THE IMPACT OF HEALTHCARE PROVIDER COLLABORATIONS ON PATIENT OUTCOMES: A SOCIAL NETWORK ANALYSIS APPROACH

by

Mina Ostovari

A Dissertation

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Doctor of Philosophy

School of Industrial Engineering

West Lafayette, Indiana

May 2019

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Denny Yu, Chair

School of Industrial Engineering

Dr. Paul Griffin

School of Industrial Engineering

Dr. Sandra Liu

College of Health and Human Science

Dr. Yuehwern Yih

School of Industrial Engineering

Approved by:

Dr. Steve Landry

Head of the Graduate Program

To My Parents, Nasrin and Jahan, may he rest in peace

ACKNOWLEDGMENTS

I would like to thank my advisor, Dr. Denny Yu, for his continuous help and support. Great appreciation for my wonderful committee, Dr. Paul Griffin, Dr. Sandra Liu and Dr. Yuehwern Yih. I am so grateful for their support, patience and understanding. A special thanks to everyone at the Regenstrief Center for Healthcare Engineering. Without their help and support, this dissertation would not be possible. I am also grateful for all the staff in the School of Industrial Engineering, specially Leza, Anita, Steve, and Cheryl. They have been my second family, always there, always ready to help.

Special thanks to my dear friends; without them, I could not do this. Last not least, my dear family. I am so grateful for all the sacrifices they made. I lost my dad in the last year of PhD to cancer. The pain of not seeing him, not being able to say goodbye killed me every second but at same time helped me move forward as I knew he was with me in every step of the way. My aunts Simin and Homa, who always supported me, without any expectations; without them I would have not been here. And finally, my mother Nasrin, there is no limit to sacrifices she made for me. I was not able to go and visit my family due to specific travel policies in the U.S. for Iranians during the past four years. My mother always supported me, encouraged me to move forward. She knew how important this journey was for me. Thank you Nasrin; this is for you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PUBLICATIONS

Refereed Journal Papers

1. Ostovari M., Yu D., Yih Y., Steele-Morris C.J. Impact of an onsite clinic on utilization of preventive services. *Journal of Environmental and Occupational Medicine*. 2017; 59 (7): 615-623. doi: 10.1097/JOM.0000000000001034.
2. Ostovari M., Steele-Morris C.J. Griffin, P., Yu D. (in press). Data-driven modeling of diabetes care teams using social network analysis. *Journal of the American Medical Informatics Association*. 2019
3. Ostovari M., Yu D. Impact of care provider network characteristics on patient outcomes: Usage of social network analysis and a multi-scale community detection. Submitted to the *BMJ Quality & Safety*.

Refereed Conference Papers

1. Ostovari M., Yu D., Xie S., Ye Q., Katare B., Adibuzzaman M., Musselman K. J., Nateghi R., Shield C.G., Yih Y. Bridging the gap between population needs and barriers into onsite clinic use. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. 2016; 60 (1):1809-1812. doi: 10.1177/1541931213601413.
2. Ostovari M., Yu D., Steele-Morris C. J. Identifying key players in the care process of patients with diabetes using social network analysis and administrative data. *AMIA Annual Symposium Proceedings*, vol. 2018, pp. 1435–1441, Dec. 2018.

ABSTRACT

Author: Ostovari, Mina. PhD
Institution: Purdue University
Degree Received: May 2019
Title: The impact of healthcare provider collaborations on patient outcomes: a social network analysis approach
Major Professor: Denny Yu

Care of patients with chronic conditions is complicated and usually includes large number of healthcare providers. Understanding the team structure and networks of healthcare providers help to make informed decisions for health policy makers and design of wellness programs by identifying the influencers in the network. This work presents a novel approach to assess the collaboration of healthcare providers involved in the care of patients with chronic conditions and the impact on patient outcomes.

In the first study, we assessed a patient population needs, preventive service utilization, and impact of an onsite clinic as an intervention on preventive service utilization patterns over a three-year period. Classification models were developed to identify groups of patients with similar characteristics and healthcare utilization. Logistic regression models identified patient factors that impacted their utilization of preventive health services in the onsite clinic vs. other providers. Females had higher utilizations compared to males. Type of insurance coverages, and presence of diabetes/hypertension were significant factors that impacted utilization. The first study framework helps to understand the patient population characteristics and role of specific providers (onsite clinic), however, it does not provide information about the teams of healthcare providers involved in the care process.

Considering the high prevalence of diabetes in the patient cohort of study 1, in the second study, we followed the patient cohort with diabetes from study 1 and extracted their healthcare providers over a two-year period. A framework based on the social network analysis was presented to assess the healthcare providers' networks and teams involved in the care of diabetes. The relations between healthcare providers were generated based on the patient sharing relations identified from the claims data. A multi-scale community detection algorithm was used to identify groups of healthcare providers more closely working together. Centrality measures of the social network identified the influencers in the overall network and each community. Mail-

order and retail pharmacies were identified as central providers in the overall network and majority of communities. This study presented metrics and approach for assessment of provider collaboration. To study how these collaborative relations impact the patients, in the last study, we presented a framework to assess impacts of healthcare provider collaboration on patient outcomes.

We focused on patients with diabetes, hypertension, and hyperlipidemia due to their similar healthcare needs and utilization. Similar to the second study, social network analysis and a multi-scale community detection algorithm were used to identify networks and communities of healthcare providers. We identified providers who were the majority source of care for patients over a three-year period. Regression models using generalized estimating equations were developed to assess the impact of majority source of care provider community-level centrality on patient outcomes. Higher connectedness (higher degree centrality) and higher access (higher closeness centrality) of the majority source of care provider were associated with reduced number of inpatient hospitalization and emergency department visits.

This research proposed a framework based on the social network analysis that provides metrics for assessment of care team relations using large-scale health data. These metrics help implementation experts to identify influencers in the network for better design of care intervention programs. The framework is also useful for health services researchers to assess impact of care teams' relations on patient outcomes.

1. INTRODUCTION

Current statistics show six in ten Americans have one chronic condition and four in ten have two or more [1]. Diabetes by itself impacts 1.5 million Americans each year [2]. Cardiovascular disease as another common chronic condition, is a major cause of health disparities and rising healthcare costs [3]. The co-occurrence of multiple chronic conditions (multi-morbidity) negatively impacts healthcare delivery quality and cost [4]. Most chronic conditions are caused by a series of risk behaviors including tobacco and excessive alcohol usage, lack of physical exercise, and poor nutrition [1]. The burden of chronic conditions has led to various efforts to initiate behavioral changes for the patients [5] and addressing rising healthcare costs by providing incentive for patients and providers [6]. Despite these efforts, multi-morbidity still remains a challenge and much work is needed to provide better interventions to address multi-morbidity and the associated patient outcomes and costs [7].

Chronic conditions negatively impact the healthcare system and the overall U.S. economy. Heart disease and stroke cause an annual cost of $126 billion in lost productivity. Obesity imposes $147 billion on U.S. healthcare system annually [8]. After hospital care with 33 percent, the highest percentage of healthcare costs comes from physicians and clinical services by 20 percent of the annual U.S. healthcare spending [9]. A significant part of chronic condition burden falls on the employers providing insurance coverages to their employees. Besides direct costs, employers have to deal with costs associated with productivity loss due to presenteeism and absenteeism of their employees [10]. The costs associated with low performance due to chronic disease problems have shown to exceed the medical treatment costs and absenteeism for employers [11].

To control increasing healthcare costs, employers have implemented various strategies including health promotion programs and onsite clinics [12]. Barriers exist to properly design and deliver services in these clinics including employees' lack of interest, and conflicting time schedule [13]. Moreover, despite the growing interest toward using onsite clinics, there is a lack of understanding about their effectiveness and return on investment for the employers [14]. In addition to these challenges, as these clinics were originally designed to provide occupational services [15], it is unclear how their other services may impact patient's wellness [16]. Another issue about these clinics is lack of understanding about the interaction of clinicians with other healthcare providers in the community and how they can provide optimized care for patients [17].

Studying the interaction of healthcare providers and associated teamwork is challenging. Most studies that addressed this issue are limited in scope to surveys and interviews [18], [19]. Although these studies are helpful to understand individual providers' perspectives, they typically include a costly design and dissemination, and have a low response rate [20]. Studying interactions of healthcare providers is essential to understand their working relations, referrals, hospital association, advice seeking, and how these relations impact their decision making process and patient outcomes [21]. New approaches are needed to better capture these relationships, the associated impact on providers' teamwork, and associated patient outcomes.

With the abundance of large-scale health data, application of quantitative approaches to these datasets has been increasing, however, most studies focus on individual patient's healthcare utilization [22],[23],[24]. To explore relations of healthcare providers from the health data, researchers have used the social network analysis which expands the scope of previous work by providing an approach and metrics to assess collaboration of healthcare providers [21],[25]. Originated from the social psychology [26] and based on the work of Moreno [27], social network analysis explores the collaboration and communication channels between the network actors (e.g. individuals, groups, organizations). Providers who have one or more patients in common form teams of healthcare providers with formal referral relations or casual discussions about the patients [21]. Researchers have used social network analysis application on large-scale claims data to assess the structure of care teams [28], [29], and the providers network characteristics impact on patient costs and outcomes [30],[31]. Previous studies that used social network analysis and patient sharing approach have limited their providers cohort to physicians and usually those associated with hospitals [32],[33],[31]. Depending on the complexity of the disease, inclusion of other providers such as pharmacists and nurses are essential to explore the information flow and activities inside the network [34].

Despite previous efforts to assess the provider role, collaboration, and impact on patient outcomes, some gaps remained to be addressed. Considering the increasing role of wellness promotion programs and worksite clinics, more effort is needed to understand where these services and providers stand in the care process (addressed in study 1). Besides the role of individual providers, data and studies about the providers involved in the care of patients with chronic conditions are limited and lacks a system-level comprehensive picture (to be addressed by study 2 & 3). Finally, the impact of the provider system level collaboration on patient outcomes requires further investigation (study 3).

The purpose of this work is to develop a novel approach to better understand patients with chronic conditions needs, provider teams involved in the care process, providers' interaction and the associated patient outcomes. In the first study, we present a framework based on large scale health claims data and machine learning techniques to understand chronic conditions preventive healthcare utilization, impact of onsite clinics on changing patterns of preventive health utilization, and patient factors associated with their decision making for utilizing preventive services. In the second study, we introduce a framework based on the social network analysis metrics and a multi-scale community detection algorithm to assess all healthcare providers involved in the care of patients with diabetes and their interactions. In the final chapter, we expand the work in chapter 3 by looking at the longitudinal network of healthcare providers, their characteristics, and associated patient outcomes. Social network analysis and community detection algorithms were used to identify providers' network and community level centrality. Generalized estimating equations models were developed to assess relations of care providers' centrality in the community and patient rates of emergency department visits, unplanned, and inpatient hospitalization. The studies in chapter 2, 3, and 4 will provide a more focused and in depth background for each topic.

The long-term goal of this research is to provide transferable metrics and measures for continuous assessment of the healthcare provider collaboration, their interaction and how those relations impact patient care process and outcomes.

2. PREVENTIVE HEALTH UTILIZATION AND ASSOCIATED FACTORS FOR PATIENTS WITH CHRONIC CONDITIONS[1]

2.1 Introduction

According to the Center for Disease Control and Prevention (CDC) six in ten Americans suffer from one chronic condition. Chronic conditions are the leading cause of death, disability and 3.3 Trillion in annual U.S. healthcare costs [1]. Due to the increasing trend of chronic conditions, the associated negative outcomes on the workforce and the increasing costs for the employers [35], companies are searching for ways to promote risk management and health promotion programs to reduce their employees' healthcare costs [36]. One of the approaches to achieve this goal is utilization of onsite wellness programs and onsite (worksite) clinics [37]. Increasing prevalence of chronic conditions in recent years has expanded scope of services offered by onsite clinics to include other services such as primary care and condition management [14]. The increased focus on preventive services and disease management programs in the workplace has been suggested to help control rising healthcare costs.

Onsite clinics are defined as "a setting where an employer offers one or more medical and wellness services, delivered by licensed providers to all or a designated portion of its active population and other eligible individuals"[38]. Providing care directly in the worksite started around 1800s, when railroad and mining companies began to provide health services to employees on production site to better manage workplace injuries and improve productivity. Healthcare services provided by onsite clinics were mostly limited to treatment of workplace injuries and acute care until the 1980s [15].

Onsite clinics could provide solutions for rising healthcare costs by providing cost-effective services in the workplace compared to offsite health providers. McCaskill et al. (2014), assessed impact of an onsite clinic in a self-insured university on healthcare utilization and healthcare costs for patients with upper respiratory tract infections and found that implementation of the onsite clinic caused a 10% increase in number of healthcare visits and also helped to save

[1] This chapter is based on: Ostovari M., Yu D., Yih Y., Steele-Morris C.J. Impact of an onsite clinic on utilization of preventive services. Journal of Environmental and Occupational Medicine. 2017; 59 (7): 615-623. doi: 10.1097/JOM.0000000000001034.

$126,619 for the university [39]. Other studies similarily observed lower costs for the employers in the period after implementation of their onsite clinics [40], [41].

Despite positive impacts from implementation of onsite clinics such as cost savings for the employers, studies have acknowledged barriers to usage and participation of employees in the wellness programs offered in the workplace such as work schedule conflicts or programs being perceived as too time consuming [42]. Bright et.al surveyed employees' attitude toward usage of a worksite health and wellness clinic in a self-insured university and found that the most common barriers to usage of services were lack of motivation, work schedule, and being too busy at work [13]. Research about wellness programs can provide insight that might be translatable to onsite clinics; however, there is a lack of understanding about motivations and barriers to usage for other services provided by onsite clinics. Moreover, understanding barriers toward usage of worksite health services, such as time pressure or working schedule, is mostly limited to cross sectional surveys that determine exposure and outcome simultaneously, impeding interpretation of the result [43], and also are limited by low response rate [20]. Recent increasing interest among employers for providing onsite health services [44] calls for more rigorous tools and frameworks to assess onsite clinic performance and addressing gaps in services.

This paper proposes a framework adapted from the Andersen Model of Healthcare Utilization, and demonstrates it to quantitatively investigate the effectiveness of an onsite clinic on improving utilization [45] of preventive services (Figure 2.1). Original model of healthcare utilization by Andersen (1968) was presented to measure usage of health services by families as a measure of analysis based on three types of factors: predisposing factors which exist before the illness occurs including age and race, enabling factors which can promote or restrain utilization such as access to health insurance, and needs which are reasons leading to usage of services.

Patient population factors are grouped to predisposing, enabling, and needs similar to the Andersen Model. An interaction has been considered between community providers, and the onsite clinic. While patient factors impact utilization of services in both the onsite clinic and offsite providers, the onsite clinic may also impact utilization of offsite providers by either referring patients or fulfilling the needs in the workplace so there would be no need for patients to seek care among offsite providers. Customer satisfaction has also been considered in the model to impact both utilization of onsite clinic and offsite providers.

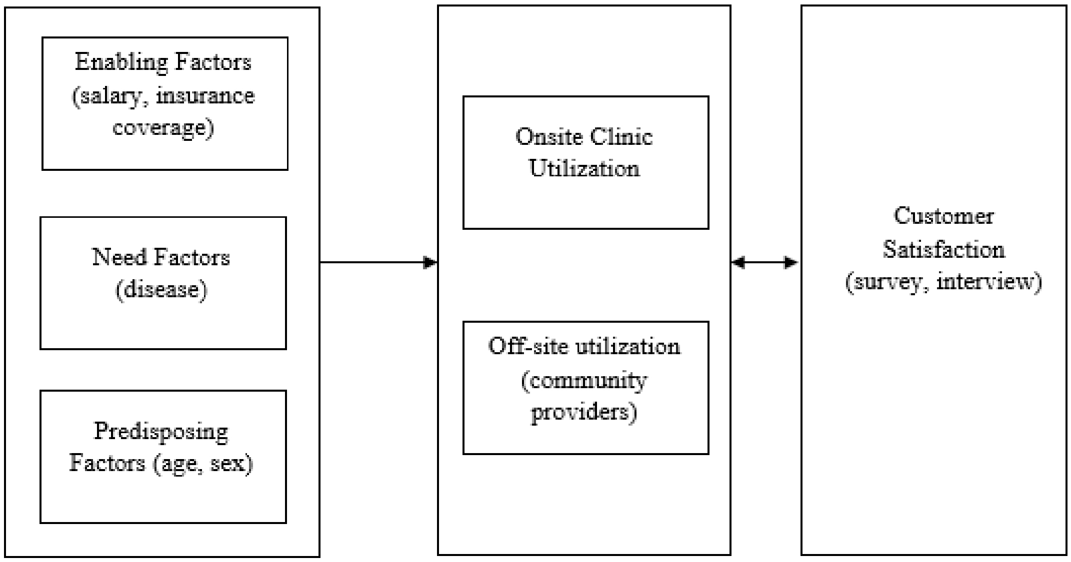


Figure 2.1 Health informatics framework for assessment of onsite clinics effectiveness

Focus of the study is on preventive care as the most common type of services offered in onsite clinics by utilizing health data to identify factors impacting usage and evaluate health status of the population and its impact on utilization of services. Future steps would leverage usage of health data by adding subjective customer assessment. The specific aims of this study are:

- Aim 1: Determine the impact of the onsite clinic on overall health services utilization before and after its implementation.
    - H1.1 (Null): Utilization of preventive services in the population does not change pre- and post- implementation of the clinic
- Aim 2: Determine patient factors that influence the onsite clinic usage
    - H2.1 (Null): Patient factors, e.g., age, sex, salary, and chronic conditions (diabetes, hypertension, and hyperlipidemia) do not impact usage of onsite clinic preventive services.

The long-term goal for this study is to develop continuous evaluations of onsite clinics performance based on the proposed framework to modify and improve services offered to the designated population aligning with changes in the population health needs using the tools and data sources suggested in the framework.

2.2 Materials and Methods

2.2.1 Study Population

This research was approved by the institutional review board (IRB). The population of this study included employees of a large public university (faculty and staff) and their dependents (partner\spouse, child\other dependents) who were continuously eligible for three consecutive years from the year 2012 that the university onsite clinic was implemented (Year1) until the year 2014 after the onsite clinic started providing service to employees and their dependents (Year3). This study excluded the student population who received health services at a student-specific health center.

2.2.2 University Onsite Clinic Description

No previous onsite clinic existed for this population. The onsite clinic had 22 staff members and provided treatments of common illnesses, such as colds and allergies, primary care and wellness, condition management and lab services to university employees and their dependents.

2.2.3 Study Design and Data Analysis

A retrospective study of fully de-identified health claim data for the study population was conducted for Year1 and Year3. The claims data included patient demographic information (e.g. sex, age), insurance coverage information, health conditions, healthcare providers, and healthcare costs. Inclusion criteria for the study was individuals who had used preventive services in Year1 and Year3. This subset was identified based on the principal diagnosis variables coded according to the International Classification of Diseases 9th edition (ICD9). ICD9 codes for preventive services were extracted from publicly available lists [46], [47]. Non-preventive services were excluded from the analysis. Data was not analyzed for the year clinic was implemented as this year was considered the learning/start-up period for both the clinic and the employee population.

2.2.4 Outcome Variables

The primary outcome variable was whether an individual used the onsite clinic in Year3. An indicator variable was defined for each individual with value of 1 if the individual had used the onsite clinic in Year3 at least once and value of 0 if the individuals had not used the onsite clinic in Year3. The secondary outcome variable, health services, was defined by ICD9 codes.

2.2.5 Independent Variables

Independent variables utilized from the dataset were age, sex (male/female), compensation type, and insurance plan. Age was defined as a categorical variable with categories 18-44, 45-64, and over 65 for employees and spouse, and less than 18, and more than 18 for the child group. The compensation variable was the employee status as hourly paid or salary-based. The type of compensation for the employee was also used for their family members meaning if the employee was categorized as hourly the family members were also categorized as hourly and similarly for salary based individuals. Three other independent variables were defined based on the information from the medical file: having diabetes, hypertension, or hyperlipidemia. Each variable had value of 1 for individuals who had been diagnosed with the condition in the three-year period and value of 0 for individuals who had not been diagnosed with the condition. Hypertension (ICD9=401.1 and 401.9), hyperlipidemia (ICD9=272.4), and diabetes (ICD9=250.00) were chosen as they were the three most common chronic diseases in the population under study.

2.2.6 Statistical Analysis

Descriptive statistics was used to describe usage of preventive services in the population. Significance of changes in the preventative care usage was assessed using McNemar’s test. Usage of preventive services between males and females was compared using effect size and Cohen’s d. Preventive visits were identified based on claim ID, place of service, and date of service; these were compared among the employees, spouse, and child groups. Visits were considered as both a preventive encounter (receiving a flu shot) and a preventive visit to a healthcare provider.

To determine population factors impacting usage of the onsite clinic, two statistical approaches were used: logistic regression and the conditional inference tree to assess impact of age, sex, compensation type, plan type, and having hyperlipidemia/hypertension/diabetes on usage of the onsite clinic. Three logistic regression models were developed separately for employees, spouse and child groups. The stepwise selection method was used to determine the best set of independent variables for each model. To further investigate significant factors impacting usage of the onsite clinic, classification trees were developed for employees, spouse and child groups using conditional inference trees (ctree package in R) to discover interdependencies between patient factors and their impact on the onsite clinic usage. For generating classification trees, observations with missing values were removed from the dataset. Output of these analyses was three classification trees for employees, spouse, and child showing significant factors as nodes and interdependencies as links between the nodes. Nodes identified groups of individuals that were significantly different from other nodes. The study data was analyzed using SAS (v 9.4, SAS Inc., Cary, NC), RStudio (version 0.99.903) and Microsoft Access (v 2013, Microsoft Corp, Redmond, WA).

2.3 Result

2.3.1 Population Demographic

The identified cohort (n=23,635) consisted of 53% females, with average age of 37±19, and 47% males with average age 36±20 years old. Forty-eight percent of the population were employees, 20% partner/spouse, and 32% child/other dependent. Percentage of individuals who had used preventive services in Year1 was 58.9% (n=13,932) and in Year3 this percentage was 64.2% (n=15,179). In Year1, 65% (n=8,091) of females in the cohort used preventive service vs. 69% (n=8,679) in Year3. For the male population, 52% (n=5,841) used preventive services vs. 58% (n=6,500) in Year3. The most common chronic conditions for the population were hypertension, hyperlipidemia, and diabetes with 23%, 18% and 9% of the cohort respectively.

2.3.2 Comparison of Services Offered Onsite vs. Offsite

Table 2.1 describes the top 10 preventive care diagnoses codes in the university onsite clinic and offsite health providers. The most common diagnosis was Influenza vaccinations (ICD9 code V0481) in the onsite clinic and Routine General Medical (ICD9 code V700) exam in the offsite providers. Among the top 10 conditions listed, six diagnoses were common between the onsite clinic and offsite providers: Routine General Medical Exam, Flu Vaccination, Routine Gynecological Examination, Unspecified Hypertension, Diabetes without Complication, and Diphtheria, Tetanus, Pertussis Vaccination. Comparisons of the selected diagnoses (Table 2.1) shows that onsite clinic services were used to address more than 80% of all observed diagnosis for Lifestyle related Problems(ICD9 code V699), Administrative Physicals (ICD9 code V703), and Counseling (ICD9 code V6549) for this population. Offsite services were used to address more than 80% of all diagnosis for Routine Gynecological Examination, Routine Child Health Exam, Hyperlipidemia, Hypertension, Screening for Malignant Neoplasm of Cervix, and Diabetes without Complication.

Table 2.1 Top preventative care diagnosis codes for onsite clinic and offsite providers based on number of patients (Year3)

Onsite Clinic				Offsite Providers			
Diagnosis(ICD9)	**# of individuals (% of total individuals with preventive claims)**	**# of claims for this diagnosis (% of total onsite preventive claims)**	**% of total claims for this diagnosis**	**Diagnosis (ICD9)**	**# of individuals (% of total individuals with preventive claims)**	**# of claims for this diagnosis (% of total offsite preventive claims)**	**% of total claims for this diagnosis**
***Influenza Vaccination (V04.81)**	2835(18%)	2859(34%)	42.5%	***Routine General Medical Examination (V70.0)**	4266(28%)	6086(14%)	87.2%
***Diphtheria, Tetanus, Pertussis Vaccination (V06.1)**	475(3%)	479(5.7%)	31.2%	***Influenza Vaccination (V04.81)**	3668(24%)	3865(8.8%)	57.5%
***Routine General Medical Examination (V70.0)**	469(3%)	895(10.7%)	12.8%	***Routine Gynecological Examination (V72.31)**	2925(19%)	3852(8.8%)	90.2%

Individuals with preventive services in Year 3 (n=15,179), Total onsite clinic claims= 8332, Total offsite providers claim=43708, *Common diagnosis between onsite clinic and offsite providers

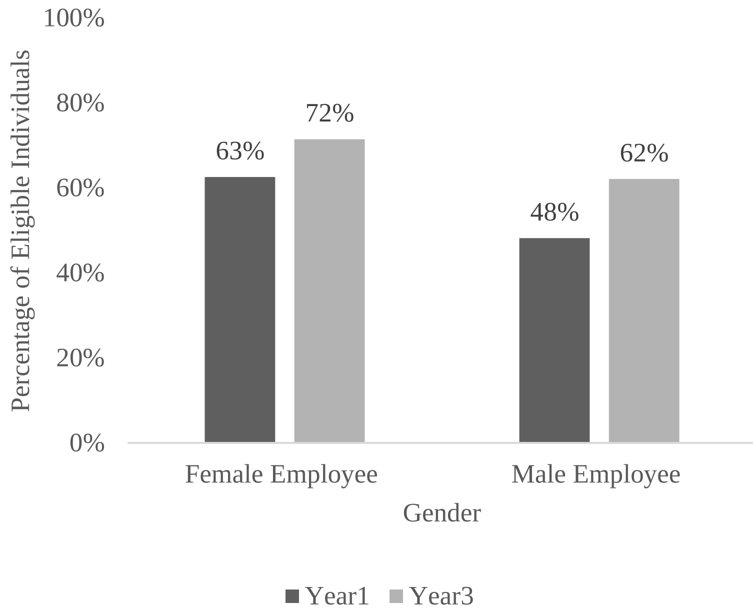
*Percentage of total claims for this diagnosis formula: $\frac{ICD9_{onsite}}{ICD9_{onsite}+ICD9_{offsite}}$

The number of preventive services visits for individuals who had used preventive services in Year3 (n=15,179) was quantified for a) employees, b) spouse, and c) child to describe the service utilization for the study population. Twenty-four percent of employees who utilized preventative services only used it once a year. For the spouse group, the percentage of people with four visits or more was observed the most at 52%. For the child group, the percentage of one visit was the highest (35%).

2.3.3 Impact of the Onsite Clinic on Health Services Utilization before and after Implementation

Potential impact of the onsite clinic on the preventive care utilization was examined by comparing preventative care utilization before and after implementation of the clinic. Figure 2.2 a, b, and c compare the utilization of preventive services for employees, spouse and child between Year1 and Year3. Usage of preventive services increased significantly after the onsite clinic implementation (McNemar's test p-value<0.0001) for female (n=6,017) and male employees (n=5,312), by 9% and 14% respectively. For the spouse group, the usage between Year1 and Year3 increased significantly (McNemar's test p-value p<0.0001) only for males (+3%). For the child groups preventative care utilization only differed significantly for the males, where usage decreased 3% in Year3. Changes in the percentage of preventive service usage was not different for female and male employees (Cohen's d 95% confidence interval (-0.003, 0.053)).

A)



B)

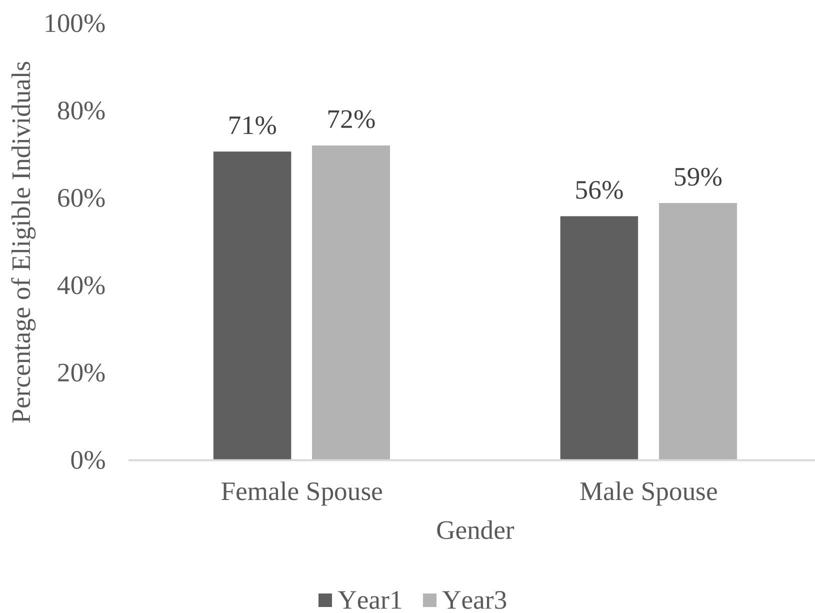
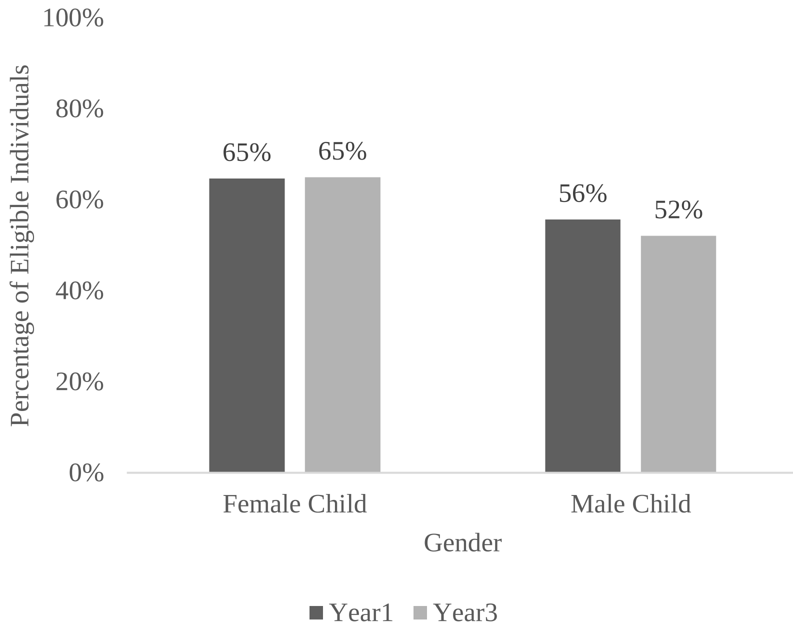


Figure 2.2 Usage of preventive services for Year1 and Year3 for a) employees b) spouse c) child

Figure 2.2 Continued

C)



2.3.4 Regression Analysis

Three multiple logistic regression models were developed for employees, spouse, and child groups (Table 2.2). For the employee group, odds of using onsite clinic preventive services for hourly-based employees was 1.13 higher compared to salary based. Odds of using onsite clinic services for individuals classified as having no coverage, or spouse Opt Out was 29 times higher compared to individuals with high deductible plans; however, individuals with low deductible plans had a lower probability for using the onsite clinic preventive service compared to high deductible individuals with odd ratio of 0.84. Interaction was significant between hypertension and hyperlipidemia factors.

Table 2.2 Regression models results for identifying factors impacting usage of onsite clinic

Regression Model for Employees

Variables		Number of individuals	Odds Ratios	95% Confidence Limits	P-Value
Compensation Type					
Salaried		4574	0.89	(0.80,0.98)	0.023
Hourly		2641	ref*	-	-
Diabetes					
Diabetes=1 (diagnosed)		831	0.84	(0.72,0.98)	0.031
Diabetes=0 (not diagnosed)		6384	ref*	-	-
Plan					
Plan a (no coverage, spouse opt out)		195	29.54	(13.85,63.03)	<0.0001
Plan b (low deductible)		1879	0.8	(0.719,0.9)	0.0001
Plan c (High deductible)		5141	ref*	-	-
Hyperlipidemia*Hypertension		-	-	-	0.0077
Hypertension=0 (not diagnosed)	Hyperlipidemia=1	921	0.73	(0.62,0.84)	
	Hyperlipidemia=0	4119	ref*	-	
Hyperlipidemia=1(diagnosed)	Hypertension=1	868	0.817	(0.67,0.99)	
	Hypertension=0	921	ref*	-	
Hyperlipidemia=0 (not diagnosed)	Hypertension=1	1307	0.59	(0.52,0.67)	
	Hypertension=0	4119	ref*	-	

Regression Model for Spouse

Variables	Number of Individuals	Odd Ratio	95% Confidence Limits	P-Value
[1]**Diabetes**				
Diabetes=1(diagnosed)	425	0.7	(0.5,0.97)	0.0345
Diabetes=0 (not diagnosed)	2579	ref	-	-
Hypertension				
Hypertension=1 (diagnosed)	1073	0.67	(0.53,0.83)	0.0004
Hypertension=0 (not diagnosed)	1931	ref	-	-

Regression Model for Child

Variables	Number of Individuals	Odds Ratio	95% Confidence Limits	P-Value
Compensation				
Salaried	3193	0.58	(0.43,0.78)	
Hourly	1243	ref	-	-
Age Group				
0<=Age<=17	3165	ref	-	-
Age>=18	1271	3.80	(2.81,5.1)	<0.0001
***Ref: Reference**				

In the spouse group, for individuals diagnosed with diabetes odds of using onsite clinic preventive services was 0.7 compared to individuals without diabetes. For individuals diagnosed with hypertension, odds of using preventive services in the onsite clinic was 0.67 compared to individuals not diagnosed with hypertension. Based on the logistic regression developed for the child group, the odds of using preventive services in the onsite clinics for individuals over 18 was higher than individuals under 18. Individuals who were classified as hourly (based on the compensation type for employee policy holder) had a higher probability of using the preventive services in the onsite clinic with an odd ratio of 3.8. No interaction term was significant.

2.3.5 Classification Tree

Classification trees were developed for employees, spouse, and child groups as a non-parametric approach to determine factors that impact onsite clinic use. Figure 2.3 shows the structure of the classification tree for the university employees. The tree started from plan as the most significant factor on usage of the preventive services in the onsite clinic- and divided to two branches based on type of coverage. Of all employees with plan type a (no coverage or spouse opt out) who had used preventive services, 96% utilized preventive services offered in the onsite clinic in contrast to 4% who used preventive services offered in offsite health providers. The second most significant factor on the usage of preventive services in the onsite clinic was hypertension. Among male employees who had used preventive services and also were diagnosed with hypertension, 31% has used preventive services offered in the onsite clinic. For the similar female employee group, the percentage was 36. The percentage of female employees who used preventive services in the onsite clinic was higher in identified classes based on gender in the tree except for males without diabetes, 53% of which used preventive services offered in the onsite clinic. Among the employees diagnosed with hyperlipidemia who had used preventive services, 45% of female employees and 33% of male employees had used preventive services offered in the onsite clinic.

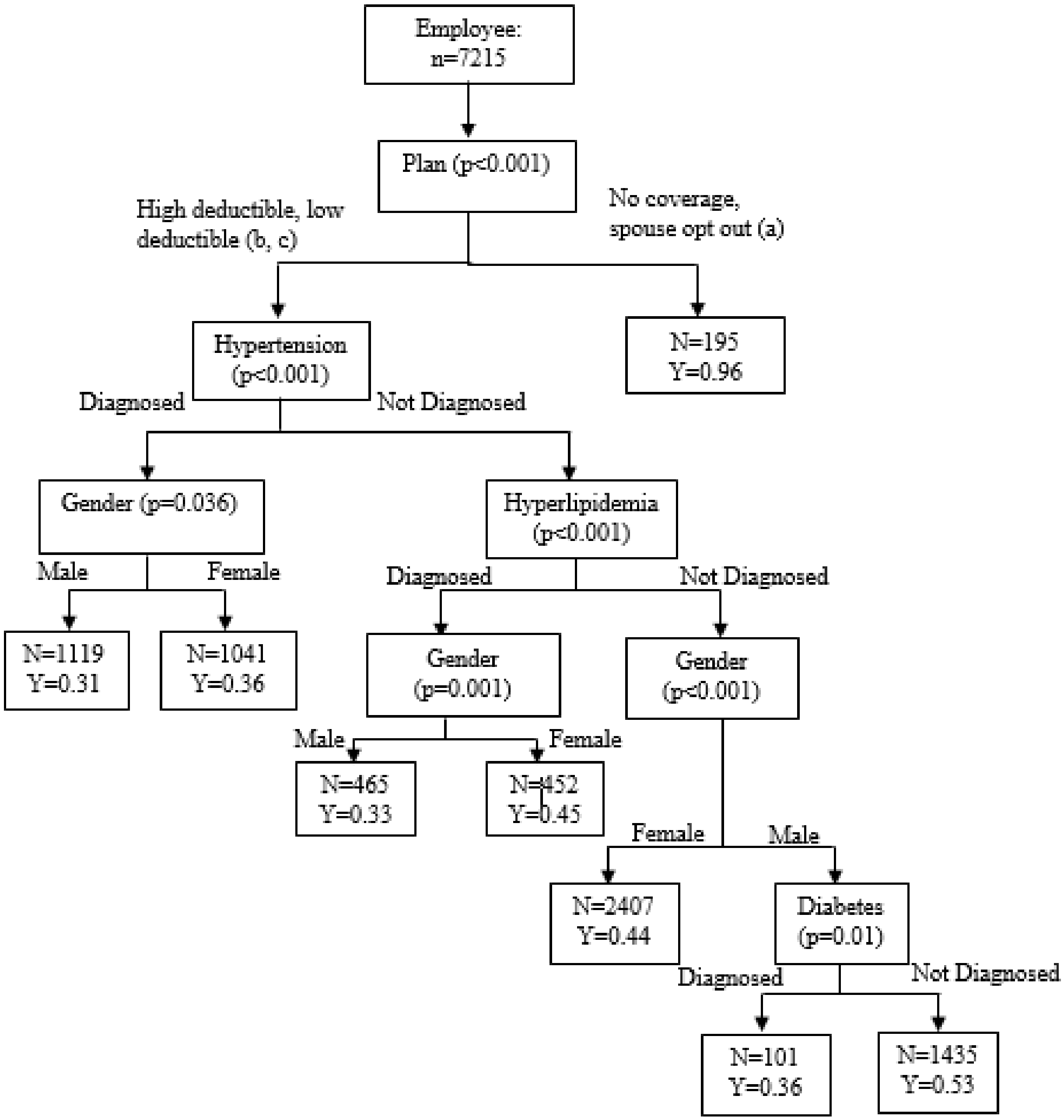


Figure 2.3 Classification tree for employee (N=number of individuals in each class, Y=proportion of individuals in the class using onsite clinic preventive services)

Based on the classification tree for the spouse group (Figure 2.4), the only significant variable for usage of the onsite clinic preventive services identified in the tree is hypertension. Of all individuals in the spouse group who were not diagnosed with hypertension and had utilized preventive services, 17.4 % had used preventive services offered in the onsite clinic vs. 82.6% who used preventive services offered by offsite health providers. Similarly, for individuals who were diagnosed with hypertension and had used preventive services, 11.8% utilized preventive services offered in the onsite clinic vs. 88.2% who utilized preventive services offered by offsite health providers.

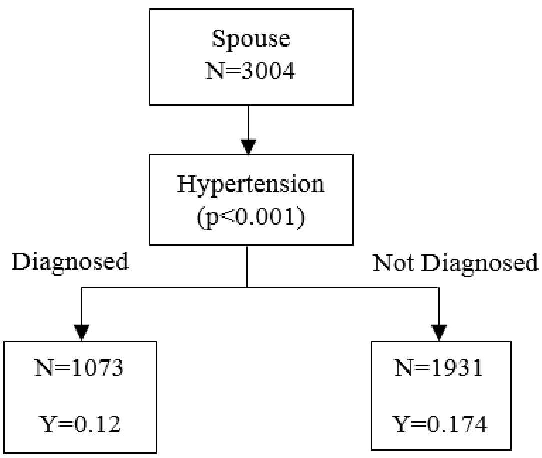


Figure 2.4 Classification tree for spouse (N=number of individuals in each class, Y=proportion of the individuals in the class using onsite clinic preventive services)

The classification tree developed for the child group (Figure 2.5) is aligned with logistic regression results. The two only significant variables for using the onsite clinic preventive services were age and compensation type (hourly vs. salary based). From individuals who had used preventive services, were above 18 and whose compensation was categorized as hourly (based on the employee compensation type which also applies to spouse and child), only 12% had used preventive services in the onsite clinic. For individuals who had used preventive services and where under 17 only, 2.5% had used preventive services in the onsite clinic. For individuals above 18 who were classified as hourly and had used preventive services, 12.7% had used onsite clinic vs. 7% who were classified as salary-based.

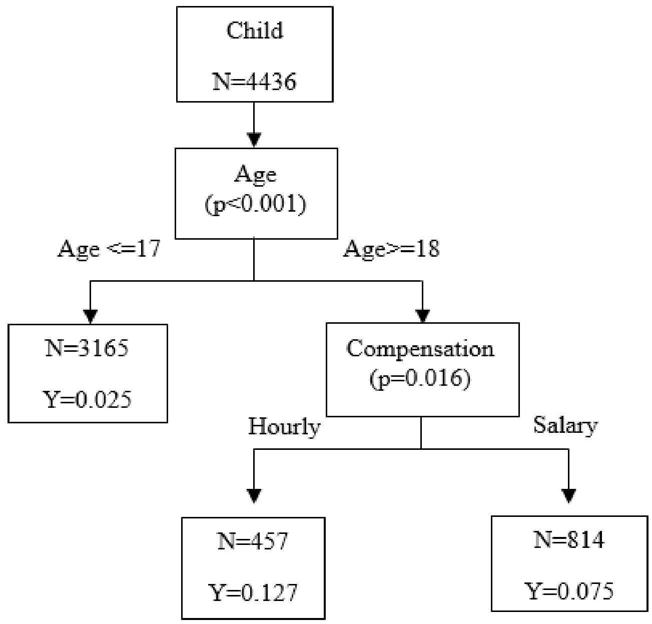


Figure 2.5 Classification tree for child (N=number of individuals in each class, Y=proportion of the class using onsite clinic preventive services)

2.4 Discussion

This study presented a framework adapted from the Andersen Model of Healthcare Utilization for assessing impact of workplace clinics on healthcare utilization patterns. The original model was developed to identify factors impacting healthcare utilization with the family as the unit of analysis and was expanded and adapted by several researchers [48],[49],[50]. Suggested framework is enhanced by combining health informatics which enables managing large amount of data into valuable information [51] and can guide analyses of healthcare utilization and patterns with claim datasets. Presented framework suggested assessing utilization by combination of longitudinal and cross sectional studies, implementing health data and tools including survey and interviews to identify factors impacting usage and tracking patterns and potential changes.

Based on descriptive analysis of the data, from the study cohort more women had visited healthcare providers for preventive care comparing to men for both Year1 (65% vs. 52%) and Year3 (69% vs. 58%). This aligns with both pervious research showing higher healthcare utilization rate for females vs. males [52] and with national statistics from the Center for Disease Control and Prevention, showing that female rate of preventive services usage is 69% higher than males. Percentage of individuals who had used preventive services had increased for both females and males from Year1 to Year3. Although the increase may not be directly linked to the onsite clinic, clinic efforts in improving awareness, and providing better access for employees may be considered as a significant factor; however, further study with subjective assessments from employees is needed to confirm the impact and clarify the interests and needs applicable to expand opportunities for improving women's or men's health.

Flu vaccination was the most common primary service used by the study population in the onsite clinic (Table 2.1). Similarly, it was the second most common service among patients who visited offsite health providers. Increasing flu vaccination among working-age population can help to address negative impacts of the disease such as absenteeism and its costs to employers [53] and hospitalization and mortality for older adults [54]. Percentage of individual who received flu vaccination in Year1 in the population under study was 11%, however this percentage increased to 27% of all eligible in Year3. Despite the marked improvement in flu vaccination utilization from Year1 to Year3 in the whole eligible population (n=23635), only 20% of individuals over 18 visited a healthcare provider (onsite clinic\offsite providers) for flu vaccination, which is lower than national health statistics for similar time period showing a 41% of vaccination for adults in the United States [55]. Various factors have been suggested for low percentage of vaccination such not considering it as effective and lack of incentives for employers to cover the cost of vaccination [56]. Increase in the vaccination rate for eligible individuals in the study population, might be attributed to onsite clinic impacts as an accessible convenient provider; however, for approaching national average and more coverage among the population, onsite clinics would be able to address those barriers and imply workplace interventions for improving vaccination rate by strategies such as providing incentives and advertisement [57].

Logistic regression and classification tree techniques were used to investigate determinants of healthcare utilization for preventive services in the onsite clinic. Although regression models could identify significant factors and their interactions, classification trees helps to further investigate relation between factors by identifying interdependencies between independent variables [58]. Moreover, in the logistic regression, impact of each predictor is considered to be uniform across all observations vs decision tree in which effect of a variable in a subset is unrelated to its effect in other subsets [59]. Three independent variables considered in the model to identify their impacts on utilization of the onsite clinic were whether individuals in the population were diagnosed with hypertension, hyperlipidemia or diabetes. This conditions were chosen as the most common chronic conditions among the cohort population with 15% of women and 18% of men diagnosed with hypertension, 12% of women and 14% of men diagnosed with hyperlipidemia and 6% of women and 6% of men diagnosed with diabetes. According to the Center for Disease Control and Prevention, 31% of men and 32% of women are diagnosed with high blood cholesterol [60], 34.1% of men and 32.7% of women are diagnosed with high blood pressure [61] and 13.6 of men and 11.2 of women in the United States are diagnosed with diabetes [62]. Comparing to national average, population of the study seems to be healthier in terms of the aforementioned chronic conditions.

Based on the logistics regression, individuals not diagnosed with diabetes had higher odds of using onsite clinic preventive services. Similar to individuals without diabetes, individuals who were diagnosed with hypertension and individuals who were diagnosed with hyperlipidemia had lower odds of using onsite clinic preventive services comparing to individuals who were not diagnosed with these conditions. Previous research has shown that underuse of preventive services is common among diabetic patients [63]. It might be argued that diabetic patients in this population may seek their preventive care in other healthcare providers despite efforts of onsite clinic in providing condition management programs with specific focus on diabetes. Perhaps the patients may have healthcare through offsite practices which are their 'medical homes which is a modernization of medical primary care that helps patients take care of all their specialty information along with traditional internal medicine or pediatric medical needs [64]. Expanding wellness activities, and advertising existing nutrition counselling, could appeal to people who think their 'medical' needs are being met elsewhere. Moreover, all of the usual preventive services should be used by people with chronic conditions at least as much as by people without chronic conditions, and since some may be actually more susceptible to suffering or complications, preventive services (for example the flu vaccination) could be extra important (for people with diabetes or autoimmune disorders). Dyslipidemia and most hypertension do not give symptoms so these could be under-diagnosed. Work schedules and thinking we're healthier than really we are, could be barriers to getting screening.

Logistic regression of child group showed that individuals over 18 have higher odds of using onsite clinic preventive services. Individuals with age between 18 to 26 can still be covered by family health plans [65] yet not have many community resource as they feel perhaps too old for their pediatric practitioners and have not established their own adult relationship with an offsite health provider. If these people are not students and they are not working, their needs can be met by onsite clinic better than the offsite providers.

Previous studies assessed onsite clinics either from a financial perspective by calculating return on investment and cost-effectiveness [39], [40], [66] or employees' attitude toward the onsite clinic services by cross sectional studies [13]. This study presented a new approach toward assessment of onsite clinics by proposing a framework implying usage of health informatics for healthcare utilization analysis. The primary focus of this research was on preventive services on a three-year time period. Future research will apply other aspects of the suggested framework by identifying financial impacts of the clinic and how clinic has impacted population health aspects including care coordination and health delivery.

2.5 Conclusion

Increasing adoption of onsite clinic in various work settings shows that more accurate methods of assessing impacts and performance of these clinics are needed to better serve and adapt to population health needs. In the proposed framework for this study we suggested usage of available claim and health record data to continuously monitor changes in the healthcare utilization and investigate onsite clinic roles in that matter. Further study is needed to assess the relation of the onsite clinic providers with other providers in the community and how they can have an efficient contribution to the care of patients. Following the patient cohort in this study, the next chapter will assess the provider team structure and collaboration for patients with chronic conditions.

3. MODELING OF DIABETES CARE TEAMS USING SOCIAL NETWORK ANALYSIS AND ADMINISTRATIVE DATA[2]

3.1 Introduction

Diabetes is the 7th leading cause of death in the United States [2]. With an aging population, diabetes prevalence is expected to increase over the next 40 years, with 1 in 3 Americans projected to have diabetes by 2050 [67]. The economic burden of the diabetes affects both individuals and the society. According to the American Diabetes Association (ADA), the total cost of diagnosed diabetes has increased from $245 billion in 2012 to $327 billion in 2017. Outside of the diabetes medication cost (30%), the next largest expenditure is the physician office visits (13% of the total costs) [68], as the management of diabetes is often complex and requires multiple visits to various types of care providers, e.g., primary care, emergency physician, specialist, and nutritionists [69]. Previous studies have highlighted the importance of healthcare providers' coordination and collaboration for improving diabetes care delivery and reducing costs [70],[71], [72], however, challenges including under- or over-utilization of healthcare providers and unclear definition of providers roles and responsibilities [73] negatively impact effective provider team-work.

General recommendations and guidelines exist for team roles in the care of patients with diabetes. The American Diabetes Association has recommended that a diabetes care team should consist of a primary care provider, nurse educator, registered dietitian, diabetes educator, endocrinologist, eye doctor, social worker, psychologist, podiatrist, pharmacist, dentist, and exercise physiologists [74]. Despite the recommendations for diabetes care teams, not all teams are assembled and structured based on those recommendations. For example, healthcare providers may be added to or removed from the care teams by patient or patient' family members, and providers' collaborations and communications may be challenged by these changes. In addition, not all providers have clear descriptions of their roles, duties, and responsibilities as part of the overall care team for patients with diabetes [75]. Moreover, the frequency at which these ADA recommendations occur in practice is unknown.

---

[2] This chapter is based on: Ostovari M., Steele-Morris C.J. Griffin, P., Yu D. (in press). Data-driven modeling of diabetes care teams using social network analysis. Journal of the American Medical Informatics Association. 2019

Some major challenges to understanding care teams are the lack of data and difficulty in collecting data that measures healthcare providers working relationships [21]. There are questionnaire tools for measuring collaborations of healthcare providers [19],[18], however, survey-based research is typically costly with a low response rate [20]. Interview methods have similar issues. Although they capture relationships from each healthcare professional's perspective, they are limited to a small number of participants [76]. Physicians and other providers form different working relationships that might be formal such as referrals or associations with a hospital [25] or informal like advice seeking [21]. New approaches are needed to better capture these relationships and the associated impact on providers' decision making and patient outcomes.

Originally developed from the social psychology [26] and based on the work of Moreno [27], social network analysis (SNA) is a powerful technique that captures hidden channels of collaboration, information flow and communication between network actors (e.g. individuals, groups, organizations). In the healthcare settings, SNA has been used to study healthcare providers' collaboration, communication, and the impact of their interaction on decision-making for the patient [34],[77],[78],[79],[80]. These studies have primarily used surveys and interviews data for SNA, however, application of this technique on the large-scale health data can further capture the physicians' relationships and their collaboration on a larger scale [21],[25],[81], [82]. Physicians who provide care to the same patients form a team of providers that may have formal referral relations or casual discussion about the patients [21]. This approach helps to go one step beyond previous utilization analysis studies that focused on individual patient health services [22],[24]. Social network analysis using the patient sharing approach among the healthcare providers allows to look at the collaboration of healthcare providers, meaning:" recurring process of working together toward common goals"[83].

SNA application on administrative data has been previously used to understand the geographic differences in the physicians' networks across the U.S. [25],[29],[84] or to assess the provider network characteristics' impacts on patient outcomes [32],[31],[30],[33],[85],[86],[87] . Landon et al. used SNA and patient sharing techniques to explore variations in the professional networks of physicians across the U.S. and showed that physicians tended to share patients with similar groups of healthcare providers that had similar patient panel characteristics [25]. Others have used a similar approach to identify groups of physicians that formed working relationships and had the potential of becoming an Accountable Care Organization (ACO) [84], physicians' regional networks and the consistency of their relationships [29]. Although applications of SNA in healthcare are increasing, current work with SNA focuses only on providers who provided direct care to patients [29],[32],[33],[88]. This approach is useful for assessing the relationships between individual providers like physicians, however, it neglects other members of the care team. Moreover, the many healthcare services that are not delivered through direct care (e.g., labs) are also overlooked.

The objective of this study is to develop a framework using social network analysis metrics and a multi-scale community detection algorithm to assess the working relationships of care teams involved in the diabetes care process. The SNA metrics identify the global structure of the network and the influencers within the network while the community detection uncovers the healthcare providers that work more closely.

3.2 Materials and Methods

3.2.1 Data Source

This study was approved by the Institutional Review Board (IRB 1511016796). The data contained de-identified administrative health claims of employees (faculty/staff) and their dependents (spouse/child) of a large university in the Midwest. The claims data included medical health utilization of the study population, insurance eligibility information, and medication usage (files) and claims. The eligibility data included demographic information such as age, sex, and type of compensation (hourly/salary). The medical file included International Classification of Diseases 9th edition code (ICD9), service date, cost of service and healthcare provider information. The medication file included medication purchase date, cost, and provider information. This study excluded the student population.

3.2.2 Study Design and Analysis

A retrospective analysis of the health claims data was performed. Table 3.1 details the variables in the dataset. Individuals with diabetes (type 1 and type 2) and their healthcare providers were tracked in 2012 (Year1) and 2013 (Year2) by using their unique patient and provider identifiers. To determine cohort population with diabetes, primary, secondary and tertiary diagnoses for the ICD9 starting by 250 were used [89] (25.00, 250.01, 250.02, 250.40, 250.41, 250.50, 250.51, 250.52, 250.60, 250, 61, 250.62, 250.80, 250.81, 250.82, 250.90, and 250.91). All health services utilization and healthcare providers' identifiers for the identified patients were extracted from the medical and medication files.

Table 3.1 Description of variables in the dataset

Health Administrative Data			
Demographic Variables	Clinical Variables	Insurance Plans	Healthcare Provider
- Sex (male/female) - Date of birth (MM/DD/YYYY) - Person ID (unique identifier) - Relationship to employee (employee, spouse, and child) - Compensation type (hourly, salary). Compensation variable for spouse and child are based on the policy holder's status	- Principal diagnosis, secondary diagnosis, and tertiary diagnosis (ICD9) - Claim ID (7-digit value)	- Coverage indicator (dental, drug, vision, hearing, Long-Term Disability, Short-Term Disability, medical, Mental Health Services Act) - Type of plan (no coverage, spouse opt out, Federal Health Plan, low deductible, high deductible)	- Provider ID (unique identifier) - Provider zip code (5-digit zip code of providers) - Provider type (e.g., primary care, dentist)

3.2.3 Constructing the Network

Network construction was based on the patient sharing relations between the healthcare providers, an approach that has been previously validated [21]. Presence of shared patients among the providers is interpreted as an information sharing relationship between the two providers [32]. Shared patients were determined from the health claims. The network nodes represented healthcare providers, and an edge between two nodes represented shared patients between the providers. The edge weights were defined by the number of shared patients, a proxy for the strength of the relationships between the providers. The network was limited to patients with diabetes and their corresponding healthcare providers. All types of healthcare providers were included in order to capture the health services needed to assess direct and indirect care and to address potential comorbidities and complications of diabetes. Previous work demonstrated that information-sharing relationships between healthcare providers were observed with link weights of two or more [21]. Thus, the network focused on providers who shared at least two patients. Excluding providers with only one shared patient has been recommended to remove information sharing and working relationships formed by chance that may not carry much information about the relationship [21],[25].

3.2.4 Constructing Communities

To further understand the structure of the network and identify the natural communities of the healthcare providers, local analysis of the network through community detection was performed [90]. Communities in this study were defined as groups of nodes that were more densely connected internally compared to their connection to the rest of the network [91]. The community detection algorithm used a multi-scale approach. Stability was the objective function optimized to find the best partitions of the network [92]. This method can identify smaller communities in the network, unlike the more common modularity-based community detection algorithms [93] which suffer from a resolution limit [94] and are unable to identify communities with edges fewer than $\sqrt{L/2}$, L being the number of edges in the entire network. Healthcare providers were assigned to single non-overlapping communities using this multi-scale method.

3.2.5 Metrics

SNA metrics were used to assess the global characteristics of the network and communities. Density measured the cohesions and frequency of collaboration among the healthcare providers [95]. Measures of centrality identified nodes with important roles in the network and greater access to other nodes [96],[97]. Three centrality measures were calculated: 1) degree centrality 2) betweenness centrality, and 3) closeness centrality. The centrality measures were applied on the largest component of the network to identify the influencer and providers with greater access and control over the flow of information in the network.

Degree centrality of a provider in the network showed the number of providers that were directly connected to that provider. The betweenness centrality measured the degree by which a node was ‘between’ pairs of other nodes in the network. A node having higher betweenness centrality indicates that it has more influence in the network for distributing information [33]. Higher closeness centrality of a node shows better access to rest of the nodes in the network [98]. In addition to network density, we identified connected components in the network in which all nodes were directly and indirectly connected [99]. Table 3.2 shows the measures of social network analysis used in the study and their definitions. The analysis was completed using SAS (v 9.4, SAS Inc., Cary, NC) and RStudio (version 0.99.903) with the igraph (version 1.1.2) [100] and devtools (version 1.12.0) [101] packages.

Table 3.2 Measures and properties of social network used in the analysis

Network Measures and Properties	Definition
Density	Proportion of edges in a graph to the maximum possible number of edges with value of 0 to 1 [99]
Degree Centrality	The number of ties a node has in a graph [99]
Betweenness Centrality	A measure of centrality based on shortest path, showing the extent to which a node is between other pairs of nodes [99]
Closeness Centrality	A measure of closeness of each node to other nodes in the network [99]
Component	A subgroup of a graph in which all nodes are connected [99]
Community	Subgroup of a network with denser internal connections compared to its connections to the rest of the network [99]
Assortative mixing[1]	A measure to assess whether nodes (physicians) with similar characteristics (e.g. patient panels) would be connected [99]
Triad closure[1]	A property among three nodes A, B, and C, such that if a strong tie exists between A-B and A-C, there is a weak or strong tie between B-C [99]
1. These measures were not included	

3.3 Results

Out of a total of 15,183 individuals with at least one medical claim in Year1, 827 patients were identified with diabetes (5.4%). The patient cohort consisted of 416 (50%) women with average age of 54±9.9 years and 411 (50%) men with average age of 57.6±10.0 years. In addition to diabetes, other common health conditions (i.e., comorbidities) identified in the cohort were hypertension (ICD9=401.1 & 401.9) with 271 (32%) patients in Year1 and 296 (35%) patients in Year2 and hyperlipidemia (ICD9=272.4) with 152 (18%) patients in Year1 and 159 (19%) patients in Year2.

The cohort with diabetes received healthcare services from 2,567 healthcare providers and 2,541 healthcare providers in Year1 and 2, respectively. A total of 1,523 of the providers (59%) present in Year1 were also present in Year2. Number of providers after the exclusion criteria of removing providers with <2 patients in common for Year1 was 896 and for Year2, 836. The provider type with the highest utilization for both years was a medical laboratory, with 418 patients in both years. The second most utilized provider was a mail-order pharmacy with 306 patients in Year1 and 337 patients in Year2.

3.3.1 Network Characteristics

The network of healthcare providers in Year1 consisted of 896 nodes and 10,200 edges with density 0.0261. The network had seven components (Table 3.2) with all nodes connected directly or indirectly [99]. The largest component had 884 nodes with 10,194 edges, while the other six components had each two nodes. These six smaller components nodes (healthcare providers) had only one shared patient with the rest of the network, but since edges with weights <2 were removed (see methods section) they resulted in separate components. The edge weights (numbers of patients shared between the providers) ranged from two to 271 (average of 4.03 and standard deviation of 7.3).

Focusing on the largest component, the median degree centrality for all the network nodes was eight, the median betweenness centrality of the networks nodes was 108.8, and the median closeness centrality was 0.0002. Table 3.3 presents the top ten healthcare providers with highest degree, betweenness and closeness centrality. Although medical laboratory was the provider type identified as having the highest degree and betweenness centrality, the actual medical laboratory provider differed in the two metrics. Specifically, Provider ID 1667 had the highest degree centrality, while Provider ID 1439 had the highest betweenness centrality. Although degree and betweenness centrality are different SNA metrics, nodes present in the top 10 for degree and betweenness centrality were similar. Specifically, 80% of providers (Provider IDs 397, 292, 2114, 107, 171, 489, 1667, and 382) ranked in the top ten for degree centrality were also in the top ten for betweenness centrality metrics (Table 3.3). Top nodes identified by the closeness centrality were also identified by degree and betweenness centrality as top providers (Provider IDs 397, 292, 2114, 107, 489,382)

Table 3.3 Year1 providers with the highest degree, betweenness, and closeness centrality. Provider IDs are random numbers generated for each provider and do not represent their real IDs.

Provider ID	**Provider Type**	**Degree Centrality**	**Provider's Unique Patients**
1667	Medical laboratory	482	418
397	Mail order pharmacy	468	306
292	Mail order pharmacy	450	306
2114	Medical laboratory	377	221
107	Health home organization	359	156
1670	Hospital	340	175
171	Hospital	338	127
489	Mail order pharmacy	321	183
382	Mail order pharmacy	308	172
126	Hospital	247	92
Provider ID	**Provider Type**	**Betweenness Centrality (Normalized Betweenness Centrality)**	**Provider's Unique Patients**
1439	Medical laboratory	47,938 (0.123)	131
2114	*Medical laboratory*	*38,786(0.099)*	*221*
397	Mail order pharmacy	36,389 (0.093)	306
107	Health home organization	35,793 (0.091)	156
292	Mail order pharmacy	28,776 (0.073)	306
489	Mail order pharmacy	23,323 (0.0598)	183
1667	Medical laboratory	23,142 (0.0594)	418
159	Hospital	19,470 (0.050)	30
382	Mail order pharmacy	19,343 (0.049)	172
171	Hospital	14,971 (0.038)	127
Provider ID	**Provider Type**	**Closeness Centrality**	**Provider's Unique Patients**
397	Mail Order Pharmacy	0.000262	306
292	Mail Order Pharmacy	0.000254	306
107	Health home organization	0.000248	156
1859	Podiatry	0.000246	8
489	Mail order pharmacy	0.000245	183
1439	Medical laboratory	0.000245	131
382	Mail order pharmacy	0.000242	172

Table 3.3 continued

928	Family Practice	0.000242	4
2114	Medical laboratory	0.000241	221
2386	Hospital	0.000241	4

The network of Year2 was connected and composed of 836 nodes (healthcare providers) and 9,722 edges (patients shared among healthcare providers) with a density of 0.0278. The edge weight range was from two to 181 (with the average of 3.97 and standard deviation of 6.04). Compared to Year1, number of nodes and of edges were reduced, but density and average edge weights were similar.

Similar to Year1, the median degree centrality was eight for all the nodes in the network, the median betweenness centrality was 107.67, and the median closeness centrality was 0.0002. In contrast to Year1, the mail order pharmacy became the provider type with the highest degree, betweenness and closeness centrality (Table 3.4). Table 3.4 presents the top ten healthcare providers identified as having the highest centrality for Year2. As with Year1, 80% of the providers ranked in the top ten for both degree and betweenness centrality metrics (Provider IDs: 301, 1628, 2064,106, 403, 170, 1630, and 1525). Mail-order pharmacy (ID 292) ranked top based on degree, betweenness, and closeness centrality. Comparing provider IDs from Table 3.3 and 3.4, 90% of the providers from Year1 (Table 3.3) also appeared as central providers based on the betweenness and degree centrality measures in Year2. Three of the top providers identified by closeness centrality (Provider IDs 292, 382, 107) were also identified as central by degree and betweenness centrality.

Table 3.4 Year2 providers with the highest degree, betweenness, and closeness centrality. Provider IDs are random numbers generated for each provider and do not represent their real IDs

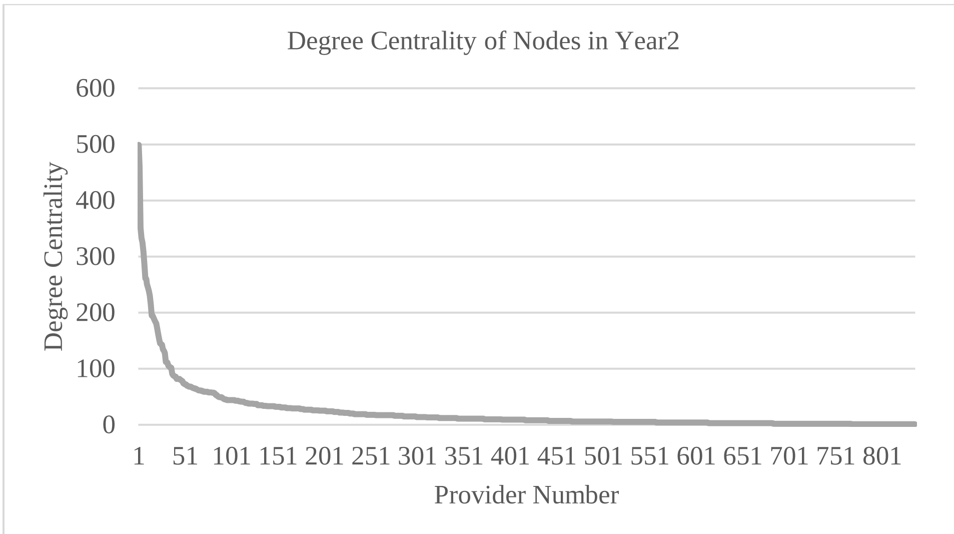
Provider IDs	**Provider Type**	**Degree Centrality**	**Provider's Unique Patients**
292	Mail order pharmacy	499	337
1667	Medical laboratory	461	418
2114	Medical laboratory	351	218
107	Health home organization	333	148
382	Mail order pharmacy	325	184
171	Hospital	309	124
1670	Hospital	287	125

Table 3.4 continued

1525	Worksite clinic	261	155
288	Retail pharmacy	260	155
2292	Pathology	251	80
Provider IDs	**Provider Type**	**Betweenness Centrality (Normalized Betweenness Centrality)**	**Provider's Unique Patients**
292	Mail order pharmacy	67,063 (0.192)	337
382	Mail order pharmacy	30,997 (0.089)	184
107	Health home organization	30,314 (0.087)	148
2114	Medical laboratory	30,032 (0.086)	218
1439	Medical laboratory	29,060 (0.083)	129
1667	Medical laboratory	25,109 (0.072)	418
170	Pathology	11,878 (0.034)	124
397	Mail order pharmacy	10,980 (0.031)	57
1525	Worksite clinic	8,553 (0.024)	155
1670	Hospital	7,969 (0.022)	125
Provider IDs	**Provider Type**	**Closeness Centrality**	**Provider's Unique Patients**
292	Mail order pharmacy	0.000286	337
382	Mail order pharmacy	0.000265	184
107	Health home organization	0.000263	148
1929	Anesthesiologist	0.000261	4
1570	Neurologist	0.000259	8
1435	Emergency Medicine	0.000256	5
2482	Gynecologist	0.000256	9
1979	Supply Center	0.000255	4
2298	Dermatologist	0.000255	11
2131	Preventive Supply Center	0.000255	8

The following graphs in Figure 3.1, provide the distribution of nodes degree (a), betweenness (b) and closeness (c) centrality for Year1 and Year2 from the highest to the lowest in each graph. The degree shows the direct connection of the nodes. As it can be seen from the graphs, a small groups of nodes have higher degree measures compared to the rest of the nodes. Similar patterns can be seen for the betweenness centrality (B). Although for betweenness centrality higher percentage of nodes have betweenness close to zero compared to degree centrality. As it can be seen from the graphs there are a small groups of node with very high centrality. These nodes with higher centralities represents the hubs in the networks [102]. Closeness centrality (C) presents a different scale (distance from other nodes) and scale are much closer for all nodes in the network.

A)



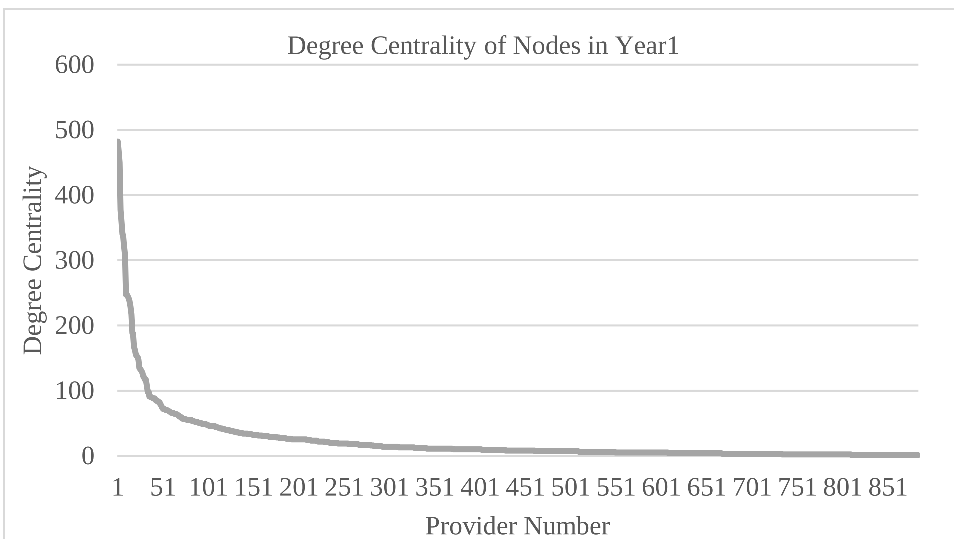
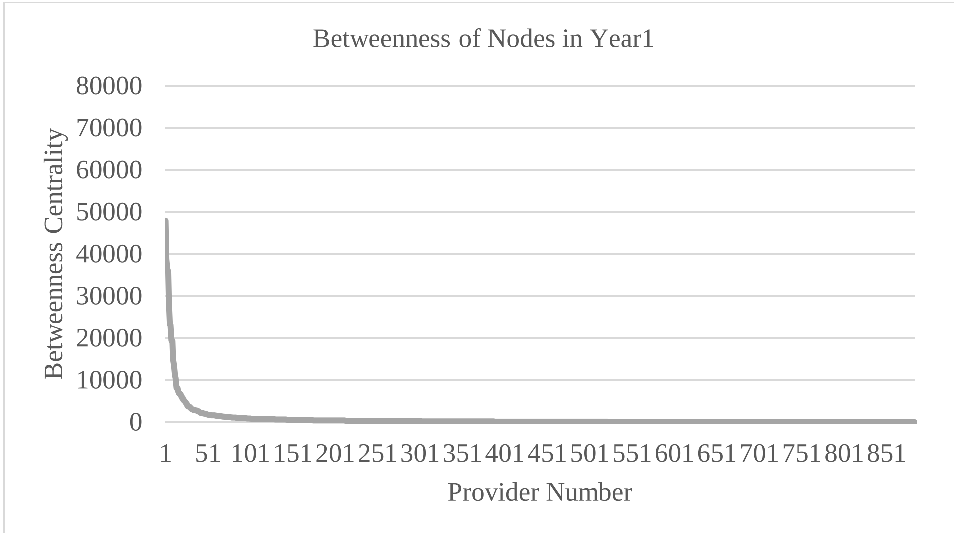


Figure 3.1 Distribution of the degree (A), betweenness (B) and closeness centrality (C) of the nodes for Year1 and Year2

Figure 3.1 Continued

B)



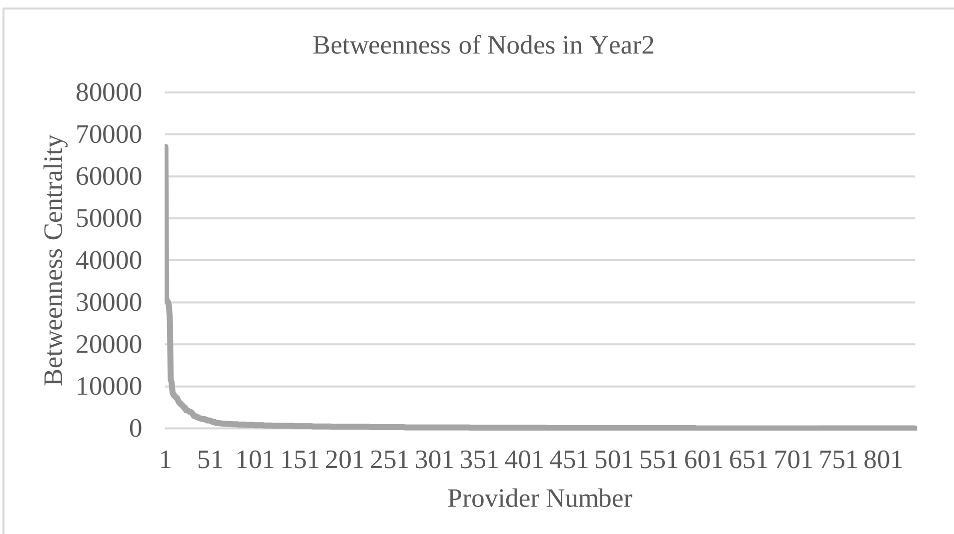
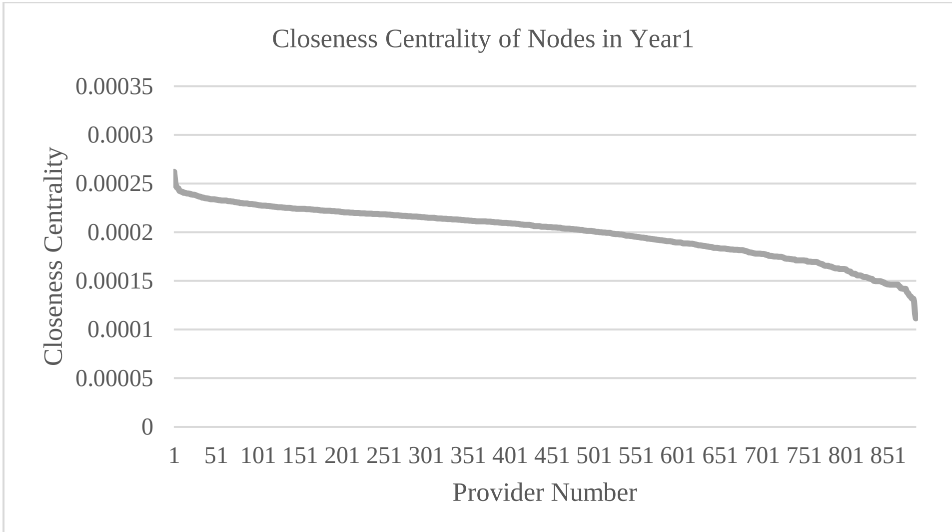
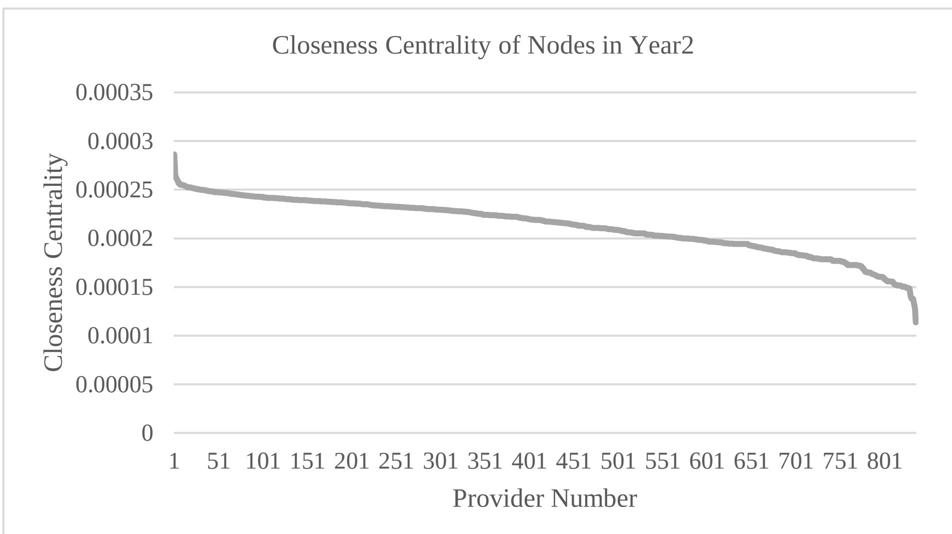


Figure 3.1 Continued

C)





3.3.2 Network Communities

Although network analysis with shared patients identified central providers and global network characteristics, it lacks the needed granularity to address the study's objective of identifying working relationships within and between care teams. As mentioned in the method section, edges between providers with less than two patients shared were removed. The main component of the network in Year1 had 884 nodes and 10,194 edges. A multi-scale community detection algorithm [92] was applied on this component to identify groups of healthcare providers more tightly linked together through the patients sharing relationships. Forty-six communities were detected for Year1 with the sizes of three to 155 members. Communities of providers served an average of 115 patients.

One of the smaller communities is shown in Figure 3.2 to illustrate community characteristics detected using this approach. The community of care providers was composed of six providers: a general surgeon (ID 501), a pharmacist (ID 556), an ophthalmologist (ID 1174), one family practice (ID 2151), an anesthesiologist (ID 2438), and a gynecologist (ID 2513). Based on the degree, betweenness, and closeness centrality applied to the community, the central node (ID 1174) was an ophthalmologist with degree of four, betweenness of eight and closeness of 0.077.

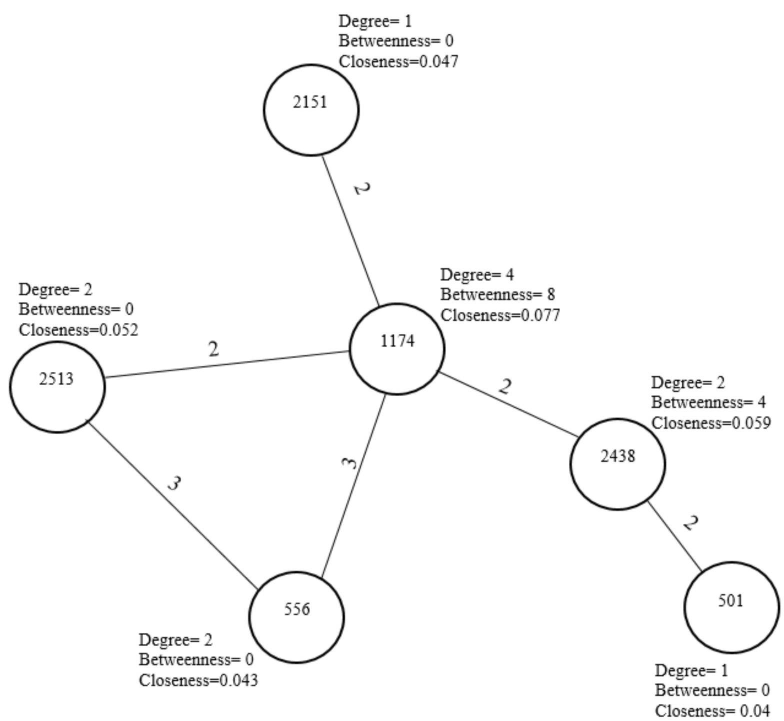


Figure 3.2 One of the communities detected in the network of healthcare providers of Year1. The node 1174 is an ophthalmologist, the node 2151 is a family practice, the node 2513 is a gynecologist, the node 556 is a pharmacist, the node 2438 is an anesthesiologist, and the node 501 is a general surgeon. Degree, betweenness and closeness centrality of each node is depicted next to it. The number on each edge represents the edge weight.

Thirty-five out of 46 (76%) detected communities included a primary care practitioner (internal medicine, family practice). Forty-five out of 46 (97%) communities included a specialist (endocrinologist, nephrologist, ophthalmologist, oncologist, gynecologist, cardiovascular specialist, podiatrist, and dermatologist). To determine the most central provider in each community, centrality metrics were used. Among the healthcare providers, pharmacists were detected as the central nodes in 24% of the communities. Radiologists and hospitals were the next most commonly central providers, observed in 13% and 11% of the communities, respectively. Specific services provided by these central providers (i.e., radiologists and hospitals) were identified using procedure codes (Current Procedural Terminology) for the patients. Chest x-rays (71020) were the most common procedure performed by central radiology providers. Procedures provided at hospitals varied widely and included screening mammogram digital (G022), fluoroscopic guidance (77003), chemistry procedure-creatinine (82565), chemistry procedure-calcium (82310), and blood count (52025).

Comparable to Year1, 45 communities were detected from the network in Year2 with the multi-scale community detection algorithm [92]. Similarly, resulting communities had a range of three to 145 healthcare providers (nodes) and served an average of 119 patients. Similar to Year1, detected communities commonly had a primary care physician (detected in 87% of the communities) and a specialist (detected in 98% of the communities). Number of primary care providers included in the network (based on 2 or more patient sharing criteria) increased from 136 in Year1 to 156 in Year2. The increase may represent the higher usage of the primary care in the second year by patients or addition of primary care providers with sharing of 2 patients or more in the network. Pharmacists were the central providers in 22% of the communities based on SNA centrality metrics. Next most common central providers were radiologists and hospitals, observed in 18% and 11% of communities, respectively. For the central radiologists, screening mammogram digital (G0202) was the most commonly performed procedure. Procedure performed by the hospital providers were diverse and included glucose monitoring (82948), screening mammogram digital (G0202), blood test (36415) and injection, fentanyl citrate (J3010).

3.4 Discussion

The study demonstrated a framework integrating SNA and state-of-the-art community detection techniques to provide insights and metrics for assessing care teams and collaborations for complex, chronic disease management. This approach leveraged big health datasets to analytically identify key players and influencers in the care teams for patients with diabetes.

Previous studies that used SNA application on large-scale data to assess coordination and working relationships between the healthcare providers only focused on physicians who provided direct care to the patients or provided care to patients with variety of health conditions [32], [31], [33]. Extending SNA to all provider types within the care system is needed to identify gaps in the services in comparison to guidelines and to quantify team relationships for complex chronic conditions such as diabetes [103]. In this study, we included all healthcare providers involved in the care of patients with diabetes and used the patient sharing approach and SNA on health administrative data to identify key stakeholders and providers in the care teams.

Potentially contrasting results were presented regarding the stability and continuity of the care provider population over the two years. Specifically, over 40% of the providers were different between the two years. Numerous reasons may cause the change in the providers, including change of the insurance plans by patient or physician' office change of policy [104],[105]. However, application of the community detection and other SNA approaches showed strong consistency between the years. Specially, number of nodes and edge weights were similar indicating that key providers of services and services needed were relatively stable. Moreover, our results showed that nine central providers based on degree and betweenness centrality from Year1 (Table 3.3) also appeared as central in Year2. Distribution of centrality measures in Figure 1, shows similar patterns for the node centralities between the two years as well. These results showed a consistency of usage of these central providers and continuation of their working relationships with other providers. Provider continuous relationship with the patient is an important aspect of the care continuity [106] which can positively impact healthcare costs, outcomes and care coordination [107]. Thus, although a global analysis of the provider turnover raises alarms regarding the patient care, our approach showed that providers with central roles in the network of providers remained consistent over the period of the study.

When expanding SNA to capture all providers, those that provided direct care were not the central nodes in the network. Based on the degree, betweenness and closeness centrality, pharmacy providers were among the central providers in Year1 (Table 3.3) and Year2 (Table 3.4) in the overall network. These findings may reflect a bias toward pharmacists as the network was designed based on the patients sharing [21] and providers with higher number of patients would be identified as central by the degree centrality [99]. Despite the possibility of this bias, because of the crucial role of the pharmacist in the diabetes management [108],[109], [110] considering the pharmacists as central providers is not unrealistic. The central role of pharmacy captured by this analysis reflects both utilization trends and intervention opportunities shown in other studies. For example, usage of mail-order pharmacies is increasing in the United States, and studies have shown that patients with diabetes who use these pharmacies have better medication adherence [111],[112]. Although they may not have the face to face consultation of retail pharmacists, the phone consulting option provides a convenient way for patients to connect with the pharmacists [113]. Moreover, usage of mail order pharmacist in the healthcare delivery level improves access to medication for chronic illnesses [114]. Our data showed a similar trend for usage of the mail-order pharmacists for the study cohort. Intervention programs for diabetes can be leveraged by the findings from this research, which identifies potential central providers other than primary care that could be effective in increasing medication adherence for the patient [112].

Eighty-seven percent of the communities detected in Year1 included primary care physicians vs. 98% that included specialists. Most communities detected had a combination of primary care, specialists, and pharmacists, but not all providers outlined as the care team by the Americans Diabetes Association [74] were identified in each community. Patients might have different types of needs based on their characteristics (e.g. age, complication), therefore, the relationships of providers and their strength may vary based on patients. The larger appearance of the specialists in the communities compared to the primary care providers may be due to various reasons. First, some patients may do self-referrals and not go through the primary care channel to see a specialist [115]. Some patients might have more complications and require specialists such as endocrinologists to manage their care [116]. The multi-scale community detection [92] is a useful tool to understand the structures of the relationships among the providers and appearance patterns for different types of physicians. SNA measures amplify this technique by identifying the central providers in each community.

This study has some limitations. As the results show, social network centrality measures might be biased toward providers with larger numbers of patients. Although we used three different measures of centrality (degree, betweenness and closeness) and community detection to address this limitation to some extent, additional refinement in the approach may be needed to better recognize central providers. Another limitation was the population cohort, which was drawn from one large employer in the Midwest. Expanding the dataset is needed to determine whether findings are generalizable to population of patients. This study is limited by provider definitions of the dataset. For example, medical labs may have pathologist consultants; however, pathologist consultants may also be labeled separately in other cases. Although the study could be impacted by the limitations of the claims data, the methodology is still effective for identifying the care team interactions and central providers. A multi-scale community detection algorithm was used to identify providers with denser connections. This algorithm assigned providers to non-overlapping communities. This might be a limitation as healthcare providers may not belong to a single team (community). Despite the limitation, the algorithm is useful for detecting providers with higher number of interactions. Another limitation of the study is identifying the teams of healthcare providers. Although the multi-scale community detection helps to identify groups of healthcare providers (nodes) more closely connected, these communities may not be the reflection of the real teams. Further investigation perhaps with qualitative studies might be needed to confirm the ability of algorithms for identifying the care teams. Finally, studies are needed to determine the associations of SNA metrics and community composition with patient health outcomes.

3.5 Conclusion

This study demonstrated a novel approach for identifying key stakeholders, working relationships, and composition of care teams for patients with chronic conditions. A multi-year analysis was performed to understand the consistency and changes of the study provider networks. The long-term goal of this research is to translate the SNA and community detection framework for designing strategies for improving provider collaboration and assessing how these relationships impact patients’ health outcomes and healthcare services costs. Although a few studies have looked at the associations between patient outcomes and the providers’ network characteristics, there is a lack of research for assessing relations of these networks to patient safety outcomes, including measures of adverse events [117]. In the next study, using the presented framework and measures from this chapter, we assess how the provider network characteristics can predict the patient outcomes.

4. IMPACT OF CARE PROVIDER CHARACTERISTICS AND ON CHRONIC PATIENTS OUTCOMES: USAGE OF SOCIAL NETWORK ANALYSIS

4.1 Introduction

Chronic conditions are the leading causes of death and disability that result in $3.3 Trillion in annual healthcare costs in the United States [118]. Six in ten Americans have at least one chronic condition, and four in ten suffer from two or more [119]. Patients with comorbidities such as hypertension, hyperlipidemia, and diabetes are at higher risk of developing additional chronic conditions including heart disease [120], [121]. Care of patients with multiple chronic conditions are often complicated with greater healthcare needs that require a larger number of healthcare providers that are typically not co-located at a single health institution [122],[123] . Collaborative approaches for management of chronic conditions are key to ensure effective healthcare delivery and prevent health deterioration and adverse outcomes for these patients [124], [125].

Although the definitions of collaboration in healthcare vary, all definitions encompass themes such as coordination, cooperation, team, shared-decision making, and partnerships [18]. Many studies have suggested that these themes impact patient care and outcomes [124], [72], [126]; however, current tools and approaches have limited ability to systematically predict the effect on patient outcomes. To measure collaboration and its effectiveness, survey tools have been the primary approach. Several tools have been validated for use to assess the collaboration and teamwork of healthcare providers [127],[128], [129]. For example, the Doctors Opinions on Collaboration (DOC) tool was designed for physicians to ask about their beliefs, quality of communication, and attitude toward working with others in the care teams. The findings showed that both specialists and general practitioners were positive toward collaboration and communication, however, compared to specialists, general practitioners had better collaborative relations [130]. Other studies that used survey tools showed the positive impact of collaboration among providers for increasing patient safety in hospitals [131] and reducing patient mortality [132]. Survey tools can capture individual providers perspectives and rapidly scale up to capture information from a large number of participants [133], however, the process of developing and disseminating the survey can be time consuming and costly, the response rate is usually low

[20], and generalizability is difficult as surveys are validated for specific healthcare settings [129]. To better assess the impact of providers collaboration on patient care and outcomes, new approaches are needed to study the healthcare provider relations, team structures, and the associated patient outcomes.

Usage of quantitative approaches on health claims data has been widespread in recent years, however, it has been limited to individuals' health services utilizations and costs [22],[24],[23]. Application of social network analysis (SNA) to large-scale health data expand previous work by providing the ability to assess the collaboration of healthcare providers [117],[77]. Recent studies have demonstrated the validity of SNA approach as a technique that can identify working relations between healthcare professionals from large-scale claims data [21],[25]. Researchers have used SNA measures to assess collaborative relations among healthcare providers and the impact on patient outcomes [117], [34]. For example, Barnett et al. used SNA to generate a network of healthcare providers based on their patient sharing relations for hospital referral regions to assess the network characteristics impact on cost and intensity of care. They found that hospitals with physicians whose patients received care from larger number of doctors (showing higher degree centrality) had higher spending compared to other hospitals [33]. Pollack et al. followed a similar approach for generating collaborative networks of healthcare providers and developed the care density measure to assess extent of patient sharing among an individual's ambulatory care providers. Patient whose providers shared higher number of patients (higher care density) had lower rates of adverse events compared to other patients [30]. While the SNA measures help to understand the characteristics of the providers in the overall network, they do not provide a granular assessment of the provider team structures.

To address this gap, researchers have modified the SNA technique by adding community detection algorithms [134], [135], [136]. Landon et al. used the SNA and a community detection algorithm to assess physician characteristics in the network and within smaller communities (i.e., sub-dividing a one large network of the dataset into smaller networks of communities). Using this approach, they showed that patients whose physicians were connected with larger number of physicians had higher spending. Patients in communities with higher proportions of primary care providers had fewer specialist visits and fewer emergency department visits [32] . Another study used the SNA technique and community detection to investigate networks of physicians associated with multiple hospitals and identified how community structure impact patient outcomes [137]. Hospitals networks with higher number of nodes in each community had higher readmission rate. Higher number of nodes in the community was interpreted as physicians having harder time getting the patient information across the community which negatively impacted patient outcomes [137]. These studies demonstrated that community detection algorithms provide novel variables to predict patient outcomes, however, most algorithms used for example the Girvan-Newman method [84], [135], [138] suffer from a resolution limit and are unable to detect smaller communities in the network. Application of other algorithms which address the resolution limit problem of commonly used techniques including the Girvan-Newman has been recommended for assessment of providers teams in the network [79]. Another limitation in previous studies is that application of the community detection techniques has been limited to mainly physicians and hospitals networks [29],[33], [32], [30], [134] Although constraining the network to focus on the key provider simplifies the model and interpretation, it limits our ability to focus on coordination of complex care teams, which is common for patients with chronic conditions. Thus, inclusion of other providers (e.g., pharmacists, nurse practitioner) is necessary to explore flow of information and activities inside the network [34].

To address the aforementioned limitations, we used the social network analysis using the patient sharing approach and a multi-scale community detection algorithm to assess the relations between the SNA measures in the community and the associated patient outcomes. The study hypotheses are as below:

H 1 (Null). SNA centrality measures (degree, betweenness, and closeness centrality) of the care provider in the community do not impact patient inpatient hospitalization rates.

H 2 (Null). SNA centrality measures (degree, betweenness, and closeness centrality) of the care provider in the community do not impact patient unplanned hospitalization rates.

H 3 (Null). SNA centrality measures (degree, betweenness, and closeness centrality) of the care provider in the community do not impact patient emergency department visits rates.

H 4 (Null). The effect of the centrality measures (degree, betweenness, and closeness) of the care provider on the patient outcomes were consistent over the period from Year 1 (2014) to Year 2 (2015).

4.2 Materials and Methods

4.2.1 Data Source

This study was approved by the institutional review board (IRB 1511016796). The study data included three years-2014, 2015, 2016- of de-identified claims data from employees (faculty/staff) of a large university in the Midwest. Student population was excluded from this study due to different health service plans and coverages. The claims data contained insurance eligibility information, medical and medication services, health services costs, healthcare providers' information, and patient outcomes based on the Johns Hopkins Adjusted Clinical Groups version 11.0 [139].

4.2.2 Study Cohort

The study cohort included patients identified with diabetes, hypertension, and hyperlipidemia. These conditions are leading risk factors for heart disease, stroke and frequently happen together [140], [121]. We tracked the cohort healthcare utilization and providers over a three-year period from 2014 to 2016.

4.2.3 Study Design and Analysis

The study cohort was identified based on the International Classification of Diseases 9th edition code (ICD9). The first, second, third, fourth, and fifth ICD9 codes were used to identify the cohort. Patients were categorized as a patient with diabetes if one of their diagnosis codes started by 250 [89], as hypertensive if one of the codes started with 401[141] and with hyperlipidemia if one of the codes started with 272 [142]. Health service utilization and associated providers of the patient cohort were identified from the medical file and medication file (e.g., pharmacists can be identified from the medication file). Records of the patients and their providers were extracted over a three-year period from 2014 to 2016.

4.2.4 Constructing the Network

Previously we showed that patient sharing among healthcare providers can be used to assess the team structure of patients with diabetes [143], [103]. A similar approach was used here to generate the network of healthcare providers and assess the impact of network characteristics on patient outcomes. The patient cohort with diabetes, hypertension, and hyperlipidemia was identified based on the ICD9 codes as described above. All providers for the patient cohort were identified for all three years of the study. Separate provider networks were generated for Year1 (2014) and Year2 (2015). The network nodes represented healthcare providers and edges represented the patients-sharing relationship. Number of patients shared between healthcare providers was interpreted as the edge weight. Only providers who shared two patients or more were included; previous validation work suggest patient sharing <2 have happened by chance without significant information sharing value [143].

4.2.5 Network Communities

To do a granular assessment of the network a multi-scale community detection algorithm was applied on the network [94]. This multi-scale algorithm was previously shown to successfully identify communities of healthcare providers closely working together [143] at better resolution compared to algorithms used in previous work as mentioned in the introduction. Specifically, previous community detection algorithms suffer from a resolution limit and cannot identify communities with edges fewer that $\sqrt{L/2}$ where L is the number of edges in the entire network [94]; a critical limitation when the goal is identifying smaller health teams from large health datasets. Using this multi-scale algorithm, communities in the network were defined as groups of healthcare providers that had denser internal connections compared to their connections with the rest of the network [91]. This approach assigned providers to distinct communities, i.e., each provider appeared once and associated with only one community.

4.2.6 Assigning Patients to Providers

Every individual patient was assigned to a “majority source of care provider,” defined as the provider who was responsible for majority of services to the patient during a one-year period based on the number of visits. This provider was chosen based on the Johns Hopkins ACG Systems which defines the majority source of care as a provider who can reasonably manage care of patients [139]. Example of eligible providers considered as majority source of care are family medicine, internal medicine, ophthalmologist, and gynecologists vs. those not considered eligible for managing the patient care including ambulance services, agencies, dentists facilities, laboratories, and medical device suppliers (Appendix). We grouped general practitioner, family practice and internal medicine as primary care. Providers such as cardiologists, urologists, and endocrinologists were grouped as specialists. Other providers identified (e.g., chiropractors, physical therapists, and diagnostic radiology) were grouped as other providers. Patients were assigned to the communities that their majority source of care belonged to. We identified the common majority source of care providers in the network of Year1 (2014) and Year2 (2015).

4.2.7 Network Measures of Interests

We focused on SNA centrality measures specifically betweenness, closeness, and degree centrality as we were interested to learn how the provider's connectedness, access, and control over the flow of the information in the community impacted the patients' outcomes. Three measures of centrality 1) degree 2) betweenness, and 3) closeness were calculated, defined as follows:

- Degree centrality of the providers showed the direct connections that providers had with other providers in the network [99].
- Betweenness centrality showed the degree to which a provider (node) was between the shortest paths connecting other nodes. Providers with higher betweenness centrality are shown to have more influence in network for dissemination of information [33] (equation 1)

$$g(X) = \sum_{B \neq v \neq t} \frac{\sigma_{st}(x)}{\sigma_{st}} \tag{1}$$

- Closeness centrality of the providers in the network: Providers with higher closeness centrality have better access to other providers in the network [98]. (equation 2)

$$C(X) = \frac{1}{\sum_y d(x, B)} \tag{2}$$

We calculated these centrality measures for all providers in the network. The degree, betweenness, and closeness centrality of the majority source of care provider assigned to each patient was identified in 1) the overall network and 2) its community. Specifically, centrality measures were first calculated for the majority source of care provider in the *entire overall network*. Then, the centrality measures were calculated at the *smaller community network level*, where the community described provider's position among other providers in the community.

4.2.8 Patient Outcomes of Interest

Patient outcomes were defined based on the Johns Hopkins Adjusted Clinical Groups [139] and included: 1) emergency department visit, 2) unplanned inpatient hospitalization, and 3) inpatient hospitalization. Unplanned in-patient hospitalization refers to all hospitalizations that were not related to a definitely planned or a potential planned procedure (e.g. cardiovascular, hip replacement). Emergency visit count considers visits to the emergency department which were not precursor to subsequent hospitalization. The emergency department visits that were followed by a hospitalization were absorbed by a hospitalization [139]. Therefor all outcome are mutually independent by definition. Patient outcomes of interests were extracted for Year2 (2015) and Year3 (2016).

4.2.9 Statistical Analysis

Descriptive statistics were used to describe the characteristics of the patient and provider cohorts. Multiple regression models with different distributions were used to find the best distribution fit for the outcome variables. For modeling outcome variables represented as count measures, the most common distributions are Poisson and negative binomial. Negative binomial extends the Poisson models by allowing the outcome variance and mean to be different [144]. In addition to the negative binomial, application of zero-inflated versions of this model is also recommended when the outcomes are skewed due to larger number of zeros compared to other numbers [145]. To decide between the zero-inflated negative binomial and negative binomial distributions, we used the Vuong' closeness test tests the null hypothesis that the two model are equally close to the true data generating process against the alternative hypothesis that on model is closer [146]. In addition to network measures of interests (degree, betweenness, and closeness), patient age, sex (male/female), and type of majority source of care providers were also entered in the model as control variables. Grouping of the majority source of care providers to primary care, specialists, and other providers helped to reduce different levels of these categorical variables and to avoid non-convergence. Due to different ranges of the network variables, we standardized all the continuous variables in the model.

To test our hypotheses about the impact of community-level centrality of the care provider on patient outcomes, we used the generalized estimating equations (GEE) models. Separate models were generated for each outcome (emergency department visits, inpatient hospitalization, and unplanned hospitalization). As GEE models do not make any assumption about the distribution of the data, they provide more robust results compared to other models, e.g., hierarchical models [32]. In addition, GEE allows to account for clustering of observations and similarity between patients that belong to the same community (patients assigned to the community of their majority source of care provider). To account for this clustering GEE fits marginal regression models with variances adjusted for clustering [147]. The predictors included community and network level centrality (degree, betweenness, and closeness), patient age, sex (male/female), and type of majority source of care providers. The interaction of the centrality measures with year were also considered to account for the longitudinal analysis. Due to different ranges of the network variables, we standardized all the continuous variables in the model.

The β coefficient for each independent variables (community and network level centrality (degree, betweenness, and closeness), patient age, sex (male/female), and type of majority source of care providers) can be interpreted as a change in the outcome variable of interest for each standard deviation change in the independent variable, representing the standardized effect size. The GEE model is described in equation 3.

Model: Negative Binomial Regression

$$g(\mu_i) = X_i^T \beta \tag{3}$$

Where

$$X_i^T \beta = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_d x_d$$

With link function

$$g(.) = \log(.)$$

$$\mu_i \sim Negative\ binomial\ distribution$$

Where i represents the i^{th} subject

The relations between independent and outcomes variables where modeled for network characteristics in Year1 and outcome in Year2 and for network variables in Year2 and outcome variables in Year3. This one year lag was considered to account the time it took for the provider collaboration impact to show in the outcome [30], [32]. The analysis was completed using SAS (v 9.4, SAS Inc., Cary, NC) and RStudio (version 0.99.903) with the igraph (version 1.1.2) [100] and devtools (version 1.12.0) [101] packages.

4.3 Results

A total of 19,247 patients were identified in Year1 (2014) with at least one medical claim out of 23,631 individuals with insurance plans (81% of the population). Out of those patients, 4,395 patients were identified with diabetes, hypertension, or hyperlipidemia. We excluded patients who were not present for all three years. The cohort received healthcare services from 2,332 providers in Year1, 2,421 providers in Year2, and 2,444 providers in Year3.

4.3.1 Network Characteristics

Out of 2,332 providers in Year1 only 38% (N=894) shared two patient or more. The network of Year1 had 894 nodes (providers) with 8,853 edges. The biggest components of the network had 890 nodes and 8,851 edges. The two other components had each 2 nodes and one edge. As the network only included providers with two patients or more in common, these smaller components were generated as those providers (nodes) only shared one patients with the providers in the bigger component of the network. Focusing on the biggest component of the network, the median degree centrality of the network nodes was 7, the median betweenness centrality was 146.86, and the median closeness centrality was 0.000193.

The network of Year2 (2015) had 930 nodes with 9,631 edges. The biggest component of this network had 924 nodes and 9,624 edges. Similar to network of Year1, two smaller components with 4 nodes and 6 edges, and 2 nodes and 1 edge were also generated. In the biggest component of the network, the median degree centrality, betweenness centrality and closeness centrality were similar to Year 1 and were 8, 132.94, and 0.000188, respectively.

4.3.2 Network Communities

To conduct a more granular assessment of the full SNA network, we applied a multi-scale community detection algorithm [92] on the biggest component of the network of Year1 and Year2 to separate the large network into smaller communities, furthering our ability to understand meaningful collaborations among providers. Nineteen communities were detected for the network of Year1 with two to 193 nodes in each community. Twenty-one communities were detected for the in the full network of Year2 with two to 197 nodes in each community.

4.3.3 Analysis of Patient Outcomes

From these networks, 212 and 214 unique providers were determined as majority source of care providers in Year1 and Year2 respectively. Of these, 167 majority source of care providers were common between the network of Year1 and Year 2. Fifty-two percent (N=86) of these providers were primary care (general practice, family medicine, internal medicine). Forty-three percent (N=72) were identified as specialists (e.g. urologist, ophthalmologist, cardiovascular), and only 5% (N=9) were identified as “other providers” (e.g. social worker, chiropractor). Patients of these providers and their outcomes were extracted for testing our hypotheses. There were 4,230 patients extracted; 2,113 females with average age of 55.78±9.53, and 2,117 males with average age of 56.37±10.8. Table 4.1 describes the outcome variables in the dataset for Year1 and Year2.

Table 4.1 Description of the outcome variables

	Inpatient Hospitalization (%)	Unplanned Hospitalization (%)	Emergency Department (%)
Year 1*	128 out of 4230 (3 %)	109 out of 4230 (2.5%)	612 out of 4230 (14%)
Year 2*	136 out of 4230 (3%)	134 out of 4230 (3%)	608 out of 4230 (14%)
* Number of patients that had that outcome at least once			

As it can be seen from the table, percentage of the non-zero values are much smaller compared to the zero values, which causes the skewness of the data. To model such outcome variables the negative binomial distribution was chosen which relaxes the Poisson assumption of equality for mean and variance of the outcome variable. Figure 4.1 shows the distribution for the three outcome variables A) inpatient hospitalization B) unplanned hospitalization, and C) emergency department visits.

A)

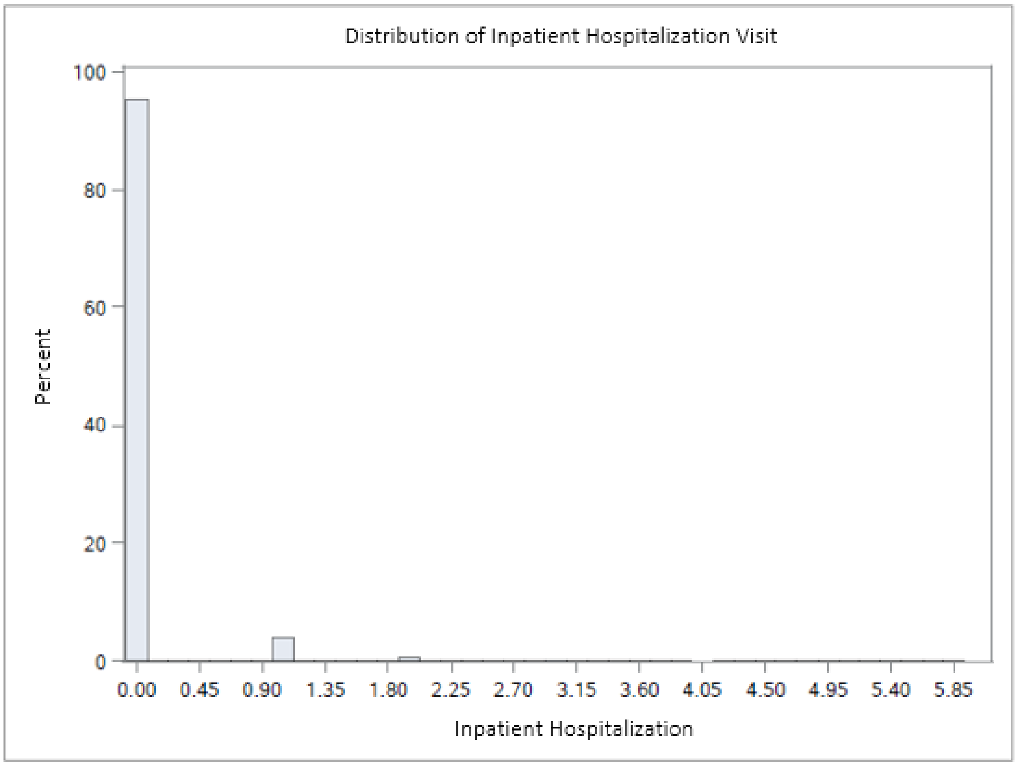
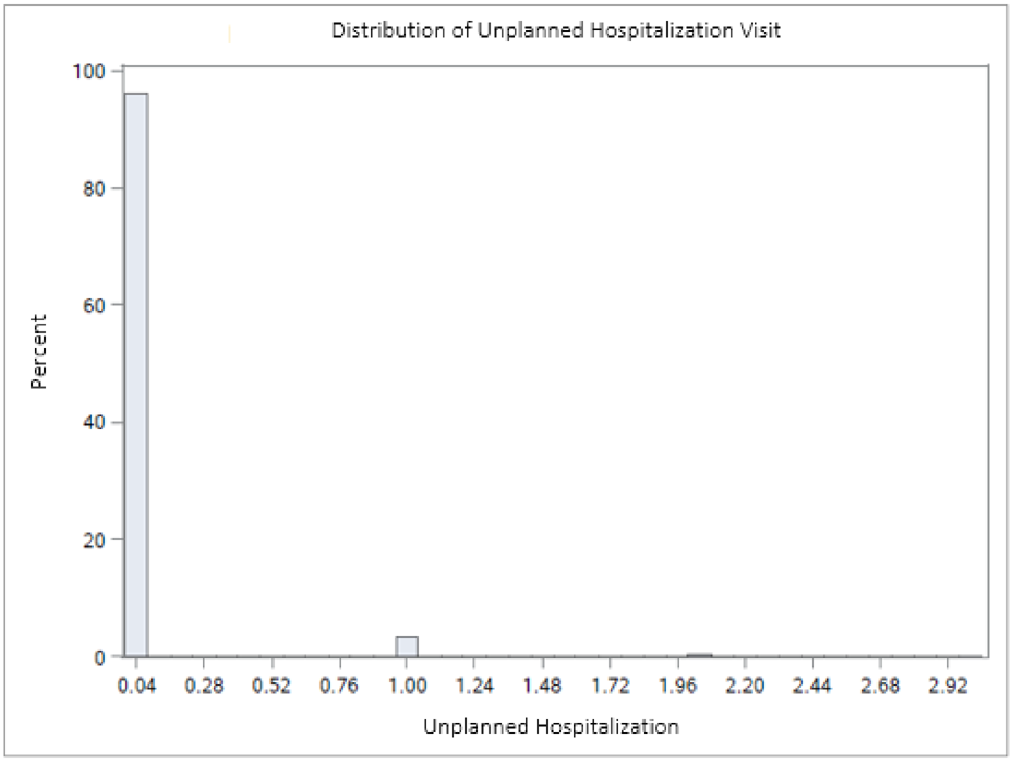


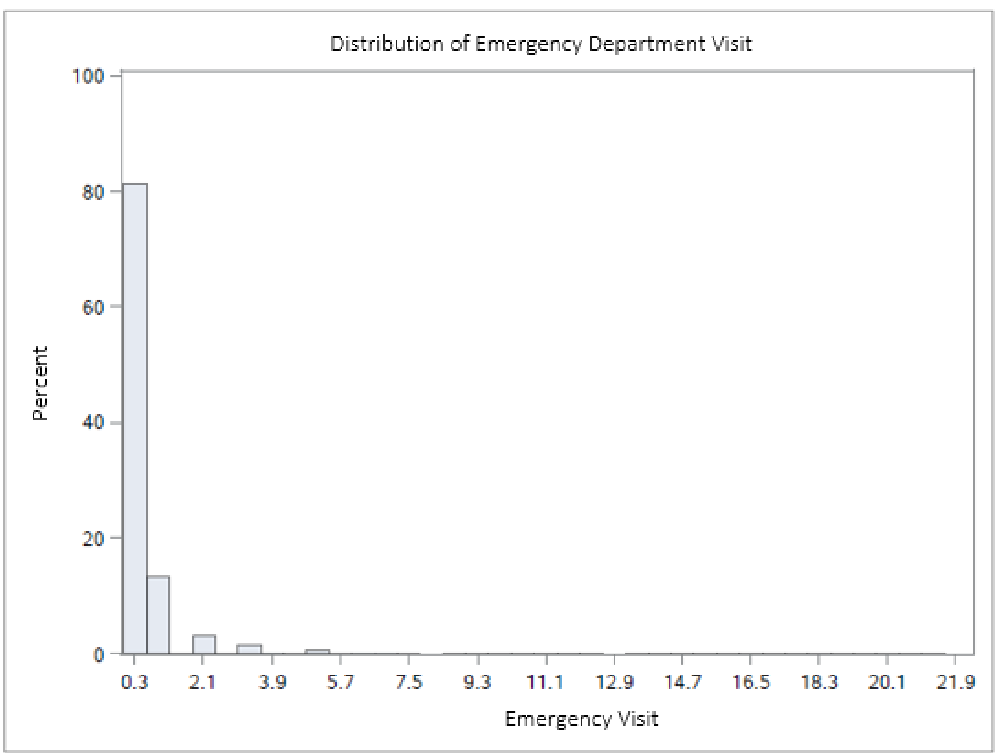
Figure 4.1 Distribution of the outcome variable used in the GEE process for A) inpatient hospitalization B) unplanned hospitalization, and C) emergency department visit. The graph shows a curve-formed distribution that matches negative binomial distribution

Figure 4.1 Continued

B)



C)



To model the relation between provider network measures of collaboration and patient outcomes, generalized estimating equations generated marginal negative binomial models (distribution determined using Vuong' closeness test) after adjusting the variance structure for patient's community clustering. Models first used both majority source of care providers' *network-level* centrality (network degree, betweenness, and closeness) and *community-level* centrality (community degree, betweenness, and closeness). *Network-level* measures were not statistically significant for any of the outcome variable models. Thus, models in the following results used only *community-level* centrality measures. Table 2 to 4 present the results of the GEE models for unplanned hospitalization, inpatient hospitalization, and the emergency department visits. The estimate of the independent variable is the difference in logarithm of the outcome variable when independent variable increases by one of standard deviation. The exponential of the estimate (Exp (estimate)) is the rate of change of the outcome variable when the independent variable increases by one standard deviation (i.e., back-transformed estimates).

Table 4.2 to 4.4 present to result of the GEE models for unplanned hospitalization, inpatient hospitalization, and the emergency department visits. The estimate is the difference in logarithm of outcome where independent variable increases by one unit of the standard deviation. The exponential of the estimate is the rate of change of the outcome variable when the independent variable increases by one unit of the standard deviation.

Table 4.2 represents the result of the GEE models for unplanned hospitalization outcome. Among control variables, patient age and sex were not significant, however, type of provider (i.e. primacy care, specialists, or other providers) was identified as significant (p-value < 0.05). Patients of providers classified as others and specialists had 2.47 and 2.59 times more unplanned hospitalization compared to patients of primary care providers. Highest number of patients with at least one unplanned hospitalization belonged to providers classified as other (N=110 for Year1 and Year2) followed by patients of specialists (N=67 for Year1 and Year2) and primary care providers' patients (N=56 for Year1 and Year2).

Table 4.2 GEE model results for unplanned hospitalization outcome

GEE Regression Models For Unplanned Hospitalization					
Variable	***Estimate***	***Exp (Estimate)***	***95% Confidence Limits***	***Z-Statistic***	***P-value***
Intercept	-4.11	0.016	(-4.470, -3.757)	-22.6	<.0001
Age	0.19	1.209	(-0.008,0.403)	1.88	0.060
Sex (female)	-0.039	0.961	(-0.197,0.117)	-0.5	0.619
Provider type (other)	**0.907**	**2.476**	**(0.352,1.462)**	**3.2**	**0.001**
Provider type (specialist)	**0.954**	**2.596**	**(0.536,1.371)**	**4.48**	**<.0001**
Community betweenness	0.112	1.118	(-0.092,0.317)	1.08	0.281
Community degree	-0.211	0.809	(-0.52,0.096)	-1.35	0.178
Community closeness	**-0.245**	**0.782**	**(-0.472,-0.018)**	**-2.12**	**0.034**
Year (year one)	**-0.142**	**0.867**	**(-0.263,-0.021)**	**-2.3**	**0.021**
*Community betweenness * year*	0.097	1.101	(-0.177,0.373)	0.7	0.486
*Community degree * year*	-0.177	0.837	(-0.45,0.095)	-1.28	0.201
*Community closeness * year*	-0.233	0.792	(-0.465,-0.001)	-1.97	0.04
Significant variables are bold					

Community closeness of the majority source of care provider was significant and higher closeness was associated with lower unplanned hospitalization. In addition to significance of closeness, there was an interaction between the provider's closeness in the community and Year. Specifically, the impact of closeness had different magnitude of effect on unplanned hospitalization in Year1 vs. Year2 (Figure 4.2). Higher provider closeness in the community was associated with lower rate of unplanned hospitalization. When closeness increased by 1 standard deviation of the closeness, the unplanned hospitalization increased by 0.78 for second year and 0.62 for the first year (exp (-0.24-0.23)). Rate of increase smaller than 1 (0.62 and 0.78) represent reduction in rate of unplanned hospitalization.

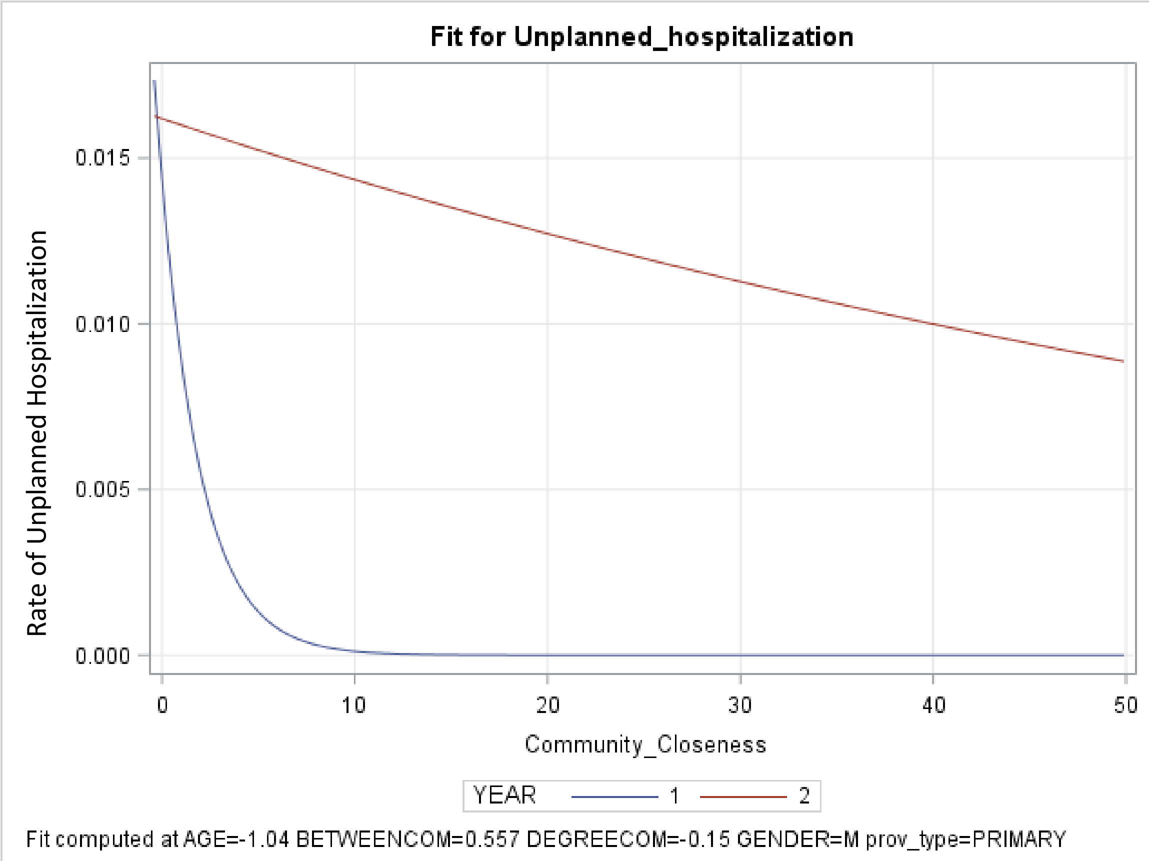


Figure 4.2 Provider closeness in the community and its effect on unplanned hospitalization in Year 1 vs. Year 2. The x-axis is the closeness of the provider in the community and the y-axis shows the predicted unplanned hospitalization rate. The graph was produced from the standardized data.

Table 4.3 shows the GEE regression results for the inpatient hospitalization outcome. Among the control variables, patient age and type of majority source of care provider (primary, specialists, or other providers) were significant (p-value<0.05). For 1 standard deviation increase in age, rate of inpatient hospitalization would increase by 1.27. Specialists and other types of majority source of care had 3.04 and 2.49 times inpatient hospitalization among their patients compared to the primary care providers. Highest number of patients with at least one inpatient hospitalization belonged to providers classified as other (N=117 for Year1 and Year2) followed by patients of specialists (N=73 for Year1 and Year2) and then primary care providers patients (N=62 for Year1 and Year2).

Table 4.3 GEE model results for inpatient hospitalization outcome

GEE Regression Models for Inpatient Hospitalization					
Variable	***Estimate***	***Exp (Estimate)***	***95% Confidence Limits***	***Z-Statistic***	***P-value***
Intercept	-4.45	0.011	(-4.408, -3.682)	-21.82	<.0001
Age	0.244	1.276	(0.031,0.457)	2.25	0.0246
Sex (F)	-0.048	0.953	(-0.189,0.093)	-0.67	0.504
Provider type (other)	**0.915**	**2.496**	**(0.301, 1.529)**	**2.92**	**0.003**
Provider type (specialist)	**1.112**	**3.040**	**(0.769,1.454)**	**6.37**	**<.0001**
Community betweenness	**0.27**	**1.309**	**(0.037,0.398)**	**2.37**	**0.018**
Community degree	**-0.347**	**0.706**	**(-0.681, -0.013)**	**-2.04**	**0.041**
Community closeness	**-0.376**	**0.686**	**(-0.648,-0.105)**	**-2.72**	**0.006**
Year	-0.075	0.927	(-0.174,0.023)	-1.49	0.35
*Community betweenness * year*	0.093	1.097	(-0.195, 0.382)	0.64	0.524
*Community degree * year*	-0.223	0.792	(-0.509,0.061)	-1.54	0.124
Community closeness * year	**-0.359**	**0.698**	**(-0.632,-0.086)**	**-2.58**	**0.010**
Significant variables are bold					

All three centrality measures were statistically significant. Higher provider degree (connectedness) was associated with reduced the inpatient hospitalization (negative estimate). As degree increases by 1 standard deviation, the inpatient hospitalization increases by 0.7. Higher provider betweenness (control over flow of information) was associated with increased inpatient hospitalization (positive estimate). As betweenness increases by 1 standard deviation, inpatient hospitalization increases by 1.31.

Similar to unplanned hospitalization, closeness of the care provider in the community was significant and interacted with Year, meaning the effect of closeness on the inpatient hospitalization in both years was significant but the effect differed from Year1 to Year2 (Figure 4.3). Higher closeness in the community was associated with lower inpatient hospitalization rate for both years (Table 4.3), however, in Year1, higher closeness reduced patient inpatient hospitalization more than higher closeness in Year2.

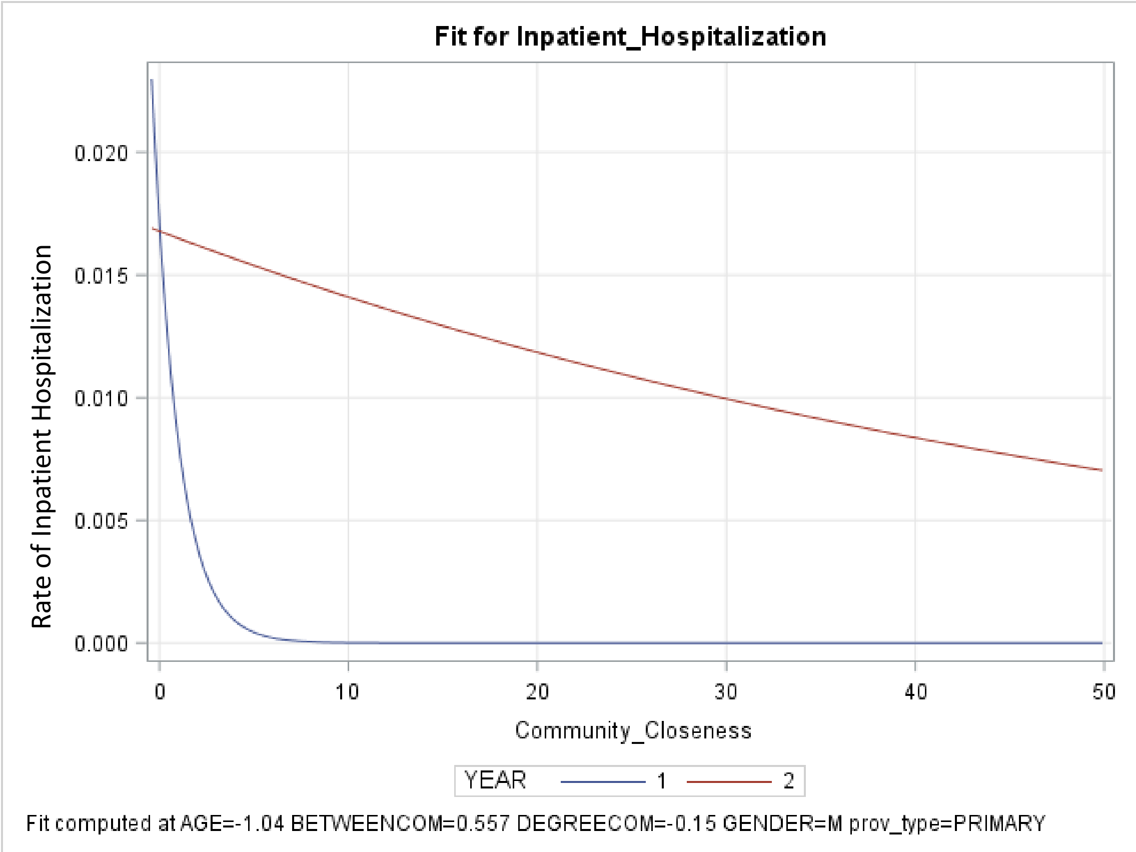
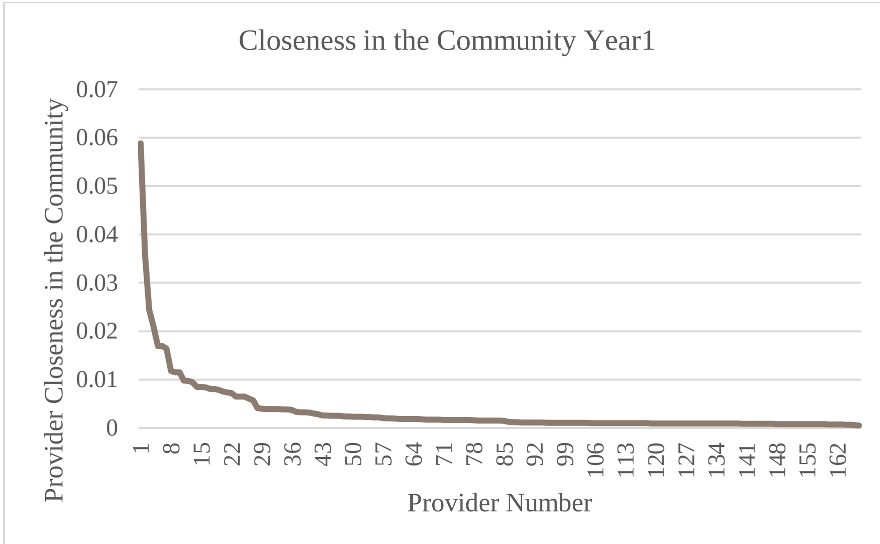


Figure 4.3 Provider closeness in the community and its effect on inpatient hospitalization in Year1 vs. Year2. The x-axis is the closeness of the provider in the community and the y-axis shows the predicted inpatient hospitalization rate. The graph was produced from the standardized data.

Figure 4.4 A and B compare the distribution of the provider closeness in the community for Year1 vs. Year2 based on unstandardized data (different scale from Figure 4.2 and 4.3) and from the highest to the lowest. Closeness of a node presents how far the node is from other nodes in the network (distance based on the shortest path). Similar patterns can be seen for both years. A smaller number or providers (<7) have much higher closeness compared to the rest of the providers that represents their higher access to other providers (nodes) in the network.

A)



B)

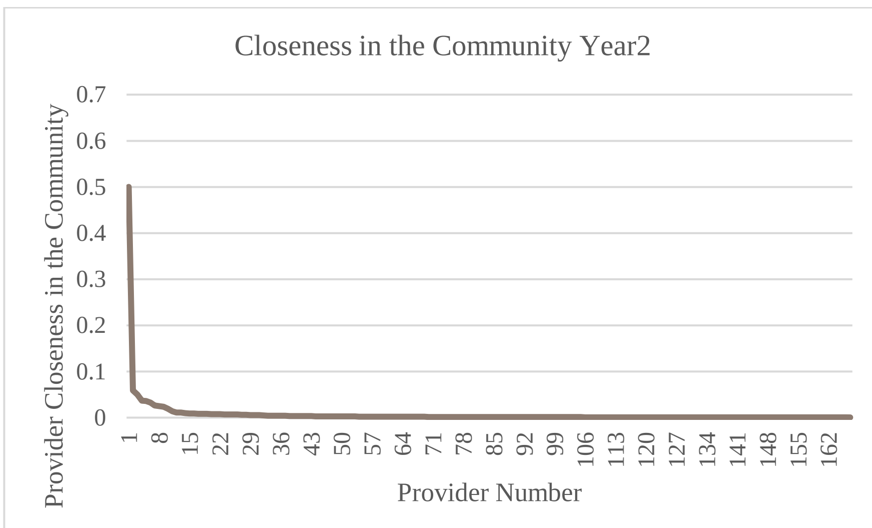


Figure 4.4 Distribution of the care manager closeness in the community for Year1 (A) vs. Year2 (B) based on unstandardized data

Table 4.4 shows the result of the GEE model for the emergency department visits outcome. Similar to inpatient hospitalization and unplanned hospitalization, type of majority source of care provider significantly impacted the outcome. Patients with specialists and other types of providers had 1.47 and 1.42 times higher emergency department visits compared to patients of primary care providers as their majority source of care. Highest number of patients with at least one emergency department visit belonged to providers classified as other (N=449 for Year1 and Year2) followed by patients of specialists (N=378 for Year1 and Year2) and primary care providers patients (N=249 for Year1 and Year2).

The only significant community measure for emergency department visits was the majority source of care betweenness in the community. When betweenness increased by 1 standard deviation of betweenness the emergency department visit rate increased by 1.1.

Table 4.4 GEE model results for emergency department visits outcome

GEE Regression Models for Emergency Department Visits					
Variable	***Estimate***	***Exp (Estimate)***	***95% Confidence Limits***	***Z-Statistic***	***P-value***
Intercept	-4.029	0.017	(-4.405,-3.653)	-21.00	<.0001
Age	-0.085	0.918	(-0.221,0.051)	-1.22	0.222
Sex (F)	-0.010	0.990	(-0.096,0.075)	-0.24	0.812
Provider type (other)	**0.353**	**1.423**	**(0.189,0.518)**	**4.21**	**<.0001**
Provider type (specialist)	**0.391**	**1.478**	**(0.143,0.639)**	**3.10**	**0.002**
Community betweenness	**0.101**	**1.106**	**(0.019,0.182)**	**2.44**	**0.014**
Community degree	-0.113	0.893	(-0.234,0.007)	-1.85	0.065
Community closeness	-0.036	0.964	(-0.098,0.025)	-1.15	0.248
Year	-0.028	0.972	(-0.087,0.030)	-0.96	0.336
*Community betweenness * year*	-0.026	0.974	(-0.109,0.056)	-0.63	0.531
*Community degree * year*	0.001	1.00	(-0.094,0.097)	0.03	0.975
*Community closeness * year*	-0.027	0.973	(-0.091,0.036)	-0.85	0.398
Significant variables are bold					

4.4 Discussion

This study presented a framework integrating social network analysis, a state-of-art community detection algorithm, and predictive modeling to provide an approach and metrics for assessment of providers' network characteristics and the impact on patient outcomes. Previous studies that used the SNA application limited their cohort of healthcare providers to mostly physicians specifically those associated with hospitals [33], [30], [134]. Depending on the type of condition, including other providers such as pharmacists, nurse practitioners, and dietitian is necessary to understand the structure of the network and information sharing patterns [34]. In this study, we focused on patients with diabetes, hypertension, and hyperlipidemia due to their similar healthcare needs and utilization. We considered the patients' care manager that provided majority source of care to the patient and assessed the provider's network characteristics on patient outcomes.

Based on the result from the GEE models, patients with specialists (e.g. cardiologist, ophthalmologists) or other providers (e.g. social workers) as their majority source of care provider had higher rate of emergency department visits, inpatient, and unplanned hospitalization compared to patients with primary care providers as their care managers. Primary care providers are identified as effective team members and typically the center of care for patients with diabetes [148], hypertension [149], and hyperlipidemia [150], therefore, they might be better than others at coordinating the team and managing the care process. This may partially explain the observed provider impact on outcomes. Another potential explanation for the higher rate of hospitalization and emergency department visits for patients with specialists as their care manager might be due to the more serious health problems that require management by specialist compared to patients with primary care as their care managers. Nevertheless, the study provides evidence that type of provider who manages the care of patients with chronic conditions can help to predict the patient outcomes.

Closeness of the majority source of care provider in the community was a significant predictor for number of planned and unplanned inpatient hospitalization of the patients; it had a consistently protective effect (negative) on poor outcomes. Higher closeness of a provider in a network represents higher access to other providers [98] and may suggest higher access to information or more familiar coordination among providers. As the care manager of the patient [139], this closeness (access) to other providers in the network and its community may be a predictive metric for care quality. For example, lack of information about patient's previous conditions, hospitalization, and emergency department visits have shown to negatively impact the patient hospitalization [151]. Thus, care managers with higher access to other providers may allow them to better coordinate and manage the care process and flow of information among other involved providers to reduce potential negative impact on patients.

Despite the important implication of provider greater closeness (access) to patient care, the impact on outcomes interacted with Year in the present study. Although this may suggest inconsistency and unreliability of closeness-centrality as a predictor of outcomes, we believe this interaction highlights the sensitivity to changes in care collaborations. Provider relationships and network characteristics are impacted by multiple factors outside of the study control variables. Specifically, factors like insurance coverages [30] can disrupt team collaborations. For the present study population, two major changes occurred in insurance coverages. First the Affordable Care Act was implemented on January 1st of Year1 which required all individuals to have insurance coverages. Second, a health savings plan was introduced in addition to the previously flexible spending accounts which allowed the insured individual's unused health funds to roll over to the next year. Potentially more people learned about the new plan options in the second year of the study which impacted their health utilization patterns and therefore the associated network structure and characteristics including provider closeness. Changes in the type of insurance plans may have impacted the network structure and provider outcome in second year. Thus, the interaction between year and closeness centrality requires further investigation. It is important to note that despite this interaction and difference in the magnitude of the impact, the direction of the closeness effect was consistently protective to poor outcomes.

Based on the regression results, providers with higher connectedness to others in the community (higher community degree) were associated with lower inpatient hospitalization. A physician higher degree indicates that the provider shares patients with higher number of providers. Previous studies have shown conflicting results about providers' degree. According to Barnet et al. higher degree of providers in the network was associated with higher patient cost and utilization of services [33]. In contrast, another study identified that provider's larger connectedness was associated with fewer adverse outcomes, and larger degree was associated with lower readmission rate after hospitalization for heart failure [152]. The present study showed that higher connectedness lowered risk for adverse events for the studied chronic conditions, which aligns with literature emphasizing the important of care collaboration and ease of dissemination of information which can positively impact the patient outcomes in chronic disease management. In addition, unlike previous work, the present study focuses on "community" connectedness while other previous work assess physician degree in the whole network [32], [33]. This suggests this relationship between connectedness and outcomes can be observed with datasets of smaller communities (identified with the multi-scale community detection algorithm).

In this study, we identified healthcare providers' centrality measures at both whole-network level and the more closely working together community-level. Previous studies that assessed relations between provider centrality (degree and betweenness) and patient outcomes explored those characteristics for the entire network [32], [33], [152]; however, the whole network centrality measures were not significant predictors for modeling of patient outcomes and only community-level metrics were significant for our dataset. This suggest the algorithm used in this study is able to identify smaller communities in the network that might be a better reflection of the care teams. The finer-scale community measures may provide better reflection of the provider characteristics and its associated impact on patients compared to those measures in the entire network. These more granular community predictors may be especially helpful for smaller datasets such as the present population.

This study has some limitations. Our dataset is limited to claims and provides limited information about the providers. We included provider type in the model, however, having other information about the providers might be helpful for more accurate assessment of their collaboration. Our network was generated based on patient sharing relationships. Although this approach is helpful for identifying working relations among healthcare providers from claims data, it might not be a reflection of real communication between the providers. We limited the network to patients with similar chronic conditions to ensure inclusion of providers with higher probabilities of working together. We used a multi-scale community algorithm to identify groups of healthcare providers more closely working together. Although, this algorithm addressed the resolution limit of commonly used algorithms in similar studies, the communities identified may not be reflections of real teams. Moreover, this algorithm assigned healthcare providers to disjoint communities, however, providers may work with different groups and belong to different communities. Overlapping community detection algorithms are complicated and may not perform well on larger networks. Developing simpler algorithms which identify overlapping communities may help to generate more accurate communities from healthcare providers' networks that are reflective of their real teams. Finally, we used a longitudinal analysis to study impact of provider's network and community centrality measures on patient outcomes. Due to limitations of claims data, we were unable to identify factors that impacted the network and community measures and their magnitude in various years. Further qualitative studies might be needed to validate factors that impact the significance of network factors and the impact on patient outcomes.

4.5 Conclusion

This study proposed a novel approach to identify network characteristics of healthcare providers involved in the care of patients with chronic conditions and the associated impact on patient outcomes. Longitudinal analysis of healthcare providers' networks helps to identify network characteristics and central providers over time. Interventions that target care managers in the network may help to improve their collaboration with other providers and the associated impact for patients.

5. DICUSSION AND CONCLUSIONS

5.1 Dissertation Overview

The objective of this dissertation was to develop new framework and metrics for 1) understanding the role of worksite clinics in the care teams for improving the population health 2) assessment of working relations among healthcare providers, and 3) analyzing the relations between healthcare provider collaborative networks characteristics and patient outcomes. Figure 5.1 describes the flow of the research and how the three studies are connected. Study 1 provided a framework for assessment of population health service utilization and the provider impact on utilization based on large scale data. Study 2 focused on the patient cohort from study 1 specifically patient with diabetes and their provider networks. A framework based on social network analysis and a multi-scale community detection was proposed to assess healthcare provider collaboration. Study 3 followed the patient cohort from study 1 and 2, focusing on the healthcare providers of patients with diabetes, hypertension, and/or hyperlipidemia. A framework based on social network analysis was presented to assess impact of provider collaboration on patient outcomes. Contribution of each study is described in the following sections of this chapter.

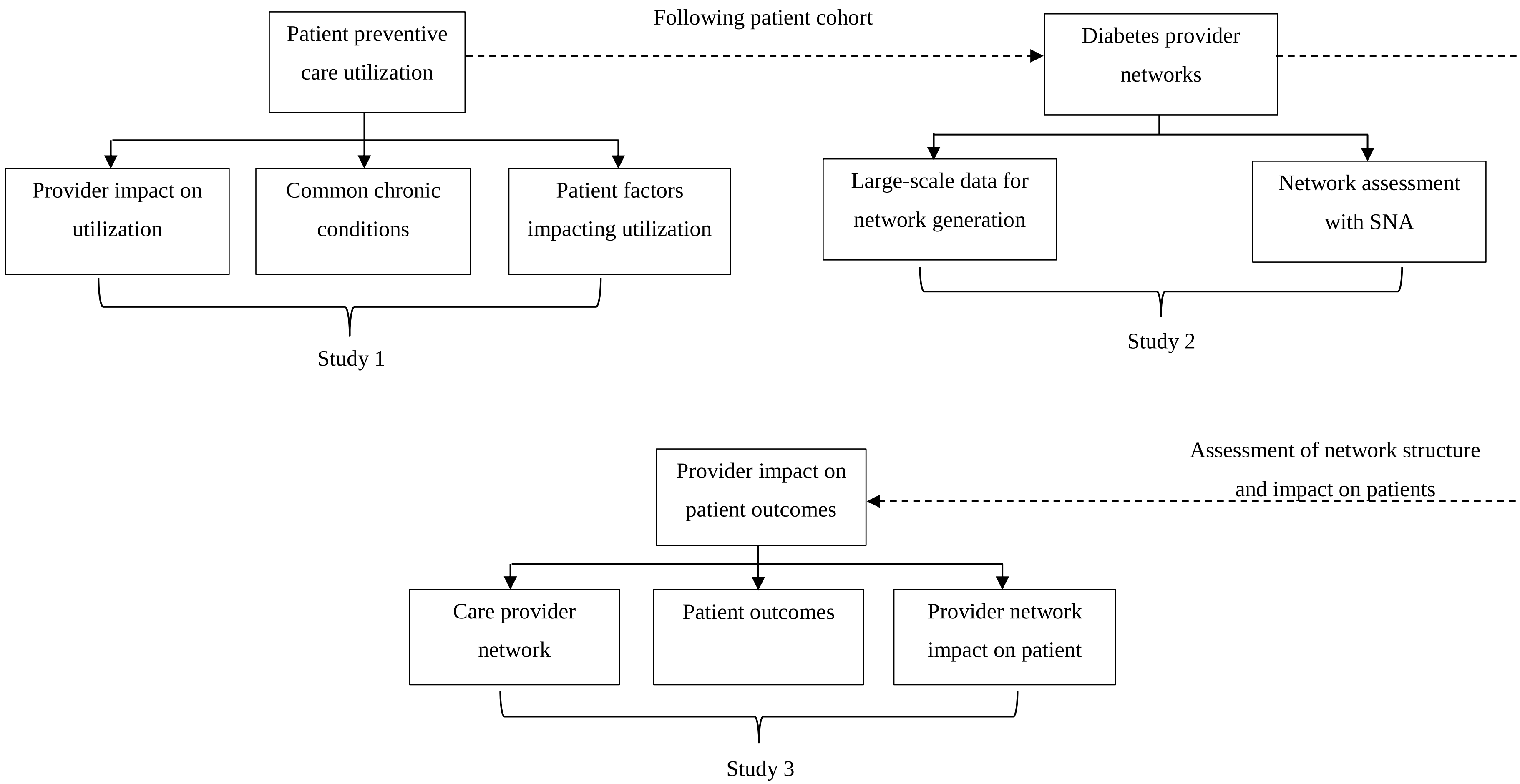


Figure 5.1 Description of the dissertation studies and their connection

5.1.1 Study 1: Preventive Health Utilization and Associated Factors for Patients with Chronic Conditions

5.1.1.1 Problem Addressed

Due to the increasing prevalence of chronic conditions and associated healthcare costs [8], [35] employers are seeking new ways for promoting healthy lifestyle and preventive care among their employees including worksite wellness programs and clinics [15]. Despite positive potential impacts from these programs and clinics, there are barriers affecting the employee utilization including work schedule conflicts or lack of alignment of services with employees healthcare needs [42], [13].Current studies assess the healthcare utilization of worksite clinics and programs using mostly surveys that are limited by low response rates and questions designed for specific setting [20] rather than the big picture and overall health service utilization patterns of the patient population.

5.1.1.2 Contribution

This study proposed a framework to quantitatively assess impact of worksite clinics as an intervention on patterns of preventive care utilization for a cohort of patients over a three-year period. Previous assessments of worksite clinic wellness are limited to financial perspective [36], [40] or employee’s attitude toward participation [13], [42] using mostly surveys and interviews with an incomplete picture of the overall population health needs and utilization. We enhanced the previous work by presenting a new framework based on large-scale claims data and machine learning techniques to provide a quantitative tool for assessment of worksite clinics utilization.

Besides assessment of worksite clinics utilization, the presented framework helps to manage large amount of data into valuable information and perspective for analysis of population healthcare needs. Although surveys exist to make inquiries from the population about their preferred types of services, retrospective analysis of healthcare data better reflects the population needs by identifying patterns of healthcare utilization and prevalence of chronic conditions. Therefore, it can leverage the impact of survey and interview tools by providing a comprehensive picture of the population and healthcare system structure. By adapting a framework based on large-scale medical claims and machine learning techniques, this study can be used for continuously monitoring changes in the population health utilization and investigating the role of worksite clinics and wellness programs.

5.1.2 Study 2: Modeling of Diabetes Care Teams Using Social Network Analysis and Administrative Data

5.1.2.1 Problem Addressed

Prevalence of diabetes is expected to increase over the next 40 years with one in three Americans projected to be diagnosed with diabetes [67]. Management of diabetes is often complex with variety of healthcare providers included in the team [69]. Although recommendation exists for diabetes care team structure and members [74], structure of teams may constantly change due to factors such as addition of new providers by the patient or patient family [29]. Moreover, the frequency that these recommended team structures happen in practice is unknown.

One of the main challenges in assessment of the care teams is difficulty in collecting data that measures these collaborations [21]. There are survey tools for assessment of these relations [18], [19] which are limited by low response rate and validation for specific populations [20]. Recently, application of social network analysis to large-scale health data is recommended to assess providers working relations [21], [25]. However, previous studies are mostly limited to physicians who provide direct care to patients and do not include providers such as pharmacists or nurse practitioners [33], [32]. Depending on the type of conditions, inclusion of other providers is essential to assess network structures and flow of information [34].

5.1.2.2 Contribution

In this study, we developed a framework based on the social network analysis and a multi-scale community detection algorithm [92] to provide a novel approach and metrics for quantifying the structures of teams and their relations in the care of patients with diabetes. Following the patient cohort from study 1, we expanded the scope of the first study from focusing on one individual provider to structure of healthcare provider networks and teams. Previous studies that used the patient sharing approach and social network analysis limited their networks to physicians usually affiliated with hospitals. Moreover, those previously used community detection algorithms suffered from a resolution limit and were unable to detect the hierarchical structure of the networks.

The proposed framework provides metrics and measure for assessment of provider collaborative relations some of which could be hidden to the researchers. For example, the SNA centrality measures identified the mail-order pharmacy as the influential provider in the network with higher connectedness to other providers. Although one may not initially consider mail order pharmacy as an effective provider in the diabetes care process. The social network analysis is leveraged by the multi-year analysis of large-scale health data which helps to understand the consistency and changes in the provider networks. Despite effectiveness of the approach for identifying provider collaborative relations and influencers in the network, the structure of the network might be impacted by external factors other than the providers for example design of the insurance coverages and plans. Further investigation of these factors using qualitative studies might be needed before implementation of the study findings.

The relations identified between the providers based on the patient sharing approach might be patient or provider driven. Patient or patient family may add or remove providers from the care team which would impact the healthcare utilization pattern and therefore the structure of provider networks. On the other hand, the provider relations might have formed based on the provider affiliation to an organization, insurance coverage plans or working in the same geographic region. Further investigations are needed to understand factors impacting the provider networks generation and how they are structured.

The study approach and framework enables health policy makers to make system level informed decision regarding the design of wellness programs and insurance coverage plans. Moreover, health service researchers can use the approach and tools proposed in the study to determine the healthcare team relationships and how the interactions impact the patient care process and outcomes.

5.1.3 Study 3: Care Provider Characteristics and the Impact on Chronic Patients Outcomes: Usage of Social Network Analysis

5.1.3.1 Problem Addressed

Prevalence of chronic conditions shows an increasing trend in the U.S. Currently six in ten Americans have at least one chronic condition while four in ten have two or more [119]. Care of patients with multiple chronic conditions is complicated with larger number of providers involved [122]. Collaborative approaches for management of chronic conditions are shown to positively impact patient care and outcomes [127]. Due to the limitation of previous tools (surveys, interviews) for assessment of provider collaboration impact on patient outcomes, application of quantitative approaches on large-scale claims data are increasing to provide a more comprehensive picture of the healthcare system and structure. Researchers have used the patient sharing approach among healthcare providers to generate networks of healthcare providers using social network analysis [25]. However, previous studies have limited the network to physicians usually with affiliation to specific health organizations like hospitals [30], [32], [33]. Moreover, the community detection algorithms generally used in this studies suffer from a resolution limit and are unable to identify the hierarchical structure of the networks [135].

In the previous chapter (study 2) we developed a framework based on the social network analysis and a multi-scale community detection for assessment of the provider collaboration and team structure. The framework provided useful measures for investigating provider network characteristics, however, it did provide tools and evidence for assessing provider collaboration impact on patient outcomes. To translate this framework into practice, in this study, we expanded our previous work to study provider network measure and their impact for prediction of patient outcomes.

5.1.3.2 Contribution

In this study we used the social network analysis to generate networks of physicians based on patient sharing relations and from the claims data. We expanded the scope of previous work by first including all provider types involved in the care of patients to provide a better picture of the network structure and information sharing patterns. Second, we used a multi-scale community detection algorithm which addressed limitations of previously used algorithms with resolution limits and inability to identify smaller communities in the network [94].

The study framework presents finer-scale metrics and assessment of the provider network characteristics and their impact on patient outcomes. We built this framework based on study 1 and 2 by expanding the measures and tools from only focusing on patient characteristics and single provider, to include provider collaborative networks and the impact on patient outcomes.

The provider metrics are identified and assessed on the community level rather than the entire network, therefore they better reflect the team interaction, structure, and the impact on patient care and outcomes. The longitudinal analysis of the provider networks is useful for monitoring the changes in the structure of the overall healthcare systems. Combined with the multi-scale community detection and multi-year analysis of the network, this approach helps to identify how the interaction of team members changes over the years. The application of these tools is leveraged by addition of patient outcomes and predictive modeling to assess how the provider network and community characteristics impact the patient care and outcomes.

5.2 Overall Contribution

As shown in Figure 5.1, in this dissertation, we present frameworks and metrics for assessment of patient needs, structure of the healthcare provider's teams, and the impact of provider collaboration on patient outcomes using the application of social network analysis, community detection algorithms, and predictive modeling on large-scale health data. With the increasing prevalence of chronic conditions and the associated healthcare costs, the structure of healthcare delivery systems need to evolve to adapt with the population needs and utilization [153]. Multiple players are involved in the patient care process including patient family, healthcare providers, and the employers. Each of these players has a critical role in the complicated care process of patients with chronic conditions.

In study 1, we focused on the role of employers and how they fit in the overall health care system. To reduce the healthcare costs and improve employee health, employers are using different tools and incentives to promote healthy living and reduce risky behaviors including worksite wellness programs and clinics. The role of onsite clinics is changing from only work-related care to a more comprehensive array of services. Despite the potential positive roles that these clinics may have for the employees, there are barriers and issues including employees schedule conflict to attend the wellness programs, different types of needs compared with what is provided in the clinic, and occasionally lack of trust in the worksite providers. Thus, more rigorous tools are needed to better design the worksite clinics structure and to improve their alignment with the population needs and utilization. The framework and tools in study 1 presents a multi-year analysis of population healthcare utilization and prevalence of health conditions. The approach is useful for need assessment of the population and informed design of worksite wellness and services which better align with population needs. Although the approach can be used to understand the patient population, it does not provide much information about the healthcare provider interaction and providers who are involved in the care process. To address this gap, the second study presents a framework and novel metrics to assess the structure of health teams involved in the care of patients.

The focus of the second study was on patients with diabetes from study 1 and their providers. We picked diabetes as one of the high prevalent chronic conditions of the patient population in study 1. A framework was developed for assessment of care teams using social network analysis and a multi-scale community detection. Although the claims data provides information in the patient level, it also provides some level of data about healthcare providers. Providers who have one or more patients in common form working relations that could be formal for example the referrals or informal, for example discussing health of patients. Using this patient sharing approach and social network analysis metrics, we generated and assessed network of healthcare providers involved in the care of patients with diabetes. While the social network analysis metrics helped to identify the central providers and influencers in the overall network, the community detection algorithm identified groups of healthcare providers more closely working together. The social network analysis and community detection framework helps to identify the structure of the healthcare systems, the influencers, and central providers in the overall network and in each community. Analysis of these networks over multi-years periods help to continuously monitor the changing structure and dynamics of the network. This would enable the health policy makers to make informed decisions regarding the design of insurance coverage plans and wellness programs. The missing piece in the study is how the provider network structure impacts the patient care and outcomes.

To address this gap, we performed the third study in which we used the social network framework from study 2 to first generate the network of healthcare providers; combined with patient outcomes data and predictive modeling, we assessed how provider social network characteristics could predict patient outcomes. While study 2 framework presented tools and metrics for assessment of provider network structure, study 3 leveraged study 2 by connecting the social network metrics with patient outcomes. The study approach is useful for health policy makers to assess the interactions of healthcare providers specifically the care managers with others in the network and how those interactions impact the patients.

5.3 Future Work

The research presented here provides new metrics and approach for assessment of patient needs and healthcare utilization, assessment of provider team structures, and the associated impact on patient outcomes. Despite the novelty of the approach and assessments, our work has some limitations that need to be assessed in the future studies.

Our data is limited to health claims which despite providing information about patients and providers, presents a limited perspective about patient outcomes and provider interaction. Usage of electronic health records can leverage the application of present framework by providing the measurable patient level outcome. The accuracy of the network generating algorithm from the claims data need to be improved so the identified networks are the true reflection of the provider working relations. This can be done by adding other elements to the patient sharing relations for example date of service or episode of care. Although longitudinal analysis of the networks and provider interactions are useful to continuously monitor the networks, the structure of the network might be impacted by factors that could not be identified from claims data for example patient or patient family role in changing the providers, or changes of insurance plans and coverages. Before implementing the perspective from this research, qualitative studies might be needed to identify other factors that could impact the interaction of providers.

5.4 Conclusion

This dissertation presented novel frameworks to assess patient population needs and utilization, impact of individual providers on healthcare utilization patterns, collaborative networks of healthcare providers, and their impact on patient outcomes. Using social network analysis, we provided metrics to measure healthcare provider collaboration from large-scale claims data. The study approach expands previous work by providing system-level collaboration metrics and measures. The long-term goal of this research is to translate the SNA and community detection framework for designing strategies for continuous improvement of provider collaboration and assessing how these relationships impact patients’ health outcomes and healthcare services costs.

APPENDIX

Majority source of care provider selection is based on the Adjusted Clinical Groups Measures and can be directly identified from the claims data by running the software [139]. Some of the eligible providers that are considered by these measures are as below:

1. Specialists: allergy and immunology, colon and rectal surgery, neurological surgery, neuro-musculoskeletal medicine, nuclear medicine, obstetrics & gynecologist, ophthalmologists, otolaryngology
2. Primary care: family practice, internal medicine
3. Other providers: nurse practitioners, physician assistant

Example of providers who are not eligible to be considered as the majority source of care (therefore care manager) are ambulance services, agencies, dental providers, facilities, psychologists, technicians, suppliers of durable medical equipment, and therapists.

REFERENCES

[1] National Center for Chronic Disease Prevention and Health Promotion (NCCDPHP), “About chronic diseases | CDC,” 19-Nov-2018. [Online]. Available: https://www.cdc.gov/chronicdisease/about/index.htm. [Accessed: 29-Jan-2019].

[2] American Diabetes Association, “Statistics about diabetes,” *American Diabetes Association*, 2018. [Online]. Available: http://www.diabetes.org/diabetes-basics/statistics/. [Accessed: 18-May-2018].

[3] G. A. Mensah and D. W. Brown, “An overview of cardiovascular disease burden in the united states,” *Health Affairs*, vol. 26, no. 1, pp. 38–48, Jan. 2007.

[4] S. M. Ornstein, P. J. Nietert, R. G. Jenkins, and C. B. Litvin, “The prevalence of chronic diseases and multimorbidity in primary care practice: a PPRNet report,” *J Am Board Fam Med*, vol. 26, no. 5, p. 518, Sep. 2013.

[5] W. J. Nilsen and D. H. Olster, “News from NIH: effective behavioral treatments for patients with multiple chronic conditions,” *Translational behavioral medicine*, vol. 3, no. 1, pp. 1–2, Mar. 2013.

[6] P. R. Orszag and P. Ellis, “Addressing rising health care costs — a view from the Congressional Budget Office,” *N Engl J Med*, vol. 357, no. 19, pp. 1885–1887, Nov. 2007.

[7] M. E. Salive, “Multimorbidity in older adults,” *Epidemiologic Reviews*, vol. 35, no. 1, pp. 75–83, Jan. 2013.

[8] “Health and economic costs of chronic disease | CDC,” 19-Nov-2018. [Online]. Available: https://www.cdc.gov/chronicdisease/about/costs/index.htm. [Accessed: 01-Feb-2019].

[9] “National health expenditures 2017 highlights.” Center for Medicare & Medicaid Services, 2017.

[10] A. B. Schultz, C.-Y. Chen, and D. W. Edington, “The cost and impact of health conditions on presenteeism to employers,” *PharmacoEconomics*, vol. 27, no. 5, pp. 365–378, May 2009.

[11] J. J. Collins, C. M. baases, C. Sharda, et al., “The assessment of chronic health conditions on work performance, absence, and total economic impact for employers,” *Journal of Occupational and Environmental Medicine*, vol. 47, no. 6, pp. 547–557, 2005.

[12] J. Reardon, “The history and impact of worksite wellness,” *Nursing Economics*, vol. 16, no. 3, p. 117, 1998.

[13] D. R. Bright, S. L. Terrell, M. J. Rush, et al., “Employee attitudes toward participation in a work site-based health and wellness clinic,” *Journal of Pharmacy Practice*, vol. 25, no. 5, pp. 530–536, 2012.

[14] B. W. Sherman and R. J. Fabius, “Quantifying the value of worksite clinic nonoccupational health care services: a critical analysis and review of the literature,” *Journal of occupational and environmental medicine*, vol. 54, no. 4, pp. 394–403, 2012.

[15] H. T. Tu, E. R. Boukus, and G. R. Cohen, “Workplace clinics: a sign of growing employer interest in wellness,” *Research Briefs*, vol. 17, pp. 1–16, 2010.

[16] V. Shahly, R. Kessler, and I. Duncan, “Worksite primary care clinics: a systematic review,” *Population Health Management*, vol. 17, no. 5, 2014.

[17] M. Ostovari, D. Yu, Y. Yih, and C. J. Steele-Morris, “Impact of an onsite clinic on utilization of preventive services,” *Journal of Occupational and Environmental Medicine*, vol. 59, no. 7, 2017.

[18] L. Morley and A. Cashell, "Collaboration in health care," *Journal of Medical Imaging and Radiation Sciences*, vol. 48, no. 2, pp. 207–216, Jun. 2017.
[19] M. Dougherty B. and E. Larson, "A review of instruments measuring nurse-physician collaboration," *The Journal of Nursing Administration*, vol. 35, no. 5, pp. 244–253, 2005.
[20] J. Kennedy and B. Vargus, "Challenges in survey research and their implications for philanthropic studies research," *Nonprofit Volunt Sector Qtly*, vol. 30, pp. 483–494, 2001.
[21] M. L. Barnett, B. E. Landon, A. J. O'Malley, N. L. Keating, and N. Christakis, "Mapping physician networks with self-reported and administrative data," *Health Services Research*, vol. 46, no. 5, pp. 1592–1609, 2011.
[22] A. Collie and K.-H. Prang, "Patterns of healthcare service utilisation following severe traumatic brain injury: An idiographic analysis of injury compensation claims data," *Injury*, vol. 44, no. 11, pp. 1514–1520, Nov. 2013.
[23] S. Hewner, Y.-P. Chang, Y. Xue, D. Somayaji, and S. Casucci, "Exploring Medicaid claims data to understand predictors of healthcare utilization and mortality for Medicaid individuals with or without a diagnosis of lung cancer: a feasibility study," *Translational Behavioral Medicine*, vol. 8, no. 3, pp. 400–408, May 2018.
[24] N. A. Elbers, P. Cuijpers, A. J. Akkermans, A. Collie, R. Ruseckaite, and D. J. Bruinvels, "Do claim factors predict health care utilization after transport accidents?," *Accident Analysis & Prevention*, vol. 53, pp. 121–126, Apr. 2013.
[25] B. E. Landon, N. L. Keating, M. L. Barnett, et al., "Variation in patient-sharing networks of physicians across the United States," *JAMA*, vol. 308, no. 3, pp. 265–273, 2012.
[26] J. Scott, "Social network analysis," *Sociology*, vol. 22, no. 1, pp. 109–127, 1988.
[27] J. L. Moreno, *Who shall survive? Foundations of sociometry, group psychotherapy and socio-drama, 2nd ed.* Oxford, England: Beacon House, 1953.
[28] E. H. DuGoff, J. Cho, Y. Si, and C. E. Pollack, "Geographic Variations in Physician Relationships Over Time: Implications for Care Coordination," *Med Care Res Rev*, vol. 75, no. 5, pp. 586–611, 2018.
[29] K. D. Mandl, K. L. Olson, D. Mines, C. Liu, and F. Tian, "Provider collaboration: cohesion, constellations, and shared patients," *Journal of General Internal Medicine*, vol. 29, no. 11, pp. 1499–1505, 2014.
[30] C. Pollack, K. Lemke, E. Roberts, and J. Einer, "Patient sharing and quality of care: measuring outcomes of care coordination using claims data," *Med Care*, vol. 53, no. 4, pp. 317–323, 2015.
[31] C. Pollack, G. Weissman, K. Lemke, P. S. Hussey, and J. Weiner, "Patient sharing among physicians and costs of care: A network analytic approach to care coordination using claims data," *J Gen Intern Med*, vol. 28, no. 3, pp. 459–65, 2012.
[32] B. E. Landon, N. L. Keating, J.-P. Onnela, A. M. Zaslavsky, N. Christakis, and A. J. O'Malley, "Patient-sharing networks of physicians and health care utilization and spending among Medicare beneficiaries," *JAMA Intern Med*, vol. 178, no. 1, 2017.
[33] M. L. Barnett, N. Christakis, A. J. O'Malley, J.-P. Onnela, N. L. Keating, and B. E. Landon, "Physician patient-sharing networks and the cost and intensity of care in US hospitals," *Med Care*, vol. 50, no. 2, 2012.
[34] E. H. DuGoff, S. Fernandes-Taylor, G. E. Weissman, J. H. Huntley, and C. E. Pollack, "A scoping review of patient-sharing network studies using administrative data," *Translational Behavioral Medicine*, vol. 8, no. 4, pp. 598–625, Jul. 2018.

[35] C. Eibner and M. S. Marquis, “Employers’ health insurance cost burden, 1996–2005,” *Monthly Labor Review*, 2008.
[36] R. Z. Goetzel, M. Tabrizi, R. M. Henke, et al., “Estimating the return on investment from a health risk management program offered to small Colorado-based employers,” *Journal of occupational and environmental medicine*, vol. 56, no. 5, pp. 554–560, May 2014.
[37] J. E. Fielding, “Worksite health promotion programs in the United States: progress, lessons and challenges,” *Health Promotion International*, vol. 5, no. 1, pp. 75–84, 1990.
[38] National Association of Worksite Health Centers, “What is an onsite clinic?,” 2017. [Online]. Available: https://www.nawhc.org/about-nawhc/what-onsite-clinic. [Accessed: 07-Sep-2017].
[39] S. P. McCaskill, L. A. Schwartz, A. L. Derouin, and A. H. Pegram, “Effectiveness of an on-site health clinic at a self-insured university: A cost-benefit analysis,” *Workplace Health & safety*, vol. 62, no. 4, 2014.
[40] D. H. Chenoweth and J. Garrett, “Cost-effectiveness analysis of a worksite clinic: Is it worth the cost?,” *AAOHN*, vol. 54, no. 2, Feb. 2006.
[41] D. H. Chenoweth, N. Martin, J. Pankowski, and L. W. Raymond, “A benefit-cost analysis of a worksite nurse practitioner program: first impressions,” *JOEM*, vol. 47, no. 11, 2005.
[42] A. L. Person, S. E. Colby, J. A. Bulova, and J. Whitehurst Eubanks, “Barriers to partcipation in a worksite wellness program,” *Nutrition Research and Practice*, vol. 4, no. 2, pp. 149–154, 2010.
[43] M. D. A. Carlson and R. S. Morrison, “Study design, precision, and validity in observational studies,” *Journal of palliative medicine*, vol. 12, no. 1, pp. 77–82, Jan. 2009.
[44] D. Barnes and J. Green, “Controlling rising health care costs with innovative employer-sponsored clinics,” *Employment Relations Today*, vol. 42, no. 3, pp. 21–28, Oct. 2015.
[45] R. Andersen and J. F. Newman, “Societal and individual determinants of medical care utilization in the United States,” *The Milbank Quarterly*, vol. 83, no. 4, pp. 10.1111/j.1468–0009.2005.00428.x, Dec. 2005.
[46] T. Owolabi and I. Simpson, “Documenting and coding preventive visits: A physician’s perspective,” *Fam Pract Manag*, vol. 19, no. 4, pp. 12–16, 2012.
[47] United Healthcare, “Coding summary for providers preventive care services – Health Care Reform Preventive Care Services – Health Care Reform,” 2012.
[48] R. Andersen, “Revisiting the behavioral model and access to medical care: Does it matter?,” *Journal of Health and Social Behavior*, vol. 36, no. 1, pp. 1–10, 1995.
[49] K. A. Phillips, K. Morrison, R. Andersen, and L. A. Aday, “Understanding the context of healthcare utilization: Assessing environmental and provider-related variables in the behavioral model of utilization.,” *Health Services Research*, vol. 33, pp. 571–596, 1998.
[50] L. Keenan, L. Marshal, and S. Eve, “Extension of the behavioral model of healthcare utilization with ethnically diverse, low income women,” *Ethnicity & Disease*, vol. 12, 2002.
[51] “Defining the basics of health informatics for HIM professionals,” *Journal of AHIMA*, vol. 85, no. 9, pp. 60–66.
[52] B. Hulka and J. Wheat, “Patterns of utilization : The patient perspective,” *Medical Care*, vol. 23, no. 5, pp. 438–460, 1985.
[53] M. Akazawa, J. L. Sindelar, and A. D. Paltiel, “Economic costs of influenza-related work absenteeism,” *Value in Health*, vol. 6, no. 2, pp. 107–115, 2003.

[54] L. A. Grohskopf, L.Z. Sokolow, K.R. Broder, et al., “Prevention and control of seasonal influenza with vaccines,” *MMWR. Recommendations and Reports*, vol. 65, no. 5, pp. 1–54, 2016.
[55] “Health, United States, 2014: With special feature on adults aged,” National Center for Health Statistics, Hyattsville, MD, 2015.
[56] Center for Disease Control and Prevention, “Barriers and strategies to improving influenza vaccination among health care personnel,” 2016. [Online]. Available: http://www.cdc.gov/flu/toolkit/long-term-care/strategies.htm. [Accessed: 04-Nov-2016].
[57] C. J. Lin, M.P. Nowalk, S.L. Toback, et al., “Importance of vaccination habit and vaccine choice on influenza vaccination among healthy working adults,” *Vaccine*, vol. 28, no. 48, pp. 7706–7712, 2010.
[58] C. A. Thiels, A. M. Abdelrahman, E. B. Habermann, S. Hallbeck, K. S. Pasupathy, and J. Bingener, “The use of patient factors to improve the prediction of operative duration using laparoscopic cholecystectomy,” *Surgical Endoscopy*, vol. 31, no. 1, pp. 333–340, 2016.
[59] W. J. Long, J. L. Griffith, H. P. Selker, and R. B. D’Agostino, “A comparison of logistic regression to decision-tree induction in a medical domain,” *Computers and Biomedical Research, an International Journal*, vol. 26, pp. 74–97, 1993.
[60] Center for Disease Control and Prevention, “High cholestrol facts,” 2015. [Online]. Available: http://www.cdc.gov/cholesterol/facts.htm. [Accessed: 06-Oct-2016].
[61] Center for Disease Control and Prevention, “High blood pressure facts,” 2015. [Online]. Available: http://www.cdc.gov/bloodpressure/facts.htm. [Accessed: 06-Oct-2016].
[62] Center for Disease Control and Prevention, “National diabetes statistics report,” 2014. [Online]. Available: http://www.cdc.gov/diabetes/data/statistics/2014statisticsreport.html. [Accessed: 06-Oct-2016].
[63] G. Beckles, M. Engelgau, M. Naarayan, W. Herman, R. Aubert, and D. Williamson, “Population-based assessment of the level of care among adults with diabetes in the U.S.,” *Diabetes Care*, vol. 21, no. 9, pp. 1432–1438, 1998.
[64] “Defining the medical home | patient-centered primary care collaborative,” 2015. [Online]. Available: https://www.pcpcc.org/about/medical-home. [Accessed: 26-Dec-2016].
[65] “Health insurance coverage for children and young adults under 26.” [Online]. Available: https://www.healthcare.gov/young-adults/children-under-26/. [Accessed: 21-Nov-2016].
[66] X. Tao, D. Chenoweth, A.S. Alfriend, et al., “Monitoring worksite clinic performance using a cost-benefit tool,” *JOEM*, vol. 51, no. 10, pp. 1151–1157, 2009.
[67] Center for Disease Control and Prevention, “CDC Online Newsroom - Press Release - Number of Americans with diabetes projected to double or triple by 2050,” 2010. [Online]. Available: https://www.cdc.gov/media/pressrel/2010/r101022.html. [Accessed: 18-May-2018].
[68] American Diabetes Association, “The cost of diabetes,” *American Diabetes Association*, 2018. [Online]. Available: http://www.diabetes.org/advocacy/news-events/cost-of-diabetes.html. [Accessed: 18-May-2018].
[69] American Diabetes Association, “Standards of medical care in diabetes—2017 abridged for primary care providers,” *Clinical Diabetes : A Publication of the American Diabetes Association*, vol. 35, no. 1, pp. 5–26, Jan. 2017.
[70] S. Rogers, “Inpatient care coordination for patients with diabetes,” *Diabetes Spectrum*, vol. 21, no. 4, p. 272, Oct. 2008.

[71] K. I. Taylor, K. M. Oberle, R. A. Crutcher, and P. G. Norton, “Promoting health in type 2 diabetes: nurse-physician collaboration in primary care,” *Biological Research For Nursing*, vol. 6, no. 3, pp. 207–215, Jan. 2005.
[72] J. S. Wrobel, M.P. Charns, P. Diehr, et al., “The relationship between provider coordination and diabetes-related foot outcomes,” *Diabetes Care*, vol. 26, no. 11, pp. 3042–3047, 2003.
[73] J. Tomasone, M. Vukmirovic, M.C. Brouwers, et al., “Challenges and insights in implementing coordinated care between oncology and primary care providers: a Canadian perspective,” *Current Oncology*, vol. 24, no. 2, pp. 120–123, Apr. 2017.
[74] American Diabetes Association, “Your health care team,” *American Diabetes Association*, 2017. [Online]. Available: http://www.diabetes.org/living-with-diabetes/treatment-and-care/whos-on-your-health-care-team/your-health-care-team.html. [Accessed: 14-May-2018].
[75] G. Goderis, l. Borgermans, C. Mathieu, et al., “Barriers and facilitators to evidence based care of type 2 diabetes patients: experiences of general practitioners participating to a quality improvement program,” *Implementation Science : IS*, vol. 4, pp. 4–41, 2009.
[76] M. Kraft, K. Blomberg, and A.-M. R. Hedman, “The health care professionals’ perspectives of collaboration in rehabilitation – an interview study,” *International Journal of Older People Nursing*, vol. 9, no. 3, pp. 209–216, Feb. 2013.
[77] J. C. Brunson and R. C. Laubenbacher, “Applications of network analysis to routinely collected health care data: a systematic review,” *Journal of the American Medical Informatics Association*, vol. 25, no. 2, pp. 210–221, Feb. 2018.
[78] S. Tasselli, “Social networks of professionals in health care organizations: A review,” *Med Care Res Rev*, vol. 71, no. 6, pp. 619–660, Nov. 2014.
[79] D. Chambers, P. Wilson, C. Thomson, and M. Harden, “Social network analysis in healthcare settings: a systematic scoping review,” *PLoS ONE*, vol. 7, no. 8, 2012.
[80] F. C. Cunningham, G. Ranmuthugala, J. Plumb, A. Georgiou, J. I. Westbrook, and J. Braithwaite, “Health professional networks as a vector for improving healthcare quality and safety: a systematic review,” *BMJ Qual Saf*, vol. 21, no. 3, p. 239, Mar. 2012.
[81] A. P. Appel, V. F. de Santana, L. G. Moyano, M. Ito, and C. S. Pinhanez, “A social network analysis framework for modeling health insurance claims data,” *arXiv:1802.07116 [cs]*, Feb. 2018.
[82] J. Herrin, P. R. Soulos, X. Xu, C. P. Gross, and C. E. Pollack, “An empiric approach to identifying physician peer groups from claims data: An example from breast cancer care,” *Health Services Research*, vol. 54, no. 1, pp. 44–51, Feb. 2019.
[83] S. Uddin, M. Kelaher, and U. Srinivasan, “A framework for administrative claim data to explore healthcare coordination and collaboration,” vol. 40, pp. 500–510, 2016.
[84] B. E. Landon *et al.*, “Using administrative data to identify naturally occurring networks of physicians,” *Medical Care*, vol. 51, no. 8, pp. 715–721, 2013.
[85] J. G. Trogdon *et al.*, “Care coordination and multispecialty teams in the care of colorectal cancer patients,” *Medical Care*, vol. 56, no. 5, 2018.
[86] H. M. K. Ghomrawi, R. J. Funk, M. L. Parks, J. Owen-Smith, and J. M. Hollingsworth, “Physician referral patterns and racial disparities in total hip replacement: A network analysis approach,” *PloS one*, vol. 13, no. 2, pp. e0193014–e0193014, Feb. 2018.
[87] S. S. Tannenbaum *et al.*, “Surgeon peer network characteristics and adoption of new imaging techniques in breast cancer: A study of perioperative MRI,” *Cancer Medicine*, vol. 7, no. 12, pp. 5901–5909, Dec. 2018.

[88] E. L. Moen, A. M. Austin, J. P. Bynum, J. S. Skinner, and A. J. O'Malley, "An analysis of patient-sharing physician networks and implantable cardioverter defibrillator therapy," *Health services & outcomes research methodology*, vol. 16, no. 3, pp. 132–153, Sep. 2016.
[89] "2014 ICD-9-CM diagnosis codes 250.* : diabetes mellitus." [Online]. Available: http://www.icd9data.com/2014/Volume1/240-279/249-259/250/default.htm. [Accessed: 20-Aug-2018].
[90] S. Gomez, P. Jensen, and A. Arenas, "Analysis of community structure in networks of correlated data," *Physical Review E*, vol. 80, no. 1, Jul. 2009.
[91] E. Le Martelot and C. Hankin, "Multi-scale community detection using stability optimisation within Greedy Algorithms," *arXiv preprint arXiv:1201.3307*, 2012.
[92] R. Lambiotte, J.-C. Delvenne, and M. Barahona, "Laplacian dynamics and multiscale modular structure in networks," *arXiv preprint arXiv:0812.1770*, 2008.
[93] M. E. J. Newman and M. Girvan, "Finding and evaluating community structure in networks," *Physical Review E*, vol. 69, no. 2, 2004.
[94] S. Fortunato and M. Barthélemy, "Resolution limit in community detection," *Proc Natl Acad Sci USA*, vol. 104, no. 1, p. 36, Jan. 2007.
[95] P. D. Patterson *et al.*, "Network analysis of team communication in a busy emergency department," *BMC Health Services Research*, vol. 13, pp. 109–109, 2013.
[96] F. Bloch, M. O. Jackson, and P. Tebaldi, "Centrality measures in networks," *arXiv preprint arXiv:1604.00696.*, 2016.
[97] G. Fattore, F. Frosini, D. Salvatore, and V. Tozzi, "Social network analysis in primary care: The impact of interactions on prescribing behaviour," *Health Policy*, vol. 92, no. 2, pp. 141–148, Oct. 2009.
[98] U. Brandes, S. P. Borgatti, and L. C. Freeman, "Maintaining the duality of closeness and betweenness centrality," *Social Networks*, vol. 44, pp. 153–159, Jan. 2016.
[99] D. Luke, *A user's guide to network analysis in R*. Switzerland: Springer, Cham, 2015.
[100] M. G. Csardi, *Package 'igraph.'* 2013.
[101] H. Wickham and W. Chang, *devtools: tools to make developing R packages easier*. 2016.
[102] M. Berlingerio, M. Coscia, F. Giannotti, A. Monreale, and D. Pedreschi, "The pursuit of hubbiness: Analysis of hubs in large multidimensional networks," *Journal of Computational Science*, vol. 2, no. 3, pp. 223–237, Aug. 2011.
[103] M. Ostovari, D. Yu, and C. J. Steele-Morris, "Identifying key players in the care process of patients with diabetes using social network analysis and administrative data," *AMIA Annual Symposium Proceedings*, vol. 2018, pp. 1435–1441, Dec. 2018.
[104] A. Dmowska, "Skip the coverage gap: Maintaining diabetes care while changing insurance," *diaTribe*, 2017. [Online]. Available: https://diatribe.org/skip-coverage-gap-maintaining-diabetes-care-while-changing-insurance. [Accessed: 18-Oct-2018].
[105] M. L. Barnett, Z. Song, S. Rose, A. Bitton, M. E. Chernew, and B. E. Landon, "Insurance transitions and changes in physician and emergency department utilization: An observational study," *Journal of General Internal Medicine*, vol. 32, no. 10, pp. 1146–1155, Oct. 2017.
[106] M. Gulliford, S. Naithani, and M. Morgan, "What is 'continuity of care'?," *J Health Serv Res Policy*, vol. 11, no. 4, pp. 248–250, Oct. 2006.
[107] D. T. Liss, J. Chubak, M. L. Anderson, K. W. Saunders, L. Tuzzio, and R. J. Reid, "Patient-reported care coordination: associations wth primary care continuity and specialty care use," *Annals of Family Medicine*, vol. 9, no. 4, pp. 323–329, 2011.

[108] C. Ponte, "Monitoring diabetes mellitus: the pharmacist's role.," *American pharmacy*, no. 1, pp. 61–65, 1992.

[109] M. Smith, "Pharmacists' role in improving diabetes Medication management," *Journal of diabetes science and technology (Online)*, vol. 3, no. 1, pp. 175–179, Jan. 2009.

[110] R. Campbell, "Role of the pharmacists in diabetes management," *Am J Health Syst Pharm*, vol. 59, no. suppl 9, p. S18, Dec. 2002.

[111] O. K. Duru, J.A. Schittdiel, W.T. Dyer, et al., "Mail-order pharmacy use and adherence to diabetes-related Medications," *The American journal of managed care*, vol. 16, no. 1, pp. 33–40, Jan. 2010.

[112] J. A. Schmittdiel, A. J. Karter, W. Dyer, J. Chan, and O. K. Duru, "The safety and effectiveness of mail order pharmacy use in diabetes patients," *The American journal of managed care*, vol. 19, no. 11, pp. 882–887, Nov. 2013.

[113] J. Wu, M. L. Davis–Ajami, and V. Noxon, "Patterns of use and expenses associated with mail-service pharmacy in adults with diabetes," *Journal of the American Pharmacists Association*, vol. 55, no. 1, pp. 41–51, Jan. 2015.

[114] C. Doyle, L. Lennox, and D. Bell, "A systematic review of evidence on the links between patient experience and clinical safety and effectiveness," *BMJ Open*, vol. 3, no. 1, Jan. 2013.

[115] Forrest CB, Weiner JP, Fowles J, and et al, "Self-referral in point-of-service health plans," *JAMA*, vol. 285, no. 17, pp. 2223–2231, May 2001.

[116] J. Mettayil, R. Quinton, and S. Wahid, "Careers in diabetes and endocrinology," *BMJ*, vol. 336, no. 7642, p. s79, Mar. 2008.

[117] S.-H. Bae, A. Nikolaev, J. Y. Seo, and J. Castner, "Health care provider social network analysis: a systematic review," *Nursing Outlook*, vol. 63, no. 5, pp. 566–584, 2015.

[118] "Chronic diseases in America | CDC," 16-Jan-2019. [Online]. Available: https://www.cdc.gov/chronicdisease/resources/infographic/chronic-diseases.htm. [Accessed: 29-Jan-2019].

[119] National Center for Chronic Disease Prevention and Health Promotion, "Promoting health during the Holidays 7 Tips to stay healthy," *Centers for Disease Control and Prevention*, 19-Dec-2018. [Online]. Available: https://www.cdc.gov/chronicdisease/index.htm. [Accessed: 29-Jan-2019].

[120] N. D. Wong, V. Lopez, S. Tang, and G. R. Williams, "Prevalence, treatment, and control of combined hypertension and hypercholesterolemia in the United States," *The American Journal of Cardiology*, vol. 98, no. 2, pp. 204–208, Jul. 2006.

[121] T. O'Brien, T. T. Nguyen, and B. R. Zimmerman, "Hyperlipidemia and diabetes mellitus," *Mayo Clinic Proceedings*, vol. 73, no. 10, pp. 969–976, Oct. 1998.

[122] G. O. Ivbijaro, Y. Enum, A. A. Khan, S. S.-K. Lam, and A. Gabzdyl, "Collaborative care: models for treatment of patients with complex medical-psychiatric conditions," *Current psychiatry reports*, vol. 16, no. 11, pp. 506–506, 2014.

[123] E. H. Wagner, R.E. Glasgow, C. Davis, et al., "Quality improvement in chronic illness care: a collaborative approach," *The Joint Commission journal on quality improvement*, vol. 27, no. 2, pp. 63–80, 2001.

[124] T. Jaarsma, "Inter-professional team approach to patients with heart failure," *Heart (British Cardiac Society)*, vol. 91, no. 6, pp. 832–838, Jun. 2005.

[125] A. Boykin, D. Wright, L. Stevens, and L. Gardner, “Interprofessional care collaboration for patients with heart failure,” *American Journal of Health-System Pharmacy*, vol. 75, no. 1, pp. e45–e49, Jan. 2018.
[126] L. S. Feldman, L.L. Costa, E.R. Feroli, et al., “Nurse-pharmacist collaboration on medication reconciliation prevents potential harm,” *Journal of Hospital Medicine*, vol. 7, no. 5, pp. 396–401, 2012.
[127] C. M. Franklin, J. M. Bernhardt, R. P. Lopez, E. R. Long-Middleton, and S. Davis, “Interprofessional teamwork and collaboration between community health workers and healthcare teams: An integrative review,” *Health services research and managerial epidemiology*, vol. 2, pp. 2333392815573312–2333392815573312, Mar. 2015.
[128] M. A. Valentine, I. M. Nembhard, and A. C. Edmonson, “Measuring teamwork in health care settings: a review of survey instruments,” *Medical Care*, vol. 54, no. 3, pp. e16–e30, 2015.
[129] S. J. Walters, C. Stern, and S. Roberston-Malt, “The measurement of collaboration within healthcare settings: a systematic review of measurement properties of instruments,” *JBI Database System Rev Implement Rep*, vol. 14, no. 4, pp. 138–97, 2016.
[130] A. J. Berendsen, W. H. G. M. Benneker, K. H. Groenier, J. Schuling, R. P. T. M. Grol, and B. Meyboom-de Jong, “DOC questionnaire: measuring how GPs and medical specialists rate collaboration,” *International Journal of Health Care Quality Assurance*, vol. 23, no. 5, pp. 516–526, 2010.
[131] J. B. Sexton Ph. D. *et al.*, “Teamwork in the operating room: frontline perspectives among hospitals and operating room personnel,” *Anesthesiology: The Journal of the American Society of Anesthesiologists*, vol. 105, no. 5, pp. 877–884, Nov. 2006.
[132] D. L. Davenport, W. G. Henderson, C. L. Mosca, S. F. Khuri, and R. M. Mentzer Jr, “Risk-adjusted morbidity in teaching hospitals correlates with reported levels of communication and collaboration on surgical teams but not with scale measures of teamwork climate, safety climate, or working conditions,” *Journal of the American College of Surgeons*, vol. 205, no. 6, pp. 778–784, Dec. 2007.
[133] T. L. Jones, M. A. J. Baxter, and V. Khanduja, “A quick guide to survey research,” *Annals of the Royal College of Surgeons of England*, vol. 95, no. 1, pp. 5–7, Jan. 2013.
[134] L. P. Casalino *et al.*, “Physician networks and ambulatory care-sensitive admissions,” *Medical Care*, vol. 53, no. 6, 2015.
[135] C. E. Pollack, G. Weissman, J. Bekelman, K. Liao, and K. Armstrong, “Physician social networks and variation in prostate cancer treatment in three cities,” *Health services research*, vol. 47, no. 1 Pt 2, pp. 380–403, Feb. 2012.
[136] M. B. Carson, D.M. Scholtens, C. N. Frailey, et al., “Characterizing teamwork in cardiovascular care outcomes: a network analytics approach,” *Circulation. Cardiovascular quality and outcomes*, vol. 9, no. 6, pp. 670–678, Nov. 2016.
[137] S. Uddin, M. Kelaher, and M. Piraveenan, “Impact of physician community structure on healthcare outcomes,” *Driving Reform: Digital Health is Everyone’s Business*, vol. 214, pp. 152–158, 2015.
[138] C. E. Pollack, H. Wang, J. E. Bekelman, et al., “Physician social networks and variation in rates of complications after radical prostatectomy,” *Value in Health*, vol. 17, pp. 611–618, 2014.
[139] Johns Hopkins, “The Johns Hopkins ACG® system version 11.0 technical reference guide,” Johns Hopkins, Bloomberg School of Public Health, 2014.

[140] B. M. Y. Cheung and C. Li, "Diabetes and hypertension: is there a common metabolic pathway?," *Current atherosclerosis reports*, vol. 14, no. 2, pp. 160–166, Apr. 2012.

[141] "2012 ICD-9-CM Diagnosis Codes 401.* : Essential hypertension." [Online]. Available: http://www.icd9data.com/2012/Volume1/390-459/401-405/401/default.htm. [Accessed: 30-Jan-2019].

[142] "2014 ICD-9-CM Diagnosis Code 272.4 : Other and unspecified hyperlipidemia," 2014. [Online]. Available: http://www.icd9data.com/2014/Volume1/240-279/270-279/272/272.4.htm. [Accessed: 30-Jan-2019].

[143] M. Ostovari, C. J. Steele-Morris, P. Griffin, and D. Yu, "Data-driven modeling of diabetes care teams using social network analysis," *Journal of the American Medical Informatics Association*, 2019.

[144] D. C. Atkins, S. A. Baldwin, C. Zheng, R. J. Gallop, and C. Neighbors, "A tutorial on count regression and zero-altered count models for longitudinal substance use data," *Psychology of addictive behaviors : journal of the Society of Psychologists in Addictive Behaviors*, vol. 27, no. 1, pp. 166–177, Mar. 2013.

[145] J. Du, Y.-T. Park, N. Theera-Ampornpunt, J. S. McCullough, and S. M. Speedie, "The use of count data models in biomedical informatics evaluation research," *Journal of the American Medical Informatics Association : JAMIA*, vol. 19, no. 1, pp. 39–44, 2012.

[146] B. A. Desmarais and J. J. Harden, "Testing for zero inflation in count models: Bias correction for the Vuong test," *The Stata Journal*, vol. 13, no. 4, pp. 810–835, 2013.

[147] J. W. Hardin and J. M. Hilbe, *Generalized estimating equations*. Chapman and Hall/CRC, 2002.

[148] M. Shrivastav, W. Gibson, R. Shrivastav, et al., "Type 2 diabetes management in primary care: The role of retrospective, professional continuous glucose monitoring," *diaspect*, vol. 31, no. 3, p. 279, Aug. 2018.

[149] K. V. Rudnick, D. L. Sackett, S. Hirst, and C. Holmes, "Hypertension: The family physician's role," *Canadian family physician Medecin de famille canadien*, vol. 24, pp. 477–484, May 1978.

[150] R. H. Nelson, "Hyperlipidemia as a risk factor for cardiovascular disease," *Primary care*, vol. 40, no. 1, pp. 195–211, Mar. 2013.

[151] A. E. Lewis Hunter, E. S. Spatz, S. L. Bernstein, and M. S. Rosenthal, "Factors influencing hospital admission of non-critically ill patients presenting to the emergency department: a cross-sectional study," *Journal of general internal medicine*, vol. 31, no. 1, pp. 37–44, Jan. 2016.

[152] A. Geva, K. L. Olson, C. Liu, and K. D. Mandl, "Provider connectedness to other providers reduces risk of readmission after hospitalization for heart failure," *Med Care Res Rev*, p. 1077558717718626, Jul. 2017.

[153] S. Lindblad, S. Ernestam, A. Van Citters, C. Lind, T. Morgan, and E. Nelson, "Creating a culture of health: evolving healthcare systems and patient engagement," *QJM: An International Journal of Medicine*, vol. 110, no. 3, pp. 125–129, 2017.