30 | | 45 | | 60 | | 30 | | | | | | | 15 | |
| **4** | 15 | 30 | 30 | 30 | | 15 | 30 | 30 | | 30 | 20 | | 15 | | | | 15 | 30 |
| **5** | 15 | | | 30 | | | 15 | | | | 20 | 30 | | | 15 | 15 | | 15 |
| **6** | 15 | 30 | 30 | 15 | | 15 | | 30 | | 45 | | | 30 | 30 | | | 15 | |
| **7** | 15 | 30 | 30 | 15 | | 15 | 15 | 45 | | 30 | | | 30 | 30 | | 15 | 15 | 15 |

Other capacity information including provider FTEs, provider schedules (working hours on each day of the week), capacity allocated for same-day appointments on each day of the week, and panel sizes were provided by the clinics.

### 4.5 Model Building

The interview, observation, and EHR data were analyzed to create input files for simulation modeling. The inputs were 1) provider capacity (capacity allocated for same-day and prescheduled appointments); 2) patient population characteristics (age, gender, race ethnicity, income insurance, health status, etc.); 3) patient demand for care; 4) scheduling horizon for each appointment type (acute, non-acute, follow-up); 5) maximum panel size for each provider; and 6) no-show prediction models and cancellation rates.

#### 4.5.1 Appointment Adherence

According to the patient flow (figure 4.2), there are 5 uncertainty nodes that the simulation model needs to make a decision. They are 1) call in vs. walk in; 2) enrollment needed vs. not needed; 3) follow up needed vs. not needed; 4) cancel vs. keep appointment; and 5) no-show vs. arrive to clinic. We described about how a simulation model might determine call ins, walk ins, enrollment visits and follow ups. No-shows and cancelations could be determined by statistical analysis of patient encounter data in the past.

No-show modeling: The patients might miss their appointments with a certain probability. Patient history, behavior, and characteristics, and provider and visit characteristics are different between no-show and arrived appointments. EHR data including patient and scheduling information were used to predict and analyze no-show behavior patients[59,110]. We discussed in chapter three that statistical or machine learning models, such as Logistic Regression, Neural Network, and Naïve Bayes Classifiers, might be used to predict patients' chance of no-showing to their next medical appointment. Multiple variables, including patient demographics (age, gender, race, insurance status), provider specialty, clinic operations (appointment reminders), and appointment characteristics (appointment type, duration, appointment day/time, season, lead time) contribute to these prediction models. In our no-show modeling study, we found that lead time, patient prior no show behavior, cellphone ownership, tobacco use, and number of days since last appointment of patient as significant predictors of appointment adherence[4].

Cancelation: The patients might cancel the appointment with a certain probability. Cancellation probabilities might either be calculated using prediction models similar to no-show modeling, or be assumed constant cancellation rates that come from the EHR data.

## 4.6 Simulation Run

Simulation modeling models the care delivery system, and identifies the impact of valid interventions on process and outcome measures. The data and information gathered in the workflow study and data collection and analysis steps were used to build the simulation models of scheduling systems. AnyLogic software is one of the powerful simulation modeling software which supports three major simulation methodologies: discrete event simulation, system dynamics, and agent-based modeling. In our study, we develop clinic specific agent based models of scheduling systems. The agent based models had four major components of patient agent, provider agent, patient flow and scheduling algorithm. Patient agent included variables such as demographics, insurance and income level, appointment variables (time, duration, etc.), compliance (arrive, no-show and cancel), health status (pregnant, complex, chronic condition, patient group etc.), primary care provider and enrollment provider, walk in, call in, or follow up, new or established, next visit (sick, well, etc.), last visit, and lead time. Provider agent included variables such as provider name, ID, appointment type, capacity, capacity per day, panel size, medical license, number of patients seen, total clinic hours and clinic hours busy. The inputs for simulation are demand and capacity distributions to create patients, providers, and their parameters. Patient flow and scheduling algorithm derived from workflow study were implemented in AnyLogic (seen in figures S.2 and S.3). The model simulated the appointment scheduling process at a tactical level to determine the impact of daily capacity allocation including total provider capacity and percentage of appointments allocated for same-day access on waiting times for appointments, no-shows, and cancellations. This model was used to test a list of potential strategies in terms of how much capacity should be allocated for same-day appointments for each provider on different days of the week.

## 4.7 Performance Measures

Performance or outcome measures are different based on research questions of studies. The main outcome measures for access to care are patient waiting time for an appointment, in-clinic waiting time on the appointment day, no-shows, cancellations, continuity of care, and resource (provider and staff) utilizations. These measures were selected based on National Committee for Quality Assurance patient-centered medical home (PCMH) recognition standards related to access and continuity of care, and process measures that reflect efficiency[111]. For example, providing same-day appointments for routine and urgent care is one of the critical factors to achieve PCMH recognition. Waiting time for an appointment is lead time between when the appointment is requested and the actual appointment. Monitoring no-shows is the other two PCMH standards related to access. No-show rate is the number of missed appointments divided by number of all scheduled appointments. Cancellations and provider utilizations are other outcome measures of the impact of reducing waiting times for appointment. Provider productivity was defined as number of hours used to see the patients divided by total number of hours allocated for patient visits. Monitoring the percentage of patient visits with selected clinician or team (continuity of care) is another PCMH standard related to continuity. We defined continuity of care as number of visits with the primary care provider divided by number of visits with all providers. In-clinic waiting times on the appointment day are also considered as an efficiency measure that affects patient satisfaction.

## 4.8 Validation

The purpose of the validation step is to validate the assumptions, input files, patient flows, performance measures and scheduling algorithms of simulation models. This step also validated and finalized potential scheduling changes, scenarios and interventions that could improve performance measures. A team of clinic staff, providers, schedulers and managers, and patients validated simulation models. This step involved all stakeholders in the development stage for continuous refinement to get the most realistic models.

Validation of assumptions: We validated assumption of scheduling algorithms, patient flow, provider FTEs, or enrollment visits. For example, If the patient cancels the appointment, the assumption was that it might occur at any time before 24 hours prior to the appointment time. Or appointments canceled within 24 hours of the appointment time were considered no-show. Another example is that enrollment appointments happen before or preferably on the same day of the medical appointments.

Validation of input data: Validation of the input data is another critical step of simulation modeling. The input data was validated using current EHR data by comparing simulation output with output calculated from real data. Validation teams were also a source to check validity of input files.

Validation of patient flow and scheduling algorithms: Validation teams verified validity of patient flow and scheduling algorithms. Clinic team verified whether the model reflects reality or current workflow in the organization. Clinic team advised on what may be missing in the simulation model such as a person, specialty, or resource. Patients validated if there is something in the patient flow that seems to be missing based on their experiences in going to providers.

Validation of performance measures: Performance measures such as no-show/cancelation rates, waiting times, productivity, and continuity of care were validated by clinic team or published guidelines. Clinic teams validated how clinics calculate their outcome measures or whether there was any other outcome measure they were interested in looking into. For example, some clinics calculate provider productivity using time allocated to see patients; whereas, in other clinics, provider productivity was calculated based on providers' FTEs.

## 4.9 Scenario Analysis

Simulation modeling was used to determine the impact of interventions identified by our research team and the clinic teams. Interventions were grouped into four categories. First, we modeled different scheduling methods to answer questions like: Does changing the number of triage appointments (open slots for same-day appointments) improve outcome measures? How does open access scheduling affect performance measures (i.e. provider productivity, continuity of care)? How does overbooking affect operational performance measures (i.e. provider productivity, overtime)? Second, we model different practice style such as creating care teams (providers and nurses working together) to improve the performance measures (i.e. provide improved access for acute visits, reduce missed opportunities for well-child visits). Third, we modeled different access modalities including after-hour, extended clinic hours or weekend hours. Fourth, we modeled different staffing levels to determine the optimal staffing levels such as number of providers or enrollment specialists. In our study, we simulated different intervention alternatives including 1) providing more same-day appointments; 2) reducing the number of appointments scheduled for more than one month in advance; 3) changing the schedules of the providers (i.e. more available hours on Mondays); 4) simplifying the appointment types; 5) standardizing the appointment durations (i.e. using 20 minute appointments instead of 15 and 30 minute appointments); and 6) using provider care teams.

### 4.9.1 Recommendations to Clinics

Simulation models tested interventions that clinic partners were willing to implement. Here, we show some of our partner clinics interested in seeing the impact of different open access scheduling alternatives including one prescheduled and 2-3 same-day appointments per hour and one provider with open schedule all day. We developed a simulation model with scheduling templates as inputs to simulate the patient flow (from arrival to discharge) throughout the day. We tested the impact of three alternatives shown in table 4.3. The simulation results show that the first alternative provides an average time in system of 45 minutes with 17.6% waiting for more than one hour. The second alternative gives an average time in system of 47 minutes with 21.6% waiting for more than one hour. The third alternative has an average time in system of 44.5 minutes with 16.8% waiting for more than one hour. The percentage of arrivals is 70% for current practice at the partner

CHC. It increases to 75% for alternative 1 and 3 and to 80% for alternative 2. We shared the results with clinicians and clinic managers for selecting the most optimal strategies to be implemented.

**Table 4.3.** Scenarios of scheduling systems and recommendations for improvement.

| Alternatives | Percentage of appointments pre-scheduled | Percentage of appointments same-day | Description | Outcomes |
|---|---|---|---|---|
| **0** | 76 | 24 | Current practice | Arrivals: ~70% |
| **1** | 50 | 50 | All providers<br>8-9am – All prescheduled (15 minute and 30 minute appointments)<br>9am-end - One pre-scheduled and 2-3 same day appointments per hour | 45 min<br>17.6% (>60min)<br>Arrivals: 75% |
| **2** | 30 | 70 | All providers<br>8-10am – All prescheduled (15 minute and 30 minute appointments)<br>10am-end - 2-4 same day appointments per hour | 47 min<br>21.6% (>60 min)<br>Arrivals: 80% |
| **3** | 50 for MDs<br>25 for NPs | 50 for MDs<br>75 for NPs | MD: 8am-12pm – All prescheduled (15 minute and 30 minute appointments)<br>12pm-end - 2-4 same day appointments per hour<br>NP: 8am-10am – All prescheduled (15 minute and 30 minute appointments)<br>10am-end – 2-4 same day appointments per hour | 44.5 min<br>16.8% (>60 min)<br>Arrivals: 75% |

Figure 4.7 and supplement figures S.4, S.5, and S.6 show results of simulation modeling of traditional, open access, and modified open access for six years. In this scenario analysis, open access was defined as allocating all clinic hours to same day appointments with maximum scheduling horizon of 2 for all appointments. Traditional scheduling was defined as scheduling patients based on acuity with maximum horizon of 30 days for acute and 90 days for non-acute and follow-up appointments. Modified open access scheduling was defined as allocating half of clinic hours to same day appointments with maximum horizon of 2 days for acute and non-acute appointments and 30 days for follow-up appointments. Full open access scheduling significantly decreased number of appointments created in the model; however, it significantly decreased the clinic no-show rates (dropped to 10% average). Traditional scheduling had the highest no-show rates (around 18%). Results show that modified open access scheduling significantly improved

access measures; however, it extremely increased provider productivity to above 95% on average. Providers might not choose this approach as they might conclude that they could be forced to see patients above their daily capacities. Parameters of modified open access scheduling should be optimized to address provider concerns.

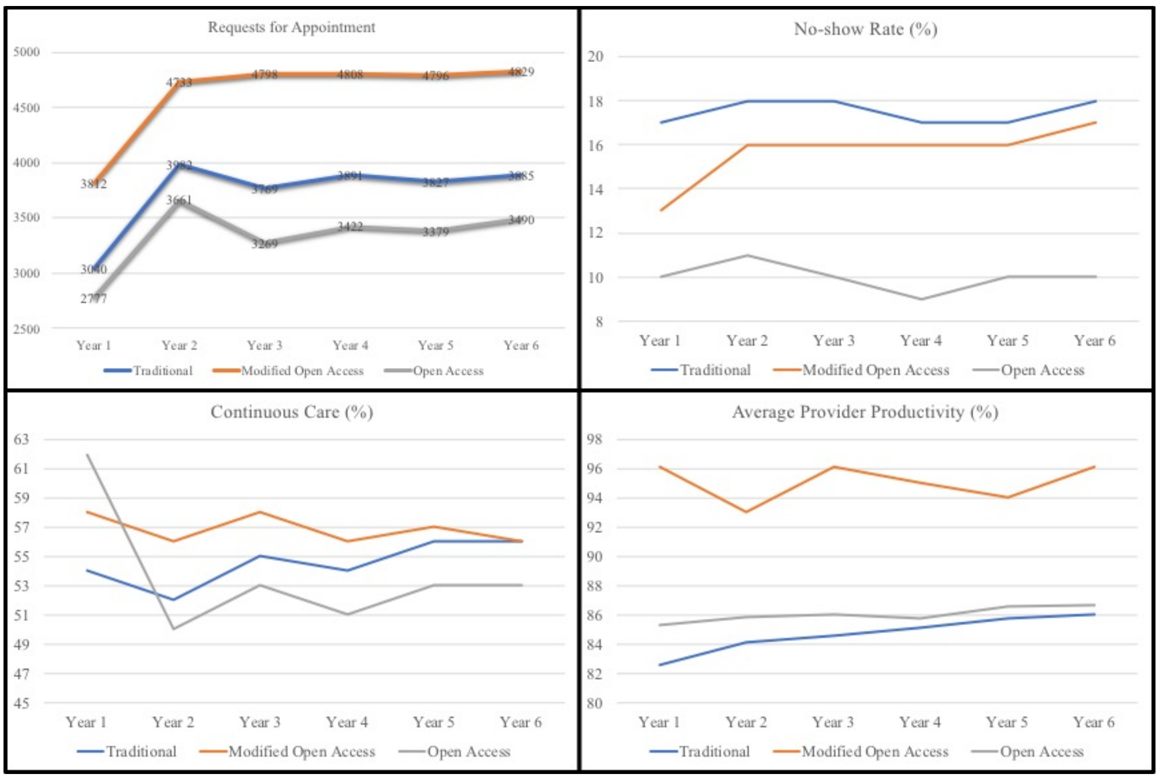


**Figure 4.7.** Outputs of simulation models of three appointment scheduling modalities for six years. Top left, number of appointments requested by patients per year; top right, average overall clinic yearly no-show rates; bottom left: continuity of care that is yearly rate of appointments scheduled to patient's PCP by all appointments; bottom right: percentage of hours used to see the patients by of hours allocated for patient visits.

## 4.10 Discussion

Agent-based simulation modeled patient flow and appointment scheduling process in CHCs. The simulation model is flexible in terms of usability by different clinics. Inputs can be through Excel files or graphical user interface, which make the model useful for any clinic. We illustrated the value of simulation models to test alternative strategies to improve access to care through scheduling. One of the major findings of this study was that an agent-based simulation model of scheduling system can try variety of "*what-if*" scenarios of modified open access. Modified open access can be clinic or provider specific that would potentially enhance willingness of providers to comply with the changes. Another finding is that patient population in different clinic varies based on the demand of the clinic. Therefore, any attempt to model a real healthcare system needs to take to account that artificial patients are generated based on reality. We found patient visit frequencies are age and gender specific, and this information can be derived from EHR data including real world evidence of practice.

A key advantage of method, agent-based simulation (ABS), proposed and implemented in this study over other techniques is that it can capture interactions between individuals[112]. This is important in simulation study of healthcare settings because interactions between patients and providers could significantly impact health outcomes[113]. Flexibility of ABS is another advantage over other simulation approaches[112]. Modelers can easily add new agents. For example, modelers can add agents such as nurses, medical or transportation staff, or advance practice providers in more complex healthcare settings such as hospitals or trauma centers.

In this study, we designed a road map and a tool to conduct workflow study and to develop simulation model of a real healthcare system. Future researchers are encouraged to leverage the workflow tool provided in this study to identify whom should be interviewed or observed, what questions should be asked, and what information should be collected. In this chapter, we provide a step by step process of developing a model of a real system. Although we focused on community health centers, findings and guide tools can be used in studies of other healthcare settings.

Our simulation modeling study has some limitations. Although we built the simulation models based on one-year retrospective appointment data, we had to change the time frame of our data analysis and the input data during the validation meetings. The changes in input data were due to the implementation of ongoing improvement initiatives such as providing more same-day appointments and extended hours on certain days of the week in the clinics, provider turnover (providers leaving the practice, new providers), and changes in provider schedules. The changes in capacity were provided by the clinics. Since we received EHR data at regular intervals (every month), we were able to analyze and observe the changes in patient data (i.e. change in insurance status, increase in number of new patients due to increased insurance coverage, waiting times for appointments, etc.), changes in scheduling practices (i.e. providing more open appointment slots), and the impact on performance measures.

Although the proposed approaches and methodologies are supported by the work during three years of the study, clinics and providers might not always fully comply with suggested strategies coming from simulation modeling, and the results of interventions might be different from simulation. Future work in this area could be to incorporate variables in the simulation model to test whether a person/system chooses to comply with interventions. Another area of future work might work on enhancing the models for improving providers and directors buy-in; thus, future researchers could include cost analysis to achieve an improved compliance. Future researcher might also consider including other interventions to make the simulation modeling findings work. For example, interventions could be adding appointment reminders to scheduling system to reduce no-show and continuously reaching out to clinicians to make sure they are implementing the findings of simulation study. Approaches discussed in this chapter can be used to develop simulation modeling in other area of healthcare, such as emergency room, trauma center, and other ambulatory care settings. For example, agent-based simulation might be used to simulate patient and provider flow in trauma centers to find what would be the most optimal number of providers, nurses, or advance practice providers. Trauma centers might be interested to specifically find out the utilization of high cost advance practice provider.

## 4.11 Conclusion

The methodologies described in this chapter are from disciplines such as statistics regression, computer science, informatics (e.g. EHR data analysis, survey and questionnaire analysis) and industrial engineering. In this chapter, we explained how to integrate informatics and health services research approaches to optimize clinical operations and processes. Core principles in our methodology are continual learning and listening to our stakeholder partners by including patient and staff perspectives, flexibility, finding win-wins,

The approaches and methodologies demonstrated in this work were used in a project to improve access to care for underserved populations through partnership with seven community health centers in Indiana. Questionnaires and interviews for understanding overall operations of the partner clinics, workflow observations and EHR data analysis can be used to build the simulation model to identify effective scheduling processes and test alternate strategies to improve timely access to care. Workflow analysis generates patient flows and scheduling algorithms shown using flow charts. A detailed step-by-step guide on how to develop a simulation model for scheduling systems in community health clinics is provided. Simulation models can be built in AnyLogic software. The results and assumptions are shared and validated by clinic staff, managers and clinicians to have the most realistic model. Simulation models are used to test different scenarios of scheduling methods and to identify how these scenarios impacted performance measures. The simulation models give the percentages that providers in a clinic should allocate to someday appointments. Agent-based simulation can model patient flow and appointment scheduling process in CHCs. The simulation model is flexible in terms of usability by different clinics.

# CHAPTER FIVE: SUMMARY OF CONTRIBUTIONS

Access to health care is fundamental for optimal health and wellness; thus, removing obstacles is not only important for improving the patient experience, but also health outcomes. Poor people, as well as racial and ethnic minorities, are more likely to receive lower quality of care and encounter more obstacles to health care access compared with other groups[114]. This research, through partnership with seven Community Health Centers (CHCs) in Indiana, constructed effective outpatient appointment scheduling systems by determining care needs of CHC patients, designing an infrastructure for meaningful use of patient health and clinic operational data, and developing prediction and simulation models for improving access to care for underserved populations. This research provides a foundation to redesign current policies and workflow to ensure that primary care is available, timely, coordinated, and effective.

Over the past two decades, machine learning, modeling, and simulation have helped teams and managers support different decision making processes. In this research we applied these techniques to answer our research questions. First, we applied word embedding, a natural language technique, to extract important patient related information from scheduler notes in EHR data that helped redesign scheduling templates. We found patient characteristics, patients' reasons for seeking health, and clinic/provider characteristics can be used to improve patient centeredness, efficiency, and timeliness of outpatient scheduling systems. Second, we utilized predictive modeling to fully understand appointment non-adherence in CHCs. We added to previously found predictors of appointment no-show and found poverty, employment, mental health status and access to health information technology are additional factors contributing to missed appointments.

We found patient engagement and tobacco concession could promote patient appointment adherence. Third, we identified the necessary steps to develop a simulation toolkit for improving scheduling systems in community health centers. Our simulation models suggest that, compared to traditional or open access scheduling, a modified open access scheduling can find a balance between clinic performance metrics and access measures.

## 5.1 Significance

In a study by Center for Disease Control (CDC) Health Disparities and Inequalities, racial, ethnic, geographical, socioeconomic, and other factors were found as barriers to healthcare[114]. Around 80% of the underserved patients in the US population have income less than 100% federal poverty level. They concluded that access to care for low-income populations is not improving. Lack of access to care can increase unmet healthcare needs, delay in getting care, inappropriate use of emergency room and cost[115]. Making use of modern technology (such as EHRs, Telehealth, etc.) have been identified an innovation for improving access[116]. This project produced a generalizable method for extracting clinical data from EHR systems and reusing them for practice-based research. Methods proposed in this research did not only make use of electronic health records, but it identified target population care needs and socio-economic status through data analytics, questionnaires, and interviews. Approaches in this study could also be used for system improvement in other areas of health care and even other industries interested in system redesign and improvement. This dissertation will potentially improve well-being of and care quality for uninsured, underinsured, and underserved people.

<u>Clinical significance:</u> The outcomes of this study helps clinicians predict appointment no-shows that can potentially reduce no-show rates in CHCs. Researchers

have reported lower no-show rates can improve clinical efficiency and utilization, reduce waste, improve provider satisfaction and lead to better health outcomes. Redesigning and testing alternate scheduling processes will help patients get appointments in a timelier manner. These more efficient and optimized scheduling systems will improve access for acute patients, increase continuity of care for chronic patients and essentially positively impact health outcomes. This project would potentially improve access to care that would lead to reduce unmet healthcare needs, decrease delay in getting care, reduce inappropriate use of emergency room, and decrease overall cost of care for health care system. This project also provided approaches to ensure scheduling systems are capable of properly meeting the populations' care needs. Although this work was confined to the state of Indiana, barriers to accessing care are common and the approaches to overcoming these barriers can likely be applied in CHCs elsewhere. Lists of solutions and strategies for implementation could be used by other CHC systems. These findings likely generalize to other patient populations given the broad diversity of patient population characteristics from those clinics and patients included in the study. Our population was diverse in race/ethnicity, age, and urban/rural location of the clinics, as well as in socioeconomic status and insurance status.

Scientific significance: This project is theoretically and empirically significant to the scientific community. The theoretical importance of this study is based on whether large patient datasets can be used to reengineer healthcare systems in a way that patients care needs are thoroughly met, and both patients and clinicians are more satisfied. This study theoretically presented the importance of data analytics, simulation, and modeling technologies as tools improve healthcare systems.

This study is empirically important to the scientific community to better understand applications of simulation and modeling in reengineering healthcare systems. This study improves the knowledge of the scientific community on what data (necessary information) and how to collect data to simulate clinical workflows and build predictive models capable of understanding the complexity of clinical appointments, patient compliance and other aspects of healthcare scheduling. This work improves awareness among the scientific community about strategies and interventions used by clinicians, clinical managers and policy makers. This study also modifies current simulation and modeling methods for healthcare industry in a way that they are more effective, accurate, and comprehensive.

## 5.2 Innovations

This project has great potential to lead to meaningful improvement in patient health, well-being, and quality of care through its focus on making use of underinsured, underserved, and uninsured patient data. In this project, we integrated the disciplines of informatics, engineering and health services to address the issue of access to care for underserved patients. The project is innovative because it provided a novel technique for CHCs to redesign and simplify their appointment types and durations. NLP techniques have been widely used for extracting medical information from free text. In this project, applying NLP techniques on patient complaint for seeking health to redesign scheduling systems is novel. Engaging patients and including clinic staff, clinicians, and managers in the process of model development, model validation, and intervention design is another factor that makes this project unique. Using large and complex real CHCs patient data and including new predictive and population-specific features improved current algorithms of predictive modeling for appointment non-adherence. Neural Networks and Naïve Bayes

Classifier were used in other areas of healthcare, for example for prediction of hospital readmission; however, application of these methods for prediction of no-show is innovative. This project expanded the research using the real electronic clinical encounters by underserved population. Agent based simulation models of scheduling processes in CHCs were built and improved by using large and complex real patient data.

### 5.3 Contributions to Health Informatics

The U.S. National Library of Medicine defines Health Informatics as "the interdisciplinary study of the design, development, adoption and application of IT-based innovations in healthcare services delivery, management and planning[117]." The methodologies used in this project are from disciplines such as statistics (e.g. logistic regression), computer science (e.g. machine learning and clustering), informatics (e.g. EHR data analysis, NLP, survey and questionnaire analysis) and industrial engineering (e.g. simulation modeling). This project integrates health informatics and health services research. Health services research approaches focusing on developing new interventions to improve health care processes help health informaticians enhance their skills and knowledge to launch evaluations of the impact of technologic innovations. In this project, use of Electronic Health Records (EHRs) data, the main vessel of health informatics, were supported by involving health services research approaches. This project also expanded the research areas of Health Informatics by using Industrial Engineering approaches such as simulation modeling to optimize clinical operations and processes. New methods of collecting and analyzing EHR data in CHC provided new methodologies for health informaticians for future research on underserved populations. Improvement of simulation and modeling methodologies for healthcare industry will equip Health Informatics researchers to expand their research areas.

## 5.4 Conclusion

The project lays the groundwork for a variety of different kinds of future research. One focus of our research was in engaging stakeholders in uncovering effective behaviors and practices to improve access to care and redesign systems. Further research is needed to compare different methodologic approaches and strategies for both uncovering potential solutions and determining how to prioritize them. Beyond supporting data analytics to support quality improvement initiatives, there are multiple clinical, operational, and population health questions that might be addressed with such data. The project provided a complex, multicomponent methodologies to improve access to care.

We uncovered many practical approaches to improving access to care by listening to patients, clinical staff, and providers who have found creative pathways through the healthcare system. These approaches could inform potential interventions in other care redesign projects and ensure that the efforts are patient-centered and appropriate for the specific population. The simulation and modeling tools developed in this project not only allowed us as researchers to assess the impact in a timely and efficient manner, they also enabled the participating health centers to look at their own data, which were previously not easily accessible.

APPENDICES

Appendix A. Supplement Tables

**Table S.1.** Associations between variables and no-show in logistic regression model. This table shows all features selected by clinic specific logistic regression models. Ranks were calculated based on number of clinics that features were selected as significant predictor in.

| Predictor | Rank | Odds Ratio | 99% Confidence Intervals | | Clinics Number |
|---|---|---|---|---|---|
| **Lead Time** | 1 | | | | All clinics |
| more than 1 month vs. next day | | 2.43 | 2.18 | 2.70 | |
| more than 1 month vs. same day | | 7.12 | 6.46 | 7.84 | |
| more than 1 month vs. within 2 weeks | | 1.70 | 1.59 | 1.82 | |
| more than 1 month vs. btw 2 weeks and 1 month | | 1.20 | 1.11 | 1.29 | |
| next day vs. same day | | 2.93 | 2.59 | 3.32 | |
| next day vs. within two weeks | | 0.70 | 0.63 | 0.78 | |
| next day vs. btw 2 weeks and 1 month | | 0.49 | 0.44 | 0.55 | |
| same day vs. within two weeks | | 0.24 | 0.22 | 0.26 | |
| same day vs. btw 2 weeks and 1 month | | 0.17 | 0.15 | 0.19 | |
| within two weeks vs. btw 2 weeks and 1 month | | 0.71 | 0.66 | 0.76 | |
| **Prior No-show Rate** | 2 | 4.86 | 4.40 | 5.38 | All clinics except Clinic2 |
| **Age Group** | | | | | All clinics except Clinic6 |
| adult vs. child | 2 | 1.56 | 1.48 | 1.64 | |
| adult vs. senior | | 3.65 | 2.92 | 4.57 | |
| child vs. senior | | 2.34 | 1.87 | 2.92 | |
| **Insurance** | 3 | | | | 1,2,3,4,8,9,10 |
| Commercial vs. Medicaid | | 0.55 | 0.50 | 0.60 | |
| Commercial vs. Self Pay | | 0.37 | 0.33 | 0.41 | |
| Marketplace vs. Medicaid | | 0.49 | 0.32 | 0.76 | |
| Marketplace vs. Self Pay | | 0.33 | 0.21 | 0.51 | |
| Medicaid vs. Medicare | | 1.60 | 1.40 | 1.82 | |
| Medicaid vs. Self Pay | | 0.67 | 0.63 | 0.72 | |
| Medicare vs. Self Pay | | 0.42 | 0.37 | 0.49 | |
| **Tobacco use** | 3 | | | | 1,2,5,7,8,9,10 |
| current every day smoker vs. current some day smoker | | 1.32 | 1.06 | 1.63 | |
| current every day smoker vs former smoker | | 1.69 | 1.49 | 1.91 | |
| current every day smoker vs never smoker | | 1.95 | 1.80 | 2.12 | |
| current some day smoker vs former smoker | | 1.28 | 1.02 | 1.61 | |
| current some day smoker vs never smoker | | 1.48 | 1.20 | 1.83 | |
| former smoker vs never smoker | | 1.16 | 1.03 | 1.30 | |
| **Days since last Appt.** | 4 | | | | 1,2,3,7,8,9 |
| between 1 and 2 weeks vs. between 6 months and a year | | 1.45 | 1.17 | 1.78 | |
| between 1 and 2 weeks vs. more than a year | | 2.18 | 1.78 | 2.67 | |
| between 1 and 2 weeks vs. new patient | | 1.41 | 1.14 | 1.73 | |
| between 1 and 3 months vs. between 6 months and a year | | 1.51 | 1.35 | 1.68 | |
| between 1 and 3 months vs. more than a year | | 2.27 | 2.07 | 2.50 | |
| between 1 and 3 months vs. new patient | | 1.46 | 1.32 | 1.63 | |
| between 2 weeks and 1 month vs. between 6 months and a year | | 1.59 | 1.37 | 1.84 | |
| between 2 weeks and 1 month vs. more than a year | | 2.39 | 2.08 | 2.75 | |
| between 2 weeks and 1 month vs. new patient | | 1.54 | 1.33 | 1.79 | |
| between 3 and 6 months vs. between 6 months and a year | | 1.36 | 1.23 | 1.51 | |
| between 3 and 6 months vs. more than a year | | 2.06 | 1.88 | 2.25 | |
| between 3 and 6 months vs. new patient | | 1.32 | 1.19 | 1.47 | |
| between 6 months and a year vs. more than a year | | 1.51 | 1.40 | 1.63 | |

| | | | | | |
|---|---|---|---|---|---|
| between 6 months and a year vs. within a week | | 0.79 | 0.65 | 0.97 | |
| more than a year vs. new patient | | 0.64 | 0.60 | 0.69 | |
| more than a year vs. within a week | | 0.53 | 0.43 | 0.64 | |
| **Cell phone ownership** | 4 | | | | 1,3,4,5,8,9 |
| No vs. Yes | | 1.61 | 1.52 | 1.71 | |
| **Visit Type** | 5 | | | | 1,4,7,8,10 |
| Acute care vs. Adult Routine/Follow up | | 0.305 | 0.28 | 0.333 | |
| Acute care vs. Behavioral Health | | 0.18 | 0.155 | 0.208 | |
| Acute care vs. Podiatry | | 0.222 | 0.175 | 0.282 | |
| Acute care vs. Pediatric | | 0.439 | 0.405 | 0.476 | |
| Acute care vs. Pregnant | | 0.3 | 0.265 | 0.34 | |
| Acute care vs. Women | | 0.343 | 0.31 | 0.378 | |
| Adult Routine/Follow up vs. Behavioral Health | | 0.589 | 0.512 | 0.678 | |
| Adult Routine/Follow vs. Podiatry | | 0.728 | 0.574 | 0.921 | |
| Adult Routine/Follow vs. Pediatric | | 1.437 | 1.344 | 1.538 | |
| Adult Routine/Follow vs. Women | | 1.123 | 1.027 | 1.227 | |
| Behavioral Health vs. Pediatric | | 2.439 | 2.129 | 2.795 | |
| Behavioral Health vs. Pregnant | | 1.669 | 1.414 | 1.969 | |
| Behavioral Health vs. Women | | 1.905 | 1.644 | 2.208 | |
| Podiatry vs. Pediatric | | 1.976 | 1.564 | 2.496 | |
| Podiatry vs. Pregnant | | 1.351 | 1.05 | 1.739 | |
| Podiatry vs. Women | | 1.543 | 1.213 | 1.963 | |
| Pediatric vs. Pregnant | | 0.684 | 0.612 | 0.764 | |
| Pediatric vs. Women | | 0.781 | 0.72 | 0.848 | |
| Pregnant vs. Women | | 1.142 | 1.008 | 1.294 | |
| **Season** | 5 | | | | 1,2,7,8,10 |
| Fall vs. Spring | | 1.13 | 1.05 | 1.22 | |
| Fall vs. Summer | | 0.90 | 0.83 | 0.98 | |
| Fall vs. Winter | | 1.22 | 1.14 | 1.31 | |
| Spring vs. Summer | | 0.80 | 0.74 | 0.86 | |
| Spring vs. Winter | | 1.08 | 1.02 | 1.15 | |
| Summer vs. Winter | | 1.36 | 1.26 | 1.47 | |
| **Using Patient Portal** | 6 | | | | 1,2,4,6 |
| No vs. Yes | | 1.41 | 1.32 | 1.51 | |
| **Employment Status** | 6 | | | | 1,2,7,8 |
| Employed full-time vs. Employed part-time | | 0.76 | 0.67 | 0.87 | |
| Employed full-time vs. Not employed | | 0.80 | 0.74 | 0.87 | |
| Employed full-time vs. Retired | | 3.05 | 2.08 | 4.47 | |
| Employed full-time vs. Unknown | | 0.36 | 0.25 | 0.53 | |
| Employed part-time vs. Retired | | 3.99 | 2.70 | 5.89 | |
| Employed part-time vs. Self-employed | | 1.78 | 1.13 | 2.81 | |
| Employed part-time vs. Unknown | | 0.47 | 0.32 | 0.70 | |
| Not employed vs. Retired | | 3.81 | 2.62 | 5.54 | |
| Not employed vs. Self-employed | | 1.70 | 1.09 | 2.64 | |
| Not employed vs. Unknown | | 0.45 | 0.31 | 0.65 | |
| Retired vs. Self-employed | | 0.45 | 0.25 | 0.80 | |
| Retired vs. Unknown | | 0.12 | 0.07 | 0.20 | |
| Self-employed vs. Unknown | | 0.27 | 0.15 | 0.47 | |
| **Race** | 7 | | | | 1,4,7 |
| Asian vs. Black | | 0.39 | 0.33 | 0.47 | |
| Asian vs. Multiple Races | | 0.51 | 0.41 | 0.63 | |
| Asian vs. White | | 0.53 | 0.44 | 0.63 | |
| Black vs. Multiple Races | | 1.29 | 1.12 | 1.49 | |
| Black vs. Native Hawaiian and other Pacific Islander | | 1.82 | 1.35 | 2.45 | |
| Black vs. White | | 1.35 | 1.28 | 1.43 | |
| Multiple Races vs. Native Hawaiian and other Pacific Islander | | 1.41 | 1.02 | 1.95 | |
| **Patient's PCP?** | 8 | | | | 3,4 |
| No vs. Yes | | 1.30 | 1.21 | 1.40 | |
| **Duration** | 8 | | | | 3,5 |
| 1 hour vs. 10 minutes | | 20.39 | 8.61 | 48.32 | |
| 1 hour vs. 15 minutes | | 2.32 | 1.77 | 3.03 | |
| 1 hour vs. 20 minutes | | 1.99 | 1.51 | 2.61 | |
| 1 hour vs. 30 minutes | | 1.44 | 1.10 | 1.89 | |
| 1 hour vs. 45 minutes | | 0.63 | 0.45 | 0.90 | |

| | | | | | |
|---|---|---|---|---|---|
| 10 minutes vs. 15 minutes | | 0.11 | 0.05 | 0.26 | |
| 10 minutes vs. 20 minutes | | 0.10 | 0.04 | 0.22 | |
| 10 minutes vs. 30 minutes | | 0.07 | 0.03 | 0.16 | |
| 10 minutes vs. 45 minutes | | 0.03 | 0.01 | 0.07 | |
| 15 minutes vs. 20 minutes | | 0.86 | 0.80 | 0.92 | |
| 15 minutes vs. 30 minutes | | 0.62 | 0.58 | 0.66 | |
| 15 minutes vs. 45 minutes | | 0.27 | 0.22 | 0.34 | |
| 20 minutes vs. 30 minutes | | 0.73 | 0.67 | 0.79 | |
| 20 minutes vs. 45 minutes | | 0.32 | 0.25 | 0.40 | |
| 30 minutes vs. 45 minutes | | 0.44 | 0.35 | 0.55 | |
| **Age x Gender Interaction** | 8 | | | | 2,7 |
| Adult Female vs. Adult Male | | 1.11 | 1.02 | 1.21 | |
| Adult Female vs. Child | | 1.60 | 1.51 | 1.69 | |
| Adult Female vs. Senior | | 3.74 | 2.99 | 4.67 | |
| Adult Male vs. Child | | 1.44 | 1.32 | 1.57 | |
| Adult Male vs. Senior | | 3.36 | 2.66 | 4.25 | |
| Child vs. Senior | | 2.34 | 1.87 | 2.92 | |
| **Marital Status** | 8 | | | | 2,7 |
| Divorced vs. Legally Separated | | 0.73 | 0.57 | 0.93 | |
| Divorced vs. Married | | 1.30 | 1.10 | 1.53 | |
| Divorced vs. Widowed | | 1.48 | 1.08 | 2.01 | |
| Legally Separated vs. Married | | 1.78 | 1.44 | 2.21 | |
| Legally Separated vs. Partner | | 1.79 | 1.06 | 3.03 | |
| Legally Separated vs. Single | | 1.28 | 1.05 | 1.56 | |
| Legally Separated vs. Widowed | | 2.03 | 1.44 | 2.84 | |
| Married vs. Single | | 0.72 | 0.66 | 0.78 | |
| Single vs. Widowed | | 1.58 | 1.20 | 2.09 | |
| **Email availability** | 9 | | | | 5 |
| No vs. Yes | | 1.22 | 1.15 | 1.23 | |
| **Translator Needed** | 9 | | | | 7 |
| No vs. Yes | | 2.06 | 1.88 | 2.54 | |
| **Ethnicity** | 9 | | | | 8 |
| Hispanic or Latino vs. Not Hispanic or Latino | | 0.57 | 0.53 | 0.61 | |
| Hispanic or Latino vs. Unspecified | | 0.41 | 0.37 | 0.46 | |
| Not Hispanic or Latino vs. Unspecified | | 0.73 | 0.66 | 0.80 | |

**Table S.2.** Guide for scheduling system workflow observation. This table shows questions that should be asked during clinic workflow observations.

| **Observation Guide - Questions or Activities to be recorded** | | | |
|---|---|---|---|
| **Schedulers Station** | **Check in Station** | **Check out Station** | **Enrollment provider or Financial Advisor Station** |
| • Does every department like medical, mental, optical, etc. have different scheduling? Or there is one center for all?<br>• What is the thought process of schedulers for decision making about timing of the appointments? First available vs. scheduling in a way to avoid crowded clinic?<br>• What type of information is asked form patient during the scheduling? SSN, DOB, Insurance, Income level, etc.?<br>• What is the scheduling process for New vs. Established patients?<br>• What is the scheduling process for child, adult, pregnant, mental, woman, etc. patients?<br>• Who is initiating the scheduling scheduler, patient, or provider?<br>• How is the scheduling process for different appointments such as acute, non-acute or follow-up?<br>• How do schedulers take patient preference into decision making? Was patient satisfied with previous visit?<br>• What are the different visit types in the EMR system?<br>• Do schedulers guide new patients to get into the healthcare system?<br>• How scheduling chronic condition patients (e.g. FU from hospital, diabetes, HTN, heart disease, asthma, well child, etc.) is different from others?<br>• How do schedulers schedule double booking? Do they have double booking time frames or days?<br>• What is the process for scheduling same-day appointments?<br>• What is the process for canceling an appointment?<br>• How do schedulers separate scheduling appointment to NP VS. MD vs. Psychiatric? | • Is the check in process different for different conditions like pregnant, mental, child, dental, optical etc.?<br>• What is the overall check in/ registration process?<br>• How check is different for new vs. established patients?<br>• What is the waiting time in lobby?<br>• How long is the paper work for new patients?<br>• How do staff handle patients arriving late?<br>• What is the check in process for walk ins vs. scheduled?<br>• What documents are required?<br>• How do staff deal with patients without insurance?<br>• Is there walk in hours and what is the waiting time for walk in patents?<br>• What are the conditions that are allowed during walk in hour? For example, flu, cough, diarrhea, etc.?<br>• How no-show is defined and entered in the EMR systems? | • Who is in charge of check out the patients? Nurse, MA or check out station staff?<br>• How are follow up appointments scheduled?<br>• What is the payment processes and methods?<br>• How are patients reminded for their next appointments?<br>• What is the referral process? Referring to specialty provider or hospital?<br>• How do they make sure patients get their medication? | • Who is in charge of enrolling patients and financial advising?<br>• How are the enrollment visits scheduled? Same day of medical appointment? Before or after the medical appointment?<br>• How Medicaid, Medicare, Marketplace or sliding fee scale enrollment are different?<br>• What does Medicaid, Medicare, Marketplace, or sliding fee scale eligibility work?<br>• What is the Medicaid application for Pregnant and Child patients?<br>• How do they handle patients with expired plans?<br>• How is the process of presumptive Medicaid?<br>• What are the different insurance delivery systems and how are they different?<br>• Is there enough enrollment provider?<br>• Can the enrollment specialist or clinic apply for Medicare?<br>• How long is the enrollment visit? And which type is the longest? For example, Disability Medicare?<br>• How does the discount/sliding fee scale systems work? |

**Table S.3.** Visit types and groups in an urban CHC. This table shows how various visit types were simplified and aggregated to patient groups based on meaning similarities.

<table>
<tr><th>Visit Types in EHR</th><th>Visit Group</th><th>Percentage</th><th>Patient Group</th><th>Percentage</th></tr>
<tr><td>Sick or Urgent</td><td>Sick or Urgent</td><td>24.65%</td><td>Sick (Acute)</td><td>24.65%</td></tr>
<tr><td>Complex Adult</td><td>Complex Adult</td><td>1.98%</td><td rowspan="4">Adult Male or Adult Female</td><td rowspan="4">18.17%</td></tr>
<tr><td>ADHD evaluation, ER, Medicine or weight management follow up for adults</td><td>Adult Follow-up</td><td>10.43%</td></tr>
<tr><td>New adult</td><td>New Adult</td><td>4.76%</td></tr>
<tr><td>Physical exam or sport for adult</td><td>Adult Physical</td><td>1.00%</td></tr>
<tr><td>Psychiatry, mental or behavioral follow up</td><td>Behavioral Follow-up</td><td>2.10%</td><td rowspan="2">Behavioral Health</td><td rowspan="2">2.43%</td></tr>
<tr><td>Psychiatry, mental or behavioral new</td><td>New Behavioral</td><td>0.33%</td></tr>
<tr><td>Complex child</td><td>Complex Child</td><td>0.33%</td><td rowspan="6">Pediatric</td><td rowspan="6">37.56%</td></tr>
<tr><td>ADHD evaluation, ER, Medicine or weight management follow up for children</td><td>Child Follow-up</td><td>6.73%</td></tr>
<tr><td>New child</td><td>New Child</td><td>4.27%</td></tr>
<tr><td>New born</td><td>New Born</td><td>0.38%</td></tr>
<tr><td>Physical exam or sport for child</td><td>Child Physical</td><td>1.06%</td></tr>
<tr><td>Risk screen, well child or well visit for adolescents</td><td>Well Child</td><td>24.79%</td></tr>
<tr><td>New OB</td><td>New OB</td><td>0.42%</td><td rowspan="3">Pregnant</td><td rowspan="3">4.80%</td></tr>
<tr><td>Postpartum</td><td>Postpartum</td><td>1.91%</td></tr>
<tr><td>Non-stress test (NST), return OB, ultrasound, or work in OB</td><td>Return OB</td><td>2.47%</td></tr>
<tr><td>Contraceptives, Depo shot, OB GYN follow up or LEEP proc.</td><td>Woman Follow-up</td><td>6.60%</td><td rowspan="3">Women</td><td rowspan="3">10.28%</td></tr>
<tr><td>New GYN, woman, or OB consult</td><td>New Woman</td><td>1.75%</td></tr>
<tr><td>Well woman or annual female</td><td>Well Woman</td><td>1.94%</td></tr>
</table>

Appendix B. Supplement Figures

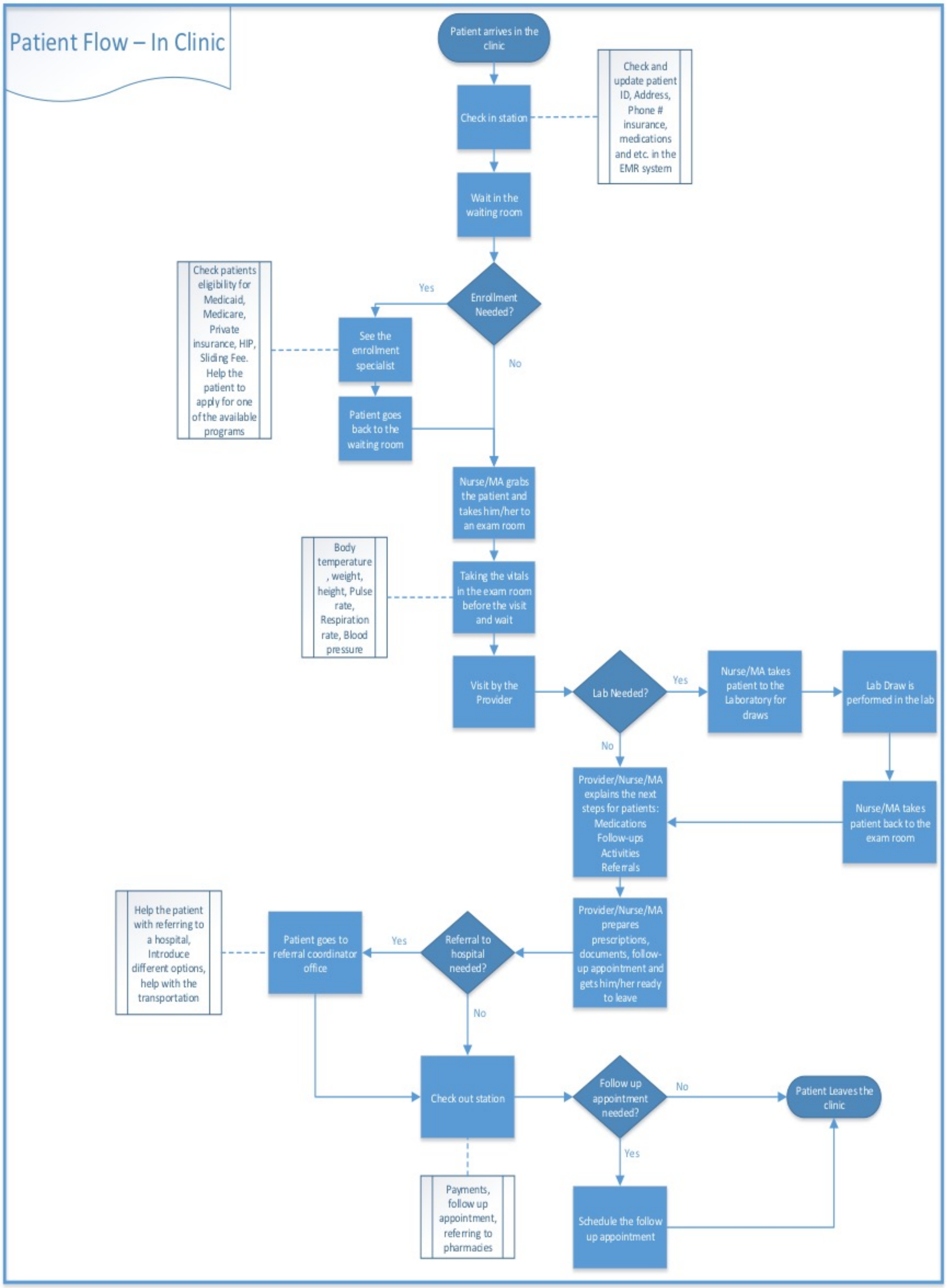



**Figure S.1.** General patient flow within outpatient clinic.

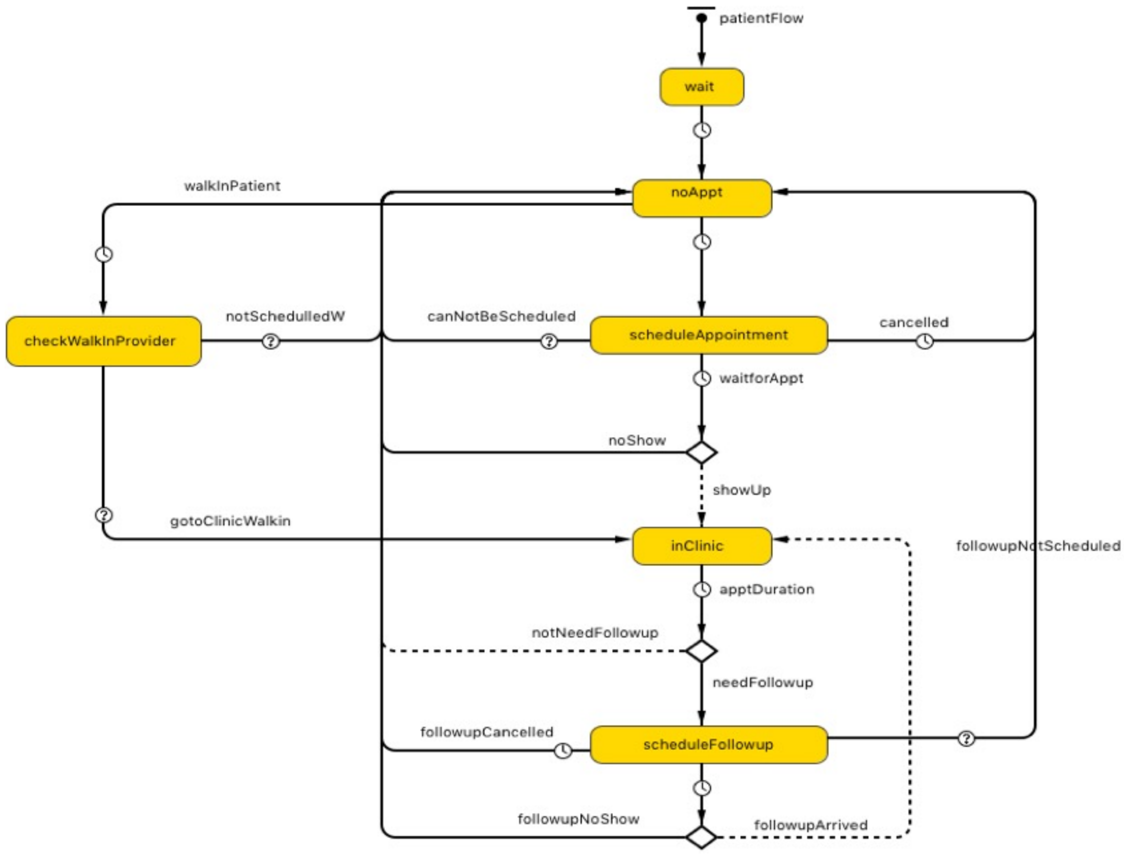
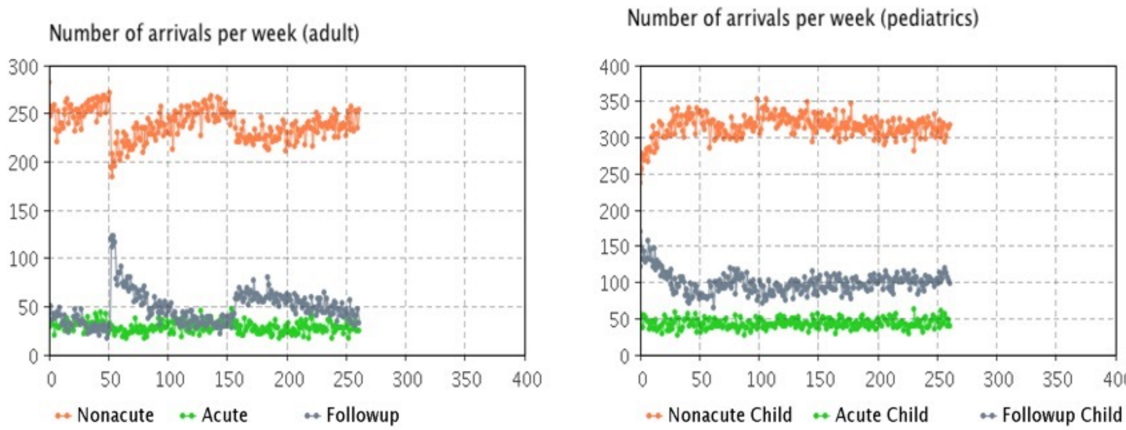



**Figure S.2.** Scheduling patient flow implemented in AnyLogic. This patient flow includes walk-ins. Not all clinics facilitate walk-ins.

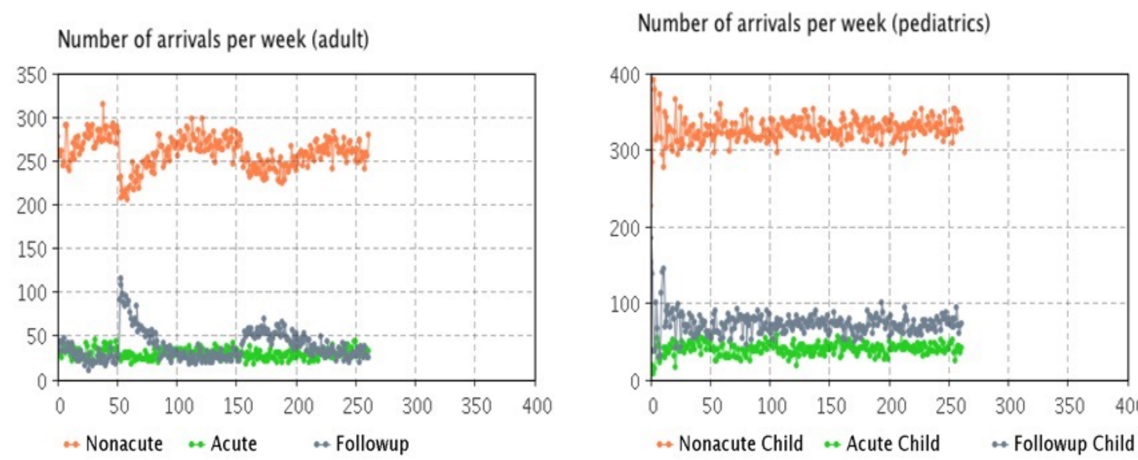

**Figure S.3.** Scheduling patient flow implemented in AnyLogic. This flow includes enrollment visit. All new patients are required to see enrollment specialist before seeing medical provider.




| ADULT | Waiting time (days) | | | |
|---|---|---|---|---|
| | Average | Std. dev. | Min | Max |
| Nonacute | 22 | 14 | 0 | 34 |
| Acute | 20 | 14 | 0 | 34 |
| Followup | 18 | 5 | 0 | 33 |

| PEDIATRICS | Waiting time (days) | | | |
|---|---|---|---|---|
| | Average | Std. dev. | Min | Max |
| Nonacute | 21 | 14 | 0 | 34 |
| Acute | 21 | 14 | 0 | 34 |
| Followup | 17 | 5 | 0 | 27 |

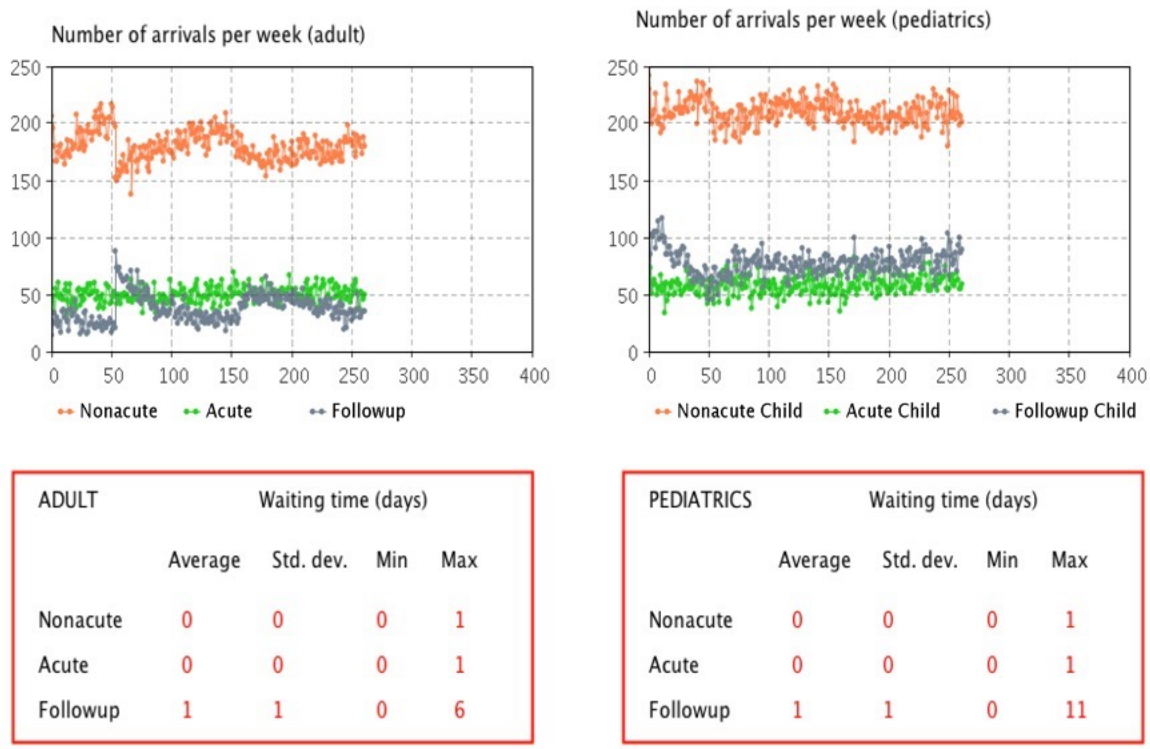
**Figure S.4.** Output of AnyLogic simulation model of traditional scheduling model for acute, non-acute, and follow-up for adult versus children for five years. Top left: number of arrivals per week for adults; top right: number of arrivals per week for children; bottom right: average, standard deviation, minimum, and maximum of waiting time to get an appointment for adult; bottom right: average, standard deviation, minimum, and maximum of waiting time to get an appointment for children.




| ADULT | Waiting time (days) | | | |
|---|---|---|---|---|
| | Average | Std. dev. | Min | Max |
| Nonacute | 19 | 11 | 0 | 29 |
| Acute | 17 | 12 | 0 | 29 |
| Followup | 13 | 5 | 0 | 38 |

| PEDIATRICS | Waiting time (days) | | | |
|---|---|---|---|---|
| | Average | Std. dev. | Min | Max |
| Nonacute | 15 | 10 | 0 | 29 |
| Acute | 15 | 10 | 0 | 29 |
| Followup | 9 | 5 | 0 | 19 |

**Figure S.5.** Output of AnyLogic simulation model of modified open access scheduling model for acute, non-acute, and follow-up for adult versus children for five years. Top left: number of arrivals per week for adults; top right: number of arrivals per week for children; bottom right: average, standard deviation, minimum, and maximum of waiting time to get an appointment for adult; bottom right: average, standard deviation, minimum, and maximum of waiting time to get an appointment for children.




| ADULT | Waiting time (days) | | | |
|---|---|---|---|---|
| | Average | Std. dev. | Min | Max |
| Nonacute | 0 | 0 | 0 | 1 |
| Acute | 0 | 0 | 0 | 1 |
| Followup | 1 | 1 | 0 | 6 |

| PEDIATRICS | Waiting time (days) | | | |
|---|---|---|---|---|
| | Average | Std. dev. | Min | Max |
| Nonacute | 0 | 0 | 0 | 1 |
| Acute | 0 | 0 | 0 | 1 |
| Followup | 1 | 1 | 0 | 11 |

**Figure S.6.** Output of AnyLogic simulation model of open access scheduling model for acute, non-acute, and follow-up for adult versus children for five years. Top left: number of arrivals per week for adults; top right: number of arrivals per week for children; bottom right: average, standard deviation, minimum, and maximum of waiting time to get an appointment for adult; bottom right: average, standard deviation, minimum, and maximum of waiting time to get an appointment for children.

## REFERENCES


1. Handler T, Overhage M, Taylor S, Underwood C. EHR definition, attributes and essential requirements. *Healthcare Information and Management Systems Society.* 2003;12.
2. Willison DJ. *Use of data from the electronic health record for health research: current governance challenges and potential approaches.* Office of the Privacy Commissioner of Canada; 2009.
3. Mohammadi I, Hashemikhabir S, Toscos T, Wu H. Health Care Needs of Underserved Populations in the City of Indianapolis. Paper presented at: Proceedings of the 7th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics2016.
4. Mohammadi I, Turkcan A, Toscos T, Wu H, Doebbeling B. Predictive Modeling for Appointment No-show in Community Health Centers Paper presented at: AMIA2016; Chicago, IL.
5. Mohammadi I, Turkcan A, Toscos T, Miller A, Kunjan K, Doebbeling BN. Assessing and Simulating Scheduling Processes in Community Health Centers. Paper presented at: AMIA; Nov. 14-18, 2014; San Francisco, CA.
6. Health Impact Assessment (HIA); The determinants of health. http://www.who.int/hia/evidence/doh/en/.
7. U.S. Census Bureau QuickFacts selected: Indiana. July 1, 2016; https://www.census.gov/quickfacts/table/PST045213/18033/embed/accessible.
8. Gusmano MK, Fairbrother G, Park H. Exploring the limits of the safety net: community health centers and care for the uninsured. *Health affairs.* 2002;21(6):188-194.
9. Shin P, Sharac J, Rosenbaum S. Community health centers and Medicaid at 50: an enduring relationship essential for health system transformation. *Health Affairs.* 2015;34(7):1096-1104.
10. Gabow P, Halvorson G, Kaplan G. Marshaling leadership for high-value health care: an Institute of Medicine discussion paper. *Jama.* 2012;308(3):239-240.
11. Institute of Medicine Committee on Quality of Health Care in A. *Crossing the Quality Chasm: A New Health System for the 21st Century*. Washington (DC)2001.
12. Quality AfHRa. *2014 National Healthcare Quality and Disparities Report.* Rockville, MD: U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES; May 2015 2014.
13. Health UDo, Services H. National healthcare disparities report 2011. *Rockville, MD: Agency for Healthcare Research and Quality.* 2012.
14. Blumenthal D, Abrams M, Nuzum R. The Affordable Care Act at 5 Years. *The New England journal of medicine.* 2015;372(25):2451-2458.
15. Aday LA, Andersen R. A framework for the study of access to medical care. *Health services research.* 1974;9(3):208-220.
16. Rhodes KV, Kenney GM, Friedman AB, et al. Primary care access for new patients on the eve of health care reform. *JAMA internal medicine.* 2014;174(6):861-869.
17. Doroudi R, Mohammadi I, Turkcan A, Toscos T, Wu H, Doebbeling B. Agent-based Simulation to Test Optimal Scheduling Scenarios and Improve Access to Care for Underserved Populations. Paper presented at: Institute of Industrial

Engineers Annual Conference and Expo; May 21 – 24, 2016, 2016; Anaheim, California.
18. Toscos T, Carpenter M, Flanagan M, Kunjan K, Doebbeling BN. Identifying Successful Practices to Overcome Access to Care Challenges in Community Health Centers: A "Positive Deviance" Approach. *Health services research and managerial epidemiology.* 2018;5:2333392817743406.
19. Levesque JF, Harris MF, Russell G. Patient-centred access to health care: conceptualising access at the interface of health systems and populations. *International journal for equity in health.* 2013;12:18.
20. Mayberry RM, Nicewander DA, Qin H, Ballard DJ. Improving quality and reducing inequities: a challenge in achieving best care. Paper presented at: Baylor University Medical Center Proceedings2006.
21. Slaveykov K, Trifonova K, Stoyanov V, Despotova-Toleva L. Electronic health records–benefits, savings and costs. *Medicine.* 2013;3(1).
22. McCarthy DB, Propp K, Cohen A, Sabharwal R, Schachter AA, Rein AL. Learning from health information exchange technical architecture and implementation in seven beacon communities. *EGEMS (Wash DC).* 2014;2(1):1060.
23. Dixon BE, Gibson PJ, Comer KF, Rosenman MB. Measuring Population Health Using Electronic Health Records: Exploring Biases and Representativeness in a Community Health Information Exchange. Paper presented at: MedInfo2015.
24. Decker SL, Kostova D, Kenney GM, Long SK. Health status, risk factors, and medical conditions among persons enrolled in Medicaid vs uninsured low-income adults potentially eligible for Medicaid under the Affordable Care Act. *Jama.* 2013;309(24):2579-2586.
25. Beal AC. Quality of Care for Underserved Populations.
26. Weinfeld JM, Davidson LW, Mohan V. Electronic health records improve the quality of care in underserved populations: a literature review. *Journal of health care for the poor and underserved.* 2012;23(3):136-153.
27. Brandenburg L, Gabow P, Steele G, Toussaint J, Tyson BJ. *Innovation and best practices in health care scheduling.* Technical report;2015.
28. Rajan B, Seidmann A. Improving Open Access Policy for Scheduling Outpatient Appointments. Paper presented at: System Sciences (HICSS), 2016 49th Hawaii International Conference on2016.
29. Cayirli T, Veral E, Rosen H. Designing appointment scheduling systems for ambulatory care services. *Health Care Manag Sci.* 2006;9(1):47-58.
30. Huang Y, Verduzco S. Appointment template redesign in a women's health clinic using clinical constraints to improve service quality and efficiency. *Applied clinical informatics.* 2015;6(2):271-287.
31. Degani N. Impact of advanced (open) access scheduling on patients with chronic diseases: an evidence-based analysis. *Ont Health Technol Assess Ser.* 2013;13(7):1-48.
32. Austin J, Manning-Courtney P, Johnson ML, et al. Improving access to care at autism treatment centers: a system analysis approach. *Pediatrics.* 2016;137(Supplement 2):S149-S157.
33. Kwong T. Patient Access: Improving Wait Times in a Specialty Clinic. *The Health Care Manager.* 2016;35(1):72-79.

34. Lynn S, Edlund BJ, Dumas BP. Open access scheduling: Improving access to rural healthcare. *Journal of Nursing Education and Practice.* 2016;6(9):p67.
35. Murray M, Tantau C. Same-day appointments: exploding the access paradigm. *Family practice management.* 2000;7(8):45-45.
36. Huang Y-L, Marcak J. Radiology scheduling with consideration of patient characteristics to improve patient access to care and medical resource utilization. *Health Systems.* 2013;2(2):93-102.
37. Huang Y-L. Appointment standardization evaluation in a primary care facility. *International journal of health care quality assurance.* 2016;29(6):675-686.
38. Huang Y-L, Hanauer DA. Time dependent patient no-show predictive modelling development. *International journal of health care quality assurance.* 2016;29(4):475-488.
39. Chowdhury GG. Natural language processing. *Annual review of information science and technology.* 2003;37(1):51-89.
40. Liddy ED. Natural language processing. 2001.
41. Liu F, Chen J, Jagannatha A, Yu H. Learning for biomedical information extraction: Methodological review of recent advances. *arXiv preprint arXiv:160607993.* 2016.
42. Wang Y, Liu S, Afzal N, et al. A comparison of word embeddings for the biomedical natural language processing. *Journal of biomedical informatics.* 2018.
43. Jagannatha A, Chen J, Yu H. Mining and ranking biomedical synonym candidates from Wikipedia. Paper presented at: Proceedings of the Sixth International Workshop on Health Text Mining and Information Analysis2015.
44. Liu S, Tang B, Chen Q, Wang X. Drug-drug interaction extraction via convolutional neural networks. *Computational and mathematical methods in medicine.* 2016;2016.
45. Xu J, Zhang Y, Xu H. Clinical abbreviation disambiguation using neural word embeddings. *Proceedings of BioNLP 15.* 2015:171-176.
46. Deerwester S, Dumais ST, Furnas GW, Landauer TK, Harshman R. Indexing by latent semantic analysis. *Journal of the American society for information science.* 1990;41(6):391-407.
47. Lund K, Burgess C. Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior research methods, instruments, & computers.* 1996;28(2):203-208.
48. Mikolov T, Yih W-t, Zweig G. Linguistic regularities in continuous space word representations. Paper presented at: Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies2013.
49. Wang Y, Rastegar-Mojarad M, Komandur-Elayavilli R, Liu S, Liu H. MayoNLPTeam at TREC 2016 clinical decision support track: an ensemble approach of clinical information extraction and retrieval. Paper presented at: Proceedings of the 2016 Text Retrieval Conference. Gaithersburg, Maryland, USA2016.
50. Liu H, Bielinski SJ, Sohn S, et al. An information extraction framework for cohort identification using electronic health records. *AMIA Summits on Translational Science Proceedings.* 2013;2013:149.

51. Pennington J, Socher R, Manning C. Glove: Global vectors for word representation. Paper presented at: Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)2014.
52. Müllner D. Modern hierarchical, agglomerative clustering algorithms. *arXiv preprint arXiv:11092378.* 2011.
53. Emanuel L, Berwick D, Conway J, et al. What exactly is patient safety? 2008.
54. Tantau C. Accessing patient-centered care using the advanced access model. *The Journal of ambulatory care management.* 2009;32(1):32-43.
55. Gupta D, Denton B. Appointment scheduling in health care: Challenges and opportunities. *IIE transactions.* 2008;40(9):800-819.
56. Penchansky R, Thomas JW. The concept of access: definition and relationship to consumer satisfaction. *Medical care.* 1981:127-140.
57. Corrigan JM. Crossing the quality chasm. *Building a better delivery system.* 2005.
58. Deyo RA, Inui TS. Dropouts and broken appointments. A literature review and agenda for future research. *Medical care.* 1980;18(11):1146-1157.
59. Turkcan A, Nuti L, DeLaurentis P-C, et al. No-show modeling for adult ambulatory clinics. *Handbook of Healthcare Operations Management*: Springer; 2013:251-288.
60. Molfenter T. Reducing appointment no-shows: going from theory to practice. *Substance use & misuse.* 2013;48(9):743-749.
61. Nguyen DL, Dejesus RS, Wieland ML. Missed appointments in resident continuity clinic: patient characteristics and health care outcomes. *Journal of graduate medical education.* 2011;3(3):350-355.
62. Hwang AS, Atlas SJ, Cronin P, et al. Appointment "no-shows" are an independent predictor of subsequent quality of care and resource utilization outcomes. *Journal of general internal medicine.* 2015;30(10):1426-1433.
63. Daggy J, Lawley M, Willis D, et al. Using no-show modeling to improve clinic performance. *Health informatics journal.* 2010;16(4):246-259.
64. Zinski A, Westfall AO, Gardner LI, et al. The contribution of missed clinic visits to disparities in HIV viral load outcomes. *Journal Information.* 2015;105(10).
65. Vulnerable populations: who are they? *The American journal of managed care.* 2006;12(13 Suppl):S348-352.
66. VanderWielen LM, Vanderbilt AA, Crossman SH, et al. Health disparities and underserved populations: a potential solution, medical school partnerships with free clinics to improve curriculum. *Medical education online.* 2015;20(1):27535.
67. DeNavas-Walt C. *Income, poverty, and health insurance coverage in the United States (2005).* Diane Publishing; 2010.
68. Torres O, Rothberg MB, Garb J, Ogunneye O, Onyema J, Higgins T. Risk factor model to predict a missed clinic appointment in an urban, academic, and underserved setting. *Population health management.* 2015;18(2):131-136.
69. Samorani M, LaGanga LR. Outpatient appointment scheduling given individual day-dependent no-show predictions. *European Journal of Operational Research.* 2015;240(1):245-257.
70. Huang Y, Zuniga P. Effective cancellation policy to reduce the negative impact of patient no-show. *Journal of the Operational Research Society.* 2014;65(5):605-615.

71. Huang Y, Hanauer D. Patient no-show predictive model development using multiple data sources for an effective overbooking approach. *Applied clinical informatics.* 2014;5(3):836-860.
72. Davies ML, Goffman RM, May JH, et al. Large-Scale No-Show Patterns and Distributions for Clinic Operational Research. Paper presented at: Healthcare2016.
73. Miller AJ, Chae E, Peterson E, Ko AB. Predictors of repeated "no-showing" to clinic appointments. *American journal of otolaryngology.* 2015;36(3):411-414.
74. Shameer K, Johnson KW, Yahi A, et al. Predictive modeling of hospital readmission rates using electronic medical record-wide machine learning: a case-study using Mount Sinai Heart Failure Cohort. Paper presented at: PACIFIC SYMPOSIUM ON BIOCOMPUTING 20172017.
75. Ottenbacher KJ, Smith PM, Illig SB, Linn RT, Fiedler RC, Granger CV. Comparison of logistic regression and neural networks to predict rehospitalization in patients with stroke. *Journal of clinical epidemiology.* 2001;54(11):1159-1165.
76. Dreiseitl S, Ohno-Machado L. Logistic regression and artificial neural network classification models: a methodology review. *Journal of biomedical informatics.* 2002;35(5-6):352-359.
77. Friedman N, Geiger D, Goldszmidt M. Bayesian network classifiers. *Machine learning.* 1997;29(2-3):131-163.
78. Ayer T, Chhatwal J, Alagoz O, Kahn Jr CE, Woods RW, Burnside ES. Comparison of logistic regression and artificial neural network models in breast cancer risk estimation. *Radiographics.* 2010;30(1):13-22.
79. Taylor LA, Tan AX, Coyle CE, et al. Leveraging the social determinants of health: what works? *PloS one.* 2016;11(8):e0160217.
80. Pickett KE, Pearl M. Multilevel analyses of neighbourhood socioeconomic context and health outcomes: a critical review. *Journal of Epidemiology & Community Health.* 2001;55(2):111-122.
81. Kaplan-Lewis E, Percac-Lima S. No-show to primary care appointments: why patients do not come. *Journal of primary care & community health.* 2013;4(4):251-255.
82. Blumenthal DM, Singal G, Mangla SS, Macklin EA, Chung DC. Predicting non-adherence with outpatient colonoscopy using a novel electronic tool that measures prior non-adherence. *Journal of general internal medicine.* 2015;30(6):724-731.
83. Liaw A, Wiener M. Classification and regression by randomForest. *R news.* 2002;2(3):18-22.
84. Health UDo, Services H. Agency for Healthcare Research and Quality.(2016). 2015 national healthcare quality and disparities report and 5th anniversary update on the national quality strategy. Rockville, MD: Agency for Healthcare Research and Quality. AHRQ Pub.
85. Dugdale DC, Epstein R, Pantilat SZ. Time and the patient–physician relationship. *Journal of general internal medicine.* 1999;14(S1):34-40.
86. Anderson K, Zheng B, Yoon SW, Khasawneh MT. An analysis of overlapping appointment scheduling model in an outpatient clinic. *Operations Research for Health Care.* 2015;4:5-14.
87. Kaandorp GC, Koole G. Optimal outpatient appointment scheduling. *Health Care Manag Sci.* 2007;10(3):217-229.

88. Murray M, Bodenheimer T, Rittenhouse D, Grumbach K. Improving timely access to primary care: case studies of the advanced access model. *Jama.* 2003;289(8):1042-1046.
89. Murray M, Berwick DM. Advanced access: reducing waiting and delays in primary care. *Jama.* 2003;289(8):1035-1040.
90. Canada CoFPo. Timely access to appointments in family practice. Mississauga, ON; 2012.
91. Turkcan A, Toscos T, Doebbeling BN. Patient-Centered Appointment Scheduling Using Agent-Based Simulation. Paper presented at: AMIA Annual Symposium Proceedings2014.
92. Sadiq SW, Marjanovic O, Orlowska ME. Managing change and time in dynamic workflow processes. *International Journal of Cooperative Information Systems.* 2000;9(01n02):93-116.
93. Law AM, Kelton WD, Kelton WD. *Simulation modeling and analysis.* Vol 2: McGraw-Hill New York; 1991.
94. Centeno MA. An introduction to simulation modeling. Paper presented at: Simulation Conference, 1996. Proceedings. Winter1996.
95. Harrison JR, Lin Z, Carroll GR, Carley KM. Simulation modeling in organizational and management research. *Academy of Management Review.* 2007;32(4):1229-1245.
96. Bhattacharjee P, Ray PK. Simulation modelling and analysis of appointment system performance for multiple classes of patients in a hospital: a case study. *Operations Research for Health Care.* 2016;8:71-84.
97. Almagooshi S. Simulation modelling in healthcare: Challenges and trends. *Procedia Manufacturing.* 2015;3:301-307.
98. Wu X-D, Khasawneh MT, Yue D-M, Chu Y-N, Gao Z-T. A simulation study of outpatient scheduling with multiple providers and a single device. *International Journal of Computational Intelligence Systems.* 2014;7(sup2):15-25.
99. Bard JF, Shu Z, Morrice DJ, Poursani R, Leykum L. Improving patient flow at a family health clinic. *Health care management science.* 2016;19(2):170-191.
100. Pan C, Zhang D, Kon AWM, Wai CSL, Ang WB. Patient flow improvement for an ophthalmic specialist outpatient clinic with aid of discrete event simulation and design of experiment. *Health care management science.* 2015;18(2):137-155.
101. Roure M, Halley Q, Augusto V. Modelling and simulation of an outpatient surgery unit. Paper presented at: Proceedings of the 2015 Winter Simulation Conference2015.
102. Zhong X, Williams M, Li J, Kraft SA, Sleeth JS. Primary care redesign: A simulation study at a pediatric clinic. Paper presented at: Automation Science and Engineering (CASE), 2014 IEEE International Conference on2014.
103. Liang B, Turkcan A, Ceyhan ME, Stuart K. Improvement of chemotherapy patient flow and scheduling in an outpatient oncology clinic. *International Journal of Production Research.* 2015;53(24):7177-7190.
104. Kougias P, Tiwari V, Berger DH. Use of simulation to assess a statistically driven surgical scheduling system. *Journal of Surgical Research.* 2016;201(2):306-312.

105. Ozen A, Marmor Y, Rohleder T, Balasubramanian H, Huddleston J, Huddleston P. Optimization and simulation of orthopedic spine surgery cases at Mayo Clinic. *Manufacturing & Service Operations Management.* 2015;18(1):157-175.
106. Maidstone R. Discrete event simulation, system dynamics and agent based simulation: Discussion and comparison. *System.* 2012:1-6.
107. Cain C, Haque S. Organizational workflow and its impact on work quality. 2008.
108. Maria A. Introduction to modeling and simulation. Paper presented at: Proceedings of the 29th conference on Winter simulation1997.
109. Ballermann MA, Shaw NT, Mayes DC, Gibney RT, Westbrook JI. Validation of the Work Observation Method By Activity Timing (WOMBAT) method of conducting time-motion observations in critical care settings: an observational study. *BMC medical informatics and decision making.* 2011;11:32.
110. Mohammadi I, Turkcan A, Toscos T, Wu H, Doebbeling BN. Predictive Modeling for Appointment No-show in Community Health Centers. Paper presented at: AMIA2016.
111. Hahn KA, Gonzalez MM, Etz RS, Crabtree BF. National Committee for Quality Assurance (NCQA) patient-centered medical home (PCMH) recognition is suboptimal even among innovative primary care practices. *Journal of the American Board of Family Medicine : JABFM.* 2014;27(3):312-313.
112. Bazghandi A. Techniques, advantages and problems of agent based modeling for traffic simulation. *International Journal of Computer Science Issues (IJCSI).* 2012;9(1):115.
113. Kaplan SH, Greenfield S, Ware Jr JE. Assessing the effects of physician-patient interactions on the outcomes of chronic disease. *Medical care.* 1989:S110-S127.
114. Control CfD, Prevention. CDC health disparities and inequalities report: United States. *http://www cdc gov/mmwr/pdf/other/su6001 pdf.* 2011.
115. Krug SE. Access and use of emergency services: inappropriate use versus unmet need. *Clinical Pediatric Emergency Medicine.* 1999;1(1):35-44.
116. Meyer PA, Yoon PW, Kaufmann RB, Centers for Disease C, Prevention. Introduction: CDC Health Disparities and Inequalities Report - United States, 2013. *MMWR Suppl.* 2013;62(3):3-5.
117. Paton C, Bamidis P, Eysenbach G, Hansen M, Cabrer M. Experience in the use of social media in medical and health education. *Yearbook of medical informatics.* 2011;20(01):21-29.

CURRICULUM VITAE

Iman Mohammadi

**Education**

- Indiana University-Indianapolis, Indianapolis, IN – PhD in Health Informatics
    - Dissertation: “Simulation and Modeling for Improving Access to Care for Underserved Populations” – December 2018
- Purdue University-Indianapolis, Indianapolis, IN – Master of Science in Biomedical Engineering
    - Research Project: “Design and Implementation of an Inexpensive Device for Biomedical Data Acquisition" – December 2012
- Azad University, Science and Research Branch, Tehran, Iran – Bachelor of Science in Biomedical Engineering
    - Final Project: “Design and Implementation of a guide system for blind people using webcams and ultrasound" – January 2009

**Research Experience and Projects**

- PhD Research. Improving Healthcare Systems – 2013 – 2018
    - Developing interviews, questionnaires and surveys to understand operations
    - Observing work and activities in healthcare systems
    - Electronic Health Record data extraction and database management (SQL)
    - Analyzing observations and Electronic Health Record data
    - Simulating and modeling workflow in healthcare systems
    - Developing no-show prediction models and statistical analysis of patients
    - Optimizing utilization of advanced practice providers in trauma center

- MS Research Projects. Medical Devices, Machine Learning, Pattern Recognition and Decision Making Processes – 2010 – 2013
    - Designing and Implementing biomedical data acquisition and amplifier circuit
    - A/D conversion of the biomedical signal and acquisition using LabVIEW
    - Pattern recognition – Developing algorithms for Handwritten Digit Recognition Using Neural Network
    - Biometrics Recognition – Developing Algorithms for Iris and Face recognition

**Healthcare Knowledge and Experience**

- Study and Research
    - Patient Centered Outcomes Research, Health Services Research, Access to Care, Quality of Care, Population Health, Clinical Research Methods, Implementation of Health System Redesign, Clinical Workflow
- Healthcare Systems
    - Community Health Centers, Specialty Outpatient Clinics, Hospitals and Trauma Centers
- Health Informatics
    - Healthcare Payment, Billing, Insurance (Medicaid, Medicare, Marketplace, Private and Self-pay), Electronic Health Records, Health IT and HIT Project Management, Clinical Decision Support Systems, Health Information Standards and Terminologies and Health Interoperability

**Data Science Skills**

- Machine Learning
  - Expert in simulation (Agent based simulation) and predictive modeling. Experienced in various supervised and unsupervised ML algorithms and their applications in healthcare and engineering domains domain (e.g. Clustering, Classification, Feature Selection, Correlation, Neural Network, PCA, LDA).
- Data Analysis Platforms
  - Expert in large data cleaning, processing and statistical analysis using SAS. Qualitative research included qualitative data analysis, interview analysis, qualitative coding and survey data analysis. Other engineering related data analysis work included image processing, ultrasonic distance detection and biomedical signal processing.

**Courses and Additional Training**

Pattern Recognition and Decision-Making Processes, Advanced Mathematics, Biostatistics, Advanced Statistics, Clinical Information Systems, Public Health Informatics, and Discrete Event Simulation.

**Software and Programming Skills**

- SAS (Professional), MATLAB (Professional), Java (Professional for data processing), Python (Professional), MSSQL (Familiar), AnyLogic (Professional), Tableau (Familiar), R (Familiar), NVivo (Familiar), LabVIEW (Professional), ANSYS (Familiar), COVENTOR (Professional)

**Work Experience**

- Data Scientist, Avalere Health. Washington, DC 01/18 – Present
    - Conducted high-level analytic modeling and statistical techniques
    - Developed solutions to key strategic business problems
    - Performed advanced-level data extraction, transformation and data management
    - Identified and resolved data issues to improve the efficiency, productivity and scalability databases
- Research Assistant and Teacher, Indiana University. Indianapolis, Indiana 08/13– 01/18
    - Helped writing grants, progress reports and papers
    - Created, updated and managed research group website (HTML)
    - Organized class sessions and graded students' assignments
    - Taught courses: Informatics Project management (Grad level), Health Informatics Standards and Terminologies (Undergrad and Grad level)
- Research Associate, Laboratory for Text information, Mining, Analysis and Prediction (TiMAP) - Indiana University. Indianapolis, Indiana 01/13– 05/2013
    - Helped writing grants and papers
    - Researched, studied and presented:
    - Biomedical data analytics
    - Text mining for unstructured data
    - Pattern discovery in longitudinal patient data

**Publications**

- Iman Mohammadi, Huanmei Wu, Tammy Toscos, Ayten Turkcan, Bradley Doebbeling. “Data Analytics and Modeling for Appointment No-show in Community Health Centers” Journal of Primary Care and Community Health (Accepted 2018)
- Amy Miller, Huanmei Wu, Iman Mohammadi, Kislaya Kunjan, Tammy Toscos, Ayten Turkcan, Mindy Flanagan, Bradley Doebbeling. “Improving Healthcare Systems for Access to Care and Efficiency by Underserved Patients”, PCORI Final Review Report (Accepted 2018)
- Iman Mohammadi, Seyedsasan Hashemikhabir, Tammy Toscos, Huanmei Wu. "Health Care Needs of Underserved Populations in the City of Indianapolis." In Proceedings of the 7th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics, pp. 512-513. ACM, 2016. Seattle, WA.
- Iman Mohammadi, Ayten Turkcan, Tammy Toscos, Huanmei Wu, Brad N. Doebbeling “Predictive Modeling for Appointment No-show in Community Health Centers”. AMIA Annual Symposium, Nov. 12-16, 2016, Chicago, IL.
- Rozhin Doroudi, Iman Mohammadi, Ayten Turkcan, Tammy Toscos, Huanmei Wu, Brad Doebbeling. “Agent-based Simulation to Test Optimal Scheduling Scenarios and Improve Access to Care for Underserved Populations”. Institute of Industrial Engineers Annual Conference and Expo. May 21 – 24, 2016. Anaheim, CA.

- Iman Mohammadi, Ayten Turkcan, Tammy Toscos, Amy Miller, Kislaya Kunjan, Brad N. Doebbeling "Assessing and Simulating Scheduling Processes in Community Health Centers". AMIA Annual Symposium, Nov. 14-18, 2015, San Francisco, CA.
- Mehrabi S, Mohammadi I, Kunjan K, Kharrazi H. "Effects of data transformation methods on classification of patients diagnosed with Myocardial Infarction". Medinfo, Copenhagen 2013.

## Honors and Awards

- Indiana Primary Health Care Association, Annual Conference Student Scholarship, 2017
- Indiana University, PhD studies Scholarship, 2013-2018