



5-2016

Geography of Health Care Access: Measurement, Analyses and Integration

Huairan Ye

University of Tennessee - Knoxville, hye3@vols.utk.edu

Recommended Citation

Ye, Huairan, "Geography of Health Care Access: Measurement, Analyses and Integration. " PhD diss., University of Tennessee, 2016.
https://trace.tennessee.edu/utk_graddiss/3757

This Dissertation is brought to you for free and open access by the Graduate School at Trace: Tennessee Research and Creative Exchange. It has been accepted for inclusion in Doctoral Dissertations by an authorized administrator of Trace: Tennessee Research and Creative Exchange. For more information, please contact trace@utk.edu.

To the Graduate Council:

I am submitting herewith a dissertation written by Huairen Ye entitled "Geography of Health Care Access: Measurement, Analyses and Integration." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Geography.

Hyun Kim, Major Professor

We have read this dissertation and recommend its acceptance:

Ronald Kalafsky, Nicholas Nagle, Madhuri Sharma, Agricola Odoi

Accepted for the Council:

Dixie L. Thompson

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

Geography of Health Care Access: Measurement, Analyses and Integration

A Dissertation Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

Huairan Ye

May 2016

Copyright © 2016 by Huairan Ye

All rights reserved.

ACKNOWLEDGEMENTS

I could not have completed my dissertation without the help and support from many people. First, I would like to express gratitude to my academic advisor, Dr. Hyun Kim. Throughout the course of PhD study, Dr. Kim's consistent support and guidance have been essential to my academic and personal development. I am truly grateful for having such a dedicated advisor on my side.

Second, I would like to thank my dissertation committee members – Drs. Ronald Kalafsky, Nicholas Nagle, Madhuri Sharma, and Agricola Odoi. My committee members offered courses that I benefited greatly from, and they have been most helpful and supportive to my research studies. Through Dr. Kalafsky's class I gained knowledge on urban and economic geography. He presented theories in such a way that's both intriguing and easy to understand. Dr. Sharma's urban social geography class taught me how to approach urban issues from a critical perspective, which is essential to becoming an independent researcher. I built up my statistics background through Dr. Nagle's classes, whose enthusiasm for the subject helps inspire his students. I expanded my knowledge and techniques on the application of Geographic Information System and statistics in health research through Dr. Odoi's class, which constitutes a vital part of my dissertation research.

Third, I would like to thank Dr. Sungkyu Lee. I worked with Dr. Lee to develop a paper in the summer of 2014, which later became the fourth chapter of my dissertation. Dr. Lee also contributed greatly during the process of getting the paper published.

Finally, my special thanks goes to my family, especially my mother Ms. Ruichun Ye – my best friend, who has always shown me great love, courage, wisdom and strength; and my husband Mr. Michael de Socio – also my best friend, who has always believed in me, and who in addition to his own career took care of much of the household while I worked hard at my studies. My family have always been there for me since the beginning. Through them I realize that love is indeed the ultimate and the highest goal to which man can aspire.

ABSTRACT

This dissertation addresses the geography of healthcare access and disparity issues in the United States using geospatial methods. Disparities in access to quality healthcare services are of great concern in the field of both public health and geography. Access is a key element within the healthcare delivery system, influenced by both spatial factors and non-spatial factors. Focusing on the spatial dimensions of access, an innovative contribution of this dissertation is the integration of spatial modeling, geo-statistics and location problems in a Geographic Information System (GIS) environment to investigate healthcare access.

Improving health access begins with developing reliable methods to measure accessibility. In health geography and social sciences literature, the term spatial accessibility is used to refer to the fusion of both availability and accessibility of health demand and supply. Thus, a major focus of this dissertation is to present an alternative set of healthcare accessibility measures – *a network-based health accessibility index method (NHAIM)* to measure accessibility and identify underserved areas. Another focus of this dissertation is to understand neighborhood factors that contribute to healthcare access – both potential and revealed access through statistical analysis. Studies have shown that social and physical environments affect individual's health status, yet less has been done on whether neighborhood factor influence health access. A final focus of this dissertation is to propose a planning method - *a Network-based Covering Location Problem (Net-CLP)* to locate healthcare facilities so as to maximize service coverage while reducing spatial disparity between healthcare supply and demand in a sustainable manner. As mentioned above, spatial accessibility relies on the geographical interactions between healthcare facilities and population in need, therefore the facility location is essential in ensuring access.

In summary, this dissertation aims to achieve following three goals: 1) develop a reliable method to measure health care accessibility and capture underserved areas; 2) investigate neighborhood factors and health care access; 3) propose a feasible planning method to locate health care facilities and improve overall access.

TABLE OF CONTENT

CHAPTER 1 INTRODUCTION	1
1.1 Introduction.....	2
1.2 The concept of health care access	3
1.3 Research goals	3
1.4 Research tasks.....	5
1.4.1 Developing an alternative measure of health care accessibility – a network-based health accessibility index method (NHAIM)	5
1.4.2 Understanding neighborhood factors and health care access	5
1.4.3 Developing feasible planning methods – location problems in healthcare facility siting	5
1.5 Organization.....	6
CHAPTER 2 MEASURING SPATIAL ACCESSIBILITY USING A NETWORK-BASED METHOD IN A GIS ENVIRONMENT.....	7
Abstract	8
2.1 Introduction.....	8
2.2 Spatial accessibility and disparity in health care systems.....	10
2.3 Analytical frameworks and study area.....	12
2.3.1 Network-Based Health Accessibility Index Method (NHAIM)	12
2.2.2 Study Area and Data	17
2.4 Results.....	18
2.5 Conclusion Remarks	21
CHAPTER 3 NEIGHBORHOOD CHARACTERISTICS AND PRIMARY CARE PHYSICIAN SUPPLY	22
Abstract	23
3.1 Introduction.....	23
3.2 Literature Review.....	24
3.3 Data and method	26
3.3.1 Study area.....	26
3.2.2 Primary care physician data	27
3.2.3 Neighborhood characteristics.....	28
3.3 Detection of Clusters of PCPs.....	28
3.4 Statistical methods	29
3.5 Results and discussions.....	30
3.6 Conclusions.....	32
CHAPTER 4 NEIGHBORHOOD CHARACTERISTICS AND HEALTHCARE UTILIZATION	34
Abstract	35
4.1 Introduction.....	35
4.2 Theoretical framework.....	36
4.3 Methods.....	37
4.3.1 Data source and sample	37
4.3.2 Study variables.....	37
4.3.3 Data analysis	38
4.4 Results.....	39

4.5 Discussion	42
CHAPTER 5 LOCATING HEALTHCARE FACILITIES: A NETWORK-BASED OPTIMIZATION APPROACH.....	45
Abstract	46
5.1 Introduction.....	46
5.2 Background	47
5.3 Motivation for the Net-CLP	49
5.3.1 Formulation of Net-SCLP	51
5.3.2 Formulation of Net-MCLP	52
5.3.3 Extension: Medical Service Capacity	52
5.4 Case study	53
5.4.1 Spatial representation of study area	54
5.4.2 Key parameters	56
5.4.3 Results and discussion	56
5.5 Conclusions.....	61
CHAPTER 6 CONCLUSIONS	64
6.1 Summary of dissertation research.....	65
Theme 1: Developing an alternative measure of healthcare accessibility	65
Theme 2: Neighborhood factors and healthcare access	65
Theme 3: Location problems in healthcare facility siting	66
6.2 Future research directions	66
REFERENCE.....	68
APPENDIX.....	79
VITA.....	83

LIST OF TABLES

Table 1. The indexes of NHAIM	13
Table 2. Selected census tract level aggregated neighborhood variables	28
Table 3. Coefficients for spatial lag regression models	31
Table 4. Correlation matrix of explanatory variables	31
Table 5. Aggregated neighborhood characteristics.....	39
Table 6. Predictors of the length of inpatient stay: Results of GLMs.....	41
Table 7. Characteristics of the Net-CLP	49
Table 8. Model behaviors of the Net-LSCP and the Net-MCLP	58
Table 9. Model behaviors of the Net-MCLP	59
Table 10. Model behaviors of the Net-LSCP-MAC and the Net-MCLP-MAC.....	62

LIST OF FIGURES

Figure 1. The procedure for calculating the Network-Based Health Accessibility Supply Index (NHA-SI)	14
Figure 2. The procedure for calculating the Network-Based Health Accessibility Demand Index	15
Figure 3. The classification of spatial units using the NHA-DP (Note: H=High, L=Low, Av=Availability, Ac=Accessibility)	17
Figure 4. The study area: distribution of hospitals in Hillsborough County, FL.....	18
Figure 5. Spatial pattern by the NHA-SI, 10 (5-a) and 30 minutes time zone buffer (5-b).....	19
Figure 6. Spatial pattern by the NHA-DI, 10 (6-a) and 30 minutes time zone buffer (6-b).	19
Figure 7. Spatial disparity by the NHA-DP, 10 (7-a) and 30 minutes time zone buffer (7-b).	20
Figure 8. Physician location map (Note: uninhabitable areas such as swamp and forest are excluded from the map. Land use data obtained from Florida Natural Areas Inventory.)...	27
Figure 9. LISA Cluster Map	30
Figure 10. Length of inpatient stay map in the United States.....	40
Figure 11. LISA cluster map.....	41
Figure 12. GWR coefficient map.....	42
Figure 13. Geo-processing to identify the covered/uncovered relationship between healthcare demand and facility location.....	50
Figure 14. Potential demand sites representation with areal interpolation	55
Figure 15. Distribution of 19 existing hospitals and 804 hexagon centroids.....	56
Figure 16. Standard Net-SCLP selected locations for 10, 20, and 30min travel time threshold ..	58
Figure 17. The Net-MCLP solution for $p_{new} = 6$, $T=20$ (a) and $p_{new}=2$, $T=30$ (b).....	60
Figure 18. Covered population percentage with p_{new} for three threshold ($T=10, 20, \text{ and } 30$).....	63

CHAPTER 1
INTRODUCTION

1.1 Introduction

Although how to pay for health care is a major dominant debate topic on the reform of health care delivery system, disparities in the spatial distribution of services have also been highlighted (Nuwer 2008a, Nuwer 2008b). Spatial analysis using Geographic Information System (GIS) can contribute greatly to understanding the geography of health care access, which is what I intend to address in my dissertation.

Maintaining health and well-being is a basic human need, which can only be achieved through quality health care. There are two major forms of health care: informal and formal health care. Informal health care lies outside the market economy. It is care provided by families and communities in a domestic or neighborhood setting (Arno, Levine and Memmott 1999). A great majority of the health care is provided informally. For example, most of the needs of elders at a later stage of life are fulfilled by informal helpers – family, friends, and neighbors (Coward and Cutler 1989). On the contrary, formal health care is provided by public, private and voluntary organizations such as hospitals and physicians, which is the main focus of this dissertation. Despite spending more per capital on medical care than other nations, the health care performance in United States is far from satisfactory (World Health Organization 2000). In the United States, access to health care is increasingly constrained by the geography of health care services and providers. A lack of access leads to widening of health disparities – defined as the differences in health status among social groups. For example, middle and high income individuals are more likely to maintain higher health status due to having better access to family physicians (Rosenberg and Hanlon 1996). Ethnic minorities are more likely to receive late-stage breast cancer diagnosis and suffer from higher mortality rates as a result of the negative influence of socioeconomic disadvantages such as low educational attainment and linguistic barriers (Meliker *et al.* 2009a, Meliker *et al.* 2009b).

The health care delivery system is undergoing tremendous transformations. Fiscal and administrative pressures, technology changes, lack of health insurance and increasing population diversity all contribute to reshaping how health care is delivered and received. Health care facilities might be closing, relocating or expanding, as well as providing different types of practices in different settings. Under this dynamic context, the geography of health care delivery became an essential issue in public health. A service delivery system is defined as “a cluster of diverse agencies within an organizational network that provides services to a common client population” (Alter 1988). The components of a health care delivery system include the patients or population in need, the health care facilities or providers, and the relationships between them. The geographical distribution of health services and geographical factors that affect health care service functions and utilizations contributes greatly to our understanding of health problems (Shannon 1980).

A powerful tool in the geography field – Geographic Information System (GIS) plays an increasingly significant role in implementing analytic methods to understand and analyze health care problems by incorporating geographical physical barriers, network-based travel time, and transportation costs required for access to health care services. (Lo and Wang 2005, Wang 2006, Yang *et al.* 2006, Maheswaran and Haining 2004). In particular, the capability of GIS necessarily highlights the spatial dimension of health care access, which will be discussed in further details in the next section.

1.2 The concept of health care access

The geography of health care access is essential to understanding health disparities. Access is a key element within the health care delivery system. It describes people's ability to use health care services when and where they are needed (Aday and Anderson 1981). According to Petchansky and Thomas (1981), access can be defined as having five dimensions: availability, accessibility, accommodation, affordability and acceptability. *Availability* defines the supply of services in relation to needs – whether there are adequate services to meet the healthcare needs. *Accessibility* refers to the geographical barriers to healthcare such as travel distance, time and cost. It incorporates the spatial interaction between the geographical locations of healthcare services and population in need. *Affordability* describes the price of services regarding people's ability to pay, which is associated with one's social economic status and insurance coverage. *Accommodation* captures the extent to which services meet patients' needs, including waiting time and hours of operation. Finally, *acceptability* describes people's sense of comfort and satisfaction when using healthcare services. It is affected by one's gender, culture, ethnicity and sexual orientation. When conceptualizing health care access, it is also important to distinguish the difference between two broad categories: *revealed accessibility* and *potential accessibility* (Joseph and Phillips 1984). Revealed accessibility focuses on the actual use of health care services, while potential accessibility emphasizes the overall supply of medical care resources available within the region.

Both spatial factors (*e.g.*, the geographical location of primary care physicians and patients, transportation system, and travel distance) and non-spatial factors (*e.g.*, social-economic status, age, gender, ethnicity, and health insurance status) are incorporated as critical determinants for health care access (Guagliardo 2004). Ideally, all should have equal access to health care. Such equal access has come to be recognized as essential in ensuring overall population health. However, the goal to reduce the inequality in health care delivery is far from being achieved. Since much of the health care delivery system is privately funded and organized, it is clear that health care resources are not evenly distributed over geographic space. In the United States, the shortage of health care resources is especially severe in rural areas and impoverished urban communities (Ye and Kim 2014, Ye and Kim 2015). Socio-economically disadvantaged neighborhoods tend to have relatively poor access to resources that promote health, such as access to health care, a healthy diet, or recreational opportunities (Macintyre 2007).

Geographers and public health researchers recognize the significance of measuring access and applied a broad spectrum of techniques to address the issue of health care access (Joseph and Bantock 1982, Higgs 2005, Gu *et al.* 2010, Busingye *et al.* 2011, Lewis and Longley 2012, Hawthorne and Kwan 2013). Within this body of literature, the application of GIS has gained great popularity. GIS necessarily highlights the geographical dimension of access, the spatial accessibility, which is a main scope of this dissertation. In health geography and social sciences literature, the term spatial accessibility is used to refer to the combination of both availability and accessibility of health demand and supply. In this dissertation, the concept of spatial accessibility is also applied when it is convenient or necessary.

1.3 Research goals

In order to understand the geography of health care access, the first goal of my dissertation is to develop an alternative measurement of spatial accessibility in a GIS environment to health care services. The spatial interaction between supply (*e.g.*, health care facilities or providers) and

demand (e.g., population in need) lies within the center of most accessibility measures, which follows a fundamental function of distance decay. The possibility of a patient to seek health care decreases with increasing travel distance, time and cost. Existing measurements include the provider-population ration, gravity models, and two-step floating catchment area method (2SFCA). Each method has its own limitations. For example, the provider-population ration fails to represent the geographical or spatial dimension of access; the gravity model doesn't take the supply side into consideration; and the 2SFCA can result in over- or under-estimation and doesn't capture the spatial mismatch between health care demand and supply. The accessibility measure developed in this dissertation aims to overcome limitations of previous methods and complement existing literature. By applying this new method, we are able to identify underserved areas and provide reference concerning where additional health care facilities and providers should be located.

My second research goal is to include a range of neighborhood characteristics in addressing health disparities, incorporating the application of GIS and statistical models. Investigating the spatial patterns of neighborhood/community characteristics is another important step in studying the geography of health care access. Health problems can be studied at both lower level and higher level units. Here it is necessary to understand the concepts of the contextual effects and compositional effects. Contextual effects refer to higher level variables contributing to differences observed at a lower level. For example, individual health status can be affected by overall neighborhood environment. Compositional effects are differences in an outcome can be attributed to the characteristics of individuals comprising the group rather than the nature of the setting (Duncan, Jones and Moon 1998). For geographers, higher levels such as neighborhoods and regional groups have spatial dimensions, and an increasing amount of research has been devoted to the relationship between people and where they live, rather than rigidly distinguishing between contextual and compositional effects (Kawachi and Kennedy 1999, Patel *et al.* 2004, Li *et al.* 2005). Previous research has confirmed that where we live affects our health status (Basta *et al.* 2008, McLafferty and Wang 2009). Nevertheless, less has been done on revealing how place contributes to the spatial distribution of health care supply – the availability dimension of potential access, as well as health care utilization – an indication of the revealed access.

The third goal of my dissertation is to consider the basic components and dimensions of the health care delivery systems, as well as to develop location models for selecting optimal solutions for siting health care facilities. As discussed above, the location of health services is a key factor that affects accessibility to health care. The supply-demand relationship is essential to health care delivery system. If we view health care facilities and professionals as the supply side, then patients and population in need would be the demand side. Needless to say, facilities should be located according to the potential demand to ensure maximum coverage as well as accessibility equity. Thus a feasible method to locate health care facilities is necessary in order to maximize service coverage while reducing spatial disparity between health care supply and demand.

In conclusion, my dissertation measures and analyzes the geography of health care access from three major perspectives: 1) developing a reliable measurement of geographical/spatial accessibility to health care services; 2) analyzing changing service distribution patterns and understanding neighborhood factors that contribute to both potential and revealed access; 3) developing a feasible planning method for locating health care services. Following section will discuss how to break down my research objectives into specific tasks.

1.4 Research tasks

As mentioned above, this dissertation addresses the health disparity issue through measurement of spatial accessibility, analysis of neighborhood characteristics and location planning. Specifically, those research goals are achieved through following tasks.

1.4.1 Developing an alternative measure of health care accessibility – a network-based health accessibility index method (NHAIM)

Improving health care access begins with developing reliable measures. To achieve this task, Chapter 2 presents an alternative set of health accessibility measures – a *network-based health accessibility index method (NHAIM)* in a GIS environment, which comprehensively evaluate both spatial dimensions of health accessibility and availability. Based on the data downloaded from the Florida Geographic Data Library Documentation and US Census Bureau, Chapter 2 applies the NHAIM to measure spatial disparity in the case of Hillsborough County, Florida. Both health accessibility and availability are measured and presented as indexes to reveal the spatial patterns of health accessibility and availability, as well as to capture the health underserved areas in a geographical context. The NHAIM is able to capture the spatial mismatch between health care supply and demand, which is the greatest contribution of this method to the current literature.

1.4.2 Understanding neighborhood factors and health care access

Chapter 3 and Chapter 4 investigate neighborhood characteristics that contribute to both potential and revealed access, respectively. In Chapter 3, the supply of primary care physicians (hereafter PCPs) is used as an indicator of potential access; while in Chapter 4, the length of inpatient stay is applied as an indicator of health care utilization to address revealed access.

Chapter 3 aims to evaluate the geographical distribution of the number of PCPs by location and investigate the relationship between neighborhood factors and the observed spatial pattern, with an empirical study in Hillsborough County, Florida. Literature highlighted that local supply of primary care physicians affects preventive healthcare service utilization directly (Continelli *et al.* 2010). Since the supply of PCPs reflects the dimension of availability in health care access, it can be applied as an indicator of an effective health care delivery system and whether the health care needs are being adequately served. Therefore Chapter 3 provides a set of spatial statistical models to assess neighborhood factors and the supply of PCPs.

In Chapter 4, revealed access is addressed through the study of length of inpatient stay – an indicator of facility utilization. Using a nationally representative U.S. sample, Chapter 4 examines the extent to which neighborhood characteristics affected length of inpatient stay in the United States. A total of 3148 U.S. counties are included in the study. Regression models are employed to examine the extent to which neighborhood characteristics affect length of inpatient stay and its spatial variation. Exploratory spatial data analysis is also conducted to explore the geographic pattern of length of inpatient stay.

1.4.3 Developing feasible planning methods – location problems in healthcare facility siting

Chapter 5 proposes a facility planning method – a *Network-based Covering Location Problem (Net-CLP)* to achieve maximum service coverage and ensure spatial equity. The Net-CLP

incorporates two models: a *Network-based Maximal Covering Location Problem (Net-MCLP)*, and a *Network-based Location Set Covering Problem (Net-LSCP)*. The Net-MCLP focuses on the supply side and its objective is to maximize covered demands with a fixed number of facilities, given spatial restrictions and the level of health care service emergency. The Net-LSCP accounts for the demand side and minimizes the total number or cost of facilities needed to cover all healthcare service demands within the network-based service range. The extensions of Net-CLP proposed consider the service capacities of healthcare facilities. By applying both models, a more comprehensive evaluation of the candidate sites can be conducted. Moreover, Chapter 5 contributes to the existing literature by demonstrating the geo-processing steps in the application of location problems with the integration of GIS.

1.5 Organization

This dissertation is organized as follows: Chapter 1 introduces research background and motivation; an alternative set of accessibility measurement is presented in Chapter 2; Chapter 3 investigates the relationship between neighborhood characteristics and the supply of PCPs, followed by the examination of the contribution of neighborhood factors to length of inpatient stay in Chapter 4; Chapter 5 proposes a facility planning method and demonstrates its application with the integration of GIS; Chapter 6 summarizes major contributions of the dissertation and discusses future research directions.

CHAPTER 2
MEASURING SPATIAL ACCESSIBILITY
USING A NETWORK-BASED METHOD
IN A GIS ENVIRONMENT

A version of this chapter is published in *International Journal of Geospatial and Environmental Research*, Volume1, Issue 1, Article 2, 2014. The use of “we” in this chapter refers to co-author Dr. Hyun Kim and myself. As the first author, I processed the data, performed the analysis and wrote the manuscript.

Abstract

In recent decades, the health care delivery system in the United States has been greatly transformed and more widely examined. Even with one of the most developed health care systems in the world, the United States still experiences great spatial disparity in health care access. Increasing diversity of class, culture, and ethnicity also has a significant impact on health disparity. The goal of this chapter is to address the spatial disparity of health care access using a *network-based health accessibility index method (NHAIM)* in a Geographic Information System (GIS) environment. Ensuring a desired level of accessibility for patients is the goal of the health care delivery system, through which health care service providers are supplied to populations in need. GIS is able to incorporate geographical physical barriers, network-based travel time, and transportation costs required for measuring access to health care services. In this study, we develop a NHAIM to examine the spatial disparity in health care access in Hillsborough County, Florida, determining the locations of registered medical doctors and facilities using data from Florida Geographic Data Library Documentation and the U.S. Census. This research reveals the spatial disparity of health care accessibility and availability in this region and provides an effective method for capturing health care accessibility surplus and shortage areas for future health care service planning.

Key words: health disparity, accessibility, GIS, Network-based Health Accessibility Index Method (NHAIM)

2.1 Introduction

Accessibility is the key element within the health care delivery system. Ideally, all should have equal access to quality health care. Such equal access has come to be recognized as being as essential to public health as individual health status (Aday and Andersen 1974, Culyer and Wagstaff 1993, Oliver and Mossialos 2004). Penchansky and Thomas (1981) described five dimensions of health access: availability, accessibility, affordability, acceptability, and accommodation. The first two are related to geographical locations and thus inherently spatial. Among them, accessibility reflects the travel impedance between population in demand and health facilities, and is usually measured in travel distance or time. Availability refers to the amount of health facilities available for population in demand to choose from. In health geography literature, the term “spatial accessibility” is used to refer to the combination of these two dimensions (Guagliardo *et al.* 2004, Luo and Wang 2003a, Luo 2004).

Generally speaking, the spatial distributions of health facilities and population in need are not matched perfectly over geographical space (Guagliardo 2004, Luo and Wang 2003b, Parker and Campbell 1998). Therefore, the goal to substantially reduce the inequality in accessing health care services is far from being achieved in the United States. According to Rosenberg and Hanlon (1996), middle and high income individuals are more likely to benefit from better access to family physicians, maintaining a higher health status and practicing preventive health care. Some other studies demonstrate that blacks are more likely to receive late-stage breast cancer diagnosis compared to whites and therefore have higher mortality rates. Additionally, African-Americans are more likely to experience the negative influence of socioeconomic disadvantages, such as low

educational attainment and linguistic barriers over late diagnosis (McLafferty and Wang 2009; Meliker *et al.* 2009a, b; Wang *et al.* 2008). Since socioeconomic and neighborhood inequalities are significantly correlated with health care accessibility, it is not surprising that the shortage of health care supply is especially severe in rural areas and impoverished urban communities (COGME 1998, 2000, Rosenblatt and Lishner 1991, Rosenthal *et al.* 2005, Shen 1998).

Thus, it is of great importance in understanding this dynamic context and exploring accessibility as a multidimensional concept contingent upon the interaction between a variety of spatial factors (*e.g.*, geographical location, travel distance) and aspatial factors (*e.g.*, socioeconomic status, age, gender, and ethnicity) (Joseph and Bantock 1984, Meade and Earickson 2000, Penchansky and Thomas 1981). Theoretically, health facilities should be located according to potential demand such as in areas with high population density to ensure maximum coverage. However, population in demand might not necessarily be covered by the service range of health facilities in reality. Shi *et al.* (2012) identified “islands” with no coverage of major cancer care facilities at a national scale. For example, the most visible high-demand area for cancer care services is located at the contact of Kansas, Missouri, Arkansas and Oklahoma, which happens to be the biggest uncovered “island” in the Midwest. This spatial mismatch between the geographical locations of health facilities and population in demand raises the following question: how do we define, measure and evaluate the accessibility to health care services?

Geographers and public health researchers recognize the significance of measuring accessibility and apply a broad spectrum of techniques to solve this issue. While some focus on mathematical modeling or statistical analysis (Field 2000, Gu *et al.* 2010, Higgs 2005, Joseph and Bantock 1982), others apply a more qualitative approach (Hanlon and Halseth 2005, Hawthorne and Kwan 2013, Kiwanuka *et al.* 2008). Within this body of literature, Geographic Information System (GIS) plays an increasingly significant role in understanding and analyzing accessibility to health care. In particular, the capability of GIS highlights the spatial dimensions of accessibility. For example, Langford and Higgs (2006) estimated ‘demand-side’ population, or potential health care client locations, by applying various spatial interpolation techniques. Yang *et al.* (2006) evaluated access to dialysis health care by using specialized gravity models. Luo and Wang (2003a) measured spatial accessibility to health care and identified health shortage areas in Chicago region. In conclusion, GIS enables researchers to store and manage sensitive yet complicated information for both patients and health service locations (Bullen *et al.* 1996, Gu *et al.* 2010, Verter and Lapierre 2002, Zhang *et al.* 2009), measure access to health services for populations in need (Curtis *et al.* 2006, Lo and Wang 2005, Wan 2006, Wang 2012), and analyze the evolving spatial distribution patterns of health facilities (Gesler and Albert 2000, Higgs 2005, Kurland and Gorr 2012, Pedigo and Odoi 2010, Ross *et al.* 1994).

In this chapter, we present an alternative set of health accessibility measures, which comprehensively evaluate both spatial dimensions of health accessibility and availability in order to address spatial disparity problems. The goal of this chapter is to measure and evaluate spatial accessibility to health care by using a *network-based health accessibility index method (NHAIM)* in a GIS environment. Based on data downloaded from the Florida Geographic Data Library Documentation and US Census Bureau, this chapter demonstrates the application of NHAIM in measuring spatial accessibility to health facilities in Hillsborough County, Florida. Both dimensions of accessibility and availability are measured and presented as indexes to reveal patterns of health disparity, as well as to capture underserved areas.

This chapter is organized as follows. In the next section, we provide a brief review of existing spatial accessibility measures. In the third section, we demonstrate the application of an

alternative method – *the network-based health accessibility index method (NHAIM)*, using Hillsborough County, Florida as a case study. The NHAIM consists of two sub-indexes to measure accessibility and availability respectively and a comprehensive index to evaluate the overall level of health disparity. The fourth section provides the analysis results of the case study, followed by conclusions in the fifth section.

2.2 Spatial accessibility and disparity in health care systems.

Spatial accessibility to health service locations is usually measured through addressing the geographical barriers like travel distance or time (Cromley and McLafferty 2012; Guagliardo 2004). The interaction between population in need and health care providers decrease with increasing travel distance, following a function of distance decay. Shorter geographical distance can lead to more frequent visits to health facilities, and eventually better health for individuals. For example, Buchmueller et al. (2006) found that increasing distances from hospitals result in higher death rates from heart attacks and unintentional injuries. Another study by Arcury et al. (2005) shows that a shorter distance between patients and physicians can increase the frequency of regular family physical exams. Other studies also confirm that early detection of disease and treatment is negatively associated with the spatial separation between medical services and patients (Campbell et al. 2000; Meyer 2012; Monnet et al. 2006; Onega et al. 2008). Distance decay is a fundamental aspect to measure spatial accessibility, and it varies for different types of medical practice and health care needs. For example, cardiovascular emergencies requires patients be delivered to an emergency center within a critical time window (Busingye et al. 2011; Hare and Barcus 2007). For routine health check-ups, there are much less restrictions over travel time or distance (Lovett et al. 2004).

Most existing measures of spatial accessibility are based on the potential interaction between health care providers (*e.g.*, primary care physicians, cancer treatment centers, hospitals, etc.) and population in need, or supply and demand (Guagliardo 2004; Higgs 2005; Wang 2012). One commonly used measure is the supply-demand ratios, or provider-population ratios, which are computed within bordered areas. The ratios are effective for gross comparisons of supply between geographical units, and are widely applied to set minimal standards for local supply and identify underserved areas (Cervigni et al. 2008; Khan 1992; Perry and Gesler 2000; Radke and Mu 2000). For example, the U.S. Department of Health and Human Services (DHHS) uses a minimum population-physician ratio to identify Health Professional Shortage Areas (HPSA). However, this basic measurement has difficulty capturing the border crossing of patients among neighborhood spatial units. Detailed variations in accessibility across space and the distance dimension of access are ignored (Guagliardo et al. 2004; Wang 2012). Another basic method is to measure average travel distance to nearest providers (Fryer Jr et al. 1999; Goodman et al. 1992). This method applies the straight line distance between the population point and the location of the health provider. However, travel routes are rarely straight lines in reality. It also cannot fully represent clusters of health providers in an urban setting and ignores the availability dimension of access.

Gravity models, initially developed for land use planning, are also utilized to account for the spatial interaction between health care supply and demand (Hansen 1959, Joseph and Bantock 1982, Shen 1998). The simplest formula for gravity-based accessibility A_i can be written as follows:

$$A_i = \sum_j^n \frac{S_j}{d_{ij}^\beta} \quad (1)$$

A_i is the index of spatial accessibility from population point i , such as a personal residence or population centroid of certain spatial unit. S_j is the service capacity of health facilities (*e.g.*, the number of hospital beds or doctors) at location j . d_{ij} is the distance or travel time between i and j , and β is the travel friction coefficient. n is the number of health facilities. Spatial accessibility improves if the number of health facilities increases, the service capacity increases, or the travel distance decreases. The improved gravity-based accessibility model proposed by Joseph and Bantock (1982) adds a population adjustment factor to the denominator. The formula can be written as:

$$A_i = \sum_{j=1}^n \frac{S_j d_{ij}^{-\beta}}{\sum_{k=1}^m P_k d_{kj}^{-\beta}} \quad (2)$$

P_k is the population at location k , d_{kj} is the distance or travel time between j and k , and the indexes n and m represent the total number of facility locations and population locations, respectively. The gravity-based accessibility model is essentially the ratio of supply to demand (Huff 1963, 2000, Luo and Qi 2009, Wang 2012). Despite its elegance in revealing geographic variation in accessibility, gravity models are not easy for public health professionals to interpret or implement. A large amount of geo-coded data for the locations of both population and health facilities are required to estimate the travel friction coefficient β . Sometimes the models also involve great effort of computation and programming (Luo and Whippo 2012, Taaffe *et al.* 1996).

Another development in spatial accessibility modeling is the two-step floating catchment area method (2SFCA) proposed by Luo and Wang (2003a, b). The fundamental assumption of 2SFCA is that availability and accessibility are not mutually exclusive and they can compensate each other. A health provider is defined as accessible if located inside the catchment, and inaccessible if located outside of the catchment. The catchment of a provider location is defined as a buffer area within a threshold travel distance or time from the provider. The 2SFCA can be implemented in a GIS environment using two steps. First for each physician location j , search all population locations k that are within the catchment area and compute the provider – population ratio R_j . Then for each population location i , search all provider locations j that are within the threshold distance from location i , and sum up R_j derived from the first step at these locations. Eventually the accessibility index A_i can be written as follows (Luo and Wang 2003a):

$$A_i = \sum_{j \in \{d_{ij} \leq d_0\}} R_j = \sum_{j \in \{d_{ij} \leq d_0\}} \frac{S_j}{\sum_{k \in \{d_{kj} \leq d_0\}} P_k} \quad (3)$$

R_j is the measurement of potential service intensity of facility j , the provider-population ratio. S_j is the service capacity of facility location j , P_k is the population in need at location k , d_{kj} is the travel distance or time between k and j , and d_0 is the threshold.

The 2SFCA has been popular and used in a number of studies (Cheng *et al.* 2012, Dai 2010, McGrail and Humphreys 2009, Ngui and Apparicio 2011, Shi *et al.* 2012, Wan *et al.* 2013, Yang *et al.* 2006). However, Luo and Wang demonstrate that their model is not fundamentally different from the gravity-based accessibility model (Luo and Wang 2003a, b). The 2SFCA overcomes the restriction of using pre-defined geographical boundaries. However, the limitation of 2SFCA is mainly found in assuming a health provider inside a catchment area is accessible and one outside the catchment area is inaccessible, which tends to be arbitrary, ignoring the possibility of overlapping areas in coverage. In addition, potential improvements may be made to account for different transportation options, as well as variable catchment sizes for different populations and

health services. While the above methods make significant contributions in revealing health disparity, we seek to complement such spatial accessibility literature by providing an alternative measure. Recognizing that spatial accessibility is a complex concept including both accessibility and availability, we seek to develop a method that can reveal and represent both dimensions respectively.

2.3 Analytical frameworks and study area

2.3.1 Network-Based Health Accessibility Index Method (NHAIM)

The concept of spatial accessibility to health care includes both dimensions of accessibility and availability. In general, *accessibility* refers to the ease to reach health services from the demand side while *availability* emphasizes choices of local service locations from the supply side. Spatial accessibility to health services is primarily dependent on the geographical locations of health care providers and population in need, as well as the travel distance/time between them (Wan et al. 2013). Since distance decay is a fundamental aspect in understanding spatial accessibility, the following questions were raised when developing our methodology: [1] how to define travel distance and reflect distance decay, [2] how to represent both health care demand and supply, and [3] how to apply the most reasonable measure for travel distance to health care services. *Network distance* has gained certain popularity in recent literature as a replacement for Euclidean distance and Manhattan distance. It is considered to be a more accurate measurement for real travel distance and time (Brabyn and Beere 2006, Cheng et al. 2012, Dai 2010, Delmelle et al. 2013, Pearce et al. 2006, Shi et al. 2012, Wan et al. 2013). However, Apparicio et al. (2008) found that Euclidean and Manhattan distances are strongly correlated with network distances. However, local variations are still observed, notably in suburban areas. Thus in those areas, network-based distance may provide more accurate results. In our study, we applied network-based distance rather than Euclidean distance and Manhattan distance, since the study site includes both urban and suburban areas.

In NHAIM, the population centroid within each spatial unit is used to represent aggregated health care demand location. When health care demand is aggregated, the true distance to health care services from each individual or household is replaced by the distance from the aggregation point (Current and Schilling 1990). The aggregation method can reduce the complexity of location and routing problems as well as protect the privacy of the individual or household by masking their individual locations, especially in sensitive research. The population centroid for each health care demand area can be obtained in a GIS environment through preprocessing.

The fundamental issue in spatial accessibility literature is addressing the potential interaction between health care providers and population in need. However, it is difficult to predict people's choices and behaviors, especially with border crossing problems. The term "edge effect" is coined to describe the possibility of accessing health providers across borders (Cromley and McLafferty 2012, Guagliardo 2004, Higgs 2005, Wang 2012). The NHAIM tries to mitigate edge effect by evaluating and integrating both dimensions of health care accessibility and availability. As summarized in Table 1, the NHAIM consists of three sub-indexes. The first sub-index, the *Network-Based Health Accessibility Supply Index (NHA-SI)*, is developed from the supply side and reveals the availability of health care providers in each spatial unit. The second one, the *Network-Based Health Accessibility Demand Index (NHA-DI)*, is developed from the demand side and evaluates health care accessibility for the population in demand residing within each spatial unit. The third sub-index, the *Network-Based Health Access Disparity Index (NHA-DP)*, is a global

index summarizing both dimensions. Ultimately, the NHAIM is designed to evaluate the interaction between health demand and supply, and present both sides of the interaction.

Table 1. The indexes of NHAIM

NHAIM	NHA-SI	Network-Based Health Accessibility Supply Index: <ul style="list-style-type: none"> • Reflects health care access from the supply side • Measures service availability in terms of health facilities
	NHA-DI	Network-Based Health Accessibility Demand Index: <ul style="list-style-type: none"> • Reflects health care access from the demand side • Measures overall health care accessibility for the population in demand
	NHA-DP	Network-Based Health Access Disparity Index: <ul style="list-style-type: none"> • Combines both the NHA-SI and NHA-DI • Measures both health care accessibility and availability

1) Network-Based Health Accessibility Supply Index (NHA-SI)

The *Network Health Accessibility Supply Index (NHA-SI)* addresses health care access problems from the supply side. The NHA-SI is an indicator quantifying the availability of health care supply within the measured spatial unit.

As illustrated in Figure 1, the NHA-SI can be achieved through the following four steps in a GIS environment. The first step (Step 1 in Fig. 1) is to represent health care demand locations using population centroids. As suggested by Current and Schilling (1990), the population centroid within each spatial unit can be used to represent the aggregated health care demand location. The demand aggregation method reduces the complexity of location problems and protects the privacy of individuals or households, especially in sensitive researches. Demand aggregation can result in over- or underestimation of true distance and health care supply coverage (Cromley and McLafferty 2012, Openshaw 1983). According to Hewko et al. (2002), aggregation error is a result of spatial separation between the distribution of individuals and the centroid of spatial unit. Thus accessibility measured for smaller units tends to be more reliable than that measured for larger spatial units. The second step (Step 2 in Fig. 1) involves calculating the network demand area of health care for each population centroid. A health care demand area is defined as a network-distance travel zone from the population centroid. Coverage is measured based on travel distances calculated using road networks. The sizes of demand areas vary according to different types of health services. For example, cancer treatment centers generally cover larger demand areas than primary care providers. Luo and Qi (2009) defined the threshold travel distance to Primary Care Physicians as 30 minute network travel distance. Wan et al. (2013) extended the travel time to 60 minute focusing on access to cancer screening and treatment facilities. Thus the thresholds for travel distances are flexibly set to reflect different types of health services. The size of a demand area expands as the threshold travel distance increases. In the third step (Step 3 in Fig. 1), we calculate the population within each demand area i generated in the previous step. The calculated population is denoted by p_i . Next we search every demand area that covers health care facility j . This can be interpreted as the health care facility j serving n demand areas ($n \geq 0$; $n =$ the number of demand areas covering facility j). In the end, we calculate P_j , the total population residing within n demand areas that facility j is serving, which is expressed by $P_j = \sum_i^n p_i$. Note that the population residing within the overlapping areas (i.e., intersections in Step 3 in Fig. 1) is only counted once

to get the most accurate result. The final step (Step 4 in Fig. 1) calculates s_j , the health care accessibility for each facility j using following formula:

$$s_j = k \frac{C_j}{P_j}, \quad 0 < s_j < 1, \quad \forall j \quad (3)$$

where

C_j : the capacity of facility j (e.g., the number of beds/rooms as a proxy for supply capacity),

P_j : the total population residing within n demand areas that facility j is serving

k : is the scalar to adjust the ratio.

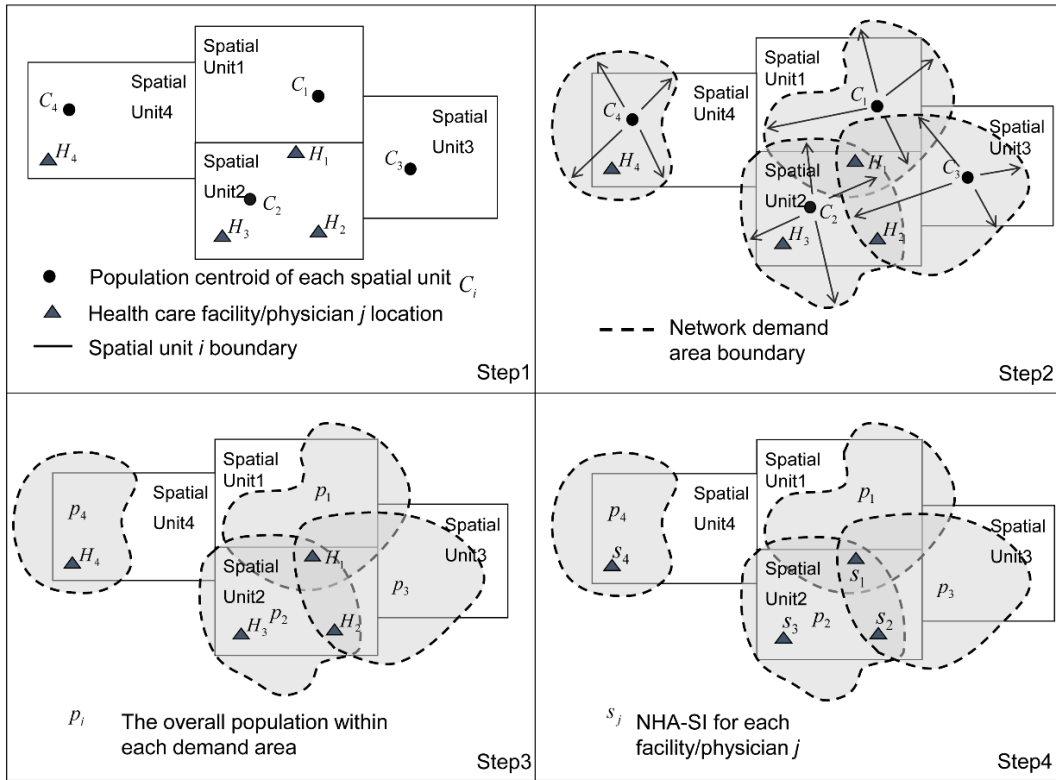


Figure 1. The procedure for calculating the Network-Based Health Accessibility Supply Index (NHA-SI)

Now, as each spatial unit i will contain n ($n \geq 0$) health facilities j with attribute s_j , the S_i , the NHA-SI for spatial unit i is calculated as follows:

$$S_i = \sum_{j=1}^n s_j, \quad j \in R \quad (4)$$

where

R : the set of facilities j located within spatial unit i .

When the facility capacity is fixed, the NHA-SI index will be smaller if a greater health care demand is identified. Both higher facility capacity and smaller population in need will result in a larger NHA-SI value, which reflects less constraints over the facilities and represents a higher level of availability in the measured spatial unit.

2) Network-Based Health Accessibility Demand Index (NHA-DI)

The *Network Health Accessibility Demand Index (NHA-DI)* evaluates the overall accessibility from the demand side by calculating the percentage of population residing within the service ranges of health care facilities in each spatial unit, as well as taking the capacities of those facilities into consideration. It reveals the general level of health care accessibility in the measured spatial unit. As illustrated in Figure 2, the NHA-DI index is achieved in the following four steps.

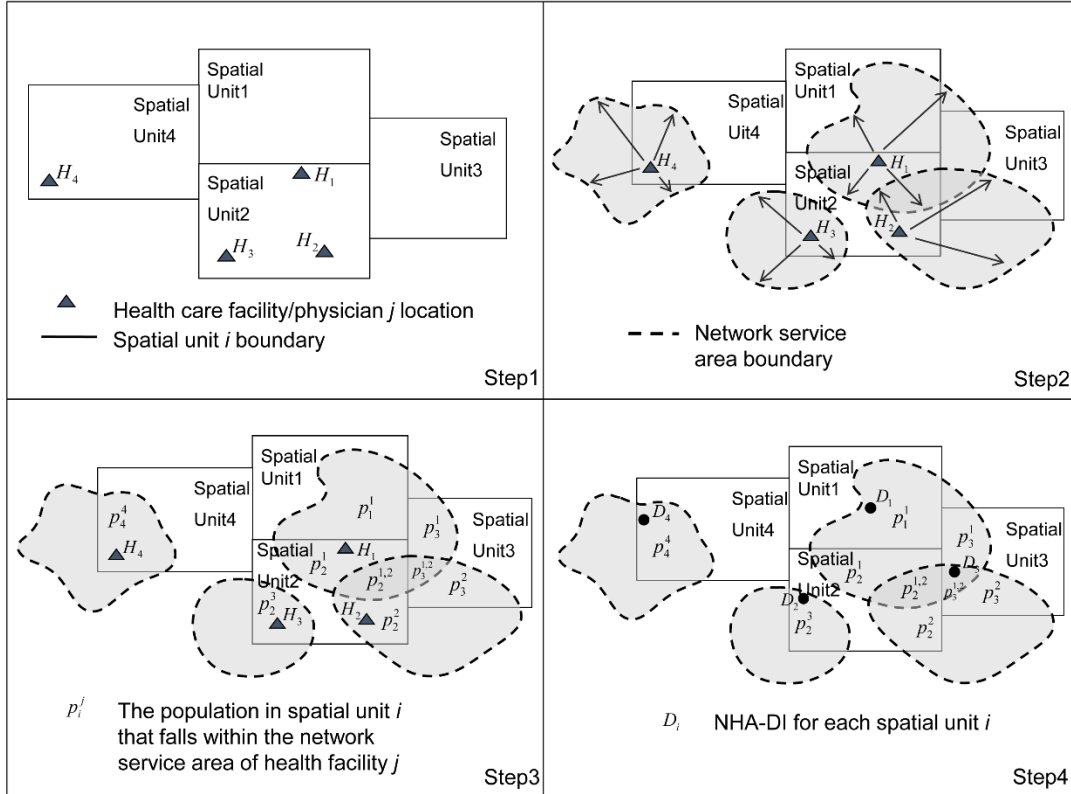


Figure 2. The procedure for calculating the *Network-Based Health Accessibility Demand Index (NHA-DI)*

In the first step (Step 1 in Fig. 2), we identify the locations of health care facilities within each spatial unit. Next, we calculate the network service area for each facility j (Step 2 in Fig. 2). The service area is defined as the network distance travel zone from facility j . Similar to the demand areas generated from population centroids when calculating NHA-SI, the sizes of service areas also vary for different types of health services according to different travel time thresholds (Luo and Qi 2009, Wan et al. 2013). In the third step (Step 3 in Fig. 2), we calculate the population ratio covered by the network service area of facility j in spatial unit i : p_i^j/p_i ($p_i^j \leq p_i$). p_i^j is the population in spatial unit i that falls within the network service area of health facility j , while p_i is the overall population residing in spatial unit i . The final step (Step 4 in Fig. 2) involves calculating D_i , the NHA-DI index for each spatial unit using the following formula:

$$D_i = k \sum_{j=1}^n C_j \frac{p_i^j}{p_i}, \quad 0 \leq D_i \leq 1, \quad \forall i \quad (5)$$

where

p_i^j/p_i : the ratio of population covered by the network service area of health facility j to the total population in spatial unit i . According to the formula, either a higher percentage of population covered by the network health service areas (p_i^j) or higher facility capacities results in a higher NHA-DI index. A higher NHA-DI index indicates that more people have access to higher capacity facilities, which is considered as having better health accessibility.

3) Network-Based Health Access Disparity Index (NHA-DP)

The *Network Health Access Disparity Index* (NHA-DP) is a comprehensive index that examines the balance between health care accessibility and availability at each spatial unit by evaluating both indexes: the NHA-SI and NHA-DI. Each spatial unit will contain two attributes: a population centroid in spatial unit i with attribute D_i and n health facilities with summed up attribute S_j . The level of spatial disparity A_i for each spatial unit i is represented as:

$$A_i = [\text{NHA-SI}, \text{NHA-DI}] = [D_i, S_i], \quad \forall i \quad (6)$$

Accordingly, a spatial unit can be categorized into one of four quadrants based on indexes A_i as illustrated in Figure 3. In detail, the first quadrant (1Q) is the area of High Accessibility and High Availability (HAc and HAv), indicating that the population in demand has high access to health care facilities and that the health care services are highly available and sufficient. This shall be the most ideal situation in the health care delivery system. The second quadrant (2Q) is for the area with Low Accessibility and High Availability (LAc and HAv). If a spatial unit falls into this category, then the area has a relatively sufficient health supply within the spatial unit but the population in demand is not well supported by access to health care facilities. This is the case of a spatial mismatch between health care demand and supply where an area has sufficient medical facilities but they are clustered in such a way that their service range does not cover well the demand within the area. The third quadrant (3Q) represents the area with Low Accessibility and Low Availability (LAc and LAv), where the population in demand has very little health accessibility, with highly constrained availability of the health care supply within the area. Moreover, such populations do not have health facilities in neighboring units. As an extreme case, this category includes areas with an extremely high demand for health care services but little to no availability, representing a significant spatial disparity in health care provision. The fourth quadrant (4Q) describes the area with High Accessibility and Low Availability (HAc and LAv) in which health care resources within the spatial unit are highly constrained to serve a large population, or there are no health care facilities located within the spatial unit. However, the population in demand is able to access health care facilities located in nearby spatial units and the accessibility index is relatively high. The shortage of health care supply within the spatial unit does not affect the health care accessibility of the population in demand due to the wide coverage of health care service by facilities. Given this categorization, the results based on the NHA-DP are straightforward to interpret and easy to apply for any geographic scale.

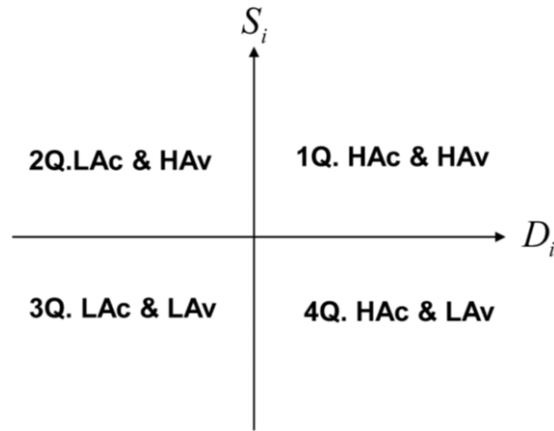


Figure 3. The classification of spatial units using the NHA-DP (Note: H=High, L=Low, Av=Availability, Ac=Accessibility)

2.2.2 Study Area and Data

To test the NHAIM, we selected Hillsborough County in Florida as the study area. Hillsborough County is located on the west coast of Florida in the Tampa-St. Petersburg metropolitan area. It is the largest county by metropolitan area and the fourth largest county in the state. Hillsborough County has a relatively even and flat landscape, which decreases the effect of geographic barriers. As shown in Figure 4, this county comprises several major cities, including Tampa, Temple Terrace, Lutz, Plant City, Brandon, Apollo Beach, Ruskin, and Sun City Center. Noticeably, many hospitals are clustered in those cities along major highways. The population data was extracted from the 2010 Census Summary File (US Bureau of Census 2010). In this case study, we used ZIP code areas as the basic spatial unit to apply the NHAIM. Population-weighted centroid was used to represent aggregated demand location, which is considered to be a more accurate representation than simple geographic centroid (Hwang and Rollow 2000). The population centroids for ZIP code areas were generated based on Census Tract level population data in a GIS environment. ZIP code areas are aggregated to develop Primary Care Service Areas (PCSAs) by the U.S. Department of Health and Human Services (DHHS), which are the basic spatial units used to identify Medically Underserved Areas/Populations (MUA/Ps) and Health Professional Shortage Areas (HPSAs). Fifty-five ZIP code areas were identified in the study area with a total population of 1,229,226 (as of 2010). The network distance between any pair of population centroids and health facility locations was measured based on the road networks for travel-time distance estimation using the 2010 Census TIGER/Line files.

The hospital data for Hillsborough County was downloaded and extracted from the Florida Geographic Data Library Documentation. The original dataset includes the addresses and capacity information of hospitals in Florida in 2010. The hospital locations were geo-coded using ArcGIS 10.1. To apply the NHAIM, we needed to define two key parameters – the travel time threshold and health care facility capacity. In the previous literature, a 30 minute travel time threshold for primary road conditions is suggested (Lee 1991). The 30 minute threshold is also used for defining rational service area and capturing HPSAs by DHHS (Wang and Luo 2005). In this case, we used both 30 minute and 10 minute travel zones for comparison. By applying different thresholds, we were able to evaluate how travel distances influence spatial accessibility. Since the information on

number of physicians is lacking, we used the number of hospital beds as a measurement for facility capacity.

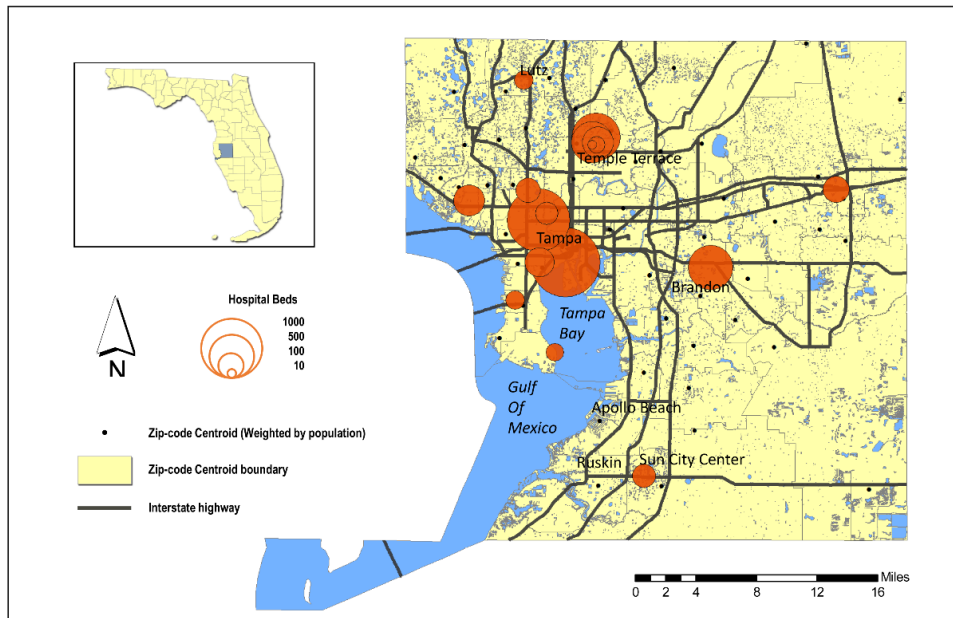


Figure 4. The study area: distribution of hospitals in Hillsborough County, FL.

2.4 Results

Figure 5 shows the NHA-SI indexes for both 10 minute and 30 minute thresholds. 75% of the ZIP code areas are identified with zero number of hospitals ($S_i=0$), and most of them are rural areas. In contrast, the NHA-SI obtains highest values within and around Tampa, followed by Brandon and Temple Terrace, which are urban areas with high population densities. The NHA-SI decreases as the threshold increases from 10 minute to 30 minute. As the travel time increases, the service range of a hospital increases. The hospital becomes less ‘available’ for it is serving a larger population while the capacity is fixed. The NHA-SI becomes smaller as the denominator – the number of population gets larger.

Compared to Figure 5, Figure 6 highlights the spatial heterogeneity of the NHA-DI. First, the highest values are observed in Tampa (West and South, in particular) and Temple Terrace, while most rural areas obtain much lower values. The high values of NHA-DI indicate the satisfaction for the demand-side. The population residing in urban areas with high NHA-DI values benefit from accessibility to local hospitals with large capacities. Second, NHA-DI is highly dependent on the threshold. As the threshold increases from 10 (Fig. 6-a) to 30 minutes (Fig. 6-b), more spatial units obtain higher NHA-DI values. Hospitals further away from the population centroid will become accessible when the threshold increases, which improves the overall level of accessibility.

The NHA-DP evaluates the interaction between both dimensions of accessibility and availability. Figure 7 shows that ZIP code areas are categorized into groups based on NHA-DP. We calculate the Means – the average values for both NHA-SI and NHA-DI, respectively to determine area’s positionality among quadrants. The availability level of a ZIP code area is *High*

Availability (HA_v) if the NHA-SI value is above the Mean and *Low Availability* (LA_v) if it's below the Mean. Similarly, the accessibility level of a ZIP code area is classified as *High Accessibility* (HA_c) if the NHA-DI is above the Mean and as *Low Accessibility* (LA_c) if the index value is below the Mean. According to the results, four quadrants of NHA-DP are identified.

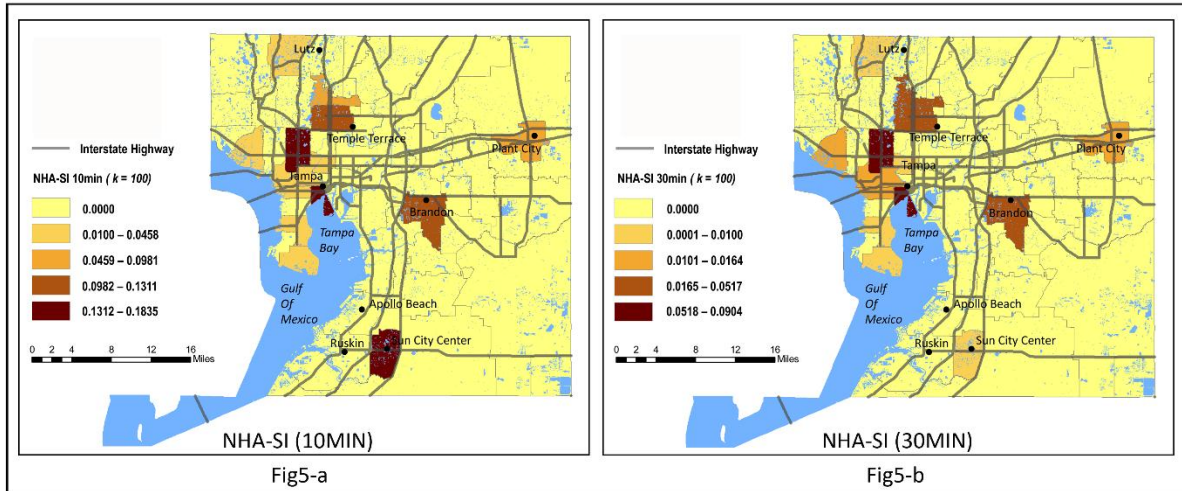


Figure 5. Spatial pattern by the NHA-SI, 10 (5-a) and 30 minutes time zone buffer (5-b).

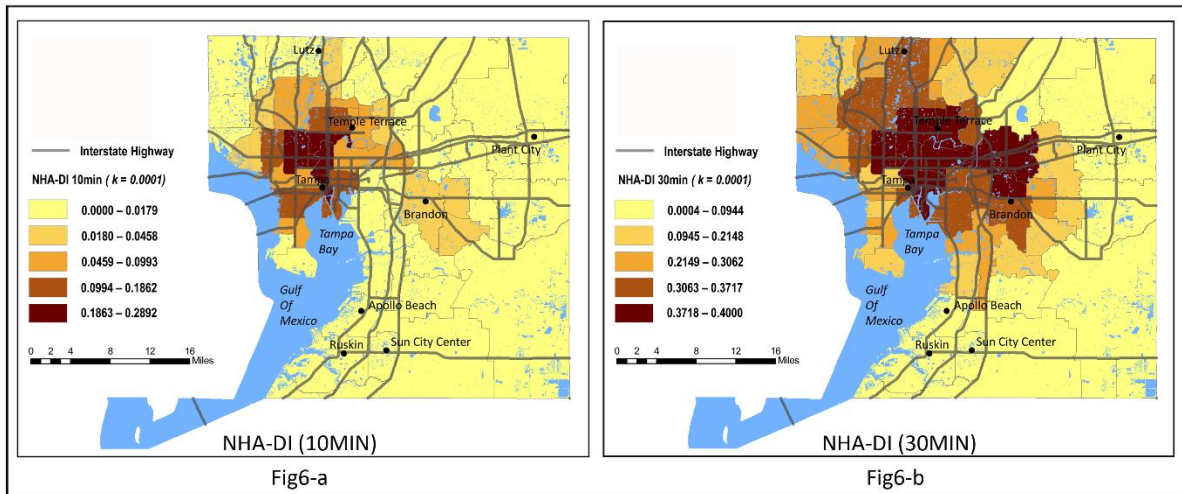


Figure 6. Spatial pattern by the NHA-DI, 10 (6-a) and 30 minutes time zone buffer (6-b).

The NHA-DP evaluates the interaction between both dimensions of accessibility and availability. Figure 7 shows that ZIP code areas are categorized into groups based on NHA-DP. We calculate the Means – the average values for both NHA-SI and NHA-DI, respectively to determine area's positionality among quadrants. The availability level of a ZIP code area is *High Availability* (HA_v) if the NHA-SI value is above the Mean and *Low Availability* (LA_v) if it's below the Mean. Similarly, the accessibility level of a ZIP code area is classified as *High Accessibility* (HA_c) if the NHA-DI is above the Mean and as *Low Accessibility* (LA_c) if the index value is below the Mean. According to the results, four quadrants of NHA-DP are identified.

- 1Q: the west and north areas of Tampa benefits from both HAv & HAc, which is considered as being ideal in terms of balance between health care demand and supply. There are two possible reasons behind this pattern. First, areas identified with HAv & HAc are recognized as being the core of the Tampa metropolitan area. It consists of a fast-growing population area called New Tampa, and the Tampa downtown area, where most hospitals are located. Second, the Tampa downtown area is the center of Hillsborough and the well-developed transportation network ensures great accessibility.
- 2Q: Several pockets with HAc & LAV are captured around Tampa and Temple Terrace. These ZIP code areas contain either none or very few hospitals, but the local population is able to access other facilities in neighboring areas. Since the level of accessibility is high, this case is considered as being acceptable.
- 3Q: The spatial mismatch between health supply and demand is captured in areas identified with LAc & HAv, such as Plant City, Brandon, and Sun City Center. Although the hospitals are considered as being ‘available’ with satisfactory capacities, the local population somehow do not obtain a high level of accessibility. This could be the result of either a low percentage of population covered by service areas or poor transportation networks. The LAc & HAv areas can potentially evolve into an ideal level, HAc & HAv. For example, when the threshold increases from 10 minute to 30 minute, Brandon is re-classified as a HAc & HAv area. Further research is needed to explore the cause of this spatial mismatch and improve the accessibility level from a health planning perspective.
- 4Q: Most rural areas in Hillsborough County are identified as having both LAc & LAV. Those areas are considered as health service shortage areas. As the threshold increases, the level of spatial accessibility in some areas improve from LAc & LAV to HAc & LAV, since population become able to access hospitals further away. Still, LAc & LAV areas require extra attentions when allocating health resources in the future.

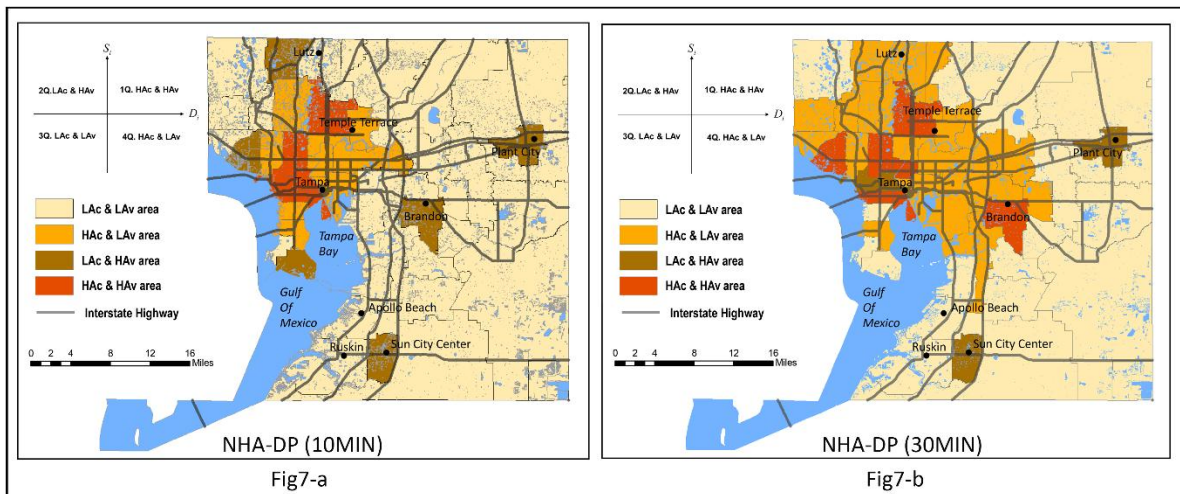


Figure 7. Spatial disparity by the NHA-DP, 10 (7-a) and 30 minutes time zone buffer (7-b).

2.5 Conclusion Remarks

This chapter aims to propose an alternative set of methodology – the NHAIM to measure spatial disparity in the health care delivery system. We demonstrate the application of the NHAIM in a case study of Hillsborough County, Florida. The NHAIM applies network distance rather than Euclidean distance, which improves accuracy in capturing distance decay. The greatest strength of the NHAIM is in measuring and representing both dimensions of accessibility and availability, respectively. Instead of using one single index to represent the overall level of spatial accessibility, the NHAIM evaluates the interaction between both dimensions of accessibility and availability, and examines the potential access to health care facilities located in neighboring spatial units. In the end, spatial units are categorized in four groups. Areas with HAc & HAv or HAc & LAV are considered as being acceptable, while areas with LAc & HAv and LAc & LAV need further investigation and improvement. The results are straightforward for health professionals and policy makers to interpret. Since applying network distances and generating population centroids can be easily achieved in GIS, the application of the NHAIM should not be intimidating for most professionals.

This research has several limitations that need to be further explored in the future. First, we only applied hospital data as a general estimate of health care resources for our analysis. The NHAIM is supposed to be applicable for different levels of health services in theory. Since the major focus of this chapter is methodology, we didn't demonstrate how NHAIM can also be applied to other levels of health care such as primary care services. We would like to expand the application of NHAIM to other levels of health care in future studies. Second, there are smaller spatial units than ZIP code areas that can be applied in the future research. Using smaller spatial units can reduce aggregation errors when applying accessibility measures. Third, we didn't include aspatial factors in this study. People with low socioeconomic status might still have no desirable access to health care despite that they are residing in areas identified with high levels of spatial accessibility. Thus, the results of the NHAIM can be complemented by qualitative analysis. Fourth, further improvement of the NHAIM can be made by applying different thresholds for urban and rural areas, as well as taking into account of multiple transportation modes. For example, network travel distances might be smaller in urban areas when the travel time is fixed considering traffic congestion.

In summary, this chapter demonstrates the application of the proposed method – the NHAIM by measuring spatial accessibility to hospitals in Hillsborough County, Florida. Some areas are identified as having a spatial mismatch between health care demand and supply, or simply being short of supply. The results provide a direct and straightforward reference for future health planning in the study area.

CHAPTER 3
NEIGHBORHOOD CHARACTERISTICS
AND
PRIMARY CARE PHYSICIAN SUPPLY

A version of this chapter is published in *Papers in Applied Geography*, Volume 1, Issue1, 2015, page 70-85. The use of “we” in this chapter refers to co-author Dr. Hyun Kim and myself. As the first author, I processed the data, performed the analysis and wrote the manuscript.

Abstract

Primary care is the key element within the health care delivery system in the United States. Primary care provides initial treatments for patients and refers them to specialists if the patients need advanced cares. However, there is an overall shortage of primary care physicians (PCPs) across the country and its spatial disparity at different geographic scales is prominent, highlighting a major challenge for public health. Thus, it is important to examine the factors that affect PCPs distribution and understand the dynamic spatial patterns of PCPs.

Two factors linked to health disparity – poverty and racial composition, have been greatly explored in health literature, concerning their significant correlations to individual health status. However, less has been done on exploring their contribution to the supply side of health care, such as the spatial distribution of PCPs at a local geographical scale. Theoretically, PCPs should be evenly distributed across space according to the demand of patients. However, some socio-economic factors may affect the spatial pattern of PCPs locations resulting in spatial disparity between demand and supply. In this chapter, we provide a set of spatial statistical models to assess the disparity and identify crucial factors such as poverty, racial composition, insurance rate and others with the case study of Hillsborough County, Florida.

Hillsborough County, Florida is known for being one of the largest counties in the United States. It has a racially diverse population and over 3,000 registered PCPs. In this study, the geographical locations of PCPs were geocoded using active PCPs’ practice addresses data in 2010. The analysis consists of three steps. First, the spatial patterns of PCPs distributions were examined to detect possible clusters. Second, to assess crucial factors that might influence PCPs distribution at a local level, both spatial lag and spatial error regression models were applied, since there was spatial dependency in both PCPs locations and regression residuals. Other than blacks and whites, two ethnic groups that were generally ignored in previous studies– Hispanics and Asians, were also included in the analysis.

Key words: health disparity, primary care physician (PCP)

3.1 Introduction

Disparities in access to quality health care services are of great concern in the field of public health (Brown *et al.* 2000). Penchansky and Thomas (1981) defined five dimensions of health care access: affordability, acceptability, accommodation, availability, and accessibility. Availability refers to the adequacy of the health care providers, which is considered as a spatial dimension of access. The supply of primary care physicians (PCPs) reflects the dimension of availability in health care access. It is essential to the health care delivery system and whether the health care needs are being adequately served (Aday and Andersen 1974, Guagliardo 2004). Both local and state level studies have firmly concluded that an adequate supply of PCPs could lead to frequent health checkups for the population in need, therefore improving the health status of individuals. A county-level study in Florida supported the conclusion that a greater supply of PCPs reduces the incidence rate and mortality rate of colorectal cancer (Roetzheim *et al.* 2001). Another study conducted in Florida found that the PCPs supply is positively correlated with early detection of

breast cancer (Ferrante *et al.* 2000). In other words, poor geographic access to primary care and screening services is linked to higher risks of late-stage cancer diagnosis (Tarlov *et al.* 2008; Wang *et al.* 2008). Other state-level studies also firmly concluded that a greater supply of PCPs leads to better population health outcomes (Shi 1992, 1994, Shi and Starfield 1999, Shi *et al.* 2002).

However, the United States is facing an overall shortage of primary care physicians (PCPs) across the country. According to AAMC (2012), without the Affordability Care Act (ACA), the U.S. would have been short of approximately 64,000 physicians by 2020 but with the implementation of the ACA, it will be 91,000 physicians short. This shortage of PCPs poses a major challenge for public health since primary care plays a fundamental role in the complex health care delivery system in the U.S. (Continelli *et al.* 2010, Cromley and McLafferty 2011). Primary care physicians are the first stop for people to seek health care, who see patients initially and provide referrals to specialists. Primary care involves the widest varieties of health care, for all different age, socio-economic status, and ethnic groups (Starfield 1994). Under this context, it is of utmost importance to evaluate the geographical distribution of PCPs and identify factors that contribute to this spatial pattern. Since much of the health care delivery system is privately funded and organized, PCPs are not evenly distributed over geographic space. While most research has focused on individual-level factors such as race, income, education, insurance status and disability (Brown *et al.* 2000, Dai 2010), there is a growing number of literature studying neighborhood-level characteristics that contribute to health care access (Kirby and Kaneda 2006, Bissonnette *et al.* 2012). Research suggested that key neighborhood social attributes such as socioeconomic status, perceptions of crime and safety, and social ties and networks play key roles in impacting individual health (Kawachi 2000, Kawachi and Glass 2000, Patel *et al.* 2004). Physical characteristics are also proved to have direct impact over the health status and health behavior of local residents (Witten *et al.* 2003, Li *et al.* 2005, Gauvin *et al.* 2008, Sallis *et al.* 2009). Within this body of literature, although more has been done to analyze how social and physical characteristics of neighborhoods contribute to the individual level of health status, there is a general gap in studying the relationship between aggregated neighborhood characteristics and the availability dimension of health care access – the supply of health care services (Kawachi and Kennedy 1999, Buka *et al.* 2003, Patel *et al.* 2003, Wen *et al.* 2006).

This chapter contributes to the body of literature by investigating associated factors to explain geographical distribution of PCPs in the selected study area – Hillsborough County, Florida. Spatial analyses were applied based on census tract level data. Our objectives are twofold. First, the geographical distribution of PCPs was explored to detect possible clusters. Next the contribution of aggregated neighborhood-level factors to the spatial heterogeneity of PCPs supply was examined using spatial regressions.

3.2 Literature Review

The availability of health care providers, in particular, PCPs, is a key dimension in health care access. Here the concept of ‘health care access’ is inherently multi-dimensional since both spatial factors (*e.g.*, the geographical location of PCPs and patients, transportation system, travel distance) and non-spatial factors (*e.g.*, social-economic status, age, gender, ethnicity, health insurance status) are incorporated as critical determinants (Penchansky and Thomas 1981, Joseph and Phillips 1984, Meade and Earickson 2000). A great number of literature on access to health care focused on the contribution of non-spatial factors to individual health outcomes (Kawachi 2000, Kawachi and Glass 2000, Patel *et al.* 2004, Kawachi *et al.* 1999, Buka *et al.* 2003). For example,

multiple studies firmly concluded that area deprivation and poverty has significant negative impact on health outcome at different geographical scales (Basta *et al.* 2008; Jones *et al.* 2004; Shouls *et al.* 1996; McLafferty and Wang 2009). Previous research also confirmed the significant role of race in health care access and disparity. Studies suggested that blacks are more likely to receive late diagnosis and experience higher mortality rate than whites (McLafferty and Wang 2009; Meliker *et al.* 2009a, b; Gebreab and Roux 2012). For example, the coronary heart disease (CHD) mortality in blacks remains significantly higher than in whites (Lloyd-Jones *et al.* 2010). CHD death rates per 100,000 populations were 161.6 for blacks and 134.2 for whites in the US (Keenan and Shaw 2011).

More specifically, aggregated neighborhood characteristics has drawn increasing attentions in public health and epidemiology. Studies have shown a strong correlation between neighborhood of residence and health outcomes, including self-rated health, chronic conditions, mental health, etc. (Kawachi 2000, Kawachi and Glass 2000, Patel *et al.* 2004, Witten *et al.* 2003, Li *et al.* 2005). Duncan *et al.* (1999) found that neighborhood socio-economic status affects health related behaviors, such as walking and smoking. Black and Macinko (2008) demonstrated that indicators of neighborhood socio-economic composition have a significant impact on the obesity risk. Dai (2010) identified a positive correlation between the rates of late-stage breast cancer diagnosis and neighborhood social-economic/racial composition. The results revealed that living in areas with greater black segregation significantly increases the risk of late-stage diagnosis. Kwag *et al.* (2011) suggested that neighborhood characteristics significantly influence the physical and mental health of Korean American older adults. Generally speaking, socio-economically disadvantaged neighborhoods tend to have relatively poor access to resources that promote health, such as access to health care, healthy diet, or recreational opportunities (Macintyre 2007). Because of the residential segregation in the United States, lack of access to health-promoting resources is associated with racial and ethnic inequity in health outcomes (Mennis *et al.* 2012).

When studying aggregated neighborhood characteristics, spatial statistics have been widely applied to account for geographical variations. For instance, Tassone *et al.* (2009) applied Bayesian spatial hierarchical modeling to evaluate the spatial heterogeneity in the relationship between racial disparity and stroke mortality in the southeastern U.S. Gebreab and Roux (2012) examined the spatial heterogeneity in black-white differences in CHD mortality across the country, as well as assessed the contribution of poverty and segregation using a geographically weighted regression approach. Chen and Truong (2012) identified areal specific correlations between neighborhood disadvantages and elevated obesity risk in Taiwan, using both multilevel modeling and geographically weighted regression.

Disparities in access to primary health care services could be in part due to differences in the supply of health care providers among neighborhoods. The accessibility to and the availability of primary health care can directly impact individual health (Kirby and Kaneda 2006). Research has demonstrated that increased travel distance from population in need to health care services results in decreased utilization of those services and therefore increased inequality in health. Arcury *et al.* (2005) found that a shorter distance between patients and physicians can increase the frequency of regular family physical exams. Buchmueller *et al.* (2006) suggested that increasing distances from hospitals result in higher death rates from heart attacks and unintentional injuries. Mennis *et al.* (2012) concluded that longer travel time to community-based psychiatric treatment suppress treatment continuity for drug-dependent patients. Other studies also confirm that early detection of disease and treatment is negatively associated with the spatial separation between medical services and patients (Campbell *et al.* 2000, Monnet *et al.* 2006, Onega *et al.* 2008, Meyer

2012). Studies also confirm that the supply of health care service may influence access and use. A better supply of primary care in a neighborhood guarantees reduced travel distances, which encourages more frequent visits and helps patients maintaining better health. For example, Andersen *et al.* (2002) found a positive correlation between the number of federally qualified health centers available and the likelihood of having a usual source of care. Thus the disparity in the geographic distribution of health care services is of great concern.

Despite the abundance of literature on neighborhood effect and health care access, studies on the relationship between neighborhood factors and the supply of health care services are generally lacking. This chapter aims to compensate this gap by examining the supply of PCPs and associated neighborhood factors with a case study of Hillsborough County, Florida. We hypothesize that aggregated neighborhood characteristics (*i.e.*, proportions of individuals living below the poverty level, proportion of individuals 40 years of age and older, and proportions of racial/ethnic groups) may be correlated with the geographical locations of PCPs, which affects the ability of local residents to seek health care.

3.3 Data and method

3.3.1 Study area

Hillsborough County, Florida was selected as our study area because it has a large racially and ethnically diverse population, which is ideal for analyzing socioeconomic, ethnic and geographic disparities in access to PCPs. Hillsborough County is located on the west coast of Florida in the Tampa-St. Petersburg metropolitan area. It is the largest county in metropolitan area and fourth largest county in the state. Note that most literature on racial disparity in health focuses on white-black inequality, and generally ignores other racial groups such as Asians and Hispanics (Wang *et al.*, 2008; Zhang and Ta, 2009; Kwag *et al.*, 2011; Kirby and Kaneda, 2013). Since Florida has the third largest Hispanic population in the States, and Asian is a major minority ethnic group in Hillsborough County, we incorporate these two racial groups as well as whites and blacks in our study. Despite the abundant research on the impact of racial disparity in individual health outcome, little work has linked the contribution of racial composition to differential distribution of health resources, such as PCPs. Identifying the contribution of racial composition to geographic heterogeneity of health resources can provide important clues for improving health equality in the future, which is also one of the contributions of this study.

According to U.S. Census Bureau in 2012, of residents who reported being of one race, 73.3% were White and 16.7% were Black or African American. Asian is a major minority group, and accounted for 3.4% of the population. American Indian and Alaska Native, Native Hawaiian and other Pacific Islander accounted for 0.6% in total. Nearly one fourth of the population are reported as being Hispanic or Latino origin. According to the Behavioral Risk Factor Surveillance Survey (BRFSS) report, 12% of Whites, 25% of Blacks and 36% of Hispanics perceived their health status as fair or poor in Hillsborough County, Florida in 2009.

Considering the dynamics of physical geography and our concern being habitable areas, we exclude areas where no people actually resides, such as swamp and forest, to represent health supply and demand areas as accurately as possible, using the land use data downloaded from Florida Natural Areas Inventory.

3.2.2 Primary care physician data

This chapter focuses on primary care physicians given their gate keeper roles in the health care delivery system. Primary care generally focuses on illness prevention, health promotion and referrals to specialists for further diagnosis and treatment. It is the first stop for the population in need to seek health care, and access to secondary and tertiary care is usually mediated through primary providers. Above all, primary care relies on an ongoing physician/patient relationship, forming an important part of the local neighborhood landscape (Bissonnett *et al.* 1996). When conceptualizing health care access, it is also important to distinguish the difference between potential and realized access (Joseph and Phillips 1984). Realized access reflects the actual use of health services, while potential access refers to the supply of health care resources (Andersen *et al.* 2001). Thus the spatial distribution of PCPs in a given area is a barometer to measure potential access (see Figure 8). Considering its role as the gateway to specialists and tertiary health care, a lack of PCPs will also affect access to higher level of health care.

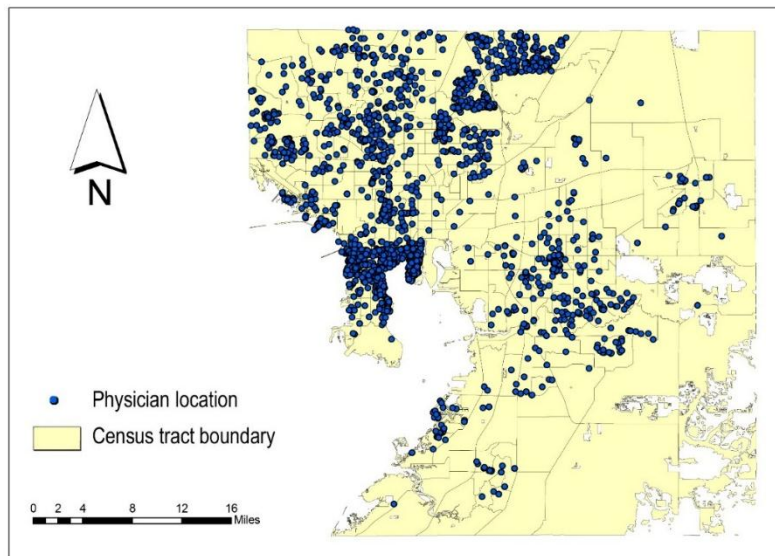


Figure 8. Physician location map (Note: uninhabitable areas such as swamp and forest are excluded from the map. Land use data obtained from Florida Natural Areas Inventory.)

The number of PCPs are aggregated in each census tract as our dependence variable for a series of regression analyses. To overcome the general lacking of the information on the supply of PCPs at the census tract or block group level, we used the physician data obtained from the Licensure Data Download of the Medical Quality Assurance Services, which consisted of a total of 3065 registered active primary care physicians in Hillsborough County in 2012. The Licensure Data Download provides the license status, mailing address, practice location address, and e-mail address for health care practitioners and establishments licensed in Florida. As preprocessing, data exploration, preparation and formatting for geocoding were performed using GIS. PCPs locations were geocoded based on their practice location addresses. After geocoding all the practice addresses of PCPs (since we did it manually, all the addresses were successfully geocoded), the numbers of PCPs are aggregated at census tract level. The number of PCPs in each census tract is considered as an indicator for the supply of primary care. Literature has confirmed that local supply

of primary care physicians affects the probability of having a primary care physician, which in turn affects preventive healthcare service utilization (Continelli *et al.* 2010). It is reasonable to assume a larger number of primary care physicians located within a census tract provides a better supply of primary care, which is advantageous for the health status of local residents.

3.2.3 Neighborhood characteristics

When studying aggregated neighborhood characteristics, the choice of spatial unit for analysis is of great importance. Despite continuous debates around the spatial or social definitions to determine the boundary and sizes of neighborhood units, a number of conceptualizations are widely accepted (De Marco and De Marco 2010). Administrative units, such as United States census tracts or block groups, are most commonly used (Wang and Luo, 2005; Wang, 2012; Pearce *et al.* 2006, Guagliardo *et al.* 2004). In this chapter, we also use census tracts as the basic unit of analysis.

For the analysis of neighborhood factors, we focus on neighborhood socio-economic and racial composition. We selected following aggregated neighborhood characteristics as explanatory variables for our analysis (see Table 2). Note that poverty rates were adopted rather than incomes as indicators for neighborhood socio-economic status, since it has been proved to be a stable variable in analyzing socioeconomic disparities (Krieger *et al.*, 2002). Data was obtained from Census 2010 data sets (U.S. Census Bureau).

Table 2. Selected census tract level aggregated neighborhood variables

Classification	Variables (variable name)	Mean	Range
Socio-economic status	Poverty (%)	17.05	0-75
	Unemployment (%)	36.10	10.9-100
Racial/ethnic composition	Non-Hispanic whites (%)	54.11	0-100
	Non-Hispanic blacks (%)	15.65	0-97.9
	Asians (%)	3.01	0-22.2
	Hispanics (%)	22.79	0-80.8
Risk age	Age 40 and up (%)	44.75	0-99.65
Urban-rural structure	Rural population (%)	1.71	0-36.04
Insurance status	Insured population (%)	78.85	0-100

3.3 Detection of Clusters of PCPs

Before analyzing the spatial patterns of clusters, the heterogeneity of the relative rates of PCPs was assessed. The concept of relative rates of PCPs is similar to that of the relative risks in epidemiology. The goal is to confirm whether there are actual differences among the different relative rates. Thus, both a Chi-square test and Potthoff-Whittinghill's test of overdispersion were conducted. The results confirmed the significant heterogeneity of relative rates of PCPs (See Appendix 1).

To detect possible spatial autocorrelation, the global Moran's I statistic was used to assess the similarity of PCP quantity among neighboring units. The question addressed here is: Do census tracts containing similar numbers of PCPs tend to be located close together, or are they randomly distributed across the study area? To assess that whether spatial autocorrelation is due to the spatial distribution of the underlying population, a Poisson constant risk parametric bootstrap assessment

(Waller and Gotway 2004) of the significant of autocorrelation of PCP counts was applied. The results confirmed the existence of significant PCP clusters, accounting for the population distribution (See Appendix 2 and 3).

The global Moran's I statistic simply answers the question that whether or not there is spatial autocorrelation in PCP quantity within the study area. To detect the location of a cluster, Besag and Newell's Statistic (1991) and Kulldorf and Nagawalla's statistic (1995) were applied as preliminary analysis to scan local PCP rates. The Besag and Newell's statistic looks for clusters of size k (*i.e.*, where the number of observed PCPs is k . In this case, $k = 100$). Then the number of neighboring regions needed to reach k cases is calculated. If the number is small, then it is marked as a cluster, since many PCPs are located in just a few regions with low expected PCP counts. On the other hand, Kulldorf's statistic identify regions within a given window, and compare the overall relative risk in the regions inside the window and that of the regions outside the window. The most likely cluster can be detected as the window with the highest value of the likelihood ratio. The preliminary scanning results can be seen in Appendix 4 and 5, both confirmed similar patterns of significant PCP clusters.

Moreover, the extent of spatial autocorrelation was determined using the local Moran's I statistic, also known as the local indicator of spatial autocorrelation (LISA) (Anselin 1995). By comparing similarities and dissimilarities among neighborhoods, LISA generates four categories of spatial clusters: high-high, low-low, low-high and high-low. A high-high PCPs cluster is one in which neighborhoods and their surrounding neighborhoods all contain high numbers of PCPs, suggesting an abundant supply of primary care providers. Conversely, a low-low cluster is one in which neighborhoods and their surrounding neighborhoods all contain low numbers of PCPs, indicating a shortage of primary care professionals. A low-high cluster is one in which neighborhoods with low numbers of PCPs are surrounded by neighborhoods with high numbers of PCPs, while a high-low cluster is one in which neighborhoods with high numbers of PCPs are surrounded by neighborhoods with low numbers of PCPs. In addition, the statistical significance of the clusters was evaluated using a Monte Carlo test, which estimates the likelihood of the clusters arising out of randomness. As a way to identify neighborhoods, we applied a queen's case spatial weight in conducting the local Moran's I test. The spatial weight counts spatial units sharing the same edges and nodes as neighbors. To make all analyses consistent, queen's case spatial weights were also applied to construct the spatial lag regression models.

3.4 Statistical methods

To determine which neighborhood factors affect the supply of PCPs, and evaluate their significance, spatial lag regression models were employed. The spatial autocorrelation in the dependence variable – the number of PCPs in each census tract – is accounted for using spatial lag regression models. The spatial lag regression model is commonly known as a diffusion model, where the value of dependence variable y is related to the values of y in neighboring locations through ρ , the spatial autoregressive parameter. For better performance of regression models, we did a log transformation on the dependent variable, *i.e.*, the number of PCPs in spatial unit i ($PCPs_i$)

$$y_i = \log(PCPs_i + 0.001) \tag{1}$$

The spatial autoregressive structure is incorporated in the spatial lag model:

$$y = \rho W y + X \beta + \varepsilon \quad (2)$$

where ρ is the spatial autoregressive parameter. W is a spatial weights matrix. X is the matrix of exogenous explanatory variables with an associated vector of regression coefficients β . ε is the vector of random error terms. By using spatial lag regression models, we were able to quantify the relationship between the explanatory variables and the number of PCPs in each census tract, while accounting for the spatial autocorrelation in the dependent variable.

3.5 Results and discussions

The value of Global Moran's I is 0.2192, with a corresponding z-score of 7.4669 ($p < 0.001$), indicating that there is significant positive spatial autocorrelation in the census tract level PCP distribution. The clustering pattern remains significant even when controlling for the heterogeneity of population distribution ($p = 0.002$). Following LISA cluster map (Figure 9) identified four categories of spatial clusters: high-high, low-low, low-high and high-low.

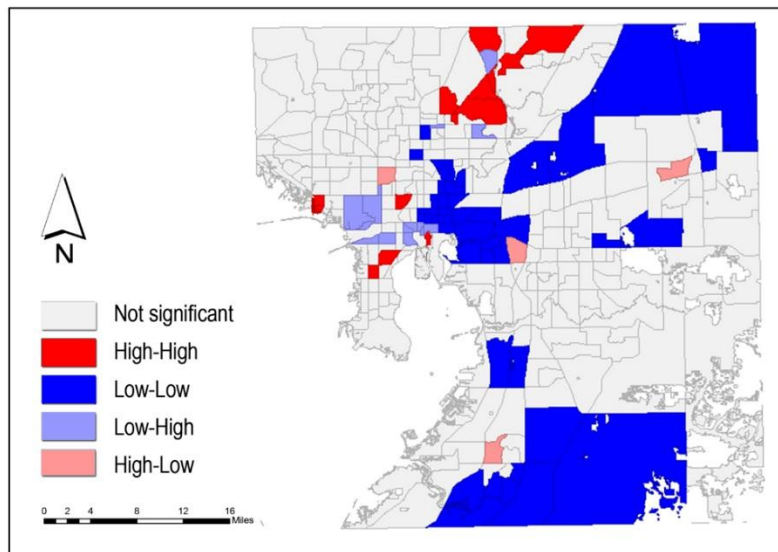


Figure 9. LISA Cluster Map

According to the selected results of spatial lag regressions listed in Table 3, the percentage of Asians has the most significant positive effect over the number of PCPs in each census tract, even when other socioeconomic variables are controlled. For example, in Model1, when the percentage of Asians increases by 10%, the number of PCPs increases by approximately 100 ($b=0.2274$, $p<0.001$). Compared to Asians, the percentage of whites in each census tract has a weaker effect over the number of PCPs. In Model1, when the percentage of whites increases by 10%, the number of PCPs increases by approximately 3.16 ($b=0.0544$, $p<0.001$). Surprisingly, both blacks and Hispanics have no significant impact over the numbers of PCPs, despite the relatively large population of Hispanics in the study area.

In summary, the racial composition of both Asians and whites is a main factor that influences the distribution of PCPs at neighborhood level. In other words, when a census tract includes more Asians or whites, there is a higher quantity of PCPs. The impact of Asians is

approximately four times more than whites while blacks and Hispanics do not seem to affect the supply of PCPs. It is no surprising that indicators of neighborhood socio-economic disadvantage such as poverty rate and unemployed rate, all have significant negative impacts over the number of PCPs, while the rate of insurance coverage has a significant positive impact over the number of PCPs. Moreover, the percentage of the age group at risk (> 40 years old) has a significant positive influence over the quantity of PCPs. However, the percentage of rural population doesn't seem to have statistically significant impact over the quantity of PCPs.

Table 3. Coefficients for spatial lag regression models

	Model1	Model2	Model3	Model4	Model5
% White	0.0544***	0.0309***		0.0272**	
% Black	0.0142				
% Asian	0.2274***	0.1986***	0.1693***	0.1515**	0.2551***
% Hispanic	0.0148				
Poverty Rate		-0.0405*	-0.0539**	-0.043**	
Insurance Rate			0.0664***		
Unemployment Rate				-0.0350**	
% Rural Population					-0.0478
% Age 40+					0.0378**
r2	0.29	0.30	0.35	0.32	0.25
AIC	1689	1682	1661	1676	1727

Note: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4. Correlation matrix of explanatory variables

	% White	% Black	% Asian	% Hispanic	% Poverty	% Insurance	% Rural	% Unemployed	% Age40+
% White		-0.64***	0.07	-0.48***	-0.54***	0.58***	0.17***	-0.17**	0.58***
% Black			-0.17**	-0.04	0.61***	-0.06	-0.14*	0.04	-0.24***
% Asian				-0.05	-0.20***	0.12*	-0.02	-0.34***	-0.16**
% Hispanic					0.34**	-0.18**	-0.05	-0.23***	-0.16**
% Poverty						-0.23***	-0.06	0.12*	-0.20***
% Insurance							0.05	-0.42***	0.53***
% Rural								-0.05	0.07
% Unemployment									0.12*
% Age40+									

Note: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4 provides a correlation matrix of the explanatory variables. Census tracts with a higher percentage of white or Asian population are found to have lower poverty rate, higher insurance rate, and lower unemployment rate, which supports the conclusion that the composition of Asians or whites has a positive impact over the quantity of PCPs. Census tracts with a higher percentage of white population are more likely to be rural areas, while census tracts with a higher proportion of Asians tend to be urban areas. This result is also consistent with the conclusion that the positive impact of Asians on the quantity of PCPs is much more significant than whites, since urban areas tend to have more hospitals and physicians in Hillsborough County, Florida (Ye and Kim 2014). Note that census tracts with a higher proportion of whites also have more 40 years and older population; while census tracts with a higher proportion of minorities have a younger

population. Poverty rate, unemployment rate, and insured population percentage are all correlated in the expected direction. It is clear that 40 years and older population turn out to have higher socioeconomic status in general, and they are most likely to be white. This result supports the assumption that the percentage of the age group at risk (> 40 years old) has a significant positive influence over the quantity of PCPs.

3.6 Conclusions

Despite the abundant research that focused on individual-level or neighborhood-level factors that contribute to individual health status (Kawachi, 2000, Kawachi and Glass 2000, Patel *et al.* 2004, Kawachi *et al.* 1999; Buka *et al.* 2003), very little has been done to study the relationship between neighborhood factors and the supply of health care services (Lo and Wang 2005; Curtis and Leitner 2006; Wang 2006; Wang 2012).

This chapter examines the possible correlation between neighborhood factors and the supply of PCPs with a case study of Hillsborough County, Florida. Spatial statistical techniques are applied to estimate the effect of neighborhood socio-economic/racial composition on the PCPs supply, which reflects the dimension of availability in health care access. First, the spatial patterns of PCPs distributions were examined and PCP clusters were detected. Second, spatial lag models were applied to assess significant neighborhood factors that might influence PCPs distribution. The spatial lag models were used to account for the spatial dependency in PCPs locations. In our study, other than blacks and whites, two ethnic groups that were generally ignored in previous studies— Hispanics and Asians, were included in the analysis.

The global Moran's I confirms the spatial autocorrelation between PCPs distribution; while local Moran's I identifies both hot and cold spots. Metropolitan areas such as Tampa and Temple Terrace are considered as hot spots – clusters of high numbers of PCPs; while certain suburban and rural areas are identified as cold spots. The results reveal the potential PCPs shortage in rural areas in Hillsborough County, Florida.

Important findings also emerged from the regression analysis. Asians are found to have a most significant positive effect on the supply of PCPs in Hillsborough County. Although one might suspect Hispanics might influence the supply of PCPs since there is a large population of Hispanics in the study area, the results of the regression analysis proved otherwise. The percentage of whites also has a positive effect on the number of PCPs, while the percentage of blacks has no correlation with the PCPs supply. It is no surprising that PCPs supply is negatively correlated with indicators of neighborhood socio-economic disadvantage. Moreover, insurance rate definitely has a great positive impact over the supply of PCPs.

This study has certain limitations. First, the results might be different if we use different sizes of neighborhoods due to modified areal unit problem (MAUP) (Openshaw, 1983). The MAUP describes the differences in empirical results that may occur when we use different spatial units for analysis. Although we have significant findings for the census tract level analysis, the result might be different for other geographical scales. Second, although the supply of PCPs reflects the availability of health care services in the study area, it does not guarantee the accessibility from the patients' side. Thus a higher supply of PCPs might not be necessarily equal to good health access.

Despite the limitations discussed above, this research contributes to knowledge on the supply of health care services in the study area by identifying significant neighborhood factors. Our findings suggest that neighborhood racial/socio-economic composition affects the supply of

PCPs. Neighborhoods that are more Asian or whites tend to have more PCPs, while neighborhood socio-economic disadvantages impede the supply of PCPs. Given that the United States is highly segregated by both race and socio-economic characteristics, this research contributes to the health care access literature by addressing the neighborhood-level characteristics and the availability of health care services.

CHAPTER 4
NEIGHBORHOOD CHARACTERISTICS
AND
HEALTHCARE UTILIZATION

A version of this chapter is published in *Social Work Research* online on March 28, 2016 (doi: 10.1093/swr/svw004). The use of “we” in this chapter refers to co-author Dr. Sungkyu Lee, Dr. Hyun Kim and myself. As the first author, I processed the data, performed the analysis and wrote the manuscript.

Abstract

Using a nationally representative U.S. sample, this study examined the extent to which neighborhood characteristics affected length of inpatient stay (LOS) in the United States. Data were obtained from the 2012 Area Health Resource Files. A total of 3,148 U.S. counties were included in the study. Generalized linear models and the geographically weighted regression model were employed to examine the extent to which neighborhood characteristics affected length of inpatient stay and its spatial variation. Exploratory spatial data analysis was also conducted to examine the geographic patterns in LOS. Hospital bed capacity was found to be the strongest predictor of LOS. Counties with a lower poverty rate, a lower uninsured rate, a higher proportion of female residents, a higher proportion of residents living in urban areas, and more diverse racial groups had a longer LOS. Significant spatial clustering pattern of LOS was also found. Findings suggest that social work professionals should be aware of spatial disparity in health care resources and develop ways of providing equitable health care for vulnerable populations in socioeconomically disadvantaged neighborhoods.

Key words: health care use; health care resources; length of inpatient stay; neighborhood characteristics; spatial disparity

4.1 Introduction

According to the U.S. national data, the average length of inpatient stay (LOS) declined from 5.7 to 4.6 days between 1993 and 2005 (HCUP 2005). Shortening length of hospital stay is one of the main strategies applied by managers and hospital administrators to cope with the increasing financial pressures, since hospitalization is the most expensive form of healthcare (Gandsas et al. 2007). The United States is facing an increasing healthcare demand and a great shortage of physician labor force (AAMC 2012). In this context, despite the ongoing debate on the relationship between LOS and healthcare expenditures, many healthcare professionals believe that reducing LOS frees up capacity to increase admissions, increase revenue, and improve healthcare quality (Herrle 2006).

Previous studies on LOS have mainly focused on individual and hospital characteristics. Martin and Smith (1996) identified several important determinants of variations in LOS, such as access to hospitals, waiting time for elective surgery, poverty status, and availability of informal care. Epstein et al. (1988) found that patients with lower socioeconomic status had longer hospital stays when the variables of age, sex, and the severity of illness were adjusted. However, another study by Ellison and Bauchner (2007) argued that socioeconomic status had no effect on LOS among children with vaso-occlusive crises in sickle cell disease. Remarkably, Gifford and Foster (2008) postulated that the length of stay is better explained by hospital characteristics than individual characteristic, and several studies also supported that hospital type is greatly associated with variations in LOS (Lee, Rothbard and Noll 2012b, Cohen and Casimir 1989, Burns and Wholey 1991).

Considering disparities in hospitalization from a geographic perspective, Ashton and colleagues (1999) confirmed that significant geographic variation existed in LOS among veterans.

Nguyen-Huynh and Johnston (2005) also found regional variation in LOS among Asians and Pacific Islanders with stroke. However, notice that these studies are limited by the use of a specific health condition (*e.g.*, stroke) and a specific hospital type (*e.g.*, VA hospital), resulting in a lack of evidence to draw a general conclusion. Thus, using a nationally representative U.S. sample, our study investigates the relationship between neighborhood characteristics and LOS. By adopting geographical perspective, our study particularly contributes to the literature on hospital utilization with an emphasis on the development and allocation of healthcare resources within different neighborhoods.

4.2 Theoretical framework

In light of ecological perspective, our study specifically focused on the effects of neighborhood characteristics on LOS. The core assumption of the perspective is that humans are dependent on their environments. To reveal the dependency, the ecological theory emphasizes on the interaction between individuals and their environments, such as family, work, and community as well as cultural and political environments at large (Chung 2012).

Previous research has shown the importance of ecological perspective in understanding the effect of environment on health. For example, a recent study by Chung (2012) highlighted the reciprocal interactions of biological, psychological, social and cultural variables on suicide attempts among Chinese immigrants in New York City. By applying the ecological system approach, Sanders et al. (2008) also identified barriers to mental health services for older adults in rural areas where resource is relatively limited (*e.g.*, lack of knowledgeable healthcare providers, funding cutbacks, and limited access to services).

The ecological perspective considers factors in both physical and social environment (Kwag *et al.* 2011). In particular, physical environment factors, recognized as geographic characteristics, such as the availability and proximity of health facilities, have drawn increasing attention. For example, Andersen et al. (2002) found a positive correlation between the number of federally qualified health centers available and the likelihood of having a usual source of care. Arcury et al. (2005) also found that a shorter distance between patients and physicians increased the frequency of regular family physical exams. Buchmueller et al. (2006) suggested that increasing distances from hospitals resulted in higher death rates from heart attacks and unintentional injuries.

Social characteristics of neighborhoods (*e.g.*, socioeconomic status, proportion of racial/ethnic groups) have also been proven to be significant predictors of an individual's health status. For example, Black and Macinko (2008) demonstrated that indicators of neighborhood socio-economic composition have a significant impact on the risk of obesity. Another study by Dai (2010) found that a higher risk of late-stage diagnosis of breast cancer was greatly associated with living in areas with greater black segregation. It has been documented that socio-economically disadvantaged neighborhoods tend to have relatively poor access to healthcare resources (Cummins *et al.* 2007). A recent study by Kwag et al. (2011) found that neighborhood characteristics, such as proportion of individuals living below poverty, proportion of individuals 65 years of age and older, and proportion of racial/ethnic minorities in the neighborhood, significantly affected physical and mental health of Korean American older adults. Due to the residential segregation in the U.S., lack of access to health-promoting resources is also associated with racial disparity in health outcomes (Mennis, Stahler and Baron 2012).

Despite the abundance of research on the influence of neighborhood characteristics on the individual's health or access to healthcare, little is known about how neighborhood characteristics affect LOS in the U.S. Thus, our study addresses the following research question: To what extent do neighborhood characteristics influence LOS in the U.S.? To answer this question, we explored (1) the geographic pattern of LOS using global Moran's I and LISA statistics; (2) significant neighborhood characteristics associated with LOS using generalized linear models (GLMs); and (3) possible spatial variations at county level using a geographically weighted regression model (GWR).

4.3 Methods

4.3.1 Data source and sample

Data were obtained from the 2012 Area Health Resource Files (AHRF), which comprise of data collected from more than 50 sources, including the American Hospital Association (AHA) annual survey of hospitals. The data set contains more than 6,000 variables associated with healthcare access and utilization at the county level (U.S. Department of Health and Human Services 2014). The sample for the current study consisted of 3148 counties across the U.S. Given that the AHRF provides county-level data only, no individual-level data were used in the current study. Based on the use of aggregated secondary data, which were also obtained from publicly available source, the University IRB has exempted the study from review.

4.3.2 Study variables

Dependent variable

All hospital utilization data in the AHRF were extracted from the AHA annual survey of hospitals. Since the county is the unit of analysis in the current study, an aggregated LOS per county was used as the dependent variable in the study. According to the AHA survey instructions (AHA 2015), LOS refers to the number of adult and pediatric days of care rendered during the entire reporting period. Those days included both medical and psychiatric short-term and long-term inpatient hospitalization regardless of the type of hospital (*e.g.*, general, non-general, community, veteran's hospital), with an exception of newborns.

Explanatory variables

Explanatory variables were based on the Andersen's behavioral model of health services use (Andersen 1968, Aday and Andersen 1974, Andersen 1995), which has been broadly used in previous studies on healthcare use (Babitsch, Gohl and von Lengerke 2012). The Andersen model specifies the role of predisposing (*e.g.*, age, gender, race), enabling (*e.g.*, income, poverty, employment, insurance status), and need factors (*e.g.*, perceived health status, medical diagnosis) in examining access and use of healthcare services. Although these variables have been used extensively to explain healthcare use, neighborhood characteristics (*e.g.*, population factors, healthcare resources in the community) may also play an important role in determining how long an individual stays in an inpatient setting (Lee, Rothbard and Noll 2012a).

Thus, two sets of explanatory variables were constructed to represent neighborhood characteristics under two categories: aggregated socio-demographic characteristics and healthcare resources. Aggregated socio-demographic characteristics included a set of eight variables: the proportion of residents 65 years of age and older, the proportion of female, the proportion of white,

the proportion of residents living in urban areas, the proportion of residents living below the poverty level, the proportion of residents who were unemployed, the proportion of residents who had no insurance, and the total number of population. In the AHRF data, urban was defined as where all territory, population, and housing units located within urbanized areas, which consist of densely developed territory that contains 50,000 or more people, and urban clusters, which consist of densely settled territory that has at least 2,500 people but fewer than 50,000 people (U.S. Department of Health and Human Services 2013). Healthcare resource variables included the number of hospitals and inpatient service unit beds. In the AHRF data, facilities with six or more inpatient beds, cribs or pediatric bassinets were considered as hospitals.

4.3.3 Data analysis

Exploratory spatial data analysis

Global Moran's I and LISA statistics were applied to explore the geographic pattern of LOS by assessing the similarity of LOS among neighboring counties. Specifically, the global Moran's I statistics were employed to examine whether or not there is spatial autocorrelation in LOS within the study area, while the extent of spatial autocorrelation was determined using the local Moran's I statistic, which is known as the local indicator of spatial autocorrelation (LISA, Anselin 1995). By comparing similarities and differences among counties, LISA generates four categories of spatial clusters: high-high, low-low, low-high, and high-low. In the context of our study, a high-high LOS cluster is one in which counties and their surrounding counties all have high values of LOS. Conversely, a low-low cluster is one in which counties and their surrounding counties all contain low values of LOS. A low-high cluster is one in which counties with low values of LOS are surrounded by counties with high values of LOS, while a high-low cluster is one in which counties with high values of LOS are surrounded by counties with low values of LOS. Statistical significance of the clusters was evaluated by a Monte Carlo test, which estimates the likelihood of the clusters arising out of randomness (Anselin 1995, Hope 1968). To identify neighborhoods, we applied a Queen's case spatial weight (Getis and Aldstadt 2004, Stetzer 1982). The spatial weight counts spatial units sharing the same edges and nodes as neighbors.

Statistical analysis

To examine the extent to which neighborhood characteristics affected LOS, we conducted generalized linear models (GLMs) with a log link, which account for positive skewness in LOS (Manninga and Mullahy 2001). GLMs generate global coefficients and assume that the relationships are constant across the study area. To capture the possible spatial variation in the relationship between LOS and covariates among counties, we also applied a geographical weighted regression (GWR). GWR represents detailed local variations, as the fitted coefficient values of a global model (*e.g.*, GLM) fail to do so (Fotheringham, Brunson and Charlton 2003, Brunson, Fotheringham and Charlton 1996),

The formula for GWR can be written as:

$$y_i = \beta_{i0}(u_i, v_i) + \sum_{k=1}^p \beta_{ik}(u_i, v_i)x_{ik} + \varepsilon_i, \quad \forall i$$

where $\beta_{i0}(u_i, v_i)$ is the intercept parameter at spatial unit (*i.e.* county) i , $\beta_{ik}(u_i, v_i)$ is the local regression coefficient for the k th independent variable at i , and (u_i, v_i) is the coordinate of the i th point in the study area (Fotheringham et al. 2003).

4.4 Results

Table 5 shows the aggregated neighborhood characteristics of 3,148 counties in the U.S. The average LOS was 75185.4 days, however, the standard deviation was 6078.9, indicating the large variation in LOS among counties. The mean proportion of residents who lived below the poverty level was 17.2%, while the mean proportion of white was approximately 82.9%. The mean proportion of older people (65 and older) was 16.2% and the mean proportion of the uninsured was 18.5%. The average number of hospital was 1.91 (SD=4.14) and the mean number of hospital bed was 285.4 (SD= 1,002.73).

Table 5. Aggregated neighborhood characteristics

	Variables	%	Mean	SD	Range
Dependence variable	Length of inpatient stay (day)		75185.4	206078.9	0-6502662
Socio-demographic characteristics	Total population (#)		92610	316349	90- 3880244
	Population density per square miles		234.2	1724	0-35369.2
	White (%)	82.89			2.70-99.20
	Female (%)	50.01			28.73-56.84
	Age 65 and older (%)	16.16			3.72-45.54
	Urban (%)	41.28			0.00-100.00
	Uninsured (%)	18.54			3.60-41.40
	Poverty (%)	17.24			2.90-49.90
Healthcare resources variables	Total hospitals (#)		1.91	4.14	0-48.00
	Hospital beds (#)		285.4	1002.735	0-9804.0

Figure 10 shows the geographic variation in LOS in the U.S. For clear comparison, our study used nine regions according to the U.S. Census Regions. The LOS was the highest among the West Pacific, South Atlantic, and Northeast coastal regions, especially within or around California, New York, and Florida (*e.g.*, Los Angeles, Cook, New York, Harris, and Maricopa County).

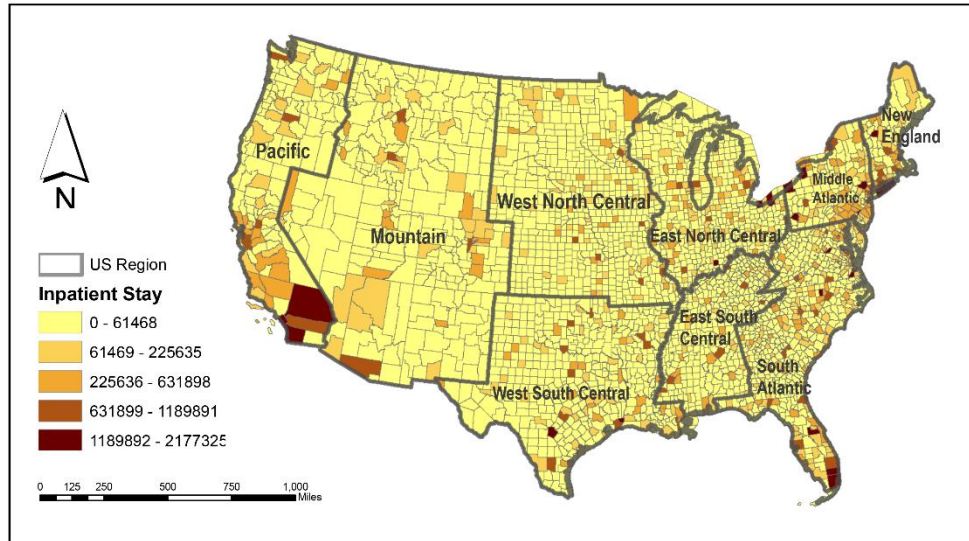


Figure 10. Length of inpatient stay map in the United States

The value of Global Moran's I was 0.1685 with a corresponding z-score of 16.6931 ($p < .001$), indicating that there is a positive spatial autocorrelation in the LOS at county level. In Figure 11, a LISA cluster map shows four categories of spatial clusters: high-high, low-low, low-high and high-low. No significant low-low clusters were identified at county level. High-high clusters were mainly located in the West Pacific, Northeastern, and South Atlantic regions. Very few high-high clusters were scattered in the West Mountain, West South Central, and East North Central, and East South Central regions. No high-high clusters were identified in the West North Central region. Noticeably, high-low clusters were mainly scattered in the Midwest and South, while very few low-high clusters were identified near the high-high clusters.

To further examine the factors that contribute to this spatial pattern of LOS, two GLMs were constructed (Table 6). First, to examine the gross effect of bed capacity – an indicator of neighborhood healthcare resources on the LOS, Model 1 included only the number of beds without controlling for covariates. The result indicates that a greater number of bed capacity increased LOS ($b=0.002, p < .001$). Second, Model 2 included both the bed capacity and other neighborhood socio-economic characteristics. After controlling for covariates, the variable of bed capacity was remained as a significant predictor of LOS ($b=0.001, p < .001$), implying that hospital capacity is a significant determinant for LOS. All other covariates were found to be associated with LOS. For example, LOS was longer in counties with a lower poverty rate, a lower uninsured rate, and a higher proportion of female. LOS was also longer in counties with a higher proportion of residents living in urban areas. Moreover, counties with more diverse racial groups and a younger population had a longer LOS.

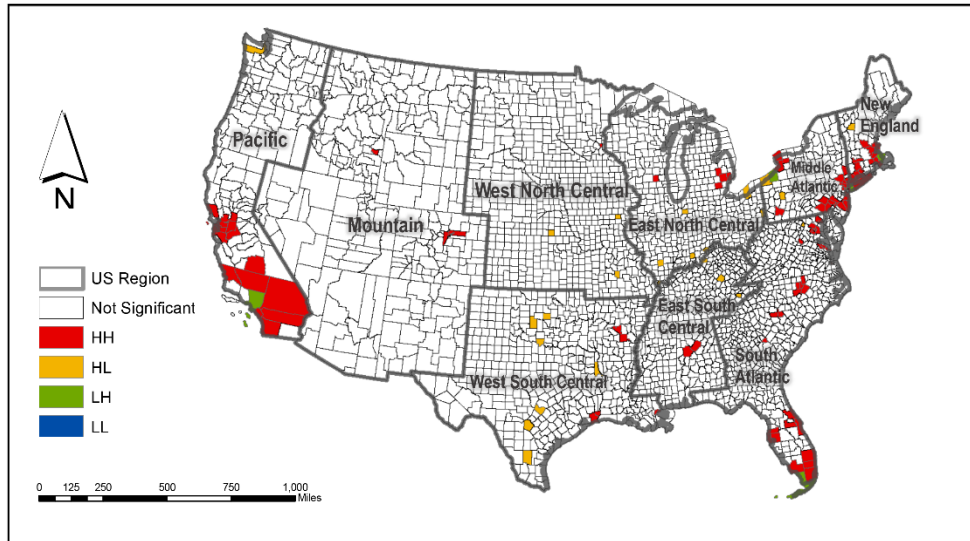


Figure 11. LISA cluster map

Table 6. Predictors of the length of inpatient stay: Results of GLMs

Variable	Model 1	Model 2
	<i>b</i>	<i>b</i>
Hospital beds	0.002***	0.001***
Poverty (%)		0.016**
Total population		-0.002**
Population density per square miles		-5.970E-7**
White (%)		-0.003*
Female (%)		5.969***
Age 65 and up (%)		-1.731**
Urban (%)		0.020***
Uninsured (%)		-0.061***
AIC	57574.860	56777.399

Note. A variable of number of hospital was excluded from Model 2 due to its high correlation with hospital bed capacity ($b=0.95, p<.001$). *** $p<.001$.

Since the bed capacity is proven to be the major determinant of LOS, we constructed a GWR model using LOS as dependent variable, and number of hospitals beds as explanatory variable. The GWR model has an adjusted r^2 of 0.99, suggesting that the model is a good fit to capture the spatial variation of LOS. The mean of hospital beds coefficients was 249.925 with a standard deviation of 9.020 (ranged from 240.200 to 274.410). In GWR model, we confirmed significant spatial variation in the relationship between LOS and hospital bed capacity by establishing more accurate local coefficients.

Figure 12 represents the visualization of the spatial variation in the relationship between LOS and hospital bed capacity. The Northeast region has the strongest positive correlation between the LOS and bed capacity ($262.6 < b < 274.4$); while the Midwest and South regions have the smallest positive correlation ($240.2 < b < 243.5$). In other words, the effect of bed capacity on LOS was greater in the Northeast region than that in the Midwest and South regions. One possible explanation behind the variation in the bed coefficients could be the different utilization patterns among populations. The population density of the northeastern states is much higher than that of the central states. The higher socio-economic status of the population as well as the more developed transportation networks contribute to better access to hospitals. Thus, the utilization rate of the healthcare facilities in the northeastern states is much higher, and the hospital beds are constantly occupied. The turnover rate is also higher. On the other hand, there might be more empty hospital beds during certain periods. As a result, when increasing the number of beds in the northeastern states, there is a greater increase of the overall length of inpatient stay.

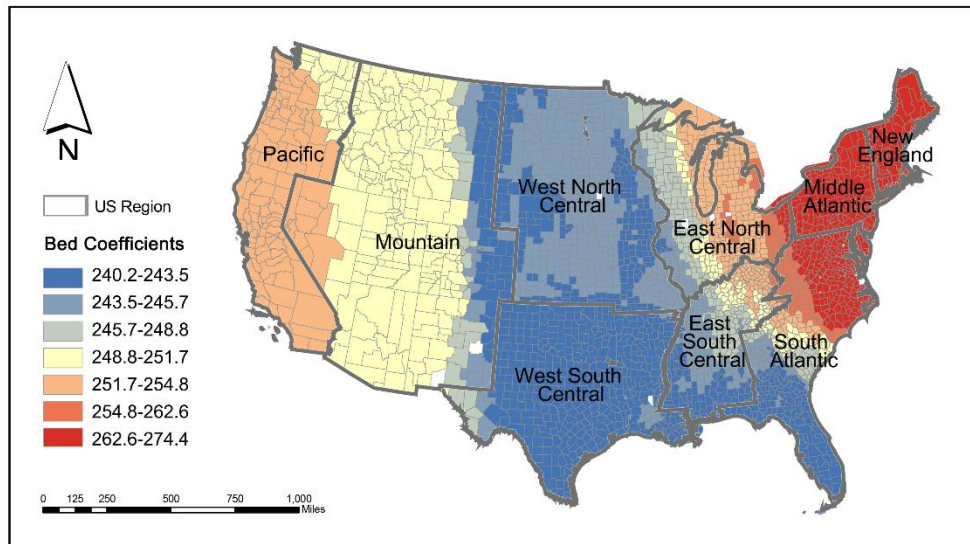


Figure 12. GWR coefficient map

4.5 Discussion

Our study confirmed the spatial clustering pattern of LOS and identified its associated neighborhood factors. Hospital bed capacity as an indicator of health care resources was found to be the major predictor of LOS. Particularly, our findings suggest that people living in urban areas tend to have better access to health care resources than their counterparts living in rural areas. Consistent with the findings of previous studies (Kwag *et al.* 2011; Martin and Smith 1996), our study also highlights the importance of neighborhood's economic status (that is, the relationship

between poverty and LOS), implying that residents in socioeconomically disadvantaged neighborhoods are likely to experience lack of health care resources, which in turn limits health care access (Macintyre, Ellaway and Cummins 2002). The spatial distributions of health care resources and population in need do not match in a desirable level across the country (Guagliardo 2004, Ye and Kim 2014), and the shortage of health care supply is especially severe in rural areas and impoverished urban communities (Campbell *et al.* 2000; Monnet *et al.* 2008; Ye and Kim 2014). Consequently, living in disadvantaged neighborhoods reduces the possibility of having a usual source of care and receiving recommended preventive services (Kirby and Kaneda 2005). Noticeably, our data show that 615 counties out of 3,184 do not have any facilities with six or more inpatient beds. Considering the importance of developing community health care resources for both inpatients and outpatients, health care policymakers should be aware of this spatial disparity and develop ways of providing equitable health care for vulnerable populations in socioeconomically disadvantaged neighborhoods.

Furthermore, given that many health care providers tend to refuse Medicaid patients due to its lower reimbursement rate than that of private insurance (Dayaratna 2012), health care policymakers should consider reforming Medicaid so that the poor have equal access to needed and quality health services. Likewise, those without insurance may be limited to have shorter length of hospital stay due to their financial burden from out-of-pocket health expenditures. Thus, the provision of affordable health insurance would be helpful for those who need to stay longer at the hospital for the necessary health services. The passage of the Patient Protection and Affordable Care Act (ACA) of 2010 (P.L. 111-148) in the United States has increased health care access for vulnerable populations, such as individuals with mental illness and substance use disorders (Donohue, Garfield and Lave, 2010), women (Johnson 2010), and low-income families (Decker *et al.* 2013). However, the ACA still does not address lack of insurance for recently arrived documented immigrants (less than five years of residence in the United States) and undocumented immigrants (González Block *et al.* 2014). Considering that delayed treatment may require costly care later in the disease process, policymakers should consider providing limited and selected health care services, instead of excluding all benefits of the ACA coverage, for this vulnerable population. Alternatively, developing community resources, which can be substituted for hospitalization, would help this vulnerable population. For example, medical homes would be a good option, especially for the neighborhoods where there is a great shortage of health care resources, such as primary physicians. The Commonwealth Fund 2006 Health Care Quality Survey indicated that when people have a medical home, their access to needed care, receipt of routine preventive screening, and management of chronic conditions improves greatly (Beal and Fund 2007). Thus, expanding medical homes for patients, particularly those who are living in the neighborhoods with a lack of primary health care resources, may improve overall health care accessibility. Considering that community health centers and other public clinics are less likely to provide medical homes for the uninsured (*ibid.*), policymakers should consider expanding medical homes for this vulnerable populations.

The proportion of white population turned out to have a negative effect on LOS. In other words, racial minority populations tend to have longer stays than their white counterparts. A study by Thompson, Neighbors, Munday, and Trierweiler (2003) found that white patients are more likely to receive a referral to aftercare, which decreases the risk of readmission. Other studies also indicated that racial minority groups are more likely to have less access to quality aftercare, suffer from long-term functional outcomes after traumatic injury, and have a higher chance of readmission and longer inpatient stays (Ball and Elixhauser 1996, Bolden and Wicks 2005, Kwag

et al. 2011; Shafi *et al.* 2007; Thompson *et al.* 2003). Considering the potential risks of longer stays without proper referrals among racial or ethnic minority groups, health care professionals should develop a comprehensive and systematic referral system in the context of culturally competent health care to reduce racial disparity in health care access and service utilization.

Several limitations should be noted. First, individual characteristics, such as severity of health condition, were not considered in the study, because the literature supports little effect of individual characteristics on LOS (Gifford and Foster 2008). Further studies are recommended to employ a hierarchical methodology, which accounts for individual, facility, and neighborhood characteristics. Second, the current study used the county as the unit of analysis in the examination of the relationship between neighborhood characteristic and length of stay. Since different boundaries or sizes of neighborhoods create modified areal unit problem (Openshaw 1983), the study is limited in its generalizability. Third, this study is a cross-sectional study. Thus, it is limited to known temporal order of study variables. For instance, we can't know whether the greater bed capacity allows patients stay longer or whether longer lengths of stay need a greater bed capacity. Therefore, to investigate this causality, further studies are recommended to replicate this study with a longitudinal design. Finally, our study findings are limited in terms of the lack of additional explanatory variables. Particularly, due to data unavailability, our current study could not control for specific diseases or illnesses (need factors proposed by the Andersen model) as covariates. To better understand comprehensive factors associated with LOS, further studies should also examine the role of need factors in health care utilization.

Nevertheless, as the first empirical study to examine the effects of neighborhood characteristics on length of stay using a nationally representative U.S. sample, our findings contribute to the literature by exploring spatial clustering pattern of LOS, identifying significant neighborhood characteristics associated with LOS, and providing implications for health care policy and practice. Social work professionals should be aware of spatial disparity in health care resources and understand the effects of neighborhood characteristics on LOS to provide continued quality health care in collaboration with community partners.

CHAPTER 5

LOCATING HEALTHCARE FACILITIES:

A NETWORK-BASED OPTIMIZATION APPROACH

A version of this chapter was revised and re-submitted to *GeoJournal* on February 20, 2016 and is currently under the second round of review. The use of “we” in this chapter refers to co-author Dr. Hyun Kim and myself. As the first author, I processed the data, performed the analysis and wrote the manuscript.

Abstract

Where should new service facilities be located is a key question in ensuring healthcare accessibility. In previous healthcare literature, most researchers applied either Maximal Covering Location Problem (MCLP) or Location Set Covering Problem (LSCP) to address the location selection problems. The MCLP tries to maximize population with access with a limited number of facilities; while the LSCP ensures full coverage with a minimum number of facilities. However, researchers rarely applied both models and compare their results. Moreover, most literature applied Euclidean distance to generate service coverages and failed to demonstrate the geo-processing steps in the application of location problems with the integration of Geographic Information System (GIS). To complement existing literature, this chapter proposes a *Network-based Covering Location Problem (Net-CLP)* building on traditional location problems. The Net-CLP incorporates two sub-models: a *Network-based Maximal Covering Location Problem (Net-MCLP)* and a *Network-based Location Set Covering Problem (Net-LSCP)*. The goal of Net-CLP in this chapter is threefold: 1) the network-based coverage is based on real world transportation networks depending on different travel thresholds; 2) addressing the location problem applying both Net-MCLP and Net-LSCP to fully evaluate candidate facility sites, considering service capabilities; 3) demonstrating the integration of GIS in location problems, with a case study of Hillsborough County, Florida.

Key words: location problems, service capability, transportation network, GIS integration, geo-processing

5.1 Introduction

The geographical proximity of healthcare facilities is crucial in ensuring spatial accessibility for population, as well as individual health outcomes (Arcury *et al.* 2005, Buchmueller, Jacobson and Wold 2006, Ferrante *et al.* 2000, Hare and Barcus 2007). Accessibility is defined as the potential interaction between the population in need and healthcare services, impeded by geographical barriers such as travel distance and cost (Guagliardo 2004). Generally speaking, the probability of visiting healthcare facilities decreases with increased travel distance, which is referred to as distance decay (Cromley and McLafferty 2012). It impedes individuals from more frequent visits to healthcare facilities for regular health checkups and timely treatments in case of emergencies. To ensure accessibility, healthcare planning should consider both the quantity of the facilities allocated and their geographical locations, when assessing service coverage in a region. If too many facilities are allocated, the operation and maintenance cost becomes excessive; if not enough facilities are allocated, service quality might not meet a satisfactory standard. Moreover, if the facilities are allocated too far away from the population in need, the increased travel distance impedes access to healthcare. Both low quantity and undesirable locations of facilities result in the underuse of the healthcare services, leading to increases in mortality and morbidity rates (Buchmueller *et al.* 2006, Burns and Wholey 1991, Daskin and Dean 2004). In summary, assessing the potential mismatch between facilities and population, as well as finding an optimal solution

when locating healthcare facilities play a significant role of ensuring access to healthcare and individual health needs.

Since the locations of healthcare facilities are essential for healthcare delivery, this research addresses the question of where to locate *new* healthcare facilities in relation to existing ones under following scenarios:

1. Minimizing the total number of new facilities to reduce cost while ensuring all the demands are covered;
2. Maximizing service coverage when locating a fixed number of *new* healthcare facilities;
3. Exploring potential locations satisfying both perspectives described in 1 and 2;
4. Considering service capability of facilities.

To address this question, this chapter employs a spatial optimization approach with an integration of Geographic Information System (GIS) to assess healthcare accessibility and support location decisions. Here we propose a *Network-based Covering Location Problem (Net-CLP)*. In order to formulate a more comprehensive evaluation of location decisions, the Net-CLP contains two sub-models building on traditional coverage location problems – a *Network-based Maximal Covering Location Problem (Net-MCLP)* and a *Network-based Location Set Covering Problem (Net-LSCP)*. Since the concepts of both spatial representation and coverage are essential to identifying facility locations in the models, the work flow of this study is summarized as: 1) design a spatial representation of the study area to reduce aggregation errors; 2) define service coverage based on a transportation network in a GIS environment; and 3) demonstrate the application of a feasible planning method – a Net-CLP that enables public health professionals to evaluate location decisions. The chapter is organized as follows: a review on healthcare facility location problems and related issues is presented in section 2, followed by a detailed description of the formulation for Net-CLP with two sub-models – Net-MCLP and Net-LSCP, in section 3. Section 4 presents the spatial representation that we developed for the model application in the case study area – Hillsborough County, Florida. Results of the analysis are presented and discussed in section 5, followed by conclusive remarks.

5.2 Background

The location problems are formulated and solved through mathematical programming consisting of objective functions and a set of constraints that are structured based on identified spatial characteristics. Other than being widely applied in the efficient siting of service-oriented facilities, location problems have gained increasing popularity in healthcare literature (Rahman and Smith 2000, Syam and Côté 2010, Shariff, Moin and Omar 2012). One of the most popular location models is the Maximal Covering Location Problem (MCLP). Introduced by Church and ReVelle (1974), the MCLP was originally employed to support location decisions for emergency service facilities, such as ambulance, police and fire stations (Daskin and Dean 2004). The objective of the MCLP is to maximize the number of covered demands served by a limited number of facilities. Limiting the number of facilities to be located is a constraint to be imposed on the supply side. In the context of healthcare accessibility, this objective can be interpreted as the maximization of overall accessibility for the population in need when locating services. Moreover, it has been extended and applied in more diverse contexts over the years. For example, Verter and Lapierre (2002) applied an adapted version of MCLP to locate preventive health care facilities, ensuring

both total coverage and the limited capacity of each facility; Moore and Revelle (1982) modified and extended the model of MCLP to consider hierarchies in service system. Another commonly applied covering model is the location set covering problem (LSCP) proposed by Toregas et al. (1971), which attempts to provide complete coverage to all areas while utilizing the fewest number of service facilities. Full coverage of all the demand locations is a constraint to be imposed on the demand side. In many circumstances, LSCP might identify a large number of facilities to ensure full coverage. As far as we know, no previous literature has applied both models to the same location selection problem and compared their results. Thus in this chapter we adopted both approaches to provide a more comprehensive evaluation of the problem at hand.

In coverage-based location models, the solution is highly sensitive to geographic representation such as the size or shape of basic spatial units (Murray and O’Kelly 2002, Murray, O’Kelly and Church 2008). The term Modifiable Areal Unit Problem (MAUP) was coined to describe both scale and aggregation effects (Openshaw 1983). In particular, major analytical differences can be found depending on the size of units used as well as the ways in which the study area is divided up. The appropriate size and shape of spatial units need to be defined to reduce uncertainties (Daskin 2011, Daskin and Dean 2004, Openshaw 1983). The general assumption for the use of location problems in health research is that the region of interest can be represented as a set of discrete spatial objects, such as points (*e.g.*, population centroids) and polygons (*e.g.*, areal units). Since a continuous space is discretized with a set of spatial objects, aggregation errors are unavoidable. In a study conducted by Murray and O’Kelly (2002), the use of alternative realizations of space was explored in the application of LSCP. Both irregularly spaced points – block centroids and regularly spaced points – three grid patterns were evaluated. The results confirmed that the region misrepresentation coverage error is low in most cases, thus the use of point-based representations is appropriate with supporting analysis. In this chapter we introduced a hexagon tessellation to represent the study area and use their centroids as candidate facility sites, in order to ensure a more equal distribution of new facilities in space. Hexagon tessellations weren’t as widely applied in previous location problem literature. With areal interpolation techniques, we were able to generate a population density surface and re-aggregate population information to designated hexagons.

Another issue, the concept of *coverage*, is essential for coverage-based location problems. The service coverage is defined by applying a threshold distance depending on different types of medical practice. A service facility is considered accessible if the travel distance between the population in need and facility location is less than or equal to the threshold distance, and inaccessible if otherwise. The most commonly used method is to generate a circular coverage using Euclidean distance. However, with the advent of the geographic information system (GIS) and transportation network data, a better coverage method can be employed using time-distance-based coverage. In the context of healthcare coverage, the network distance is a better measurement due to several reasons in practice: 1) people travel along the transportation network to access health services; 2) it is easier for people to relate to, and interpret travel distance along the transportation network; 3) there is the potential of accounting for transportation modes and traffic conditions for further analyses. It should be noted that network-based service coverage is very sensitive to the threshold travel distance depending on the level of medical practice. For example, the ideal network coverage of a cardiac or stroke center would differ from that of a primary care physician. The recommended treatment windows for myocardial infarction (MI) and stroke are 90min and 180min respectively, while there are no time window limits for regular health check-ups (Busingye, Pedigo and Odoi 2011). Since the threshold travel distance between a patient and a

facility should be defined based on the facility type, the impact of coverage is explored with a range of travel thresholds in this chapter.

5.3 Motivation for the Net-CLP

The Net-CLP was employed to balance between the service efficiency of facilities and the number of covered demands. Like the name indicates, the Net-CLP is network-based and applied in a GIS environment. The structure of the Net-CLP can be summarized in Table 7. The Net-CLP contains two models – the Net-LSCP and the Net-MCLP, and tackles the facility siting problem from two different perspectives. Considering full coverage of the study area, the Net- LSCP attempts to minimize the cost or number of the facilities to cover all the demands. The objective is to ensure the total number of covered demands rather than to distinguish the sizes of demand nodes. In contrast, the objective of the Net-MCLP is to maximize covered demand when the number of facilities to be added in the area is known. The model considers the size of the demand (*e.g.*, the total population of a population centroid) and seeks to cover larger demands.

Table 7. Characteristics of the Net-CLP

Net-CLP	Net-SCLP	<p>Network-based set covering location problem</p> <ul style="list-style-type: none"> • Minimize the number of facilities in order to cover all the demand nodes regarding spatial efficiency of equity • Impose constraints on the demand side • Existing facility location and its covered demands are considered • Facility coverage is defined based on transportation network
	Net-MCLP	<p>Network-based maximal covering location problem</p> <ul style="list-style-type: none"> • Maximize covered demand given limited number of potential facilities • Impose constraints on the supply side • Existing facility location and its covered demands are considered • Facility coverage is defined base on transportation network

As mentioned above, the notion of network coverage is essential to the Net-CLP. In the Net-CLP, coverage is delineated by applying threshold travel distance based on a transportation network. The location of the population is covered if the network travel distance to healthcare facilities within proximity is less than or equal to the threshold travel distance, and not covered if otherwise. As GIS measures coverage areas based on real world transportation networks, a population location in the street network is identified as the anchor point and the network-based demand areas are calculated according to the threshold travel distance. Accordingly, the shape of the network coverage area is usually irregular and smaller than the circular buffer area calculated using Euclidean distance. Although the Euclidean and Manhattan distances are correlated with network distances, network distances are proven to be much more accurate in defining service coverage, especially in areas where suburban areas are sparsely connected with road networks (Apparicio *et al.* 2008).

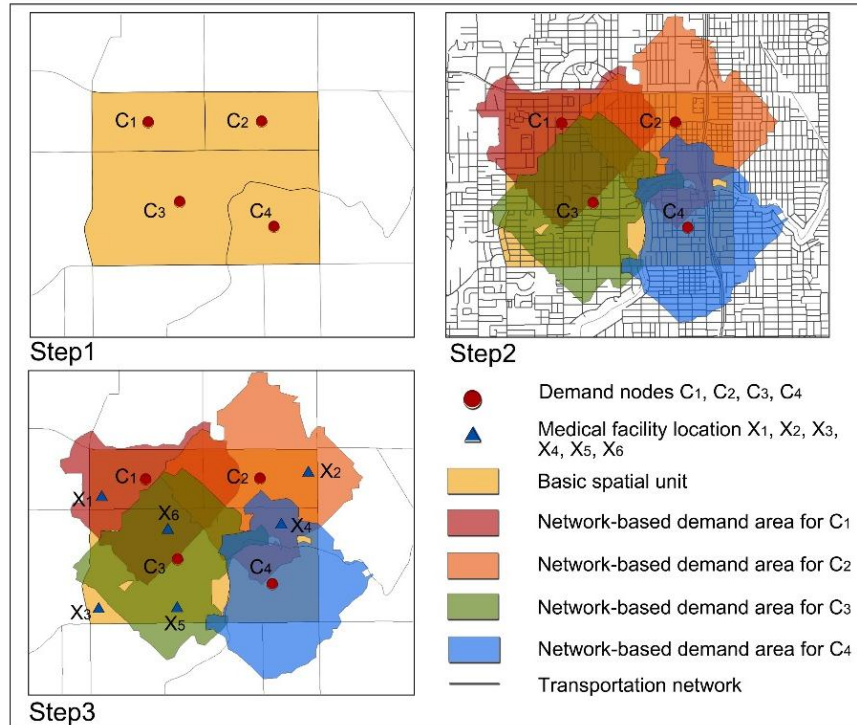


Figure 13. Geo-processing to identify the covered/uncovered relationship between healthcare demand and facility location

Other than defining the network-based coverage, it is crucial to identify the covered and uncovered relationship between healthcare demand and facilities. Most traditional location problem literature didn't incorporate the application of GIS in this process. The Net-CLP identifies the relationship between demand node and facility location (*i.e.*, covered or not covered) with a set of geo-processing steps in a GIS environment (Figure 13). *Step 1* defines the basic spatial unit of analysis (*e.g.*, census tract) and generates the aggregated population center, such as the population centroid. Each center contains the information on the overall population within the spatial unit. *Step 2* generates the network-based demand area from each population centroid within a certain travel time threshold. The network demand area defines the boundary that people can travel to within a certain time threshold. *Step 3* identifies the facilities (*e.g.*, the location of hospitals, clinics, primary care physicians, etc.) that are located within each demand area. If a facility is found to be within a demand area, the population has access to the facilities and is considered covered. For the formulation of the Net-CLP, a binary variable a_{ij} is used to describe the relationship between demand node i and facility location j : $a_{ij}=1$ if demand node i is covered by healthcare facility j ; $a_{ij}=0$ if otherwise. In other words, if the shortest network travel distance between the demand node i (C_1 , C_2 , C_3 and C_4 in Step 1) and the facility location j (X_1 , X_2 , X_3 and X_4 in Step 3) is smaller than the threshold time distance, the demand node is covered by the facility service area. For example, facilities X_1 and X_6 are within the travel threshold distance from the population demand node C_1 ; thus we consider C_1 covered by the service range of X_1 and X_6 . Notice that one facility might be accessible by several demand nodes. For example, facility X_6 is located within the demand areas of both C_1 and C_3 ; thus C_1 and C_3 can both access X_6 .

5.3.1 Formulation of Net-SCLP

It should be noted that the traditional LSCP doesn't ensure spatial dispersion of facilities to cover as many population as possible, resulting in considerably large overlapping areas of facility service coverage. For example, the traditional LSCP might generate a solution that facilities are placed near existing facilities. This phenomenon raises the potential problem of "spatial inefficiency of service" in locating healthcare facilities. If we assume all demands are covered by a fixed number of facilities, then we expect that those facilities are dispersed over sufficient space to avoid unnecessary competition. To make sure the most optimal solution is selected, a spatial weight w_j is included in the objective function to ensure "spatial efficiency". The formula for Net-LSCP can be written as follows:

$$\text{Minimize: } \sum_{j \in J} w_j X_j \quad (1)$$

$$\text{Subject to: } \sum_{j \in J} a_{ij} X_j \geq 1 \quad \forall i \quad (2)$$

$$\sum_{j \in N_e} X_j = n \quad (3)$$

$$X_j = \{0, 1\} \quad \forall j \quad (4)$$

T : total sum of healthcare demand areas, which equals to the total study area

p_j : total sum of population located within the service range of facility j

w_j = service efficiency weight, the ratio between T and P_j ($w_j = k * T/p_j$); k is a scalar ($k = 10,000$ in this application)

I = set of demand nodes i

J = set of potential healthcare facility locations j ; both demand nodes and existing healthcare facilities are considered as potential sites

N_e = set of existing healthcare facility locations; $N_e \in J$, $N_e = \{j \mid d_{ij} \leq R\}$;

n = number of existing facilities (i.e. $n = |N_e|$)

$a_{ij} = 1$ if demand at location i covered by potential healthcare facility j ; 0 otherwise

d_{ij} = travel distance from location i to potential healthcare facility j

R = effective coverage distance of a healthcare facility

$X_j = \begin{cases} 1 & \text{if potential site } j \text{ is selected} \\ 0 & \text{Otherwise} \end{cases}$

The objective function (1) minimizes the total number of facilities that are selected to cover all demands. Notice that the objective function includes the weight w_j to facility location X_j , which is an adjusted ration between total study area and population covered by facility j . Since the total area and population located within the service range of facility j are constant, the weight w_j would be constant for each facility j . This treatment is important as a model specification when the best configuration of facilities should be explicitly identified, because multi-feasible optimal solutions may exist in the traditional Covering Location Problems (Wei 2015, Niblett and Church 2015). Constraints (2) state that each demand node i must be covered by at least one healthcare facility. The left-hand side gives the number of located facilities that can cover demand node i . Constraint (3) considers the situation in which there are n existing facilities. Constraints (4) are the integrality restriction. The decision variable for facility location X_j is defined as $X_j = 1$ if facility location j is selected; $X_j = 0$ if otherwise.

5.3.2 Formulation of Net-MCLP

The decision variable for the Net-MCLP needs to be defined: $Z_i = 1$ if demand node i is covered by at least one of the facilities; $Z_i = 0$ if otherwise. With this notation, the Net-MCLP can be formulated as follows:

$$\text{Maximize: } \sum_i h_i Z_i \quad (5)$$

$$\text{Subject to: } Z_i \leq \sum_j a_{ij} X_j \quad \forall i \quad (6)$$

$$\sum_j X_j = p \quad (7)$$

$$\sum_{j \in N_e} X_j = n \quad (8)$$

$$X_j = \{0, 1\} \quad \forall j \quad (9)$$

$$Z_i = \{0, 1\} \quad \forall i \quad (10)$$

I = set of demand nodes i

J = set of new healthcare facility locations j , including both new and existing healthcare facilities

N_e = set of existing healthcare facility locations; $N_e \in J$

n = number of existing facilities (i.e., $n = |N_e|$)

h_i = the size of demand at node i

$a_{ij} = \begin{cases} 1 & \text{if demand node } i \text{ can be covered by a facility at candidate site } j \\ 0 & \text{if not} \end{cases}$

$Z_i = \begin{cases} 1 & \text{if demand node } i \text{ is covered} \\ 0 & \text{if not} \end{cases}$

p = total number of facilities

The objective function (5) maximizes overall covered demand – the total population covered by the service range of healthcare facilities. Note that the objective considers the overall covered demand rather than simply the number of covered demand nodes. Notice that since the main concern of the Net-MCLP is to maximize covered demand, efficiency weight w_j is replaced with h_i . Constraints (6) state that demand node i is counted as covered when it is within the service range of at least one facility. Constraints (7) state that a total number of p facilities are to be located. Note that the value of p includes both existing and new facilities that are to be located. Constraint (8) considers the situation that are n existing facilities. Constraints (9) and (10) are integrality restrictions.

5.3.3 Extension: Medical Service Capacity

Other than the number of facilities and their locations, the performance of health service also depends on the capability of those facilities. Thus, consideration of the medical service capacity (MAC) in the model is crucial for a more effective healthcare delivery system. A facility should have enough capacity to accommodate patients as well as to ensure that the minimum capacity requirements are met. On the one hand, if the patients visiting the facility exceed the threshold limit of the medical service capacity, there would be adverse effects such as reduced consultation hours, delay of service, late transferring in case of emergency, and so forth. On the other hand, the operation of the healthcare service depends on having a sufficient number of patients. Having not enough patients may cause discontinuation or relocation of the services. In general, MAC can be measured by indicators such as the number of beds, number of medical personnel, or other similar indexes. For example, the concept of carry capacity for emergency service was introduced by

Narasimhan et al. (1992). They consider multi-service levels of facilities to determine their location. Using a different set of capacity for each service level, the model maximizes the amount of covered demand while considering the total demand assigned to each facility. Although the model takes the form of the p -median problem, it is adapted from the covering models with capacity constraints, ensuring that the population visiting a facility does not exceed its capacity.

In this chapter, we define MAC using the ratio between population in need and the number of beds to indicate the service level. The computation of MAC is as follows:

$$MAC = k \left(\frac{\text{the total amount of population of region}}{\text{the total number of beds in the region}} \right)$$

Here k is a scalar, which is used to adjust the ratio when it is used as a constraint MAC^U or MAC^L (see constraints 11 and 12 below). A lower MAC indicates less burden on the supply side since the facilities are able to accommodate more people, while there is less population in need in the regions; a higher MAC indicates otherwise. According to the Global Health Facts (2013), the average MAC in the United States is 333, and the MAC in Hillsborough County is 170. We applied these two values ($MAC = 170$ and 333) for analyses for comparison. For the Net-LSCP-MAC and Net-MCLP-MAC, the following constraints (11) and (12) are prescribed:

$$\text{Subject to: } \sum_{j \in J} g_j X_j \leq MAC^U \quad \forall j \quad (11)$$

$$\sum_{j \in J} g_j X_j \geq MAC^L \quad \forall j \quad (12)$$

g_j = the availability of medical service at facility j (i.e. $g_j = k \left(\frac{\text{total population covered by facility } j}{\text{the number of beds in facility } j} \right)$)

MAC^U = the maximum medical service capacity for a facility

MAC^L = the minimum medical service capacity for a facility

Net-CLP can incorporate constraints (11) or (12), or both if necessary. In the following case study, we only apply MAC^U , the constraints (11). Since the study area has a large and diverse population, our main concern is to prevent facilities from exceeding their service capacities.

5.4 Case study

To demonstrate the application of the Net-CLP in making healthcare facility location decisions, we selected Hillsborough County, Florida as the study area. Hillsborough County is located on the west coast of Florida in the Tampa-St. Petersburg metropolitan area. It is ideal as the study site for exploring the spatial interaction between healthcare supply and demand for the following three reasons: 1) the area has a large and diverse population distributed across enough space to fully represent the healthcare demand; 2) the landscape is flat and mostly urban, thus the developed road network can generate larger network-based coverage; 3) the information from the set of existing hospitals helps to demonstrate the application of the Net-CLP, which takes the existing facilities into account. Three data sets were applied in the analysis: 1) population distribution data extracted from the 2012 Census Summary File (US Census Bureau 2012); 2) existing healthcare facilities data including location and capacity information collected from the Florida Geographic Data Library Documentation (FGDL 2010); 3) the land use data downloaded from Hillsborough Community Atlas. In order to reduce aggregation, Low-density and isolated residential areas as

well as non-residential areas were excluded from the analysis to better represent health care demand areas.

5.4.1 Spatial representation of study area

In covering location problems, the solution is highly sensitive to how the study area is represented (Murray 2005). Thus, selecting an appropriate spatial unit for analysis is critical for minimizing aggregation errors (Apparicio *et al.* 2008). For example, the people requiring services are usually grouped together by residential locations, and distances to the healthcare providers are calculated from the aggregated centroids to the service locations, instead of from the individual residential locations (Cromley and McLafferty 2012). In some cases, the true distance exceeds the modeled distance; in other cases, the true distance is shorter than the modeled distance. This will result in under- or overestimation of distance or coverage. In most healthcare research, disaggregated location data on patients is not available, which has brought forth the alternative use of centroids for a given spatial unit. Thus aggregation errors arise from the distribution of population points around the population centroids (Hewko, Smoyer-Tomic and Hodgson 2002).

Figure 14-a presents the pattern of residential areas in the county. In this case study, if we use census units such as block groups as basic units of analysis, the irregular shapes of those units will result in great variability in aggregation errors. For example, aggregation errors will be especially large for census units that contain very little population and cover larger geographical areas. To ensure a more equal spatial representation of the study area, we used a regular tessellation of hexagons. Applying regular tessellation provides an alternative way of spatial representation. Conventional wisdom suggests that representation errors reduce when the smaller grid cells are applied. However, a study by Murray and O'Kelly (2002) showed that a smaller grid pattern doesn't necessarily enhance representation accuracy. Nevertheless, to emphasize on a more equal spatial distribution of facilities and reduce computational efforts, the potential healthcare demand areas are represented by the regular size of hexagon units (1 mile in height and length), as illustrated in Figure 14-b. Note that the distances from all the edges to the hexagon centroid are equal, thus the aggregation errors will be more consistent.

Since the original census units contain population information while the hexagon units do not, we applied spatial interpolation techniques to generate a population density surface from the census data and reassign the population information to the hexagon units (Figure 14-c). Areal interpolation is a kriging-based interpolation method applied to process polygons with various shapes. Predictions and standard errors can be generated for all points between and within polygons. In this case, the population for a continuous surface of points was estimated. Then the population predictions are re-aggregated back to a new set of polygons. With this geo-processing step, the population predictions generated from 18366 block group units within the residential area are re-aggregated to 804 hexagon units (Figure 14-d). These homogeneous hexagons helped to ensure the quality of solutions by reducing model complexity when mixed linear programming was applied.

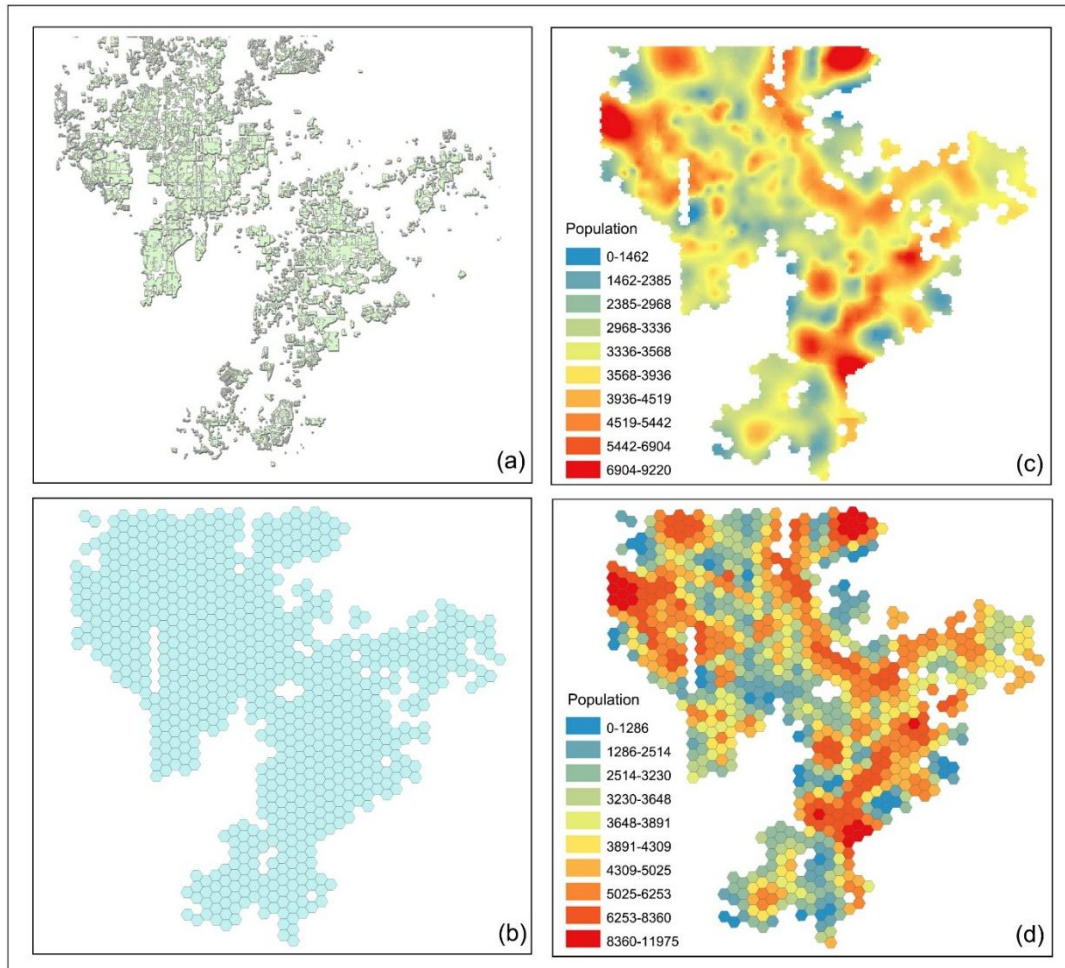


Figure 14. Potential demand sites representation with areal interpolation

Note that in the application of the Net-CLP, the centroids of the hexagons serve as both the demand nodes and the potential locations for new facilities, since in theory the facilities should be located as close to the population locations as possible. As displayed in Figure 15, the population is clustered around urban centers, especially those within and around the Tampa metropolitan area. There are also greater population residing around the Westchase, Brandon, Plant City, Apollo Beach and Sun City centers, where major highway networks are well developed. When trying to locate a new healthcare facility, it is reasonable to assume that it should be located within or near one of these population clusters. The health facilities addressed in this case are hospitals, which offer tertiary care – the highest level of healthcare, and their location should be strategically decided. In Hillsborough County, the distribution of hospitals matches the distribution of the population in general (Figure 15). However, previous literature has identified an overall shortage of healthcare resources in this area, raising the critical question of where new hospitals should be located to improve healthcare access (Ye and Kim 2014).

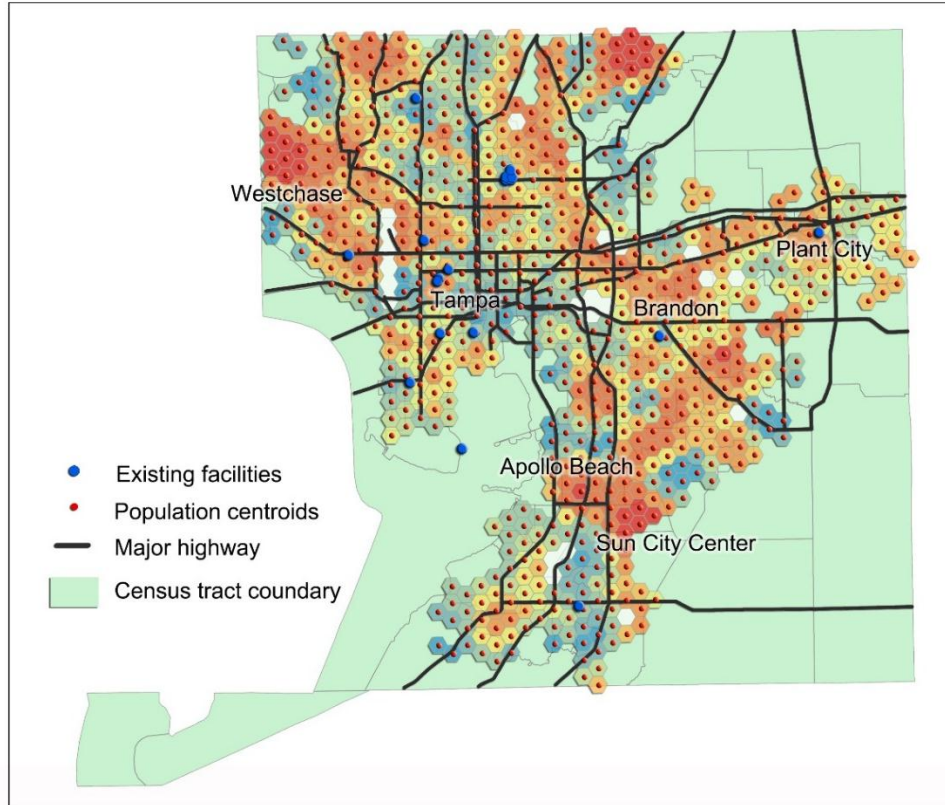


Figure 15. Distribution of 19 existing hospitals and 804 hexagon centroids

5.4.2 Key parameters

To apply the Net-CLP, two key parameters need to be defined: the travel time threshold and spatial weight w_j . In previous literature, a 30-minute travel time threshold for the primary road conditions was suggested for tertiary care (Lee 1991). The 30-minute threshold was also used for defining rational service areas and capturing Health Professional Shortage Areas (HPSAs) by the Health Resources and Service Administration (HRSA) (Luo and Wang 2003). Moreover, we set up 10- and 20-minute travel thresholds and compared the results to that of the 30-minute travel threshold. To define the w_j , total sum of healthcare demand areas to be covered was measured as the numerator ($T = 617 \text{ mile}^2$) and the total population covered by facility j was calculated as the denominator.

The goal of the Net-CLP was to improve overall healthcare access for the population in need. In this specific case, two scenarios were of primary concern: 1) what is the minimum number of new hospitals needed to cover all the re-delineated population centroids (*i.e.*, hexagon centroids) and where should they be located, when the travel time thresholds are 10-minute, 20-minute and 30-minute respectively? 2) Where should we locate a certain number of new hospitals so as to maximize the overall covered population, considering the population size at each centroid, for 10-minute, 20-minute and 30-minute travel time thresholds respectively?

5.4.3 Results and discussion

The application results of the standard Net-LSCP and Net-MCLP are reported in Table 8. All instances were conducted with an Intel *i*-3 core (2.13 GHz) on Window 7 with 8 GB memory.

They were solved to achieve optimality within a few seconds using CPLEX 12.1. The results were imported into ArcGIS for visualization and analyses. First, Table 8 summarizes the solutions of the standard Net-LSCP for the three thresholds of travel distance ($T=10, 20$ and 30 -minute). Here, p_{new} is the number of new facilities needed to achieve 100% coverage. The overlapped coverage means the percentage of the total area that is covered by both the new facilities and existing facilities. Not surprisingly, the number of facilities needed for 100% coverage is reduced with increased T as these facilities can serve a larger area. For example, the number of needed facilities to ensure 100% coverage is 18 for 10-minute travel threshold and 2 for 30-minute travel threshold. The objective function representing service efficiency is also considerably decreased. This implies a tradeoff between improving the service coverage of a medical facility and increasing the number of facilities with smaller service coverage. It is reasonable to assume that a more developed and accessible transportation system would help increasing service coverage and enhancing the service efficiency for the study area.

The total overlapped coverage generally increases with increasing travel thresholds. Medical facilities face more competitions when more of their service areas overlap with one another, while the population residing within the overlapped service areas has more choices and opportunities to access healthcare services. Figure 16 shows the service areas of existing and new facilities when the standard Net-LSCP is applied. As the travel time threshold increases, the service range of a new healthcare facility is increased to cover more demands. Note that the most efficient solution is ensured with the application of the spatial weight w_j . With the integration of GIS, we were also able to visualize the most optimal solution. Most selected new facility locations are close to the highways. For $T=10$ (Figure 16-a), the previous shortage areas in the Northwest and Southeast are covered by the new facilities. For $T=20$ (Figure 16-b), the shortage areas around Westchase and Sun City center with high population density are now within the service range of new facilities. For $T=30$ (Figure 16-c), two new facilities are sited in the upper-north and southeast area to cover previous shortage areas.

The results from the standard Net-MCLP are presented in Table 9. In general, the total covered area increases with a larger number of new facilities (p_{new}). For $T=10$, 18 facilities are required to ensure 100% coverage, but only 6 facilities are needed for $T=20$ and 2 for $T=30$. However, the percentage of total covered area does not increase linearly with increasing p_{new} . With the same number of new facilities, the results can vary depending on the travel threshold when applying Net-MCLP. Also, applying Net-MCLP and applying Net-LSCP generate different results when ensuring 100% coverage, since they have different objective functions. Net-LSCP seeks to cover more demand locations and minimize cost with the efficiency weight w_j , while the Net-MCLP seeks to cover higher demand locations and maximize economic return. As shown in Figure 5, the spatial pattern of new facilities for $T=20, p_{new}=6$ (Figure 17-a) and $T=30, p_{new}=2$ (Figure 17-b) generated by Net-MCLP are different from those in Figure 16-b and 16-c respectively.

Table 8. Model behaviors of the Net-LSCP and the Net-MCLP

Threshold	p_{new}	Objective	Solution (facility location ID)	Overlapped coverage
10 min	18	70831.6	25, 151, 158, 230, 271, 300, 320, 331, 459, 527, 546, 596, 611, 634, 683, 687, 702, 724	29.1%
20min	6	4010.5	38, 194, 298, 327, 623, 757	59.9%
30min	2	37.0	71, 326	45.0%

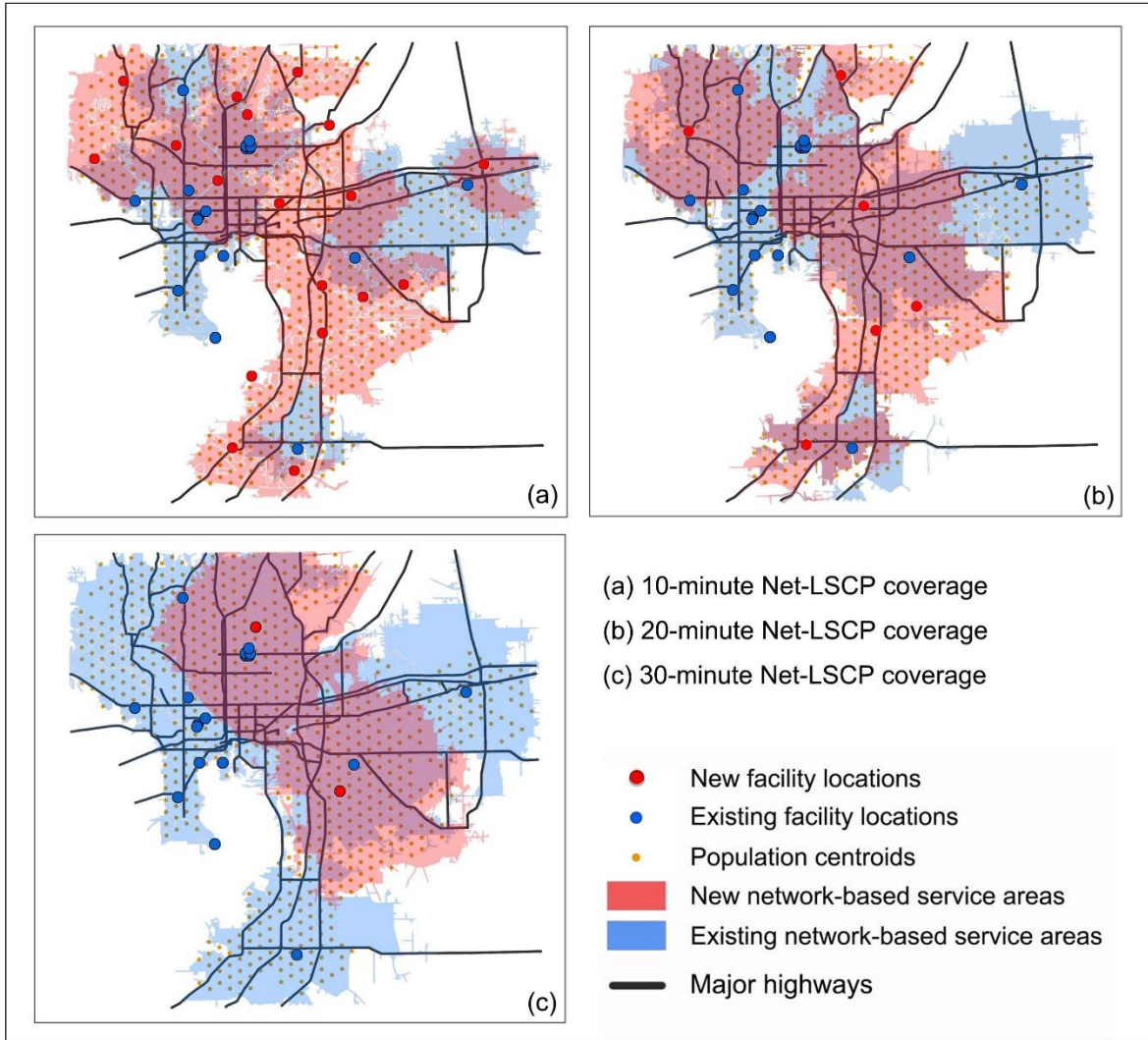


Figure 16. Standard Net-LSCP selected locations for 10, 20, and 30min travel time threshold

Table 9. Model behaviors of the Net-MCLP

Threshold	p_{new}	Objective	Solution (facility location ID)	Covered areas
10min	1	746,939	298	69.4%
	2	809,716	25, 298	75.2%
	3	863,635	25, 298, 641	80.3%
	4	906,939	25, 298, 641, 320	84.3%
	5	932,376	25, 298, 641, 320, 700	86.6%
	6	956,738	25, 298, 641, 320, 700, 724	88.9%
	7	979,902	25, 224, 298, 641, 320, 700, 724	91.1%
	8	999,297	25, 271, 320, 300, 331, 700, 724, 641	92.9%
	9	1,016,784	271, 300, 335, 459, 519, 320, 641, 724, 700	94.6%
	10	1,030,805	151, 271, 300, 320, 332, 459, 515, 641, 700, 724	95.8%
	11	1,042,921	151, 271, 300, 320, 332, 459, 515, 641, 700, 724, 783	96.9%
	12	1,051,933	151, 271, 300, 320, 332, 459, 481, 515, 641, 683, 700, 724	97.7%
	13	1,057,934	151, 234, 271, 300, 320, 332, 459, 481, 519, 641, 700, 724, 751	98.3%
	14	1,063,560	151, 234, 271, 300, 320, 332, 459, 481, 519, 611, 641, 700, 724, 783	98.8%
	15	1,068,515	151, 234, 271, 300, 320, 332, 459, 481, 515, 611, 646, 669, 700, 724, 751	99.3%
	16	1,072,110		99.6%
	17	1,074,278	25, 151, 158, 224, 271, 300, 320, 331, 459, 527, 546, 596, 611, 683, 687, 702, 724	99.8%
	18	1,076,124	69, 71, 120, 151, 158, 234, 253, 271, 298, 335, 459, 481, 515, 611, 646, 700, 724, 783	100.0%
20min	1	988,075	298	91.8%
	2	1,016,750	298, 764	94.5%
	3	1,036,735	298, 729, 764	96.3%
	4	1,051,691	298, 515, 646, 764	97.7%
	5	1,066,331	224, 254, 360, 653, 702	99.1%
	6	1,076,124	134, 194, 253, 515, 647, 702	100.0%
30min	1	1,075,594	263	99.9%
	2	1,076,124	13, 757	100.0%

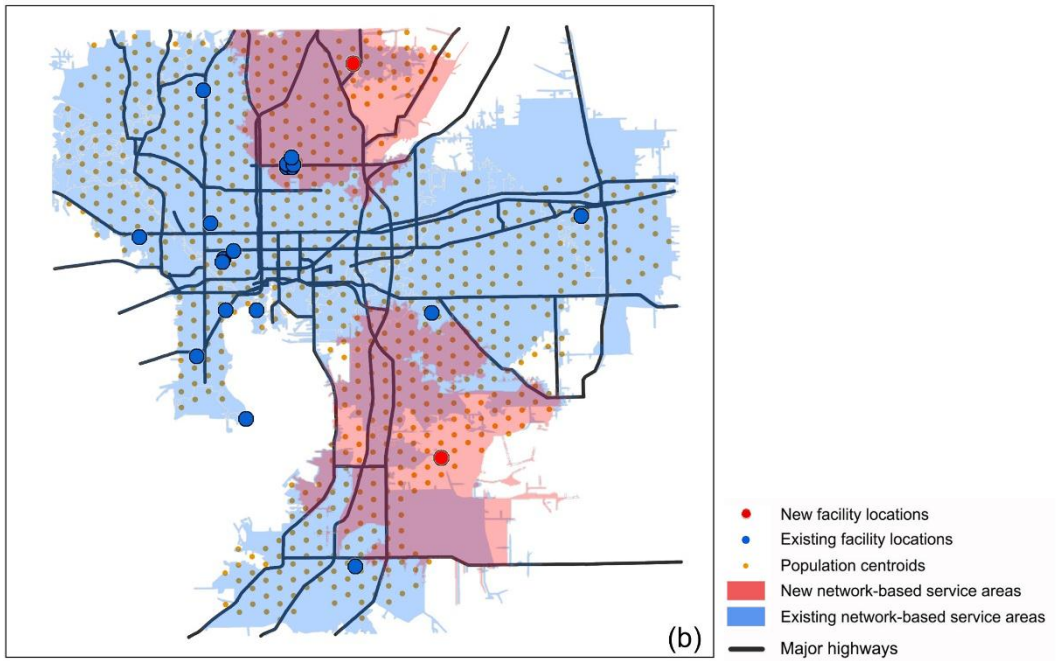
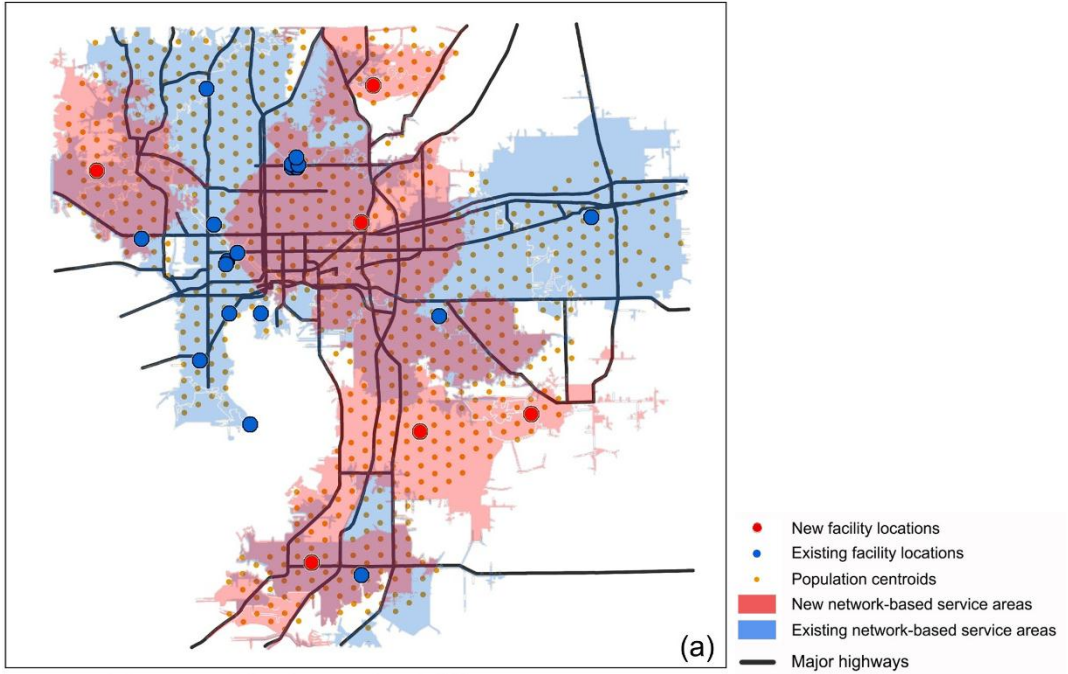


Figure 17. The Net-MCLP solution for $p_{new}=6, T=20$ (a) and $p_{new}=2, T=30$ (b)

Table 10 summarizes the model behaviors of the Net-LSCP-MAC and the Net-MCLP-MAC. To compare the results and examine the model behaviors, we used two sets of MACs (i.e., $MAC^U = 170$ and 333) for the Net-LSCP-MAC, which represent the current MAC for Hillsborough County and the average of the United States, respectively. For the Net-MCLP-MAC, we added another small MAC value = 50 to examine the model's response. Overall, the behavior of the Net-LSCP-MAC is more sensitive to the change of the medical service capacity than that of the Net-MCLP-MAC. For example, for the solutions when $T=10$, different facility locations are selected for MACs = 170 and 333. On the contrary, the solutions of the Net-MCLP-MAC are more consistent despite the change of MACs. In Table 10-a, smaller service capacity in general increases the objective function. For example, when $T=10$ and $MAC^U = 170$, more facilities are needed to achieve full coverage due to limited service capacities. In Table 10-b, the objective function increases with increasing MAC^U before achieving full coverage. However, it should be noted that the solutions by the Net-MCLP-MAC can help to identify locations that are selected regardless of service capacity level. For example, location 298 is included in most instances, and 263, 699 and 764 are also frequently selected in many instances, indicating that these locations are more desirable than others.

Figure 18 presents the coverage percentage change in population associated with solving the Net-MCLP for $T=10, 20$ and 30 . As displayed in Figure 18, the covered population percentage for $T=10$ is increasing rapidly until $p_{new}=7$ (91%), but increasing slows down afterwards until reaching 100% coverage with $p_{new}=18$. Based upon the increasing pattern, it is indicated that $p_{new}=7$ could be the critical transitional point. Thus, serving the area with 7 new facilities is an effective strategy for the 10-minute threshold. For $T=20$, the increasing of coverage percentage starts to slow down at $p_{new}=4$ where the coverage percentage is nearly 98%, implying that adding 4 healthcare facilities would be effective enough to cover demands. In addition, for the $T=30$, adding 1 facility would be sufficient since most of the demand area (99%) can be covered.

In summary, the main advantage of running the Net-CLP with two sub-models is balancing spatial equity and economic efficiency when assessing potential facility locations. For example, when we evaluate candidate facility locations applying Net-CLP, those locations are categorized into 4 categories: 1) locations selected by both the Net-LSCP and the Net-MCLP (for example, the locations at 25 and 298) are regarded as the most critical locations since they satisfy both spatial equity and economic efficiency; 2) locations selected only by Net-LSCP are able to cover larger areas to ensure 100% spatial coverage; 3) the locations selected only by the Net-MCLP can cover larger population given limited resources; 4) locations selected by neither the Net-LSCP nor Net-MCLP are not optimal choices for healthcare planning.

5.5 Conclusions

The main focus of this research is to propose a Net-CLP when locating healthcare facilities. The application of this method was demonstrated with a case study of Hillsborough County, Florida. The Net-CLP applies a network-based distance with an integration of GIS, which greatly improves accuracy in representing actual travel distance in order to capture distance decay and generate service coverage.

To address the balance between healthcare demand and supply, the model consists of two sub-models – the Net-LSCP and Net-MCLP. It considers the situations in which all the demand nodes should be covered, as well as the situations in which there are only a limited number of new facilities. Thus Net-CLP can identify the locations that satisfy both spatial equity and efficiency,

Table 10. Model behaviors of the Net-LSCP-MAC and the Net-MCLP-MAC.

(a) Solutions of Net-LSCP-MAC

Threshold T	MAC^U	p_{new}	Objective	Solution (facility location ID)	Coverage
10 min	170	19	77618.6	69, 133, 151, 158, 224, 230, 253, 295, 320, 404, 459, 546, 572, 596, 611, 687, 700, 724, 751	100%
	333	18	71880.5	151, 158, 224, 230, 243, 298, 300, 320, 404, 459, 524, 546, 596, 611, 683, 687, 702, 724	100%
20min	170	6	4195.6	38, 194, 298, 305, 515, 757	100%
	333	6	4010.5	38, 194, 298, 327, 623, 757	100%
30min	170	2	37.02	71, 327	100%
	333	2	37.02	71, 326	100%

(b) Solutions of Net-MCLP-MAC

Threshold T	MAC^U	p_{new}	Objective	Solution (facility location ID)	Coverage
10min	50	1	720,326	263	66.9%
		2	754,953	263, 313	70.1%
		3	776,355	263,313,549	72.1%
	170	1	746,939	298	69.4%
		2	809,716	25, 298	75.2%
		3	863,635	25, 298, 641	80.3%
	333	1	746,939	298	69.4%
		2	809,716	25, 298	75.2%
		3	863,635	25, 298, 641	80.3%
20min	50	1	968,089	263	90.0%
		2	986,490	263, 549	91.6%
		3	995,816	263, 412, 549	92.5%
	170	1	988,075	298	91.8%
		2	1,016,750	298, 764	94.5%
		3	1,036,735	298, 729, 764	96.3%
	333	1	988,075	298	91.8%
		2	1,016,750	298, 764	94.5%
		3	1,036,735	298, 729, 764	96.3%
30min	50	1	1,075,594	263	99.9%
		2	1,076,124	263, 699	100%
	170	1	1,075,594	133	99.9%
		2	1,076,124	13, 699	100.0%
	333	1	1,075,594	133	99.9%
		2	1,076,124	13, 699	100.0%

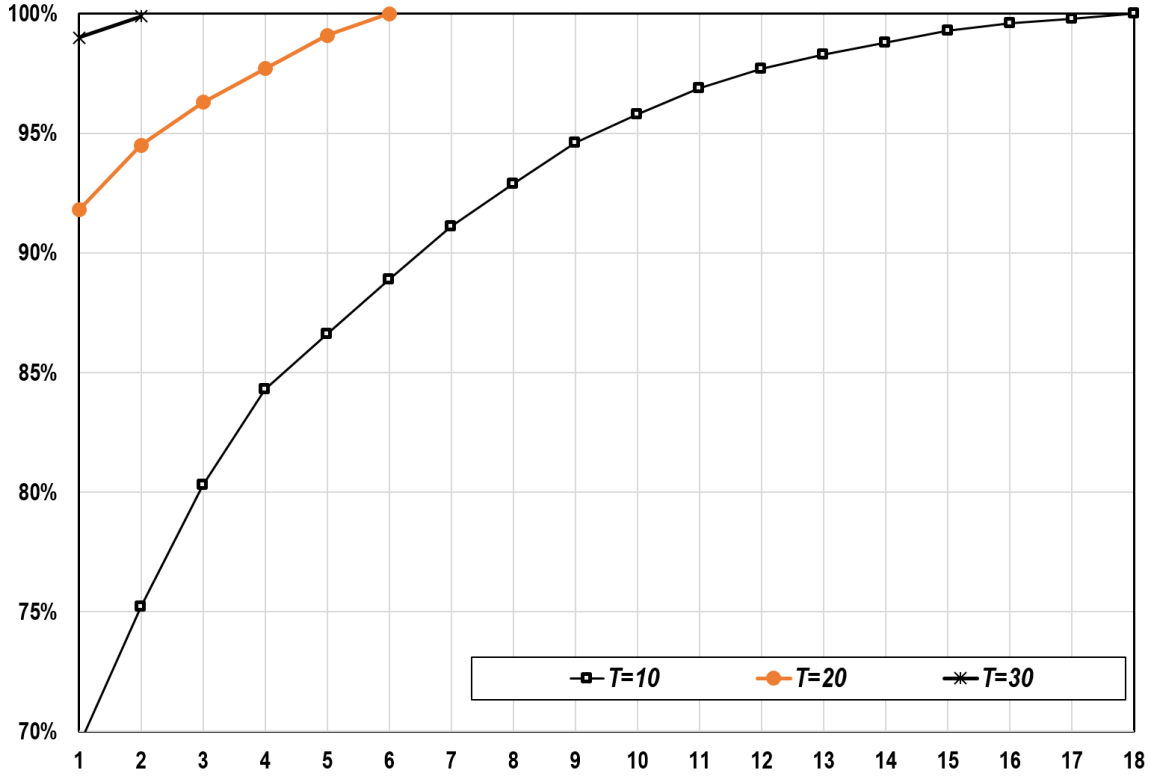


Figure 18. Covered population percentage with p_{new} for three threshold ($T=10, 20, \text{ and } 30$).

by comparing the results of the Net-LSCP and Net-MCLP. Moreover, Net-CLP can be extended with constraints on the service capacities, ensuring that there is enough capacity to accommodate patients while minimum capacity requirements is met to sustain operation cost. Using the set of results given by the Net-CLP, we were able to evaluate potential facility locations with a relatively small amount of computation and programming. In many cases, the equity of service for the healthcare demand conflicted with the economic efficiency of the healthcare facility supply. Thus, taking these two perspectives into consideration is necessary when making location decisions.

The Net-CLP is easy to implement and straightforward to interpret. It can be applied for different kinds of medical practices by considering different travel time thresholds and service capacities. The location decisions made can be used for better healthcare planning and improvement of accessibility in underserved areas. However, the Net-CLP needs to be expanded in future studies. One major issue of current application is that the model does not account for non-spatial factors when identifying facility locations. For example, in a low socio-economic neighborhood, people might not be able to afford healthcare despite a new hospital having been assigned to the community. In a more practical sense, the location selected might not be feasible if the geographical features are not suitable for the establishment of a new hospital or if the cost is too high due to factors such as land rent. Since the model can be easily extended by changing the objective function and adding constraints sets, there is great potential to incorporate more socio-economic factors in the formulation. Moreover, travel distance estimations as well as transportation modes might differ for urban and rural areas. For example, network travel distance might be shorter in urban areas than in rural ones or vice versa, especially when considering traffic conditions and commuting times.

CHAPTER 6
CONCLUSIONS

6.1 Summary of dissertation research

This dissertation addresses the geography of healthcare access and disparity issues in the United States using geospatial methods. Disparities in access to quality healthcare services are of great concern in the field of both public health and geography. Influencing by both spatial factors (*e.g.*, the geographical location of primary care physicians and patients, transportation system, and travel distance) and non-spatial factors (*e.g.*, social-economic status, age, gender, ethnicity, and health insurance status) (Guagliardo 2004), the essence of access is widely recognized. Thus, an innovative contribution of my work is the integration of spatial modeling, geo-statistics and location problems in a Geographic Information System (GIS) environment to investigate healthcare access. Specifically, my dissertation is comprised of three major themes.

Theme 1: Developing an alternative measure of healthcare accessibility

Improving health access begins with developing reliable methods to measure accessibility. I adopted the concept of spatial accessibility in this dissertation, which refers to the fusion of both availability and accessibility of health demand and supply. Spatial accessibility to healthcare is usually measured through addressing the geographical relationship between population in need and health services. Building on previous theoretical frameworks, my dissertation presents an alternative set of healthcare accessibility measures – *a network-based health accessibility index method (NHAIM)* to complement the existing literature. The NHAIM comprehensively evaluates both spatial dimensions of health accessibility and availability, in order to address the spatial disparity of healthcare access in a GIS environment. Both health accessibility and availability are measured and presented as indexes to reveal the spatial patterns of health accessibility and availability, as well as to capture underserved areas in a geographical context. Based on the data downloaded from the Florida Geographic Data Library and the US Census Bureau, I demonstrated the application of the NHAIM to measure spatial disparity and capture spatial mismatch between healthcare supply and demand in the case of Hillsborough County, Florida. The results confirmed the lack of access in certain areas. The result index maps shall provide reference for locating new health care services in the future.

Theme 2: Neighborhood factors and healthcare access

Another focus of my dissertation is to understand neighborhood factors that contribute to healthcare access – both potential and revealed access. For statistical analysis, the supply of primary care physicians (PCPs) is used as an indicator of potential access, and the length of inpatient stay is applied as an indicator of revealed access. The supply of PCPs reflects the dimension of availability in healthcare access. It is essential to an effective healthcare delivery system and whether the healthcare needs are being adequately served. Literature highlighted that local supply of primary care physicians affects preventive healthcare service utilization directly (Continelli et al., 2010). It is reasonable to assume a larger number of primary care physicians located within a census tract provides a better supply of primary care, which is advantageous for the health status of local residents. Thus my dissertation aims to evaluate the geographical distribution of the number of PCPs by location and investigate the relationship between neighborhood factors and the observed spatial pattern. To address this research goal, I conducted an empirical study in the case of Hillsborough County, Florida. The geographical distribution of quantities of PCPs was explored. The contribution of aggregated neighborhood-level factors (*i.e.*,

proportions of individuals living below the poverty level, proportion of individuals 40 years of age and older, and proportions of racial/ethnic groups) to the spatial heterogeneity of PCPs supply was examined using spatial regressions. The results indicate a strong correlation between percentage of Asians in a neighborhood and PCP supply.

On the other hand, revealed access can be addressed through the study of length of inpatient stay – an indicator of healthcare service utilization. I conducted another research which examines the extent to which neighborhood characteristics affected length of inpatient stay in a national scale, which includes a total of 3148 U.S. counties. Generalized linear models and geographically weighted regression models were employed to examine the extent to which neighborhood characteristics affected length of inpatient stay and its spatial variation. The geographic pattern of length of inpatient stay was also examined. The results show that the number of hospital beds is the strongest indicator of length of inpatient stay across the county.

Theme 3: Location problems in healthcare facility siting

A final focus of my dissertation is to locate healthcare facilities so as to maximize service coverage as well as to reduce spatial disparity between healthcare supply and demand. Facility location is essential in ensuring health accessibility. Facilities should be located according to the potential demand to ensure maximum coverage as well as accessibility equity. To address this problem, a *Network-based Covering Location Problem (Net-CLP)* is proposed. The objectives of Net-CLP are: 1) to maximize covered demands with a fixed number of facilities given spatial restrictions and the level of health care service emergency and 2) to account for the demand side and minimizes the total number or cost of facilities needed to cover all healthcare service demands within the network-based service range. It allows for overlapped coverage measured based on threshold network distance. A demand is considered covered if it falls within the threshold network distance from a facility, and uncovered if otherwise. Population centroids are aggregated to represent potential healthcare demand at spatial units. To better demonstrate the potential of using Net-CLP in siting healthcare facilities, Hillsborough County, Florida was selected for a case study. The objective of Net-CLP in this specific case is to identify the optimal locations for new healthcare facilities considering existing ones. A design of optimal locations is proposed and used for comparison and evaluation of existing healthcare facility locations in the study area, as well as providing reference to future healthcare planning.

6.2 Future research directions

The completion of my dissertation is just a beginning. Building on my dissertation, I will continue my research on health care access and disparity issues from four major perspectives.

First, I plan to apply my proposed models to address a specific subgroup of the population to a specific medical practice in regional studies. For example, the NHAIM proposed in Chapter2 can be applied to measure the spatial accessibility to hospitals and clinics that provide Spanish language services for Hispanic population. The Net-CLP proposed in Chapter5 can be used to locate a specific type of facility (e.g., cancer treatment center) in a specific underserved area such as a Black neighborhood.

Second, I will continue to focus on ensuring equity in service distribution by applying location problems. To ensure equal spatial accessibility, sufficient service area and healthcare demand, clustering of public facilities should be avoided. Disbursing public resources in a way that they can be equitably distributed has attracted increasing attention (Batta, Lejeune and Prasad

2014). Despite the fact that the dispersion models have been widely applied in siting obnoxious facilities and business franchises, little has been done in the application of dispersion models in public facility siting such as healthcare facilities. To fill in this gap, I intend to incorporate another perspective in my future research – ensuring equity in the spatial distribution of healthcare resources through application of dispersion models in a GIS environment. Since service areas vary depending on the type of healthcare, a multi-level dispersion problem will be considered.

Third, I intend to incorporate the usage of social media data such as Twitter in the study of healthcare access from a geographical perspective. User messages in social media provide a tremendous amount of information. Researchers have been analyzing social media data such as Twitter messages to address population characteristics, including public health issues. For example, Twitter messages can be used for tracking diseases, revealing sentiment and emotions, and spreading health information and news (Paul and Dredze 2011). Twitter should have more impact on revealing spatial patterns of health access and behavior rather than just disease tracking. Thus my research goal is to extract information on healthcare access and behavior from Twitter, and investigate how social media can promote provider-patient interaction.

Last, I plan to investigate the role of transportation in the study of health behavior and outcomes. Transportation is essential in promoting opportunities in health care access, especially for the vulnerable segments of the population such as low income social groups and the aging population. Thus it is important to evaluate public transport investment and improve accessibility in disadvantaged areas. Moreover, the choice of transportation mode also impacts individuals' health status. For example, walking and bicycling are ways of transportation which are environmental and promote individual health. Thus it is necessary to investigate how to promote healthy modes of transportation through policy and urban planning.

REFERENCE

- AAMC. (2012). AAMC Physician Workforce Policy Recommendations. *Association of American Medical College*.
- Aday, L. A. & R. Andersen (1974) A framework for the study of access to medical care. *Health services research*, 9, 208.
- Alter, G. (1988). *Family and the female life course: the women of Verviers, Belgium, 1849-1880*. Univ of Wisconsin Press.
- Andersen, R. M., H. Yu, R. Wyn, P. L. Davidson, E. R. Brown & S. Teleki (2002) Access to medical care for low-income persons: how do communities make a difference? *Medical care research and review*, 59, 384-411.
- Anselin, L. (1995) Local indicators of spatial association-LISA. *Geographical analysis*, 27, 93-115.
- Apparicio, P., M. Abdelmajid, M. Riva & R. Shearmur (2008) Comparing alternative approaches to measuring the geographical accessibility of urban health services: Distance types and aggregation-error issues. *International Journal of Health Geographics*, 7, 7.
- Arcury, T. A., W. M. Gesler, J. S. Preisser, J. Sherman, J. Spencer & J. Perin (2005) The effects of geography and spatial behavior on health care utilization among the residents of a rural region. *Health services research*, 40, 135-156.
- Arno, P. S., C. Levine & M. M. Memmott (1999) The economic value of informal caregiving. *Health Affairs*, 18, 182-188.
- Ball, J. K. & A. Elixhauser (1996) Treatment differences between blacks and whites with colorectal cancer. *Medical care*, 34, 970-984.
- Batta, R., M. Lejeune & S. Prasad (2014) Public facility location using dispersion, population, and equity criteria. *European Journal of Operational Research*, 234, 819-829.
- Beal, A. C. & C. Fund. 2007. *Closing the Divide: How Medical Homes Promotes Equity in Health Care*. Commonwealth Fund New York.
- Berman, O. & R. Huang (2008) The minimum weighted covering location problem with distance constraints. *Computers & Operations Research*, 35, 356-372.
- Besag, J., & Newell, J. (1991). The detection of clusters in rare diseases. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 143-155.
- Bissonnette, L., K. Wilson, S. Bell & T. I. Shah (2012) Neighbourhoods and potential access to health care: The role of spatial and aspatial factors. *Health & place*, 18, 841-853.
- Black, J. L. & J. Macinko (2008) Neighborhoods and obesity. *Nutrition reviews*, 66, 2-20.
- Bolden, L. & M. N. Wicks (2005) Length of stay, admission types, psychiatric diagnoses, and the implications of stigma in African Americans in the nationwide inpatient sample. *Issues in Mental Health Nursing*, 26, 1043-1059.
- Brandeau, M. L., F. Sainfort & W. P. Pierskalla. 2004. *Operations research and health care: a handbook of methods and applications*. Springer.

- Bronfenbrenner, U. 1979. *The Ecology of Human Development*. Cambridge, MA: Harvard University Press.
- Brown, E. R., R. Wyn & S. Teleki (2000) Disparities in health insurance and access to care for residents across US cities. *UCLA Center for Health Policy Research*.
- Brunsdon, C., A. S. Fotheringham & M. E. Charlton (1996) Geographically weighted regression: a method for exploring spatial nonstationarity. *Geographical analysis*, 28, 281-298.
- Buchmueller, T. C., M. Jacobson & C. Wold (2006) How far to the hospital? The effect of hospital closures on access to care. *Journal of health economics*, 25, 740-761.
- Bullen, N., G. Moon & K. Jones (1996) Defining localities for health planning: a GIS approach. *Social Science & Medicine*, 42, 801-816.
- Burns, L. R. & D. R. Wholey (1991) The effects of patient, hospital, and physician characteristics on length of stay and mortality. *Medical care*, 29, 251-271.
- Busingye, D., A. Pedigo & A. Odoi (2011) Temporal changes in geographic disparities in access to emergency heart attack and stroke care: Are we any better today? *Spatial and spatio-temporal epidemiology*, 2, 247-263.
- Campbell, N., A. Elliott, L. Sharp, L. Ritchie, J. Cassidy & J. Little (2000) Rural factors and survival from cancer: analysis of Scottish cancer registrations. *British Journal of Cancer*, 82, 1863-1866.
- Carter-Pokras, O. & C. Baquet (2002) What is a "health disparity"? *Public health reports*, 117, 426.
- Cervigni, F., Y. Suzuki, T. Ishii & A. Hata (2008) Spatial accessibility to pediatric services. *Journal of community health*, 33, 444-448.
- Cheng, Y., J. Wang & M. W. Rosenberg (2012) Spatial access to residential care resources in Beijing, China. *Int J Health Geogr*, 11, 32.
- Chung, I. (2011) Sociocultural study of immigrant suicide attempters: An ecological perspective. *Journal of Social Work*.
- Church, R. L. (1999) Location modelling and GIS. *Geographical information systems*, 1, 293-303.
- Church, R. L. & J. L. Cohon. 1976. *Multiobjective location analysis of regional energy facility siting problems*. Brookhaven National Lab., Upton, NY (USA).
- COGME. 1998. *Physician distribution and health care challenges in rural and inner-city areas*. ed. C. o. G. M. Education. Washington, DC: US Department of Health and Human Services, Health Resources and Services Administration.
- Cohen, C. I. & G. J. Casimir (1989) Factors associated with increased hospital stay by elderly psychiatric patients. *Psychiatric Services*, 40, 741-743.
- Continelli, T., S. McGinnis & T. Holmes (2010) The effect of local primary care physician supply on the utilization of preventive health services in the United States. *Health & place*, 16, 942-951.

- Cordasco, K., N. Ponce, M. Gatchell, B. Traudt & J. Escarce (2011) English Language Proficiency and Geographical Proximity to a Safety Net Clinic as a Predictor of Health Care Access. *Journal of Immigrant and Minority Health*, 13, 260-267.
- Coward, R. T. & S. J. Cutler (1989) Informal and formal health care systems for the rural elderly. *Health Services Research*, 23, 785.
- Cromley, E. K. & S. L. McLafferty. 2011. *GIS and public health*. Guilford Press.
- Culyer, A. J. & A. Wagstaff (1993) Equity and equality in health and health care. *Journal of health economics*, 12, 431-457.
- Cummins, S., S. Curtis, A. V. Diez-Roux & S. Macintyre (2007) Understanding and representing 'place' in health research: a relational approach. *Social science & medicine*, 65, 1825-1838.
- Current, J. R. & D. A. Schilling (1990) Location modeling: perspective and overview. *Geographical Analysis*, 22, 1-3.
- Curtis, A. J., J. W. Mills & M. Leitner (2006) Spatial confidentiality and GIS: re-engineering mortality locations from published maps about Hurricane Katrina. *International Journal of Health Geographics*, 5, 44.
- Dai, D. (2010) Black residential segregation, disparities in spatial access to health care facilities, and late-stage breast cancer diagnosis in metropolitan Detroit. *Health & Place*, 16, 1038-1052.
- Daskin, M. S. 2011. *Network and discrete location: models, algorithms, and applications*. John Wiley & Sons.
- Daskin, M. S. & L. K. Dean. 2004. *Location of health care facilities*. In *Operations research and health care*, 43-76. Springer.
- Daultrey, H., C. Gooday & K. Dhatariya (2011) Increased length of inpatient stay and poor clinical coding: audit of patients with diabetes. *JRSM Short Reports*, 2.
- De Marco, A. & M. De Marco (2010) Conceptualization and measurement of the neighborhood in rural settings: a systematic review of the literature. *Journal of Community Psychology*, 38, 99-114.
- Diez Roux, A. V. (2001) Investigating neighborhood and area effects on health. *American journal of public health*, 91, 1783-1789.
- Duncan, C., K. Jones & G. Moon (1999) Smoking and deprivation: are there neighbourhood effects? *Social Science & Medicine*, 48, 497-505.
- Ellison, A. M. & H. Bauchner (2007) Socioeconomic status and length of hospital stay in children with vaso-occlusive crises of sickle cell disease. *Journal of the National Medical Association*, 99, 192.
- Epstein, A. M., R. S. Stern, J. Tognetti, C. B. Begg, R. M. Hartley, E. Cumella Jr & J. Z. Ayanian (1988) The association of patients' socioeconomic characteristics with the length of hospital stay and hospital charges within diagnosis-related groups. *The New England journal of medicine*, 318, 1579-1585.
- Erkut, E., C. ReVelle & Y. Ülküsal (1996) Integer-friendly formulations for the r-separation problem. *European Journal of Operational Research*, 92, 342-351.

- Ferrante, J. M., E. C. Gonzalez, N. Pal & R. G. Roetzheim (2000) Effects of physician supply on early detection of breast cancer. *J Am Board Fam Pract*, 13, 408-14.
- Field, K. (2000) Measuring the need for primary health care: an index of relative disadvantage. *Applied Geography*, 20, 305-332.
- Fields, D., E. Leshen & K. Patel (2010) Analysis & commentary driving quality gains and cost savings through adoption of medical homes. *Health Affairs*, 29, 819-826.
- Fotheringham, A. S., C. Brunsdon & M. Charlton. 2003. *Geographically weighted regression: the analysis of spatially varying relationships*. John Wiley & Sons.
- Fu, A. Z., J. Z. Jiang, J. H. Reeves, J. E. Fincham, G. G. Liu & M. I. Perri (2007) Potentially Inappropriate Medication Use and Healthcare Expenditures in the US Community-Dwelling Elderly. *Medical Care*, 45, 472-476 10.1097/01.mlr.0000254571.05722.34.
- Gandsas, A., M. Parekh, M. M. Bleech & D. A. Tong (2007) Robotic Telepresence: Profit Analysis in Reducing Length of Stay after Laparoscopic Gastric Bypass. *Journal of the American College of Surgeons*, 205, 72-77.
- Gebreab, S. Y. & A. V. D. Roux (2012) Exploring racial disparities in CHD mortality between blacks and whites across the United States: A geographically weighted regression approach. *Health & place*, 18, 1006-1014.
- Gesler, W. & D. Albert (2000) How spatial analysis can be used in medical geography. *Spatial analysis, GIS and remote sensing applications in the health sciences*, 11-38.
- Getis, A. & J. Aldstadt. 2010. *Constructing the spatial weights matrix using a local statistic*. In *Perspectives on spatial data analysis*, 147-163. Springer.
- Gifford, E. & E. M. Foster (2008) Provider-level effects on psychiatric inpatient length of stay for youth with mental health and substance abuse disorders. *Medical Care*, 46, 240-246.
- Greene, R. P. & J. B. Pick. 2011. *Exploring the urban community: A GIS approach*. Pearson Prentice Hall.
- Gu, W., X. Wang & S. E. McGregor (2010) Optimization of preventive health care facility locations. *International Journal of Health Geographics*, 9, 17.
- Guagliardo, M. F. (2004) Spatial accessibility of primary care: concepts, methods and challenges. *International journal of health geographics*, 3, 3.
- Guagliardo, M. F., C. R. Ronzio, I. Cheung, E. Chacko & J. G. Joseph (2004) Physician accessibility: an urban case study of pediatric providers. *Health & place*, 10, 273-283.
- Hare, T. S. & H. R. Barcus (2007) Geographical accessibility and Kentucky's heart-related hospital services. *Applied Geography*, 27, 181-205.
- Hawthorne, T. L. & M.-P. Kwan (2013) Exploring the unequal landscapes of healthcare accessibility in lower-income urban neighborhoods through qualitative inquiry. *Geoforum*, 50, 97-106.

- HCUP. 2005. *HCUP Facts and Figures, Statistics on Hospital-Based Care in the United States*. ed. H. C. a. U. Project. Agency for Healthcare Research and Quality: U.S. Department of Health & Human Services.
- Herrle, G. (2006) Reducing inpatient length of stay: the time has come to revisit this discarded strategy. *Milliman: Consultant's Corner*.
- Higgs, G. (2004) A literature review of the use of GIS-based measures of access to health care services. *Health Services and Outcomes Research Methodology*, 5, 119-139.
- Hope, A. C. (1968) A simplified Monte Carlo significance test procedure. *Journal of the Royal Statistical Society. Series B (Methodological)*, 582-598.
- Hox, J. J. 2002. *Multilevel analysis: Techniques and applications*. Psychology Press.
- Huff, D. L. (1963) A probabilistic analysis of shopping center trade areas. *Land economics*, 81-90.
- Huff, D. L. (1964) Defining and estimating a trading area. *The Journal of Marketing*, 34-38.
- Huff, D. L. (2000) Don't misuse the Huff model in GIS. *Business Geographies*, 8, 12.
- Hwang, H. L., & Rollow, J. (2000). *Data processing procedures and methodology for estimating trip distances for the 1995 American Travel Survey (ATS) (No. ORNL/TM-2000/141)*. Oak Ridge National Lab., TN (US).
- John-Baptiste, A., G. Naglie, G. Tomlinson, S. M. Alibhai, E. Etchells, A. Cheung, M. Kapral, W. L. Gold, H. Abrams & M. Bacchus (2004) The Effect of English Language Proficiency on Length of Stay and In-hospital Mortality. *Journal of general internal medicine*, 19, 221-228.
- Lee, R. C. (1991). Current Approaches to Shortage Area Designation. *The Journal of rural health*, 7(s4), 437-450.
- Joseph, A. E. & P. R. Bantock (1982) Measuring potential physical accessibility to general practitioners in rural areas: a method and case study. *Social science & medicine*, 16, 85-90.
- Joseph, A. E. & D. R. Phillips. 1984. *Accessibility and utilization: geographical perspectives on health care delivery*. Sage.
- Katz, V. S., A. Ang & R. Suro (2012) An Ecological Perspective on U.S. Latinos' Health Communication Behaviors, Access, and Outcomes. *Hispanic Journal of Behavioral Sciences*, 34, 437-456.
- Kawachi, I. & B. P. Kennedy (1999) Income inequality and health: pathways and mechanisms. *Health services research*, 34, 215.
- Kirby, J. B. & T. Kaneda (2005) Neighborhood socioeconomic disadvantage and access to health care. *Journal of Health and Social Behavior*, 46, 15-31.
- Kirby, J. B. & T. Kaneda (2006) Access to Health Care: Does Neighborhood Residential Instability Matter? *Journal of Health and Social Behavior*, 47, 142-155.
- Krieger, N., J. T. Chen, P. D. Waterman, M.-J. Soobader, S. Subramanian & R. Carson (2002) Geocoding and monitoring of US socioeconomic inequalities in mortality and cancer incidence:

- does the choice of area-based measure and geographic level matter? The Public Health Disparities Geocoding Project. *American journal of epidemiology*, 156, 471-482.
- Kuby, M. J. (1987) Programming Models for Facility Dispersion: The p-Dispersion and Maximum Dispersion Problems. *Geographical Analysis*, 19, 315-329.
- Kulldorff, M. & Nargarwalla, N. (1995). Spatial disease surveillance using a scan statistic. *Journal of the Royal Statistical Society A*, 164, 61-72.
- Kurland, K. S. & W. L. Gorr. 2012. *GIS Tutorial for Health*. Esri Press.
- Kwag, K. H., Y. Jang, S. H. Rhew & D. A. Chiriboga (2011) Neighborhood effects on physical and mental health: A study of Korean American older adults. *Asian American Journal of Psychology*, 2, 91.
- Larsen, K. & J. Merlo (2005) Appropriate assessment of neighborhood effects on individual health: integrating random and fixed effects in multilevel logistic regression. *American journal of epidemiology*, 161, 81-88.
- Lee, R. C. (1991). Current Approaches to Shortage Area Designation. *The Journal of rural health*, 7(s4), 437-450.
- Lee, S., A. B. Rothbard & E. L. Noll (2012) Length of inpatient stay of persons with serious mental illness: Effects of hospital and regional characteristics. *Psychiatric Services*, 63, 889-895.
- Lewis, D. J. & P. A. Longley (2012) Patterns of patient registration with primary health care in the UK National Health Service. *Annals of the Association of American Geographers*, 102, 1135-1145.
- Li, F., K. J. Fisher, R. C. Brownson & M. Bosworth (2005) Multilevel modelling of built environment characteristics related to neighbourhood walking activity in older adults. *Journal of epidemiology and community health*, 59, 558-564.
- Lo, L. & L. Wang. 2005. *A GIS approach to examining the delivery of immigrant settlement services in Toronto*. In Annual Meetings of the Association of American Geographers, Denver, USA, April.
- Lovett, A., G. Sünnerberg & R. Haynes (2004) Using GIS to Assess Accessibility to Primary Healthcare Services. *GIS in public health practice*, 187.
- Luo, W. (2004) Using a GIS-based floating catchment method to assess areas with shortage of physicians. *Health & place*, 10, 1-11.
- Luo, W. & Y. Qi (2009) An enhanced two-step floating catchment area (E2SFCA) method for measuring spatial accessibility to primary care physicians. *Health & place*, 15, 1100-1107.
- Luo, W. & F. Wang (2003) Measures of spatial accessibility to health care in a GIS environment: synthesis and a case study in the Chicago region. *Environment and Planning B*, 30, 865-884.
- Luo, W. & F. Wang (2003) Spatial accessibility to primary care and physician shortage area designation: a case study in Illinois with GIS approaches. *Geographic Information Systems and Health Applications*, 260-278.

- Macintyre, S., A. Ellaway & S. Cummins (2002) Place effects on health: how can we conceptualise, operationalise and measure them? *Social science & medicine*, 55, 125-139.
- Maheswaran, R. & R. Haining. 2004. *Basic issues in Geographical Analysis*. IN R. Maheswaran and M. Craglia (Eds.) GIS in Public Health Practice. CRC Press, Boca Raton.
- Martin, S. & P. Smith (1996) Explaining variations in inpatient length of stay in the National Health Service. *Journal of Health Economics*, 15, 279-304.
- McLeroy, K. R., D. Bibeau, A. Steckler & K. Glanz (1988) An Ecological Perspective on Health Promotion Programs. *Health Education & Behavior*, 15, 351-377.
- Meade, M. S. 1988. *Medical geography*. Wiley Online Library.
- Meliker, J. R., P. Goovaerts, G. M. Jacquez, G. A. AvRuskin & G. Copeland (2009) Breast and prostate cancer survival in Michigan. *Cancer*, 115, 2212-2221.
- Meliker, J. R., G. M. Jacquez, P. Goovaerts, G. Copeland & M. Yassine (2009) Spatial cluster analysis of early stage breast cancer: a method for public health practice using cancer registry data. *Cancer Causes & Control*, 20, 1061-1069.
- Mennis, J., G. J. Stahler & D. A. Baron (2012) Geographic barriers to community-based psychiatric treatment for drug-dependent patients. *Annals of the Association of American Geographers*, 102, 1093-1103.
- Meyer, S. P. (2012) Comparing spatial accessibility to conventional medicine and complementary and alternative medicine in Ontario, Canada. *Health & place*, 18, 305-314.
- Monnet, E., C. Ramée, A. Minello, V. Jooste, D. Carel & V. Di Martino (2008) Socioeconomic context, distance to primary care and detection of hepatitis C: a French population-based study. *Social Science & Medicine*, 66, 1046-1056.
- Moon, G. & R. Gillespie. 1995. Society and health: an introduction to social science for health professionals. *Psychology Press*.
- Moon, I. D. & S. S. Chaudhry (1984) An analysis of network location problems with distance constraints. *Management Science*, 30, 290-307.
- Murray, A. T. (2005) Geography in coverage modeling: exploiting spatial structure to address complementary partial service of areas. *Annals of the Association of American Geographers*, 95, 761-772.
- Murray, A. T. & R. L. Church (1996) Analyzing cliques for imposing adjacency restrictions in forest models. *Forest Science*, 42, 166-175.
- Nutting, P. A., B. F. Crabtree, W. L. Miller, K. C. Stange, E. Stewart & C. Jaén (2011) Transforming physician practices to patient-centered medical homes: lessons from the national demonstration project. *Health affairs*, 30, 439-445.
- Nuwer, M. R., Esper, G. J., Donofrio, P. D., Szaflarski, J. P., Barkley, G. L., & Swift, T. R. (2008). Invited Article: The US health care system Part 1: Our current system. *Neurology*, 71(23), 1907-1913.

- Nuwer, M. R., Barkley, G. L., Esper, G. J., Donofrio, P. D., Szaflarski, J. P., & Swift, T. R. (2008). Invited Article: The US health care system Part 2: Proposals for improvement and comparison to other systems. *Neurology*, 71(23), 1914-1920.
- Oliver, A. & E. Mossialos (2004) Equity of access to health care: outlining the foundations for action. *Journal of epidemiology and community health*, 58, 655-658.
- Onega, T., E. J. Duell, X. Shi, D. Wang, E. Demidenko & D. Goodman (2008) Geographic access to cancer care in the US. *Cancer*, 112, 909-918.
- Openshaw, S. 1983. *The modifiable areal unit problem*. Geo Books Norwich.
- Openshaw, S. & S. Openshaw. 1984. *The modifiable areal unit problem*. Geo Abstracts University of East Anglia.
- Page, A. C., G. R. Hooke & J. Rampono (2005) A Methodology for Timing Reviews of Inpatient Hospital Stay. *Australian and New Zealand Journal of Psychiatry*, 39, 198-201.
- Patel, K. V., K. Eschbach, L. L. Rudkin, M. K. Peek & K. S. Markides (2003) Neighborhood context and self-rated health in older Mexican Americans. *Annals of epidemiology*, 13, 620-628.
- Paul, M. J., & Dredze, M. (2011). You are what you Tweet: Analyzing Twitter for public health. In *ICWSM* . 265-272.
- Pearce, J., K. Witten & P. Bartie (2006) Neighbourhoods and health: a GIS approach to measuring community resource accessibility. *Journal of epidemiology and community health*, 60, 389-395.
- Pedigo, A. S. & A. Odoi (2010) Investigation of disparities in geographic accessibility to emergency stroke and myocardial infarction care in East Tennessee using geographic information systems and network analysis. *Annals of epidemiology*, 20, 924-930.
- Penchansky, R. & J. W. Thomas (1981) The concept of access: definition and relationship to consumer satisfaction. *Medical care*, 19, 127-140.
- Penhollow, T. M. & K. E. Rhoads (2014) Preventing Obesity and Promoting Fitness: An Ecological Perspective. *American Journal of Lifestyle Medicine*, 8, 21-24.
- Perry, B. & W. Gesler (2000) Physical access to primary health care in Andean Bolivia. *Social Science & Medicine*, 50, 1177-1188.
- Roetzheim, R. G., E. C. Gonzalez, A. Ramirez, R. Campbell & D. J. Van Durme (2001) Primary care physician supply and colorectal cancer. *Journal of Family Practice*, 50, 1027-1031.
- Rollow, J. & H. Hwang. 2000. Data processing procedures and methodology for estimating trip distances for the 1995 American Travel Survey (ATS).
- Rosenberg, M. W. & N. T. Hanlon (1996) Access and utilization: a continuum of health service environments. *Social Science & Medicine*, 43, 975-983.
- Rosenblatt, R. A. & D. M. Lishner (1991) Surplus or shortage? Unraveling the physician supply conundrum. *Western journal of medicine*, 154, 43.
- Ross, C. E. (2000) Walking, exercising, and smoking: does neighborhood matter? *Social Science & Medicine*, 51, 265-274.

- Roux, A. V. D., S. S. Merkin, D. Arnett, L. Chambless, M. Massing, F. J. Nieto, P. Sorlie, M. Szklo, H. A. Tyroler & R. L. Watson (2001) Neighborhood of residence and incidence of coronary heart disease. *New England Journal of Medicine*, 345, 99-106.
- Sanders, G. F., M. A. Fitzgerald & M. Bratteli (2008) Mental Health Services for Older Adults in Rural Areas: An Ecological Systems Approach. *Journal of Applied Gerontology*, 27, 252-266.
- Shafi, S., C. M. de la Plata, R. Diaz-Arrastia, K. Shipman, M. Carlile, H. Frankel, J. Parks & L. M. Gentilello (2007) Racial disparities in long-term functional outcome after traumatic brain injury. *Journal of Trauma-Injury, Infection, and Critical Care*, 63, 1263-1270.
- Shannon, G. W. (1980). Editorial comment. Social Science & Medicine. Part D: *Medical Geography*, 14(1), 1-2.
- Shen, Q. (1998) Location characteristics of inner-city neighborhoods and employment accessibility of low-wage workers. *Environment and planning B: Planning and Design*, 25, 345-365.
- Shi, L., B. Starfield, B. Kennedy & I. Kawachi (1999) Income inequality, primary care, and health indicators. *The Journal of family practice*, 48, 275-284.
- Shi, L., B. Starfield, R. Politzer & J. Regan (2002) Primary Care, Self-rated Health, and Reductions in Social Disparities in Health. *Health Services Research*, 37, 529-550.
- Shi, X., J. Alford-Teaster, T. Onega & D. Wang (2012) Spatial access and local demand for major cancer care facilities in the United States. *Annals of the Association of American Geographers*, 102, 1125-1134.
- Shier, D. (1977) A min-max theorem for p-center problems on a tree. *Transportation Science*, 11, 243-252.
- Smith, C. J. 1977. Geography and mental health. Association of American Geographers Washington, DC.
- Starfield, B. (1994) Is primary care essential? *The Lancet*, 344, 1129-1133.
- Stetzer, F. (1982) Specifying weights in spatial forecasting models: the results of some experiments. *Environment and Planning A*, 14, 571-584.
- Taaffe, D. R., Pruitt, L., Pyka, G., Guido, D., & Marcus, R. (1996). Comparative effects of high- and low-intensity resistance training on thigh muscle strength, fiber area, and tissue composition in elderly women. *Clinical Physiology*, 16(4), 381-392.
- Tassone, E. C., L. A. Waller & M. L. Casper (2009) Small-area racial disparity in stroke mortality: an application of Bayesian spatial hierarchical modeling. *Epidemiology (Cambridge, Mass.)*, 20, 234.
- Thompson, E. E., H. W. Neighbors, C. Munday & S. Trierweiler (2003) Length of stay, referral to aftercare, and rehospitalization among psychiatric inpatients. *Psychiatric services*, 54, 1271-1276.
- Waites, C. (2013) Examining the Perceptions, Preferences, and Practices That Influence Healthy Aging for African American Older Adults: An Ecological Perspective. *Journal of Applied Gerontology*, 32, 855-875.

- Waller, L. A., & Gotway, C. A. (2004). *Applied spatial statistics for public health data* (Vol. 368). John Wiley & Sons.
- Walsh, C. M. & J. S. Coldiron (1993) Enhanced Length of Stay Management through Monitoring of Discharge Planning Parameters. *American Journal of Medical Quality*, 8, 128-133.
- Wan, N., F. B. Zhan, B. Zou & J. G. Wilson (2013) Spatial access to health care services and disparities in colorectal cancer stage at diagnosis in Texas. *The Professional Geographer*, 65, 527-541.
- Wang, F. 2006. *Quantitative methods and applications in GIS*. CRC Press.
- Wang, F. (2012) Measurement, optimization, and impact of health care accessibility: a methodological review. *Annals of the Association of American Geographers*, 102, 1104-1112.
- Wang, F. & W. Luo (2005) Assessing spatial and nonspatial factors for healthcare access: towards an integrated approach to defining health professional shortage areas. *Health & place*, 11, 131-146.
- Wang, F., S. McLafferty, V. Escamilla & L. Luo (2008) Late-stage breast cancer diagnosis and health care access in Illinois. *The Professional Geographer*, 60, 54-69.
- Wang, L. (2007) Immigration, ethnicity, and accessibility to culturally diverse family physicians. *Health & place*, 13, 656-671.
- Wang, L., M. Rosenberg & L. Lo (2008) Ethnicity and utilization of family physicians: A case study of Mainland Chinese immigrants in Toronto, Canada. *Social science & medicine*, 67, 1410-1422.
- Wei, R. (2015) Coverage Location Models Alternatives, Approximation, and Uncertainty. *International Regional Science Review*, 0160017615571588.
- White, J. M., & Klein, D. M. 2002. *Family Theories (2nd ed.)*. Thousand Oaks, CA: Sage.
- Wilper, A. P., Woolhandler, S., Lasser, K. E., McCormick, D., Bor, D. H., & Himmelstein, D. U. (2009). Health insurance and mortality in US adults. *American Journal of Public Health*, 99(12), 2289.
- World Health Organization. 2000. *The world health report 2000: health systems: improving performance*. World Health Organization.
- Yang, D.-H., R. Goerge & R. Mullner (2006) Comparing GIS-based methods of measuring spatial accessibility to health services. *Journal of Medical Systems*, 30, 23-32.
- Ye, H. & H. Kim (2014) Measuring Spatial Health Disparity Using a Network-Based Accessibility Index Method in a GIS Environment: A Case Study of Hillsborough County, Florida. *International Journal of Geospatial and Environmental Research*, 1, 2.
- Ye, H. & H. Kim (2015) Analyzing the effect of neighborhood characteristics on the spatial pattern of primary care physician locations in Hillsborough County, Florida. *Papers in Applied Geography*, 37, 86-94.
- Zhang, W. & V. M. Ta (2009) Social connections, immigration-related factors, and self-rated physical and mental health among Asian Americans. *Social Science & Medicine*, 68, 2104-2112.

APPENDIX

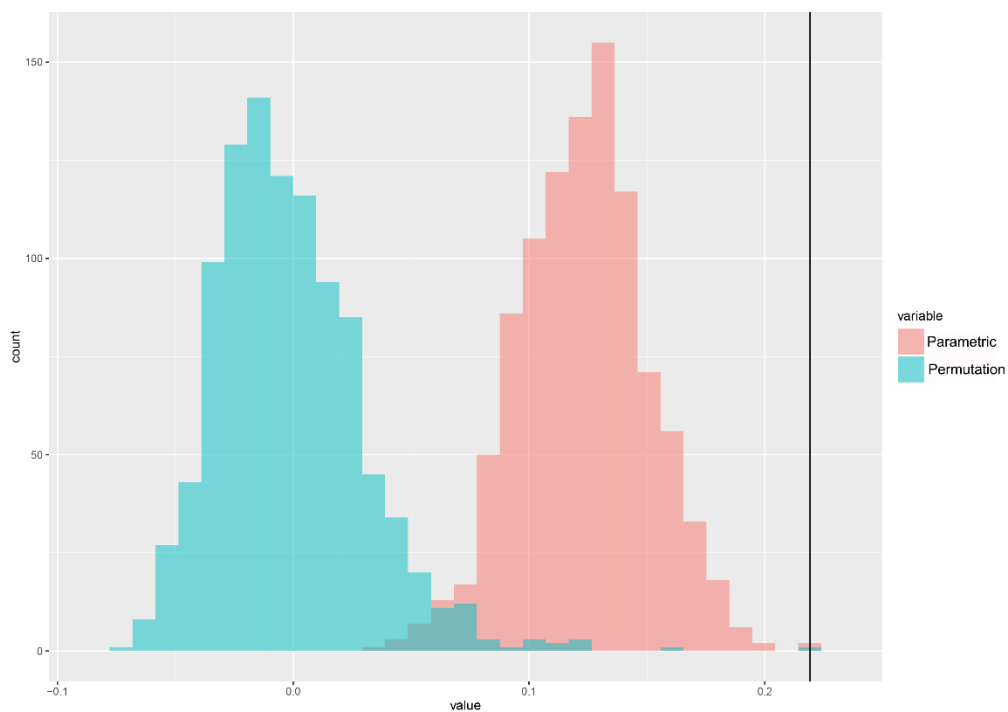
Appendix1. Test results of heterogeneity of the relative PCP rate

	Type of boots	Model	Simulations	<i>p-value</i>
Chi-square test	Parametric	Poisson	999	0.001
Potthoff-Whittinghill's test	Parametric	Poisson	999	0.001

Appendix2. The results of Global Moran's I

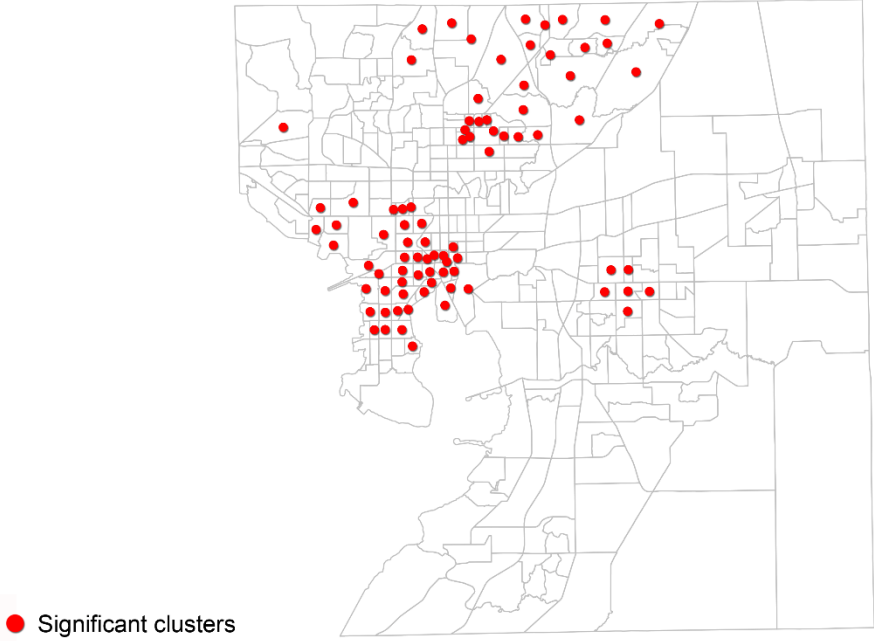
	Moran's I statistic	Simulations	p-value
Permutation bootstrap	0.219	999	< 0.001
Parametric bootstrap		999	0.002

Appendix3. Histograms of simulated values of Moran' I under random permutations of the data and parametric samples from the expected constant PCP rate. The observed values of Moran's I are marked by vertical lines. (Note that the distribution of Moran's I shifts rightwards with the parametric simulations, since the impact of population distribution is taken into consideration.)



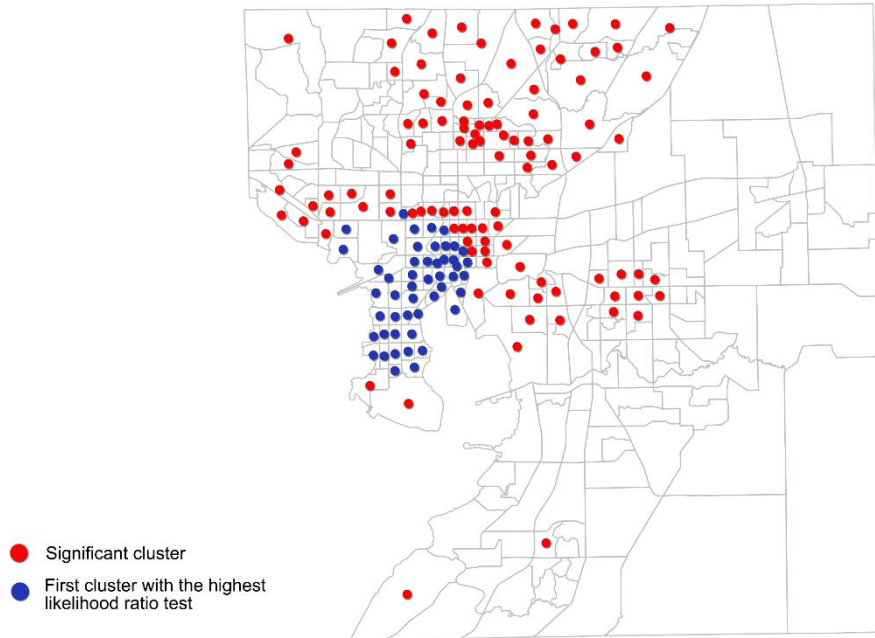
Appendix4. Results of Besag and Newell's statistic ($k = 100$). The dots represent the center of the clusters.

Besag and Newell Approach



Appendix5. Results of Kulldorff's statistic. The dots represent the center of the clusters. The blue dots show the most likely clusters.

Kulldorff and Nagarwalla's method



VITA

Huairan Ye was born and raised in Fujian, the People's Republic of China. She received her Bachelor of Architecture from Xiamen University in 2009, and then received her Master of Architecture from Zhejiang University in 2012. She came to the United States in fall 2012 to pursue her Doctor of Philosophy degree. She has maintained an active research agenda and published multiple papers in peer review journals. Her areas of research focus include health geography, urban and regional planning, Geographic Information Science and statistics.