

2015

Delineating Hospital Service Areas in Florida Based on Patients' Travel Patterns

Peng Jia

Louisiana State University and Agricultural and Mechanical College, jiapengff@hotmail.com

Follow this and additional works at: https://digitalcommons.lsu.edu/gradschool_dissertations



Part of the [Social and Behavioral Sciences Commons](#)

Recommended Citation

Jia, Peng, "Delineating Hospital Service Areas in Florida Based on Patients' Travel Patterns" (2015). *LSU Doctoral Dissertations*. 4032.
https://digitalcommons.lsu.edu/gradschool_dissertations/4032

This Dissertation is brought to you for free and open access by the Graduate School at LSU Digital Commons. It has been accepted for inclusion in LSU Doctoral Dissertations by an authorized graduate school editor of LSU Digital Commons. For more information, please contact gradetd@lsu.edu.

DELINEATING HOSPITAL SERVICE AREAS IN FLORIDA BASED ON PATIENTS'
TRAVEL PATTERNS

A Dissertation

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

in

The Department of Geography and Anthropology

by

Peng Jia

B.Eng., Nanjing Agricultural University, 2007

M.S., University of Chinese Academy of Sciences, 2010

M.S., University of Florida, 2012

December 2015

Acknowledgements

I would like to propose a toast for my respectable mentor, Dr. Fahui Wang, to express my deepest appreciation. I have been greatly benefiting from his guidance and advice both academically and personally. He always has students he mentors and teaches as his priority. Even when he was in bad health at one time, he reviewed my work at home and reminded me of all the deadlines. I feel grateful for having chances to be under his guidance, and having long enough time to understand his fatherly care and thought. His natural generosity and his great tolerance and support have provided me room to grow and explore my infinite possibilities, and made it possible to develop this amazing dissertation research project.

Furthermore, I would like to thank my committee members, Dr. Robert V. Rohli, for his generous care and support for more than four years, and Dr. Imam M. Xierali, for his understanding and guidance. I feel fortunate to develop great trust and friendship with them through working together. The success of this work also includes the contribution from Dr. Eugene Kennedy in the School of Education.

Special thanks to my best classmate, colleague and friend, Dr. Timothy Andrew Joyner, the assistant professor in the Department of Geosciences at East Tennessee State University. He has been providing me countless advice, help, and support from multiple aspects, just like his advisor (Dr. Robert V. Rohli) does. I am extremely happy to see him live a wonderful life with his wife, their adorable daughter, and new coming baby boy!

Rome wasn't built in one day. I wish to extend my sincere appreciation to my previous mentors, Dr. Youliang Qiu, Dr. Andrew Tatem at the University of Southampton, Dr. Yueping Nie at the Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, and Dr.

Corene J. Matyas at the University of Florida. Without their kind help and support, I would not have reached this point.

Thank you also to Mr. Luke Driskell for IT support, and Ms. Dana C. Sanders, Ms. Elizabeth Honeycutt and Ms. Linda Strain, who provided me much advice and ease during my three years' study.

Last but not least, I would like to show my biggest appreciation to my parents, Xiu Li and Hongwei Jia, who have been teaching me how to be a decent and generous person, which ultimately leads me to find my kindly and caring fiancé, Xinyu Shi, and selfless love and unconditional support from her patients, Rongying Zhang, and Huiwen Shi. I feel life is always full of goodness whenever

Limited space keeps me from listing all other relatives and friends I want to thank here, especially my drinking buddies that make me enjoy being a human being. Given my straightforwardness, I am 100% sure you all know exactly how important you are to me. I wish all the best to you all.

Table of Contents

Acknowledgements.....	ii
List of Tables	vi
List of Figures.....	vii
Abstract.....	ix
Chapter 1 Introduction	1
Chapter 2 Literature Review.....	5
2.1 Central place theory and hierarchical health care systems.....	5
2.2 Hospital service areas (HSAs) and hospital referral regions (HRRs).....	10
2.3 Delineation of other functional areas	14
2.4 Distance decay in health care studies.....	17
2.5 Summary	20
Chapter 3 Study Area and Data	22
Chapter 4 Evaluating the Dartmouth HSAs.....	26
4.1 Localization index	26
4.2 Effectiveness of the Dartmouth HSAs	28
4.3 Effectiveness of the Dartmouth HRRs	31
4.4 Summary	35
Chapter 5 Analysis of Patients' Distance Decay Behavior in Hospital Visits.....	37
5.1 Distance decay effects in hospital visits.....	37
5.2 Comparison of distance decay functions.....	39
5.3 Travel patterns of subpopulations	42
5.4 Summary	48
Chapter 6 Delineating the HSAs by the Huff Model.....	50
6.1 The Huff model approach	50
6.2 Discussion on Huff-based HSAs.....	51
6.3 Travel time within the HSAs.....	56
6.4 Comparison of the HSAs.....	58
6.5 Summary	61
Chapter 7 Constructing Hierarchical HSAs.....	68
7.1 Travel patterns of patients for specialized cares	68
7.2 Huff-based HRRs and hierarchical HSAs.....	73
7.3 Summary	78
Chapter 8 Conclusions	81

References.....	85
Vita.....	91

List of Tables

Table 1. Descriptive statistics of natural log-transformed localization index of the 1992-1993 and 2011 Medicare-derived HSAs.....	28
Table 2. Results of <i>t</i> -test for the difference of localization index of the 1992-1993 and 2011 Medicare-derived HSAs.....	29
Table 3. Descriptive statistics of natural log-transformed localization index of the 2011 Medicare-derived and overall-derived HSAs	31
Table 4. Results of <i>t</i> -test for the difference of localization index of the 2011 Medicare-derived and overall-derived HSAs.....	31
Table 5. Optimal parameters and assessment of the four candidate functions for modeling the interactions from postal zones to hospitals	37
Table 6. Parameters and assessment of the log-logistic functions modeling the travel patterns of subpopulations (Unit: minutes for Time_all and Time_60)	45
Table 7. Candidate influential factors for patients' average travel time within the HSAs	57
Table 8. Coefficients and significance of the candidate influential factors for patients' average travel time within the HSAs.....	58
Table 9. Comparisons of the mean values of different parameters within the Huff-based HSAs and 2011 overall-derived HSAs (minimum value — maximum value).....	59
Table 10. Results of <i>t</i> -test for the differences of the localization index (LI) values of the 2011 overall-derived and Huff-based HSAs.....	60
Table 11. Optimal parameters and assessment of the four candidate functions based on the specialized (cardiovascular and neurological), cardiovascular (Cardio), neurological (Neuro), and overall inpatients (All)	70
Table 12. Optimal parameters and assessment of the four candidate functions for modeling the interactions between the HRRs (from origin HRRs to destination HRRs).....	76

List of Figures

Figure 1. Distribution of the included hospitals with the number of beds represented by the size of circles.....	25
Figure 2. Boundaries of the 1992-1993 and 2011 Medicare-derived HSAs.....	29
Figure 3. Boundaries of the 2011 Medicare-derived and overall-derived HSAs	32
Figure 4. Boundaries of the 1992-1993 Medicare-derived and 2011 overall-derived HRRs.....	34
Figure 5. A flowchart of generating the 2011 overall-derived HRRs.....	35
Figure 6. Curves of the four distance decay functions with the optimal sets of parameters	40
Figure 7. Comparison of the <i>pseudo-R</i> ² and Akaike information criterion (AIC) of four distance decay functions	41
Figure 8. Travel patterns of the patients across racial/ethnic subgroups	46
Figure 9. Travel patterns of the patients across socioeconomic subgroups	47
Figure 10. Travel patterns of the patients across socioeconomic subgroups.....	48
Figure 11. A flowchart of generating the Huff-based HSAs	52
Figure 12. Boundaries of the Huff-based HSAs and counties in Florida	53
Figure 13. The number of the population within the Huff-based HSAs.....	54
Figure 14. The ratio of hospital beds to population within the Huff-based HSAs	55
Figure 15. The average travel time of the patients to hospitals within the Huff-based HSAs	56
Figure 16. Comparison of the boundaries between the 2011 Huff-based and overall-derived HSAs in Florida	62
Figure 17. Comparisons of proportion and cumulative probability of the localization index (LI) values between the 2011 Huff-based and overall-derived HSAs in Florida.....	63
Figure 18. Comparison of the <i>pseudo-R</i> ² and Akaike information criterion (AIC) of four distance decay functions for specialized patients	71
Figure 19. Comparison of the travel patterns of the specialized and overall inpatients by log-logistic function	72

Figure 20. A flowchart of generating the Huff-based HRRs	74
Figure 21. Population within the preliminary HRRs	75
Figure 22. Huff-based hierarchical HSAs-HRRs.....	78
Figure 23. Boundaries of the 2011 overall-derived HRRs and Huff-based HRRs.....	79
Figure 24. Population within the Huff-based HRRs.....	80

Abstract

Longer travel times to hospitals and other medical resources have been shown to be associated with decreasing health outcome and increasing mortality risk. Many countries have actively responded to the call of the World Health Organization (WHO) on creating local environments for the provision of and residents' access to health resources. The hospital service areas (HSAs) and hospital referral regions (HRRs) have been proposed as more proper functional units for analysis of performance of healthcare market. However, the widely used Dartmouth HSAs and HRRs were produced solely based on two-decade-old Medicare hospitalization records. In this study, the effectiveness of the Dartmouth HSAs and HRRs are first evaluated. Based on the recent overall hospital discharge data, this study then examines the travel patterns for the overall population as well as subgroups (e.g., by race/ethnicity, socioeconomic status, and urbanicity) to identify the best-fitting distance decay function after comparing different models. The study goes on to delineate the new HSAs by the Huff model using the defined distance decay function. Finally, the travel patterns of patients for more specialized services are further modeled, and the results are utilized to define the new HRRs, which are further adjusted so that the HSAs are nested within the HRRs. Built upon the travel patterns of inpatient visits by different groups, the study enhances the theoretical foundation for defining an integrated and hierarchical system of HSAs-HRRs, and provides a scientific sound and easy-to-replicate research framework for delineation of HSAs that can benefit a variety of health care studies.

Keywords: hospital service areas (HSAs), hospital referral regions (HRRs), distance decay function, Huff model, hierarchical system of HSAs-HRRs, Florida

Chapter 1 Introduction

Longer travel times to hospitals and other medical resources have been shown to be associated with decreasing health outcome and increasing mortality risk, especially when considering emergency services (Nicholl et al. 2007). For example, a longer travel time to the dialysis unit was negatively associated with patients' health-related quality of life (Moist et al. 2008). Likewise, increasing travel time to the nearest hospitals providing surgery, chemotherapy, or radiotherapy for cancer patients made patients less likely to receive those services (Jones et al. 2008).

Many countries, including the United States, New Zealand, Hungary, Canada, Australia, Ireland, and the United Kingdom, have actively responded to the call of the World Health Organization (WHO) on creating local environments for the provision of and residents' access to health resources (Goodman et al. 1997; Shortt & Moore 2006). Of these nations, the United States hosts the most expensive health care system worldwide with annual health care expenditures at \$8,508 per capita (United Press International 2013). The analysis of local hospital use and resource allocation may result in approaches for reducing costs and expenditures related to long travel distance for hospitalization (Basu & Friedman 2007). In addition to the retention of revenue (Basu & Cooper 2000), a more localized hospital atmosphere can promote a favorable patient-doctor relationship (Wilbush 1974).

The most common spatial units for reporting local hospital use or hospitalization rates are administrative units, such as at state and county levels. However, even within the same county, hospitals vary considerably in terms of type, size, and location, and additionally, patients often cross over state or county boundaries for hospital services. Thus, state and county boundaries

may not be the best spatial units for analyzing trends in patient-to-hospital flows and representing underlying local patterns of hospital visits (Lembcke 1952).

Alternative functional units such as hospital service areas (HSAs) and hospital referral regions (HRRs) have been proposed for analysis of local hospitalization and for studies of performance of health care market. HSAs are delineated so that most of local hospitalization occurs within the service areas (Klauss et al. 2005). The HRRs represent an aggregated level of analysis based on the HSAs, wherein patients are referred for major cardiovascular surgical procedures and for neurosurgery within their own service regions (Center for Evaluative Clinical Sciences 1999). The present HRRs and HSAs in the U.S. were created by Dartmouth almost two decades ago as part of the Dartmouth Atlas of Health Care project to avoid problems with geopolitical boundaries, such as counties and states, and to allow analysis of local hospitalization at a finer spatial resolution than at the state and county levels (Klauss et al. 2005).

However, changes to infrastructure, political regulations, and insurance policies influence the spatial patterns of the patients seeking health care (Klauss et al. 2005), so the boundaries of the HSAs may have shifted over time. Also, the currently used HSAs were produced solely based on Medicare hospitalization records that specifically target ‘legal residents aged 65 and older, and younger people with disabilities, people with end stage renal disease, and persons with Amyotrophic lateral sclerosis’ (Medicare.gov 2014). But others who acquired health care were not considered in the delineation of HSA boundaries. Thus, it is necessary to assess the appropriateness of the Medicare-derived HSAs before applying them to demographic groups other than the seniors and the people with other insurances.

Understanding the travel patterns of patients for seeking medical care at the hospitals is fundamental for health resource allocation and planning, including delineating the HSAs and

HRRs (Joseph & Phillips 1984). Patients' travel behavior and utilization of hospital resources are to a large extent influenced by the distance to the hospitals, normally decreasing with distance due to more spatial barriers (O'Neill 2004). Where a weak distance decay effect exists, patients are more likely to travel farther for hospital services. In such cases, a larger HSA/HRR is needed to contain most of the patient-to-hospital travel flows within it. On the other hand, a relatively small HSA/HRR is able to capture most of the patient-to-hospital flows where the distance decay effect is strong. However, the relationships between patients' travel behavior and delineation of HSAs/HRRs have been disregarded among the majority of the existing approaches (Center for Evaluative Clinical Sciences 1999; Klauss et al. 2005).

This gap is bridged in this study by bringing the Huff model (D. L. Huff 1964) into the delineation of the HSAs and HRRs, given that the Huff model includes a built-in distance decay function capturing the travel patterns of the customers (patients in health care settings), and has been successfully applied to delineating trade areas of various business entities (Dramowicz 2005). Furthermore, distance decay functions enable us to model travel patterns of the patients seeking general and specialized care. The distinctive travel behavior of patients for different levels of health care services provides a solid theoretical foundation for defining a hierarchical HSA-HRR system.

The State Inpatient Database (SID), assembled and edited by the Agency for Healthcare Research and Quality (AHRQ) as part of the Healthcare Cost and Utilization Project (HCUP), provides key insights into the travel patterns of patients for seeking hospital services. Three major questions are addressed in this study on Florida: 1) Have the boundaries of the Medicare-derived HSAs significantly changed over the last two decades, and do they represent the patterns of local hospitalization for the overall population? 2) Which distance decay function best

captures the travel behavior of the patients? 3) Can the Huff model help us construct a hierarchical HSA-HRR system? This study contributes to the research and public-policy-making communities by revealing the travel patterns of the patients seeking general and specialized services from large datasets, and incorporating the results into the Huff model to delineate the HSAs and HRRs.

The remainder of the dissertation is structured as follows. Chapter 2 reviews hierarchical healthcare systems, in particular the delineation of the HSAs and HRRs, and general approaches of regionalization, especially the Huff model approach and applications of distance decay functions in the field of health care. Chapter 3 introduces the study area and data in detail. Chapter 4 evaluates the effectiveness of the Dartmouth Medicare-derived HSAs and HRRs. The travel patterns of the overall patients are thoroughly examined in Chapter 5, and integrated with the Huff model for delineating the HSAs in Chapter 6. Chapter 7 specifically examines the travel patterns of the tertiary patients, which are then used to produce the new HRRs that, in conjunction with the Huff-based HSAs, form the hierarchical HSA-HRR system. Chapter 8 concludes the dissertation with major findings of the study, points out the limitations, and discusses future work.

Chapter 2 Literature Review

This chapter reviews four primary topics related to the HSAs and HRRs. The first section provides a broad overview of hierarchical healthcare systems including the theoretical linkage to the Central Place Theory. The second section reviews an array of research contexts for the HSAs and/or HRRs and the evolution of methods for their delineation. The third section covers the approaches for delineating other types of services areas. The fourth section focuses on the Huff model, which is adopted for this study. Since the Huff model utilizes a built-in distance decay function capturing the travel behavior of patients for seeking health care services, Section 2.4 also surveys the applications of distance decay functions in the field of health care studies.

2.1 Central place theory and hierarchical health care systems

The HSAs and HRRs form a two-tiered hierarchical health care system. The Central Place Theory (CPT) provides a theoretical framework for the hierarchical structure in the spatial organization of various human activities. Hence, it is necessary to review the classic CPT and some key features of a general hierarchical system before moving to hierarchical health care systems.

The CPT was originally developed by German geographer Christaller (1933). The CPT partitions a whole region as a set of hierarchical hexagonal nets/lattices, and each hexagon at a level serves as a market area of a service center at that level. Two concepts, *threshold* and *range*, are critical to help us understand market areas. Under the assumption of homogeneous population distribution, the *threshold* refers to a minimum population/area needed for a business in that market area to be sustainable, while the *range* represents a maximum distance customers are willing to travel to purchase its goods/services. In the CPT, various types of goods are differentiated from one another by *range*, or the spatial size of its market area. Under the

assumption that only one supply location in the center of each market area serves all population there equally, the goods with an identical *range* are provided in a bundle at that supply location. The congruent and regular market areas lead to two distinctive geometric features in this model. One is that market area size increases from the lowest to the highest level by a constant factor, without either absence of a particular size or existence of an intermediate size. The other is that the center of a market area at any level is also the center of the lower-level market area(s) served by that center, which allows centers to provide multiple levels of goods/services and minimizes the total number of centers required in the system (Parr 1978). The latter feature makes this theoretical model similar to the local hospitalization pattern that is desirable in many countries, where each hospital provides services only within its own catchment area, and a larger range for a particular service corresponds to a higher-level service center (hospital) that provides this unique higher-level service in addition to all other lower-level services.

Three major principles are noted by Christaller (1933). They are the marketing ($K=3$), transportation ($K=4$), and administrative principles ($K=7$), where K represents the aforementioned constant factor by which a market area size increases from any to a higher level. Each principle forms a specific arrangement of central places. In a $K=3$ system, centers at any level (except the highest level) are located at the corner of consecutive higher-level hexagons. However, lower-level centers are not on the straight line between higher-level centers, which makes the transport network less efficient. This situation is eliminated in a $K=4$ system, where centers at any level (except the highest level) are located at the middle of the edges of consecutive higher-level hexagons, as well as on the main transport routes connecting the higher-level centers. Nevertheless, from administrative perspectives, $K=3$ and $K=4$ systems are similarly inconvenient as each of most lower-level centers (any level except the highest one) is always

served by multiple consecutively higher-level hexagons. Different from those two systems, a $K=7$ system has each market area at any level (except the highest level) exclusively and completely enclosed within its consecutive higher-level hexagon. Therefore, $K=7$ is usually adopted for structuring the HSA-HRR systems so that the HSAs are completely embedded within the HRRs.

Health care systems in most countries are organized as hierarchical (Rahman & Smith 2000). For example, the U.S. has a 3-hierarchical health care system (Reutstein 1971; Schultz 1970), where three types of health facilities collectively deliver services: 1) neighborhood health centers (or reception centers), providing primary diagnosis and outpatient treatment (e.g., checkups, screening tests, and auxiliary treatment); 2) general hospitals (or community hospitals), providing general treatment (e.g., acute and chronic care, ambulatory services, psychiatric care, and preventive care); and 3) central hospitals (or large medical centers), providing general treatment, specialized services, and complex diagnosis and procedures (e.g., cardiovascular surgery and neurosurgery). Normally, patients first visit primary care physicians (PCPs) in neighborhood health centers. They may be referred to general or central hospitals for needs beyond PCPs' capabilities, and the patients going to general hospitals may be referred to central hospitals as well if specialized services are needed but not provided in general hospitals. A similar 3-hierarchical hospital system exists in England, composed of rural, district, and regional hospitals, from low to high (Llewelyn-Davies & Macaulay 1966).

Based on the relationships between various levels of hierarchies, three major types of hierarchies are suggested (Rahman & Smith 2000): 1) successively inclusive hierarchy, where providers at any level offer all their available goods/services to all locations, such as the current U.S. health care system in which health facilities at any level offer all their available medical

services to all patients visiting them; 2) locally inclusive hierarchy, where providers at any level offer all their available goods/services only to the customers in their own service areas, and offer only their highest-order goods/services to the patients in other locations, as described in Christaller's Central Place model; and 3) successively exclusive hierarchy, where providers at any level exclusively offer their own level of goods/services (no other high-/lower-level services) to all locations. Despite difficulty finding a perfect successively exclusive hierarchy in reality, the system of the U.S. government (federal, state, and local) loosely provides an example in which each level of government has its own separation of powers and checks and balances, and provides a corresponding portion of the fabric of all services. Although currently functioning as a successively inclusive hospital hierarchy, the ideal model of a health care system, as well as the ultimate goal of all relevant efforts, is a locally inclusive hospital hierarchy with all levels of services obtained from the closest hospitals providing corresponding services.

The descriptive classifications of hierarchical systems usually focus on the following aspects: 1) single-flow or multi-flow, where single-flow customers pass through all levels to the highest level, while multi-flow can be from any level to any other level; 2) referral or non-referral, where referral (non-referral) systems do (do not) promote patient referrals between levels, normally upward; 3) nested or non-nested, where nested (non-nested) higher-level facilities do (do not) provide all the services offered by lower-level facilities; 4) coherent or non-coherent, where coherent (non-coherent) customers assigned to the same lowest-level facilities are assigned to the same (different) higher-level facilities. According to those descriptions, the HSA-HRR systems should be considered as multi-flow, referral, nested, and coherent.

Although appropriate approaches had been badly needed by hospital administrators and policy-makers for optimal decision-making regarding the number, size and location of health

care facilities, the earliest study was not available in the U.S. until 1970, i.e., the first use of the CPT in health care planning (Schultz 1970). Schultz found more intensive distance decay patterns for lower order of services than higher orders of services. The objective of his model was to maximize the *per capita* net social benefit for the population served, calculated by subtracting *per capita* service and travel cost from *per capita* benefits. As for location planning, the geometric structure of regular hexagons in theory is not a realistic representation of the real world.

Dökmeçi (1973) found that lower-level facilities were decentralized due to decreasing cost with an increasing number, while upper-level facilities were centralized due to increasing cost with an increasing number. This phenomenon is also described by the CPT. Accordingly, a series of the successively inclusive hierarchical systems was structured. While the overall value of the CPT diminished due to frequent incapability of describing and analyzing actual hierarchical systems satisfactorily, most concepts and principles have guided and been validated in subsequent studies (Dökmeçi 1973).

An effective and efficient hierarchical system should maximize social benefit from three major aspects: user (e.g., patients' travel cost), operator (e.g., hospital finance and staff's travel cost), and community sectors (e.g., environmental and economic impacts on communities) (Calvo & Marks 1973). A detailed framework of the social benefit, including all the factors in each sector, is described elsewhere (Calvo & Marks 1973), and a set of surrogate measures has admittedly been recognized due to difficulty of direct quantification for some of the original attributes, such as convenience and comfort of patients.

A wide array of location-allocation models have been used to determine or optimize locations of facilities for making higher-level facilities effectively and efficiently serve lower-

level facilities, while maximally satisfying all or part of the set of universal criteria. The criteria primarily include 1) minimization of distance or travel time, 2) minimization of user costs, 3) maximization of demand, 4) maximization of utility, and 5) minimization of penalty paid in adopting the optimal rather than a sub-optimal solution (Calvo & Marks 1973; Dökmeçi 1977; Schultz 1970). Comprehensive reviews on those models have been conducted in some studies (Farahani et al. 2014; Şahin & Süral 2007), in particular their applications in health service planning (Rahman & Smith 2000). However, they are not involved in this study due to three main reasons in different aspects. Firstly, the assumptions, all or part of which underlie those models, are rarely valid in reality, primarily including 1) the population is distributed homogeneously across the region, 2) each customer goes to or receives services from the nearest facility, and 3) each provider has an infinite capacity to serve customers (at a time). Second, most of those models are developed for finding optimal sites of facilities; in reality, relocating an existing system may be infeasible both politically and economically, and disadvantages could outweigh possible benefits resulting from relocating if only for optimization purposes. Third, most previous models do not take into account realistic data, which make them fail to present the current patterns of the healthcare market.

2.2 Hospital service areas (HSAs) and hospital referral regions (HRRs)

The HSAs, within which most of local hospitalization occurs, have been considered as basic spatial units for local hospital use or hospitalization studies (Center for Evaluative Clinical Sciences 1999). The HSAs offer promise for small area analyses (Gittelsohn & Powe 1995; Wennberg & Gittelsohn 1973), such as comparison of hospitalization practices and surgical procedure rates among different hospitals in various regions (Ashton et al. 1999; Fisher et al. 1994). Nevertheless, not all the hospitals are equipped to deliver a complete array of health

services, especially some tertiary care (e.g., cardiac surgery and neurosurgery). Hence, the HSAs with most of their patients going to the same hospital for specialized services can be further aggregated into a *Hospital Referral Region (HRR)*, representing a regional health care market for tertiary care (Center for Evaluative Clinical Sciences 1999). The HRRs have been chosen as analysis units in a growing body of large-scale studies, such as nationwide geographic variation in *per capita* physician supply (Goodman & Fisher 2008), access to care or preventive care in particular (Radley & Schoen 2012), performance of hospitals (Jha et al. 2005), patients' experiences in hospitals (Jha et al. 2008), Medicare drug spending (Zhang et al. 2010), and health care spending, utilization, and quality by the Institute of Medicine (Newhouse et al. 2013).

According to the CPT, there are fewer central hospitals (centers of the HRRs) than general hospitals (centers of the HSAs) due to relatively sparser demand. Hence, an HRR normally needs to deliver specialized care to a larger geographical area than an HSA, in order to have enough patients (customers) to sustain. On the other hand, travel can also be expected to increase with complexity of the procedure. In other words, patients are more likely to travel longer in general for highly technical or specialized services than routine care, and these distinctive travel patterns of the patients seeking different levels of care may underlie a hierarchical health care system. In reality, plenty of factors, such as travel distance, hospital size, quality of services, and health insurance policy, intertwined with one another to impact on decision-making of patients about which hospital to visit. Unlike the past when hospital attendance data were difficult to obtain even in developed countries (Hodgson 1988), availability of hospital attendance/discharge data has been dramatically growing in the recent two decades. Therefore, these data hold great values to be incorporated into traditional models for examining the realistic patterns of the HSA-HRR system.

The HSAs in the U.S. were defined in the Dartmouth Atlas of Health Care project (Center for Evaluative Clinical Sciences 1999) through a three-step process:

- 1) all acute care hospitals identified from the American Hospital Association (AHA) and Medicare provider files were assigned to the town or city in which they are located;
- 2) each postal zone was assigned to the town or city containing the hospital that the majority of residents in that postal zone visited, based on all 1992 and 1993 Medicare hospitalization records;
- 3) a visual examination was undertaken to ensure the geographic contiguity of all zip codes in one HSA, with the disconnected re-assigned to an adjacent HSA on a locational basis.

As a result, 3,436 HSAs were produced for the 50 states and the District of Columbia. This is also referred to as the Dartmouth approach, representing the earliest effort to develop the HSAs based on actual travel flow data. However, the data used in delineating the Dartmouth HSAs have been outdated and unrepresentative.

The Swiss approach is an improved version of the Dartmouth approach and first introduced to define the HSAs in Switzerland, where patient and hospital locations are recorded as census regions that are aggregated by postal zones. The HSAs in Switzerland were created by a similar three-step algorithm:

- 1) each hospital was assigned to a census region by location, referred to as hospital region, and each census region was assigned to the hospital region where the highest number of discharges in that census region occurred;
- 2) a visual examination was undertaken to ensure the geographic contiguity of all census regions in one HSA, with the disconnected re-assigned to an adjacent HSA on a locational basis;

3) each HSA with more residents admitted to the hospital within another HSA than their own HSA were merged into that HSA, referred to as the plurality rule (Center for Evaluative Clinical Sciences 1999).

The two approaches above delineate the HSAs only based on the actual hospital discharge data during a specific period of time, which are subject to the temporal variation in patient-to-hospital flows (Shortt et al. 2005). Jia et al. (2014) brought the Huff model into the delineation of the HSAs in Florida through the following steps:

- 1) using the power function with pre-assumed parameters to fit the discharge data in 2011 under each assumed threshold of travel distance;
- 2) selecting the model with an optimal set of parameters with which the minimum difference between theoretical and actual hospital visits was produced;
- 3) using that model to calculate the value of attractiveness of each hospital to each postal zone;
- 4) assigning each postal zone to the hospital with the largest attractiveness to that postal zone, and aggregating all the postal zones assigned to the same hospitals into the HSAs.

Nevertheless, that process has failed to 1) evaluate the whether the power function is an appropriate form of function to fit the actual data, and 2) objectively select an optimal set of parameters for minimizing the difference between theoretical and actual hospital visits.

The HRRs represent regional health care markets for tertiary care. The only effort to develop the HRRs is in the Dartmouth Atlas of Health Care project, where the HSAs with most of their patients going to the same city for cardiovascular surgical procedures and neurosurgery could be further aggregated into an HRR (Center for Evaluative Clinical Sciences 1999). The *Dartmouth HRRs* in conjunction with the *Dartmouth HSAs* is also known as the *Dartmouth*

*HSA*s-*HRR*s (Center for Evaluative Clinical Sciences 1999). The Dartmouth *HRR*s were defined through the following steps:

- 1) assigning each *HSA* to the city where the greatest proportion of major cardiovascular surgical procedures was performed;
- 2) undertaking a visual examination to ensure the geographic contiguity of all the *HSAs* in one *HRR*, with each disconnected *HSA* re-assigned to the *HRR* surrounding that *HSA*;
- 3) ensuring a minimum population size of 120,000 and at least 65% of their residents' hospitalizations occurring within each *HRR*.

The process resulted in 306 *HRR*s aggregated from 3,436 *HSAs* within the contiguous U.S. Similar to the *Dartmouth HSAs*, the outdated and unrepresentative Medicare hospitalization data may undermine the usefulness of the *HRR*s. Moreover, being built based on the *HSAs* instead of the travel patterns of the patients seeking specialized services makes the *HRR*s lack in theoretical foundation, and also difficult to update directly.

2.3 Delineation of other functional areas

A multitude of methods have been proposed for demarcating and analyzing service areas for different purposes, such as trade areas (*TAs*), labor market areas (*LMAs*), and housing market areas (*HMA*s). The goal of demarcation is to produce a set of self-contained functional areas so that most of spatial interactions occur within rather than between the functional areas.

One of the simplest and most intuitive approaches is the *ring-based approach* (Patel, Fik & Thrall 2008), in which a circle is drawn around a provider to capture a specified number or percentage of customers. This approach is easy to implement and interpret by assuming that the customers are evenly distributed around the providers while adjusting the radius uniformly toward all directions. Similarly, the *patient origin method* captures a certain percentage of

customers or a number of the area units closest to the provider. It can be easily extended by replacing Euclidean distance with network distance or travel time. With more accurate travel distance (or time), the buffer zone within a certain travel distance (or a certain period of travel time) can be used to represent the service area of a provider, such as 15 miles (or 30 minutes) (Luo & Wang 2003). A disadvantage of both ring-based and patient origin approaches is that the specified number or percentage is subjective, varying by analyst, such as 60% (Garnick et al. 1987), 75%, and 85% (Shortt et al. 2005; Shortt & Moore 2006). More importantly, as the customers living in one region may have different choices, multiple service areas may overlap with each other and not be mutually exclusive.

The *wedge-based approach* adds the component of directionality to the origin-destination (O-D) distance by dividing the region into a certain number of sectors or wedges and specifying an incremental distance, according to the analysts' experiences. The procedure starts from a small core region around a provider. In each iteration, only the wedge that captures the greatest number of customers by its incremental extension is extended. The iteration stops when a specified number or percentage of customers is reached or no additional increment is gained by the extension in all directions (Patel et al. 2008). This approach identifies the spatial heterogeneity of travel willingness and patterns in different directions. However, subjectivity in multiple steps leads to the introduction of errors and the inability to replicate results, such as the determination of how many sectors are needed and how long an incremental distance is. Moreover, if a large number of customers are clustered farther than one specified incremental distance from a wedge's current radius, no customer is able to be captured by only one incremental extension, and the wedge will not be allowed to extend. Those customers will not be

detected and included in the service area of that provider, which could create holes in coverage, also referred to as artificial discontinuity.

The *proximal area method* is another simple geographic approach (Ghosh & McLafferty 1987), which only considers travel distance (or time). Customers are assigned to their nearest providers, and all customers assigned to the same providers constitute the trade areas. This assumes that customer always choose their nearest service providers.

Another group of approaches defines service areas based on the flows of customers that are outcomes of the interaction between the supply and demand (Brown & Hincks 2008). To measure the independence of one service area from others, the *self-containment* for a given service area is defined as the residents (customers) interacting with the providers within that area, expressed as a percentage of all residents within that area. Sometimes the measurement unit of self-containment is a unit of interaction rather than a person. For example, each hospital visit for one person with multiple visits is counted only once. A degree of self-containment is set for all flow-based approaches, below which a service area should not be regarded as independent and must be merged with another for a higher self-containment than previously defined. The threshold of self-containment, although the precise level of which remains unanswered (Jones 2002), has been suggested as 50% (Jones 2002) or 70% (Pieda 2004) in some studies.

The approach utilized to delineate and analyze the labor market areas (LMAs) in Sweden uses the municipalities as basic units and the intensity of commuting flows from living to working places within and between municipalities to combine municipalities into LMAs (Carlsson et al. 1993). Two major steps are involved in this approach. In the first step, some municipalities form initial LMAs by themselves if 1) more than 80% of employed residents commute to work within their municipalities; and 2) the percentage of employed residents

commuting to any other individual municipality is less than 7.5%. The second step assigns each of the remaining municipalities to an LMA to which most of their employed residents commute, until all municipalities are assigned. In some cases, the second step needs to be executed iteratively because a municipality to which most of its employed residents commute from a different municipality may have not, but will become a part of an existing LMA in subsequent steps. However, a threshold of 80%, plus 7.5% for the maximum percentage of the commuters going outside of the service area, will lead to a small number of LMAs. In my case, an appropriate approach for delineating the HSAs should produce as many service areas as possible in order to preserve as much resolution of analysis unit as possible.

The *Synthetic Data Matrix* (SDM) is devised to integrate the service areas derived from a variety of datasets, in order to propose an adequate set of locality boundaries, which can consistently serve multiple purposes, such as local institutions, demography, economy, facilities, and landscape (Coombes 2000). This approach focuses on the common instead of specific features of different locality issues, which, however, is difficult to realize owing to large conceptual differences in those notions in nature.

2.4 Distance decay in health care studies

The *Huff model* (Huff 1964) has been widely successful in delineating trade areas of various business entities (Dramowicz 2005). The discussion is started by introducing the concept of gravity (spatial interaction) model (Haynes & Fotheringham 1984). For instance, the spatial interaction T_{ij} between a service provider and its customers is estimated as:

$$T_{ij} = P_i P_j / d_{ij} \quad (1)$$

where P_i = scale of provider i , P_j = scale of potential customer group j , and d_{ij} = distance between them. Adding the friction of distance effect, the improved measure of spatial interaction T_{ij}^* is expressed as:

$$T_{ij}^* = P_i P_j / d_{ij}^\beta \quad (2)$$

where β = distance decay friction factor.

Using the gravity model to delineate the service areas applies a process of identifying a threshold of distance from the provider, up to which most of the customers are inclined to travel (Patel et al. 2008). A grid with the specified cell size is overlaid on the study area to aggregate the numbers of customers falling into each cell, referred to as customer intensity (Patel et al. 2008). The centroid of each cell is extracted and the distance between each centroid and the provider is calculated. A distance decay function is fit to all pairs of customer intensity and their distance to the provider as described in the following formula

$$I_k = \mu d_k^\beta \quad (3)$$

where I_k = the customer intensity in the grid cell k , μ is a parameter, d_k = the distance from the centroid of the grid cell k to the provider, and β = the distance decay friction factor. Least square regression is used to estimate the parameters μ and β , and a distance threshold d^* is calculated by replacing I_k with the lower limit of a 95% (or 99%) confidence interval of the mean customer intensity over the entire study area.

With the distance friction coefficient β defined, the Huff model is expressed as the probability of selecting a given provider by customers, such as:

$$P_{ij} = S_j d_{ij}^{-\beta} / \sum_{k=1}^n S_k d_{ik}^{-\beta} \quad (4)$$

where P_{ij} = probability of selecting provider j by customer i , S_j = attraction of provider j , d_{ij} = distance between customer i and provider j , β = distance decay friction factor, S_k = attraction of

provider k , d_{ik} = distance between customer i and provider k , and n represents the number of providers accessible to customer i . It can be improved by adjusting the influence of providers:

$$P_{ij} = S_j^\alpha d_{ij}^{-\beta} / \sum_{k=1}^n S_k^\alpha d_{ik}^{-\beta} \quad (5)$$

where α = elasticity of hospital capacity. The probability of selecting each accessible provider by each customer (or cluster of customers, like a block or zip code area) is computed and each customer/cluster of customers is assigned to the service area of a provider with the highest probability.

Comparing all the approaches above, the Huff model provides a solid theoretical foundation for producing the HSAs and HRRs, as a built-in distance decay function enables the model to capture the spatial behavior of the patients for seeking different levels of health services. Not only does this feature allow integrating the realistic hospital attendance/discharge data with the model, but also there is flexibility to choose a function that best fits the realistic data, which is the key of configuring the Huff model. The following reviews the applications of distance decay functions in health care studies.

Also referred to as the ‘distance decay effect’, Tobler’s *first law of geography* states that ‘everything is related to everything else, but near things are more related than distant things’ (Tobler 1970). Similarly, many health studies reveal that the number of patients or frequency of visits gradually declines with distance from health care providers.

The degree of the distance decay varies by context. In a health care setting, it may vary by type of illness or illness severity, level in a service hierarchy, and population characteristic (Arcury et al. 2005). The distance decay effects of various subpopulations could be compared visually by the distance decay curve and statistically by the distance decay friction factor. For example, a recent study in Florida revealed that, in general, blacks were more likely to seek

hospitalization within the HSAs compared to whites, which might imply that blacks have a sharper distance decay effect than whites (Jia & Xierali 2015).

From a broader view, a distance decay function can be used to analytically capture the extent of any decreasing trend in response to increasing travel distance/time from service providers. The built-in function form in the earliest Huff model is a power function, which was also used to describe the spatial behavior of hospital visits in Florida (Jia, Xierali & Wang 2014). Hodgson (1988) used a negative exponential function to explore the rural accessibility to health care in a developing country. Delamater et al. (2013) adopted a log-logistic function to describe the distance decay of hospital utilization in Michigan.

Different types of patient-physician interactions can be conceptualized by different distance decay functions (Wang 2012). The question of which function delineates the actual travel patterns of patients most accurately needs to be addressed on a case-by-case basis by analyzing the actual hospital attendance/discharge data.

2.5 Summary

A few observations can be made from the above review:

- 1) Most previous approaches to delineating the HSAs and HRRs have not utilized the actual hospital attendance/discharge patterns of all population.
- 2) All current approaches heavily rely on O-D flow data of patients in hospital visits, which may not be available for many studies.
- 3) The Huff model is an appropriate approach for producing theoretically sound and practical HSAs and HRRs.
- 4) Different distance decay functions need to be compared in order to determine the best-fitting function that accurately captures patient's travel patterns.

Chapters 4 to 7 will respond to the above issues by presenting an integrated research framework for delineating HSAs-HRRs. Prior to that, Chapter 3 explains the study area and data processing.

Chapter 3 Study Area and Data

Florida is situated in the southeastern U.S. with three facets bordered by water: the Gulf of Mexico to the west, the Florida Straits between the U.S. and Cuba to the south, and the North Atlantic Ocean to the east. Excluding seasonal residents returning to their original states and permanent residents traveling an unusually long distance to other states for health care seeking, only those in North Florida counties bordering Georgia and Alabama are more likely to cross the state boundary for hospital services. Compared to any other stand-alone state, the potential influences from adjacent states to mix the travel patterns of patients should be theoretically (geographically) minimized in Florida. Therefore, Florida is an ideal state for a statewide study of patients' travel patterns for hospitalization.

The Healthcare Cost and Utilization Project (HCUP) is a suite of health care databases and related software tools and products developed through a Federal-State-Industry partnership, which was created with the purpose of providing a large-scale source for national, state, and all-payer health care data, and enhancing nationwide comparability among independent health outcomes in different states. The HCUP State Inpatient Database (SID) of the Agency for Healthcare Research and Quality (AHRQ) (Healthcare Cost and Utilization Project 2011) was used in this study, which includes individual discharge records from all hospitals in Florida during 2011. Each record includes a wide range of attributes: primary and secondary diagnoses and procedures, admission and discharge date and status, patient demographics, expected source of payment (e.g., health insurance type), length of stay, zip code of residence, hospital of presentation, and total charges for each hospitalization. Using the zip code of residence of the patients, individual discharge records can be geocoded at the zip code spatial scale.

The dataset of hospital discharge records consists of 2,656,249 records from 281 hospitals in Florida, of which 268 hospitals were identified and geocoded. These hospitals were linked to 2013 American Hospital Association's (AHA) survey files for information on hospitals, such as hospital type and number of beds. A total of 12,745 records (0.5% of the dataset) were excluded because they were from 13 unidentified hospitals. Another 23,046 (0.9%) had missing residence zip codes, 21,598 (0.8%) others indicated that the patients lived outside of the contiguous U.S., and 123,035 (4.6%) others were from psychiatric, rehabilitation, children's, women's, and other specialty hospitals. The remaining 2,475,825 records are included in this study.

The patients were discharged from 221 general hospitals in Florida, including 22 acute long-term care hospitals and 199 general medical and surgical hospitals. The distribution of the hospitals, with the number of beds represented by the size of circles, is shown in Figure 1. A total of 123 out of 221 hospitals included are located in the four metropolitan areas of Florida with a population of over 1 million: 57 in Miami-Fort Lauderdale-Pompano Beach (5,564,480 persons in 2010), 34 in Tampa-St. Petersburg-Clearwater (2,778,478 persons), 18 in Orlando-Kissimmee-Sanford (2,127,209 persons), and 14 in Jacksonville (1,340,760 persons). A total of 67 hospitals are located in the 16 metropolitan areas with a population of at least 50,000 but less than 1 million, ranging from 93,454 (Palm Coast) to 700,803 (North Port-Bradenton-Sarasota). Sixteen hospitals are in ten micropolitan areas with a population between 10,000 and 50,000, and the remaining 15 hospitals not located within any metropolitan or micropolitan area are considered as rural hospitals in this study.

All the included discharges were geocoded according to inpatients' personal residential zip codes, as well as by hospitals that the inpatients visited. Of 2,475,825 inpatient records,

2,376,743 (96%) were geocoded within 983 postal zones in Florida, of which 1,069,369 (45%) discharges were paid by Medicare. The boundaries of 1992-1993 Medicare-derived HSAs produced in the Dartmouth Atlas of Health Care project were downloaded from the Dartmouth Atlas of Health Care (<http://www.dartmouthatlas.org/>). Other datasets used include the 2010 U.S. postal zones boundaries, and 2010 U.S. Census block boundaries with the number of total population within each block.

A major issue in the SID is general to most studies, which is the unknown exact address of patients due to confidentiality laws. It is worth noting that Florida is a special case in terms of population composition within the U.S., where a number of seasonal residents originally from other states only spend their winter in Florida, especially the elder people that are more vulnerable to many health risks than other populations. Nevertheless, the seasonal residents that postpone their health care seeking until returning to their permanent states are not included in this study. If they choose to seek hospitalization in Florida hospitals, they are assumed to make decisions similarly as permanent residents do. Though the attributes of the SID do not allow telling seasonal apart from permanent residents, there is a time stamp marking the calendar quarter in which each hospital visit occurred (January-March, April-June, July-September, or October-December), which enables to examine the variations in patients' travel pattern among four quarters in 2011.

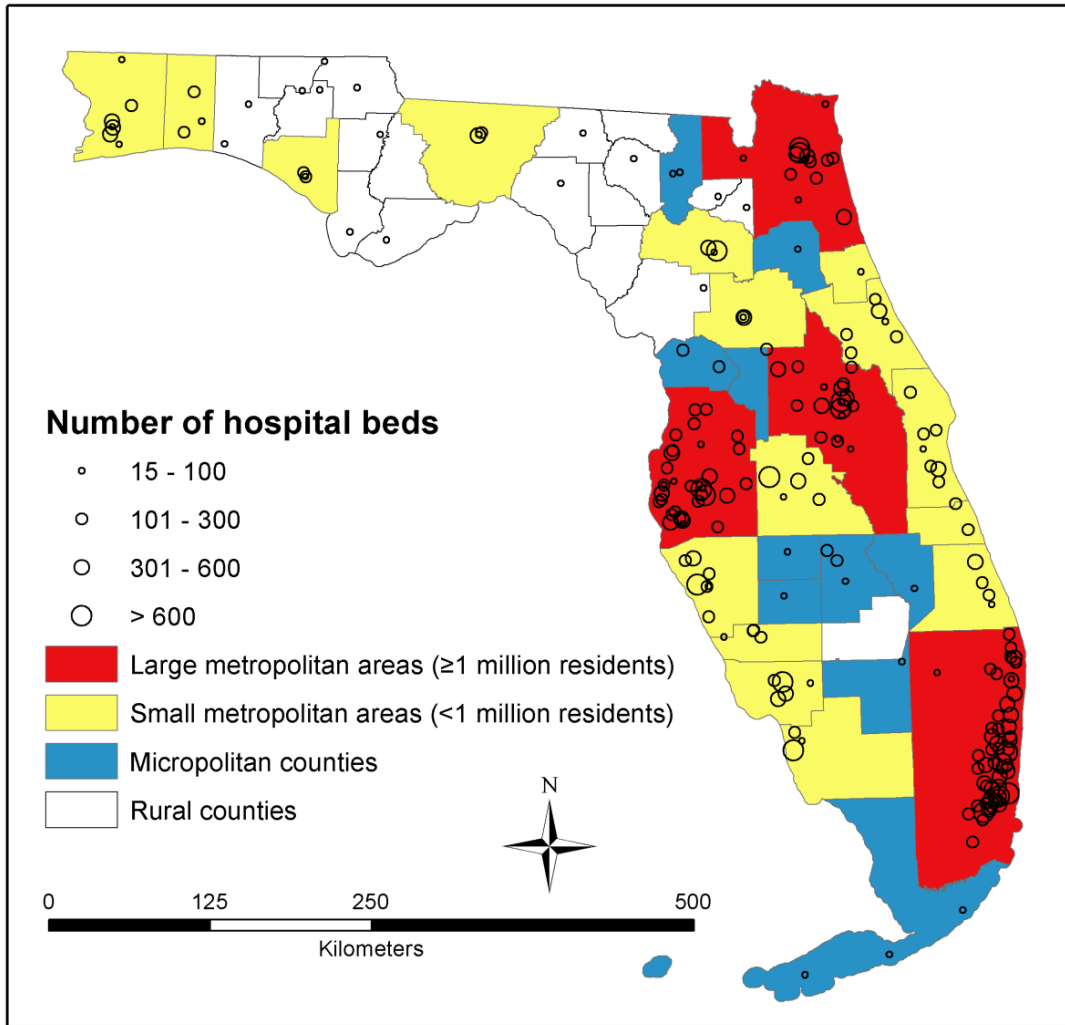


Figure 1. Distribution of the included hospitals with the number of beds represented by the size of circles

Chapter 4 Evaluating the Dartmouth HSAs

This chapter evaluates effectiveness of the most widely used Dartmouth HSAs/HRRs.

The Dartmouth method is replicated on a basis of the hospital discharge data in 2011 to construct the updated HSAs and HRRs for examining two questions. Are the HSAs/HRRs derived by the Dartmouth method outdated? Do the Dartmouth HSAs/HRRs based on the Medicare data represent the pattern of hospital visits by the general population?

4.1 Localization index

The localization index (LI) refers to the fraction of discharges of HSA residents that occurred within their own HSA, and is an important index used in the method for measuring the degree of localization of hospital care (Guagliardo et al. 2004; Klauss et al. 2005). The LI of a given HSA is calculated by dividing the number of the HSA discharges from the hospitals within that HSA by the total number of the discharges within that HSA. When all hospital care for residents is provided within their HSA, $LI = 1$. An LI value above 0.5 indicates that more admissions occur within the HSAs, and is thus a desirable threshold for the definition of HSAs.

All individual discharge records paid by Medicare in 2011 are geocoded to 2011 zip codes where inpatients lived at the time of admission. The sum of the discharges within each zip code is aggregated. To demonstrate the temporal variation of the Medicare-derived HSAs, the 2011 zip codes need to be matched to the 1992-1993 Medicare-derived HSAs. Due to the changes of postal zone boundaries between 1992-1993 and 2011, the postal zone centroids are used to represent all Medicare discharges within the 2011 postal zones, and then reassigned to the 1992-1993 Medicare-derived HSAs by location. The LIs are calculated within the 1992-1993 Medicare-derived HSAs.

The Dartmouth method is replicated to construct the new HSAs at a finer scale based on the 2011 Medicare patient-to-hospital flow data, referred to as the *2011 Medicare-derived HSAs*, where hospitals are assigned to postal zones instead of towns/cities. The implementation also uses the plurality rule in the Swiss method, and thus may be considered a *Dartmouth-Swiss hybrid method*. The percentage of Medicare discharges from each hospital within each postal zone is computed, and each postal zone is assigned to the hospital discharging the highest percentage of inpatient records within that postal zone. Each contiguous collection of postal zones assigned to the same hospital, with that hospital located within the cluster, forms an initial HSA. Geographic contiguity and the plurality rule are enforced to revise the divisions of the 2011 Medicare-derived HSAs. Repeatedly, each HSA with an $LI < 0.5$ is merged into an adjacent HSA that either geographically encircled that HSA or discharged the highest percentage of inpatient records from that HSA, until the LI of all HSAs is ≥ 0.5 . Similarly, the LIs in the final 2011 Medicare-derived HSAs are calculated, and the two sets of LIs are compared for difference by a *t*-test for statistical significance. A natural log transformation is performed prior to each *t*-test to reduce the influence of the skewed distribution of LI values.

For evaluating the representativeness of the Medicare-derived HSAs, all individual (not just Medicare patient) discharge records in 2011 are aggregated to zip codes and matched to the 2011 Medicare-derived HSAs. The LIs are again updated within each 2011 Medicare-derived HSA. Based on the data of all discharges, the similar procedures are followed for deriving the HSAs, referred to as the *2011 overall-derived HSAs*. The percentage of discharges from each hospital within each postal zone is re-computed according to the updated patient-to-hospital flows. Each postal zone is assigned to the hospital discharging the highest percentage of inpatient records within that postal zone, and each contiguous collection of postal zones assigned to the

same hospital forms an initial HSA. Geographic contiguity and the plurality rule are enforced to ensure the LI of each HSA ≥ 0.5 . The LIs are then calculated within the 2011 overall-derived HSAs, and compared with the LIs using the overall discharges but in the 2011 Medicare-derived HSAs by a *t*-test after a natural log transformation.

4.2 Effectiveness of the Dartmouth HSAs

There are 114 divisions in 1992-1993 Medicare-derived HSAs and 104 divisions in 2011 Medicare-derived HSAs across Florida. An overlay of the boundaries of the 1992-1993 Medicare-derived HSAs on those of the 2011 Medicare-derived HSAs (Figure 2) reveals statewide changes over the period. The decreased number and larger areas of the HSAs in 2011 imply that nowadays Medicare patients travel farther for hospitalization than two decades ago. The LIs for 2011 Medicare inpatients have a larger mean and narrower range within 2011 Medicare-derived HSAs than within 1992-1993 Medicare-derived HSAs (mean: 0.66 versus 0.6, range: 0.5 to 0.96 versus 0 to 0.95). It can be seen that the LI resulting from matching 2011 Medicare patient-to-hospital flows to 1992-1993 Medicare-derived HSAs is generally not as high as that within the 2011 Medicare-derived HSAs, which is also confirmed by *t*-test (Table 1 and 2) at the 99% confidence level. The boundaries of 1992-1993 Medicare-derived HSAs are significantly subject to temporal variation, and hence need to be updated using more recent data.

Table 1. Descriptive statistics of natural log-transformed localization index of the 1992-1993 and 2011 Medicare-derived HSAs

HSA	Number	Mean of LI	Mean	Std. Deviation	Std. Error Mean
1992-1993 Medicare-derived	114	0.60	-0.526	0.340	0.032
2011 Medicare-derived	104	0.66	-0.424	0.166	0.016

Table 2. Results of *t*-test for the difference of localization index of the 1992-1993 and 2011 Medicare-derived HSAs

	Levene's Test		<i>t</i> -test for Equality of Means						
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Lower	95% Upper
Equal variances assumed	24.860	0.000	-2.786	212	0.006	-0.103	0.037	-0.176	-0.030
Equal variances not assumed			-2.835	159.879	0.005	-0.103	0.036	-0.174	-0.031

*95% confidence interval

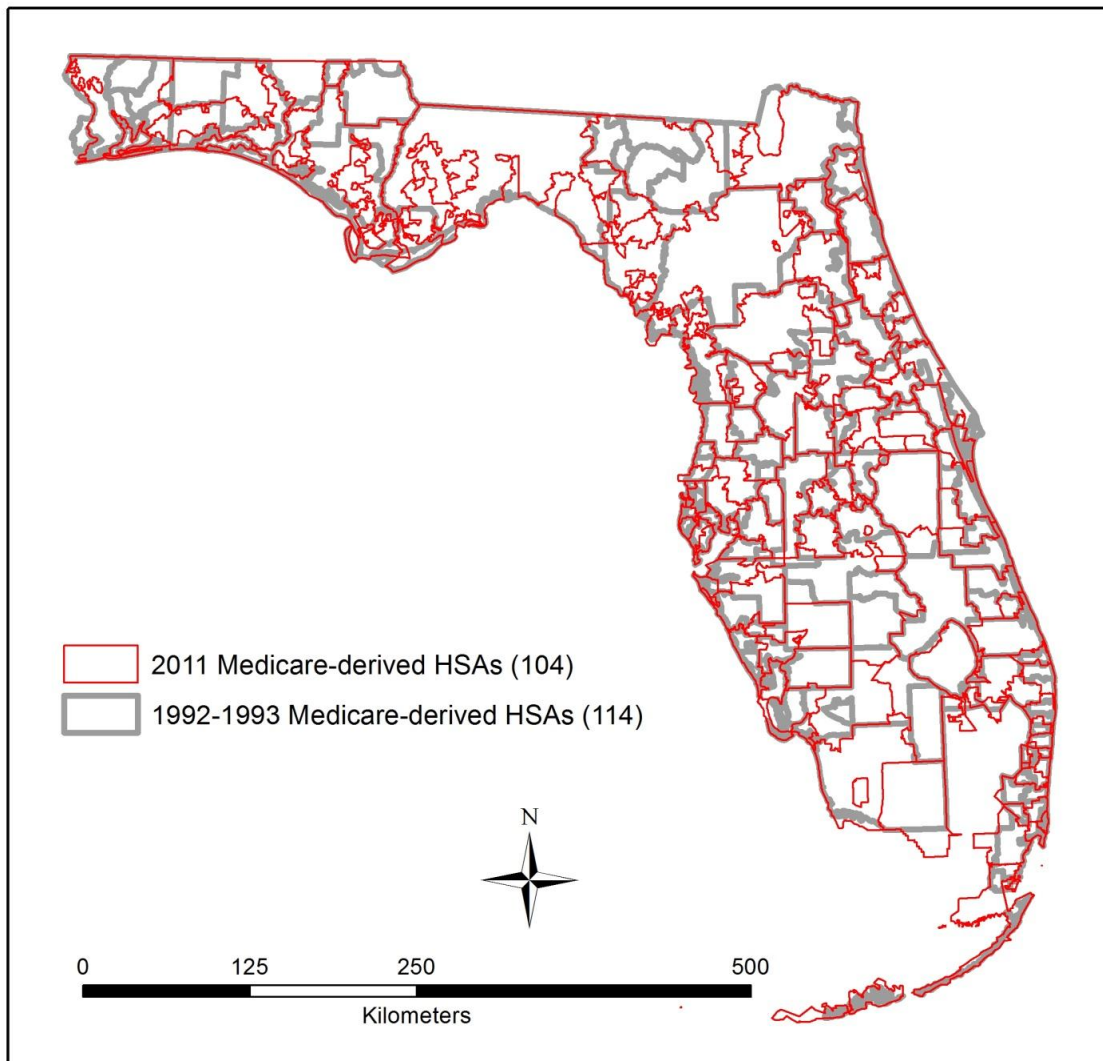


Figure 2. Boundaries of the 1992-1993 and 2011 Medicare-derived HSAs

There are 78 divisions in the 2011 overall-derived HSAs, compared to 104 divisions in the 2011 Medicare-derived HSAs. The boundaries of the 2011 Medicare-derived HSAs are overlaid with those of the 2011 overall-derived HSAs (Figure 3), where more local service areas can be observed within 2011 Medicare-derived HSAs, particularly in three metropolitan areas, Miami-Fort Lauderdale-Pompano Beach, Tampa-St. Petersburg-Clearwater, and Orlando-Kissimmee-Sanford. Many Medicare-derived HSAs have been aggregated to larger service areas when more discharges are paid independently of Medicare, such as by Medicaid, private insurance, self-pay, and even no charge. It implies that the overall patients covered by various types of health insurance have traveled longer distances on average than Medicare patients for hospitalization services.

The LIs for 2011 overall discharges are calculated within both 2011 Medicare-derived and overall-derived HSAs. The LIs for 2011 overall discharges have a larger mean and a smaller range within 2011 overall-derived HSAs than within 2011 Medicare-derived HSAs (mean: 0.65 versus 0.59, range: 0.5 to 0.93 versus 0.32 to 0.96). It can be seen that the LIs resulting from matching 2011 overall patient-to-hospital flows to 2011 Medicare-derived HSAs are generally not as high as those within 2011 overall-derived HSAs, which is also confirmed by *t*-test at the 99% confidence level (Table 3 and 4). It implies that the boundaries of Medicare-derived HSAs are not as appropriate to the overall inpatients as to Medicare inpatients. Therefore, the boundaries of appropriate HSAs need not only to draw on more recent data, but also to take into account the discharges paid by entities other than Medicare, to be more representative of the travel behavior of the overall population.

Table 3. Descriptive statistics of natural log-transformed localization index of the 2011 Medicare-derived and overall-derived HSAs

HSA	Number	Mean of LI	Mean	Std. Deviation	Std. Error Mean
2011 Medicare-derived	104	0.59	-0.559	0.236	0.023
2011 overall-derived	78	0.65	-0.446	0.177	0.020

Table 4. Results of *t*-test for the difference of localization index of the 2011 Medicare-derived and overall-derived HSAs

	Levene's Test		<i>t</i> -test for Equality of Means						
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Lower	95% Upper
Equal variances assumed	4.701	0.031	-3.535	180	0.001	-0.113	0.032	-0.175	-0.050
Equal variances not assumed			-3.679	179.998	0.000	-0.113	0.031	-0.173	-0.052

*95% confidence interval

4.3 Effectiveness of the Dartmouth HRRs

There are 19 divisions in the Dartmouth HRRs in Florida, also referred to as the *1992-1993 Medicare-derived HRRs* somewhere in this study, among which the four northern ones cross the northern state boundary due to being produced at a national scale (Figure 4).

Aggregated from the Dartmouth HSAs, the value of the Dartmouth HRRs may be diminished by the outdated and underrepresented HSAs.

In order to evaluate the effectiveness of the Dartmouth HRRs, the specialized (i.e. cardiovascular and neurological) discharge records paid by all types of insurance are extracted from the 2011 SID (in detail described in Section 7.1), and geocoded by inpatients' residential zip codes and the hospitals they visited, forming specialized patient-to-hospital travel flows that are subsequently matched to the Dartmouth HRRs. The LIs calculated within the HRRs range from 0.34 to 0.98, which implies that the Dartmouth HRR boundaries need to be updated as well. The LI of the AL-Dothan HRR is only 0.34 as most of the area of that HRR is located in

Alabama. This phenomenon also exists in other three boundary-crossing HRRs (FL-Pensacola, Tallahassee, and Jacksonville), but is not as influential on the LI as it is in the AL-Dothan.

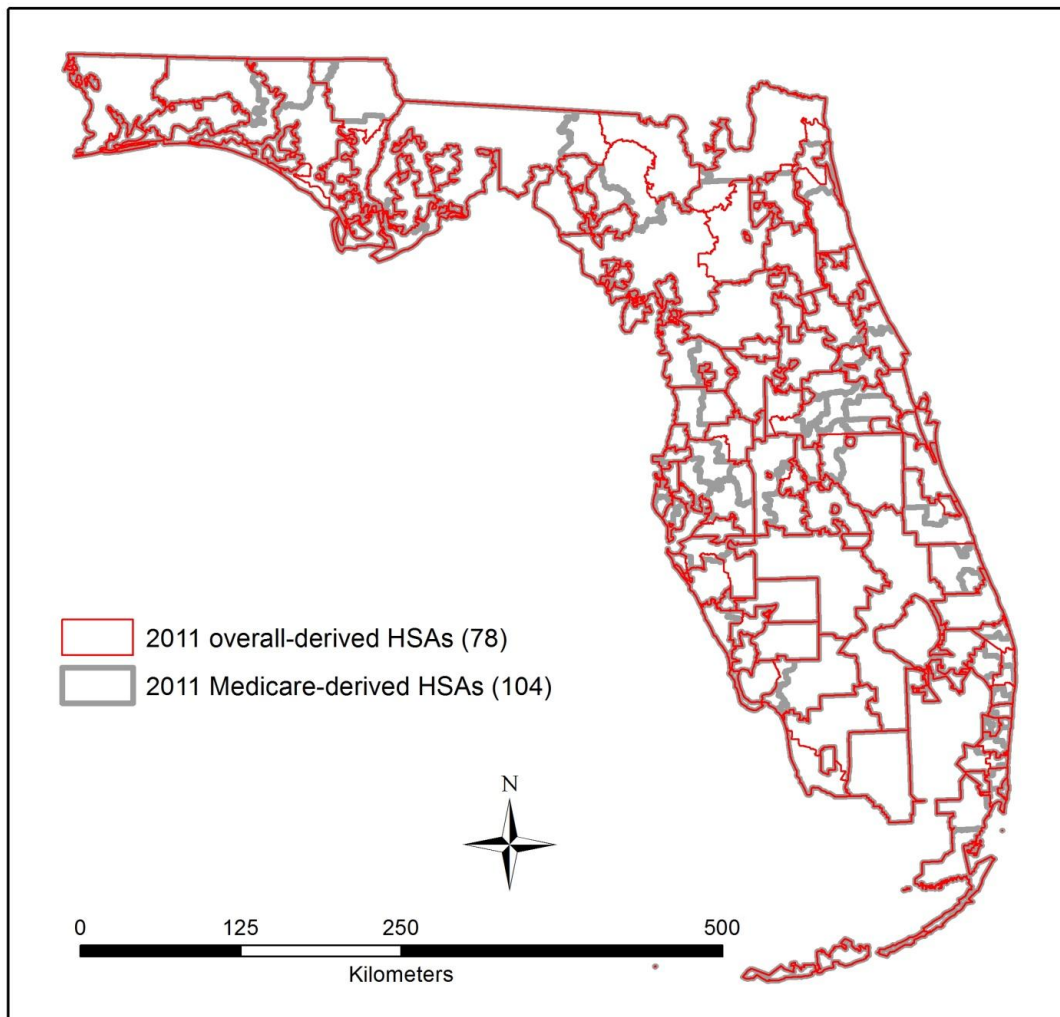


Figure 3. Boundaries of the 2011 Medicare-derived and overall-derived HSAs

To compare with the Dartmouth HRRs, the Dartmouth aggregation method is replicated at a finer scale to construct the *2011 overall-derived HRRs* based on the specialized discharge data in 2011, where hospitals are assigned to postal zones, hence HSAs are as well assigned to postal zones (i.e. hospitals) instead of cities. In addition, the Dartmouth HRR eligibility rules are also applied where both a minimum population size of 120,000 and $LI \geq 0.65$ need to be satisfied

within each HRR. A flowchart of the aggregation and adjustment processes is provided (Figure 5) and also described as follows:

- 1) all the 2011 overall-derived HSA units including at least one *specialized hospital* providing both cardiovascular surgical procedures and neurosurgeries form HRR cores;
- 2) all individual specialized discharge records are geocoded to HSAs or HRR cores and converted into the HSA-to-hospital flows (i.e. the patient-to-hospital flows at the HSA level), with the numbers of discharges from each HSA to each HRR core computed;
- 3) each HSA is assigned to the HRR core discharging the highest percentage of specialized records within that HSA, and each collection of HSAs assigned to the same HRR core forms an initial HRR;
- 4) any disconnected HSA without hospital(s) enclosed is re-assigned to the neighboring initial HRR discharging the highest percentage of specialized records from that HSA, until all the HRRs are continuous geographically;
- 5) each HRR with an $LI < 0.65$ is merged with the other HRR that discharged the highest percentage of specialized records from that HRR, until all the HRRs possess an $LI \geq 0.65$;
- 6) each HRR with a population size $< 120,000$ is merged with the other HRR that discharged the highest percentage of specialized records from that HRR, until all the HRRs possess a population number $\geq 120,000$.

The 2011 overall-derived HRRs include 34 divisions, with LIs ranging from 0.66 to 0.97. The LI values within the Dartmouth HRRs are on average larger than within the 2011 overall-derived HRRs (0.85 versus 0.82), but insignificant at the 95% confidence level ($p = 0.394$ by *one-tailed t-test*). Compared to the city-level Dartmouth HRRs, some metropolitan areas include more than one eligible HRR, according to the Dartmouth HRR eligibility rules. For example, the

Orlando HRR includes completely or partly ten eligible HRRs; the Fort Lauderdale and Fort Myers HRRs include five eligible HRRs, separately. Therefore, the traditional approach suffers from the data limitation, and is outdated because of the increasing number of specialized hospitals and enhanced resolution of hospital records.

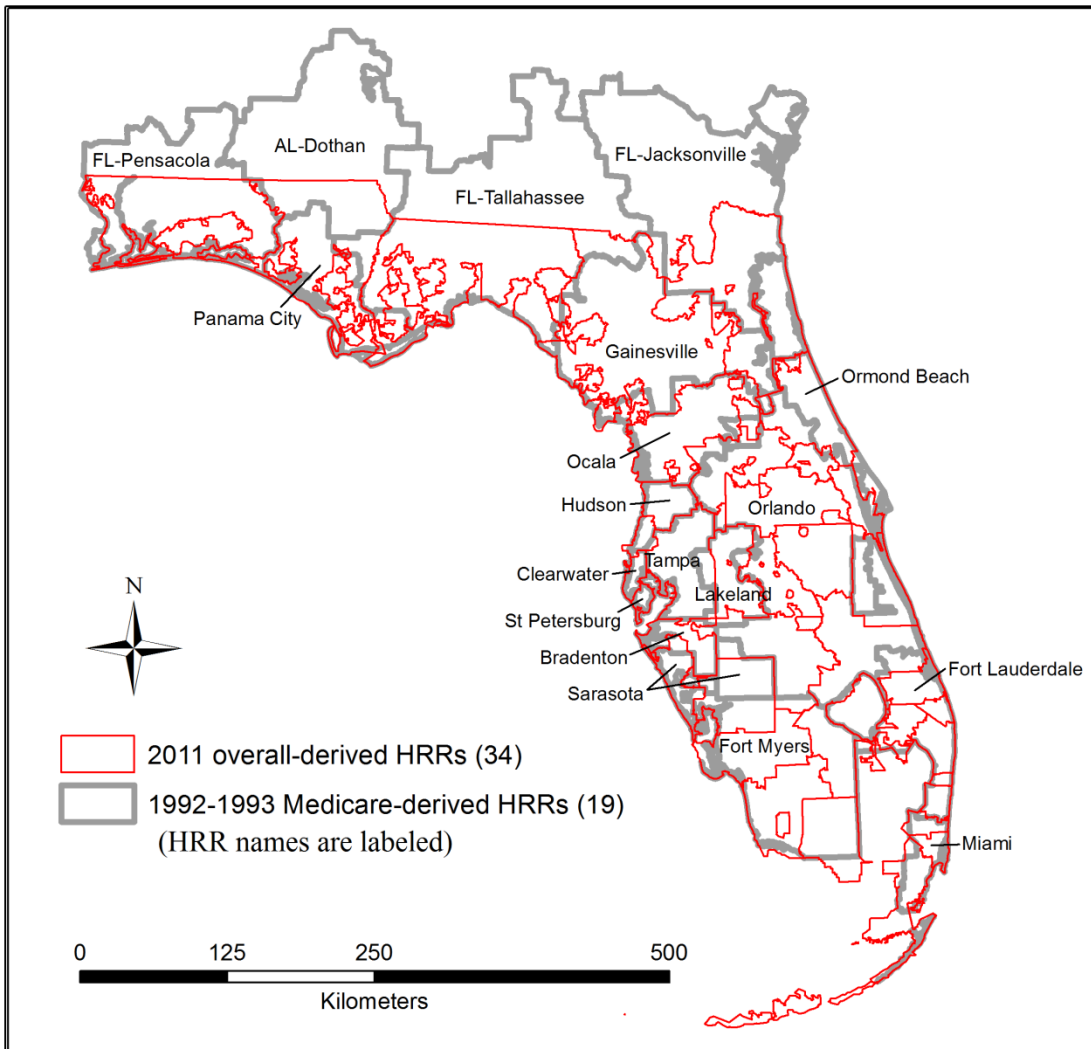


Figure 4. Boundaries of the 1992-1993 Medicare-derived and 2011 overall-derived HRRs

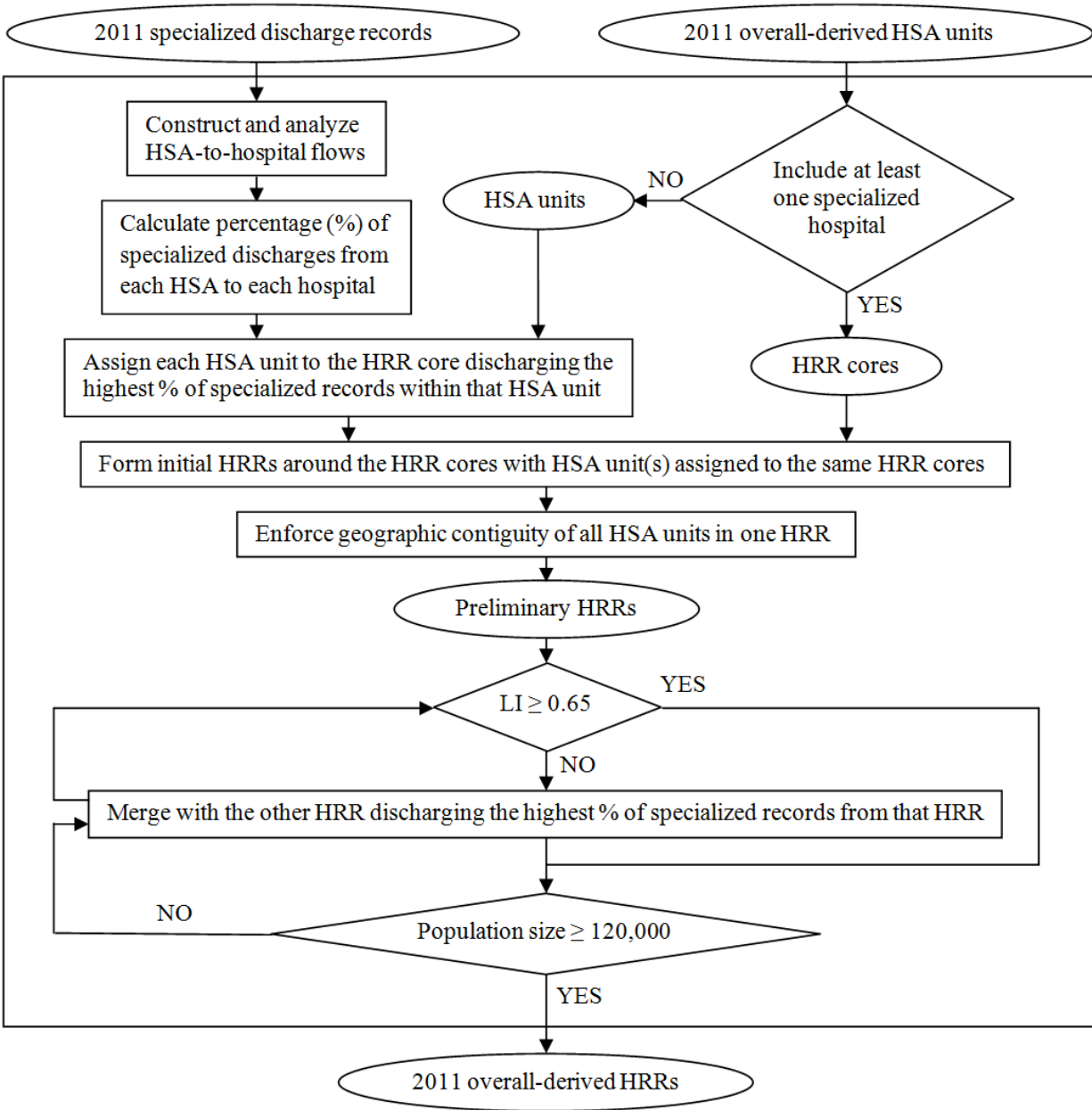


Figure 5. A flowchart of generating the 2011 overall-derived HRRs

4.4 Summary

In summary, three key conclusions are obtained from the analysis in this chapter. First, HSA boundaries have significantly changed over the last two decades. Second, Medicare-derived HSA boundaries are not representative of the travel patterns of the overall population. Third, boundaries of the HRRs aggregated from HSAs are impacted by the changes of the HSA

boundaries. Therefore, both HSAs and HRRs need to be updated regularly based on the up-to-date data of hospital visits to keep them useful, and an annual surveillance is suggested to find an appropriate interval for updating the HSAs.

The Dartmouth approach for producing the HRRs assigns each HSA to a city instead of an HSA or other finer-scale spatial unit, such as zip code. Therefore, it is unsurprising to find that the LI values calculated based on the recent discharge data are on average higher within the 1992-1993 Medicare-derived HRRs than within the 2011 overall-derived HRRs. Due perhaps to the data or methodological limitations two decades ago, the spatially conservative implementation of the Dartmouth approach over-satisfies the Dartmouth eligibility rule of $LI \geq 0.65$. Due to a lack of HSA/HRR managers, the HSAs/HRRs including multiple hospitals make it difficult to assign responsibility for variation among HSAs/HRRs to the specific hospital(s). The most immediate result is the difficulty in making the research findings directly serve the managerial and policy recommendations (Shwartz et al. 2011). Moreover, there is evidence showing that the competition among hospitals reduces the quality of hospital care (Propper, Burgess & Green 2004). Therefore, each HSA/HRR should include as fewer hospitals as possible. In another word, the preferred number of the HSA/HRR units within a given area should be as many as possible on the premise of $LI \geq 0.65$. The replication of the Dartmouth approach on the 2011 data at a fine scale, where each HSA is assigned to a hospital instead of a city, provides a more proper output (the 2011 overall-derived HRRs) as a reference than the Dartmouth HRRs in the rest of this study.

Chapter 5 Analysis of Patients' Distance Decay Behavior in Hospital Visits

This chapter analyzes the travel patterns of overall population and also by subpopulations (e.g., race/ethnicity, socioeconomic status (SES), urbanicity, and discharge quarter). Different distance decay functions are compared, and the one with the highest fitting power is identified. The best-fitting function will be used in the Huff model for delineation of HSAs and HRRs in later chapters.

5.1 Distance decay effects in hospital visits

The interaction between patients and hospitals in a geographic context, represented by either the volume of the patients within the postal zone i visiting the hospital j , or, conversely, the volume of the discharges from the hospital j to the postal zone i (T_{ij}), is formulated as

$$T_{ij} = p_i^\alpha b_j^\sigma f(d_{ij})$$

where p_i is total population in postal zone i , b_j is number of beds in hospital j , α and σ are parameters describing the effects of the numbers of zip code population and hospital beds upon the interaction respectively, d_{ij} is travel time from zip code i to hospital j in minutes, $f(d_{ij})$ is a distance decay function that can be defined in various forms, such as power, exponential, Gaussian and log-logistic functions (Table 5), where β is the distance decay friction factor and θ is another parameter (only present in the log-logistic function) to be estimated (Wang 2015).

Table 5. Optimal parameters and assessment of the four candidate functions for modeling the interactions from postal zones to hospitals

Function	$f(d_{ij})$	α	σ	θ	β	$pseudo-R^2$	AIC
Power	$d_{ij}^{-\beta}$	0.45	0.43	-	0.56	0.2788	319,604.2
Exponential	$e^{-\beta d_{ij}}$	0.53	0.38	-	0.14	0.4742	312,474.0
Gaussian	$e^{-(d_{ij}/\theta)^2/2}$	0.51	0.35	7.54	-	0.4503	313,479.4
Log-logistic	$1/(1 + (d_{ij}/\theta)^\beta)$	0.52	0.37	6.10	2.05	0.4744	312,467.1

Population data are only available in census blocks, block groups, and census tracts in the 2010 U.S. census, not in zip code areas, which are the smallest unit the hospital discharge data are geocoded to. Hence, population at the census block level is transformed to the zip code level by the areal weighting method (AWM) (Goodchild & Lam 1980). The AWM assumes that population is uniformly distributed within an areal unit such as block. The block layer is spatially overlaid with the postal zone layer to split postal zones into intersecting zones. If a given postal zone is completely located within a block, the population in the postal zone would be the product of the block population and the area ratio of the postal zone to the block. If a given postal zone contains or/and intersects more than one block, each subzone or/and intersecting zone within the postal zone would be assumed to have population proportional to the area ratio of itself to the block to which it belonged. The total population in that postal zone is the sum of the population in all subzones and intersecting zones within it.

The travel time d_{ij} must be measured between points. Population-weighted centroids are adopted to represent of the zip code locations (Luo & Wang 2003), such as:

$$x_q = \sum_{i=1}^{n_q} p_i x_i / \sum_{i=1}^{n_q} p_i \quad (1)$$

$$y_q = \sum_{i=1}^{n_q} p_i y_i / \sum_{i=1}^{n_q} p_i \quad (2)$$

where x_q and y_q are the x and y coordinates of the population-weighted centroid of a given postal zone q , respectively; x_i and y_i are the x and y coordinates of the geometric centroid of the i th subzone (block or part of block) within that postal zone, respectively; p_i is the population of the i th subzone (block or part of block) in that postal zone, and n_q is the total number of subzones in that postal zone. Therefore, d_{ij} is calculated from the population-weighted centroid of each postal

zone (x_q, y_q) to each hospital, along the shortest path of the road network with speed limits by the Network Analyst Tools in *ArcGIS*.

With all discharge records aggregated by postal zone and hospital, the volume of the discharges from hospital j to postal zone i in 2011 is used to measure T_{ij} , i.e., the patient-to-hospital flow. With p_i , b_j , d_{ij} , and T_{ij} all assigned or calculated, the parameters α , σ , β , and θ are estimated using the non-linear least squares estimators available in *R* (Development Core Team 2011). The *pseudo-R*² is similarly defined as the portion of dependent variable's variation explained by a nonlinear regression model (Wang 2015). Due to varying complexities of the four functions (one parameter in power, exponential, and Gaussian function vs. two parameters in log-logistic function), the Akaike information criterion (AIC) is also used to measure the relative quality of the models. The *pseudo-R*² and AIC of the four functions are calculated for comparing their performance in modeling the travel patterns of Florida inpatients in 2011. The model with the maximum *pseudo-R*² and minimum AIC is selected as the best-fitting model.

It is argued that data of the flows with more discharges are more reliable than the ones with fewer discharges. The sensitivity is tested by repeatedly calculating and comparing the *pseudo-R*² and AIC values of four models as the flows with small discharges (<10) are gradually excluded from the analyses. The best-fitting model is expected to perform best across the whole spectrum.

5.2 Comparison of distance decay functions

A total of 217,243 (983×221) patient-to-hospital flows are formed from 983 postal zones to 221 hospitals, where only 37,216 actual flows had at least one discharge in 2011. Among the 37,216 flows, 14,650 flows (39.4%) had only one discharge. Considering the flows with few discharges unreliable, an arbitrary decision is made to exclude travel flows with only one

discharge. Four functions are fitted separately based on the remaining flows with at least two discharges (Table 5). Keeping p_i and b_j constant (set both as 1), four functions with the optimal set of parameters are drawn (Figure 6). To compare the performance among four functions, the $pseudo-R^2$ and AIC are repeatedly calculated as the minimum number of discharges in one flow increased from two to ten progressively (Figure 7).

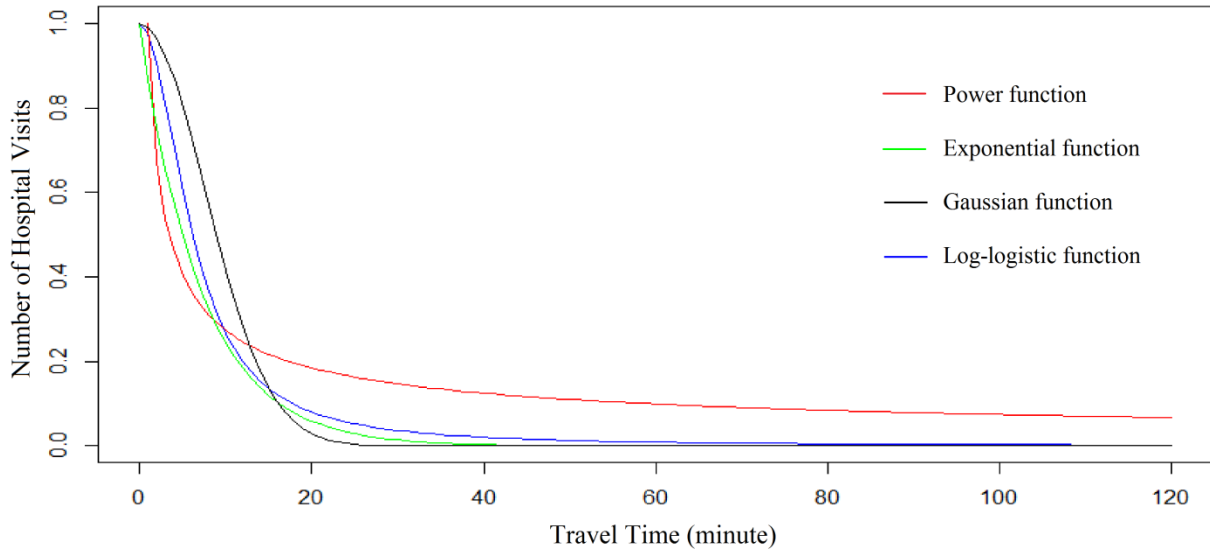


Figure 6. Curves of the four distance decay functions with the optimal sets of parameters

It is indicated that, by large, the $pseudo-R^2$ and AIC of the log-logistic function remain higher and lower (respectively) than the exponential function, and superiority has become more visually distinguishable for the $pseudo-R^2$ since the minimum number of discharges in one flow increases to five (Figure 7). The Gaussian function generates lower $pseudo-R^2$ and higher AIC than the log-logistic and exponential functions, and the power function produces even significantly lower $pseudo-R^2$ and higher AIC than the Gaussian function. Therefore, the log-logistic function is used as the best-fitting function in this study for fitting all patient-to-hospital flows including at least two discharges. The optimal combination of parameters produced is $\alpha = 0.52$, $\sigma = 0.37$, $\theta = 6.1$ and $\beta = 2.05$ (Table 5). Therefore, the expected interactions between postal zone i and hospital j (T'_{ij}) could be formulated as

$$T'_{ij} = p_i^{0.52} b_j^{0.37} / (1 + (d_{ij}/6.1)^{2.05})$$

where the parameter values are all statistically significant ($p < 2 \times 10^{-16}$).

Additionally, according to the log-logistic function, 622 out of 983 postal zones have the largest expected interactions with the hospitals discharging the most patients to those zip codes in 2011, followed by the exponential (620), Gaussian (604), and power functions (495). This consistent order further validates the selection of the best-fitting model.

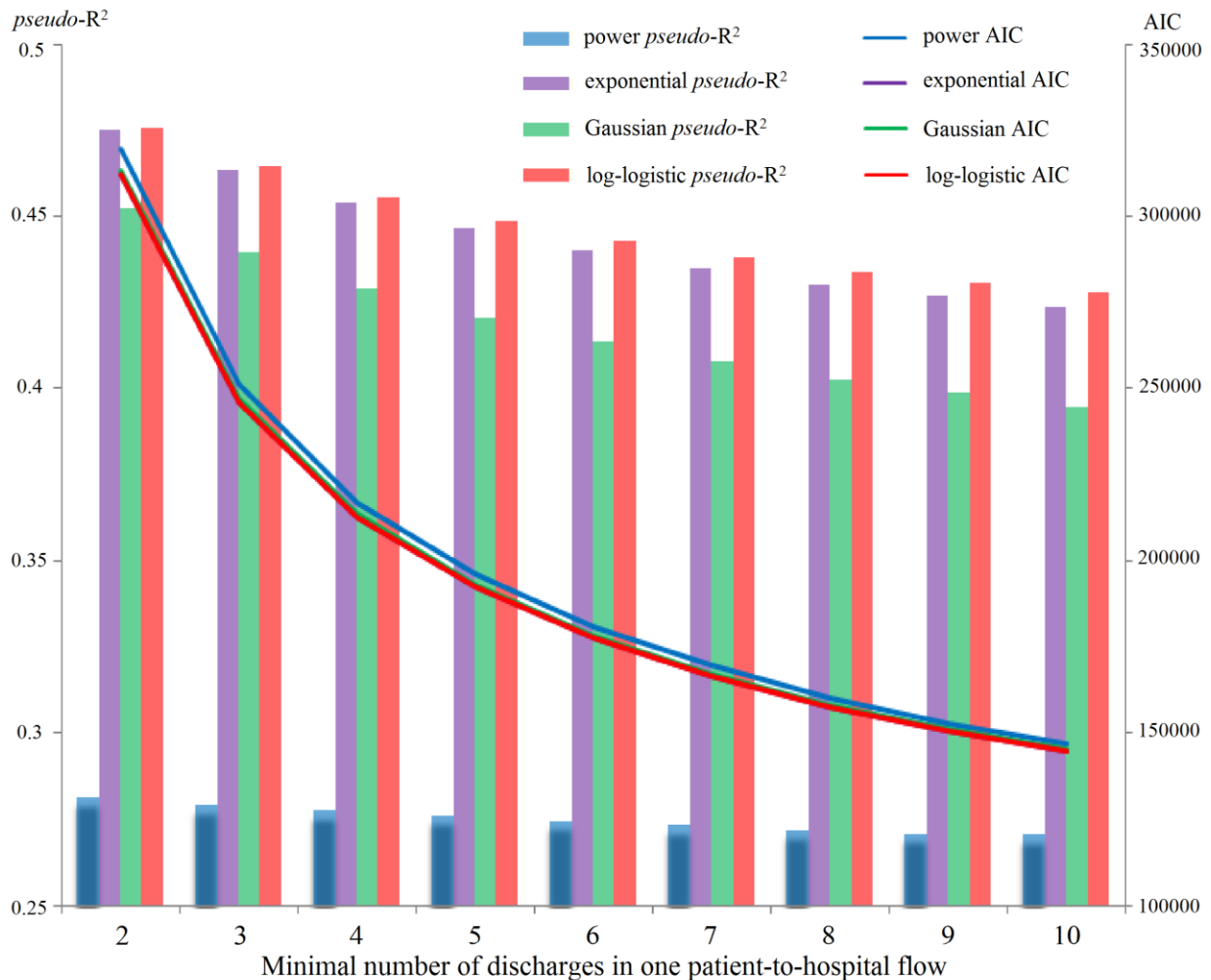


Figure 7. Comparison of the *pseudo-R*² and Akaike information criterion (AIC) of four distance decay functions

5.3 Travel patterns of subpopulations

Patients' travel patterns may vary by age, gender, race, socioeconomic status, health insurance, location of residence, and so on (Basu & Cooper 2000; Biello et al. 2010; O'Neill 2004). Disparities in travel pattern stem from varying responses by different groups of patients to increased spatial barriers to hospital care, which could have an indirect and long-term influence on the eventual outcomes for patients (O'Neill 2004). Moreover, travel behavior determines how reliant different groups of patients are on local hospitals, which makes them to different degrees vulnerable to the changes in local hospital setting (e.g., reduced number of beds, hospital closures) (Escarce & Kapur 2009).

The available information from the U.S. census and HCUP data allows examining the travel patterns by subpopulations in terms of race/ethnicity, socioeconomic status (SES), urbanicity, and discharge quarter. After coordinating the racial/ethnic categories in the U.S. census and SID, six primary categories are used to represent the race/ethnicity of the residents and patients: white, black, Hispanic, Asian, Native American, and other. Assuming that the population in each race/ethnicity is evenly distributed within blocks, the AWM is implemented to estimate the total population in each race/ethnicity over postal zones. The discharge records in each race/ethnicity can be extracted from the SID by race/ethnicity of inpatients and aggregated by travel flows. With the travel time and number of hospital beds known, and the total population over postal zones estimated and actual volume of hospital visits over travel flows calculated for each race/ethnicity, the travel pattern of each racial/ethnic subgroup can be modeled.

The median household income is used as an indicator for SES, and a national quartile classification of the estimated median household income of residents within postal zones is used

to mark each Florida postal zone exclusively as quartile 1 ($< \$39,000/\text{year}$), 2 ($\$39,000\text{--}47,999$), 3 ($\$48,000\text{--}62,999$), or 4 ($\geq \$63,000$). As for urbanicity, each postal zone is also exclusively assigned, on a basis of location, to large metropolitan (≥ 1 million residents), small metropolitan (50,000—1 million residents), micropolitan (10,000—49,999 residents) or rural areas. Therefore, simpler than the racial/ethnic subgroups that need to be disentangled from the mixed population within postal zones, the travel patterns of SES, urbanicity, and quarterly subgroups can be easily modeled without additional efforts to calculate the total population of the subgroups within each postal zone. The SID is correspondingly collapsed by patient's postal zone into four mutually exclusive SES and urbanicity subsets, with the volume of actual hospital visits over travel flows calculated for each subgroup. The travel patterns of quarterly subgroups are even more easily modeled, as the total population of the subgroups within each postal zone is the same as that of the overall population calculated by the AWM. Therefore, the SID is collapsed into four subsets by calendar quarter, and the travel pattern of each subgroup is modeled based on the corresponding subset.

All four distance-decay functions are used to model the travel pattern of each subgroup of discharges, and compared with one another with respect to *pseudo-R*² and AIC. The log-logistic function consistently outperforms the other three functions across all the subgroups. With p_i and b_j set as 10,000 and 100 respectively, an optimal log-logistic curve for each subgroup is drawn, which amounts to modeling the number of hospital visits from a zip code with 10,000 residents to a hospital with 100 beds. Each of four parameters independently impacts the shape of the curve in a different way. An increase in α or/and σ , as an exponent of the population within postal zones and number of hospital beds respectively, represents a larger number of hospital visits from a given postal zone to a hospital. An increase in the parameter θ similarly reflects

more hospital visits. However, as β increases, the distance decay effects become stronger with a more rapid decline in the number of hospital visits with travel time. If keeping the other three parameters constant, an increase in β can lead more patients to travel shorter and less patients to travel longer. The synergetic effects caused by the respective changes of four parameters are more complex than their independent impacts, which can be observed through the comparison among the fitting curves of the subgroups.

Blacks have stronger distance decay effects ($\beta = 2.35$) than Hispanics ($\beta = 1.94$) and whites ($\beta = 1.92$) (Table 6). Due to a relatively small number of discharges and *pseudo-R*² (not well fitted) for Asians, the smallest distance decay friction factor ($\beta = 0.96$) here might not be sufficient to reflect the actual travel pattern of Asians relative to other races/ethnicities. Although arbitrary, a constraint of 60 minutes for most patients traveling from patient to hospital has been justified by previous research (Delamater et al. 2013). Therefore, the average travel time of both all patients and those spending 60 minutes or less on their ways to hospitals is calculated for each subgroup (Table 6). For the overall patients, whites spend the longest travel time to hospitals on average (19.2 minutes), followed by Asians (16.7 minutes), blacks (14.7 minutes), and Hispanics (14 minutes). Moreover, a considerable number of whites travel to the hospitals more than 30 minutes away from their residence, while the numbers of patients traveling that far in other subpopulations have fallen to a trivial level at around 25 minutes (Figure 8). For those traveling 60 minutes or less, whites (14.2 minutes) and Asians (13.4 minutes) consistently travel longest, but Hispanics (11.6 minutes) conversely travel longer than blacks (11.5 minutes) on average, which corresponds to the comparison of the distance decay effects. Only a few discharges of Native Americans are included in 2011 Florida SID, so they are left out from sound conclusions.

Table 6. Parameters and assessment of the log-logistic functions modeling the travel patterns of subpopulations (Unit: minutes for Time_all and Time_60)

	N	Time_all	Time_60	α	σ	θ	β	<i>pseudo-R</i> ²	AIC
Race									
White	1,489,589	19.2	14.2	0.62	0.21	6.27	1.92	0.4677	249,238.5
Black	400,428	14.7	11.5	0.50	0.33	6.28	2.35	0.4860	98,804.4
Hispanic	399,395	14.0	11.6	0.58	0.19	4.44	1.94	0.4756	101,609.0
Asian	19,837	16.7	13.4	0.39	0.09	7.03	0.96	0.2775	17,153.1
Native American	3,343	23.9	14.2	-	-	-	-	-	-
SES									
0-25 th	827,281	16.4	11.4	0.53	0.35	6.53	2.47	0.5334	99,278.0
26 th -50 th	718,694	17.7	13.7	0.49	0.45	6.44	2.09	0.4762	86,750.3
51 st -75 th	617,952	18.3	14.5	0.53	0.33	5.98	1.92	0.4337	82,989.1
76 th -100 th	207,852	20.0	15.2	0.54	0.27	6.31	1.83	0.4213	37,724.2
Urbanicity									
Large city	1,497,608	13.9	12.1	0.52	0.37	5.50	2.05	0.4901	184,927.8
Small city	719,150	20.5	14.3	0.52	0.40	11.17	2.36	0.6766	88,827.8
Micropolitan	109,831	34.2	20.2	0.62	0.27	15.56	3.11	0.7744	21,747.0
Rural	50,154	50.9	28.0	0.45	0.59	9.23	1.50	0.3241	10,745.0
Quarter									
Jan-Mar	608,851	17.5	13.3	0.43	0.32	5.54	1.82	0.4405	146,648.3
Apr-Jun	583,737	17.7	13.3	0.41	0.33	5.63	1.84	0.4499	146,206.3
Jul-Sep	588,641	17.7	13.3	0.41	0.34	5.66	1.84	0.4494	146,504.3
Oct-Dec	595,514	17.7	13.3	0.42	0.32	5.75	1.85	0.4417	147,733.5

Time_all: Average travel time of all patients

Time_60: Average travel time of the patients spending 60 minutes or less to hospitals

A gradual decrease in distance friction factor and increase in average travel time are observed from the poorest to richest SES quarters (Table 6), which may imply two scenarios in which the better-off are more likely to travel longer than the poor. First, the better-off neighborhoods are usually at a farther distance away from hospitals than are inner-city ones. A plethora of wealthy neighborhoods do not have any hospital within their own zip code, and even within their neighboring postal zones. In this case, the better off have to bear with this spatial barrier. Second, the better off can afford traveling longer for better services, instead of being limited to choose only local hospitals. It can be observed apparently through the comparison of the curves that, despite a larger number of hospital visits occurring within 15 minutes in the

poorest quartile, fewer of them occur beyond 15 minutes from their residence relative to other quartiles (Figure 9).

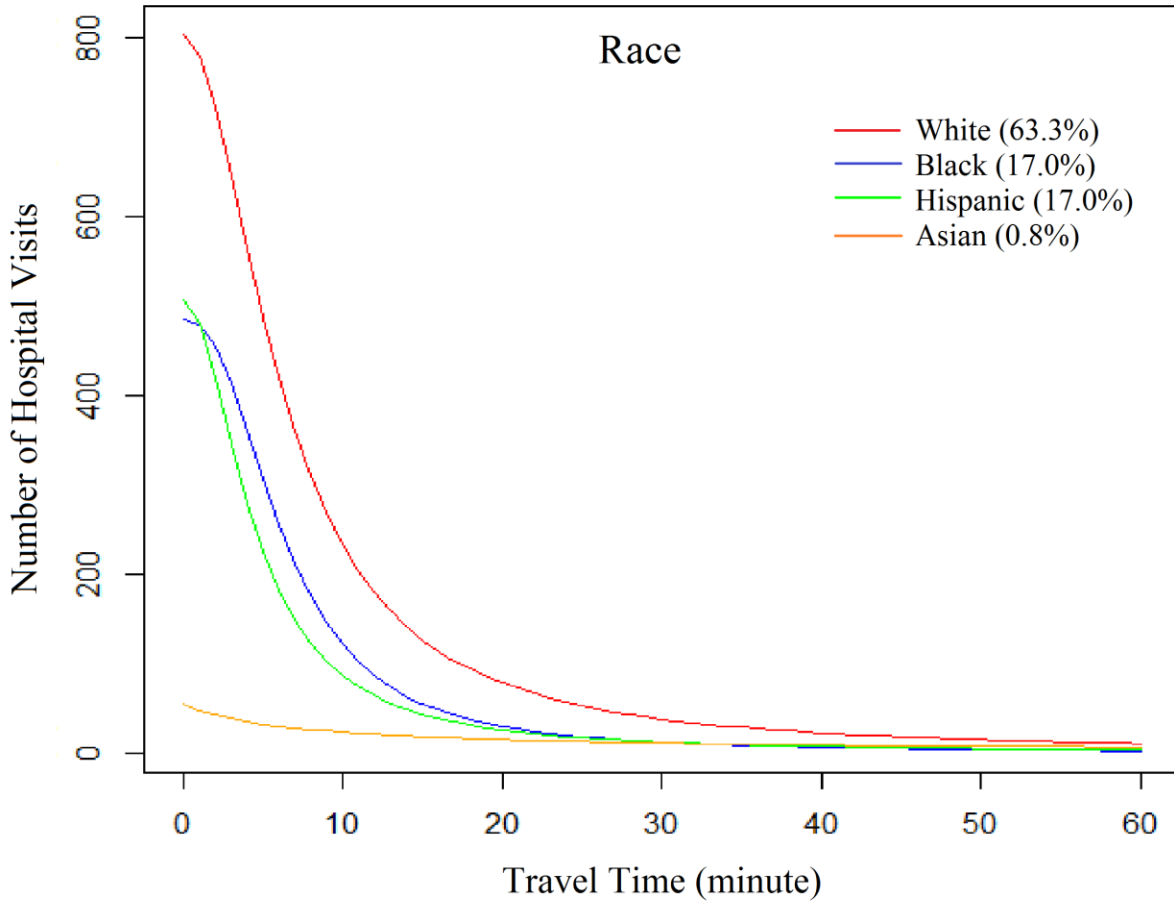


Figure 8. Travel patterns of the patients across racial/ethnic subgroups

In addition, geographic disparities in travel pattern are found, where the average travel time gradually increases with the residential location becoming less urbanized, from large metropolitan, small metropolitan, micropolitan, to rural areas (Table 6). Distinct from the trends in SES subgroups, different intervals of travel time are primarily dominated by various subgroups (Figure 10). Most hospitalizations in large metropolitan areas occurred within 10 minutes, especially within 5 minutes in which few hospital visits occur in the other three subgroups. Most patients in small metropolitan areas spend 5-20 minutes traveling to hospitals, while most micropolitan patients spend 5-30 minutes where the numbers of hospital visits

decline most rapidly with time ($\beta = 3.11$). The numbers of hospital visits decline more rapidly in small metropolitan areas ($\beta = 2.36$) than in large metropolitan areas ($\beta = 2.05$). The rural patients have the weakest distance decay effects ($\beta = 1.50$), with more hospital visits occurring at a distance of beyond 30 minutes than the other three subgroups. The soundness of this comparison, however, is hurt to some extents by an apparently smaller *pseudo-R*² in the rural subgroup, which implies an inadequate fitting for the travel pattern of rural patients.

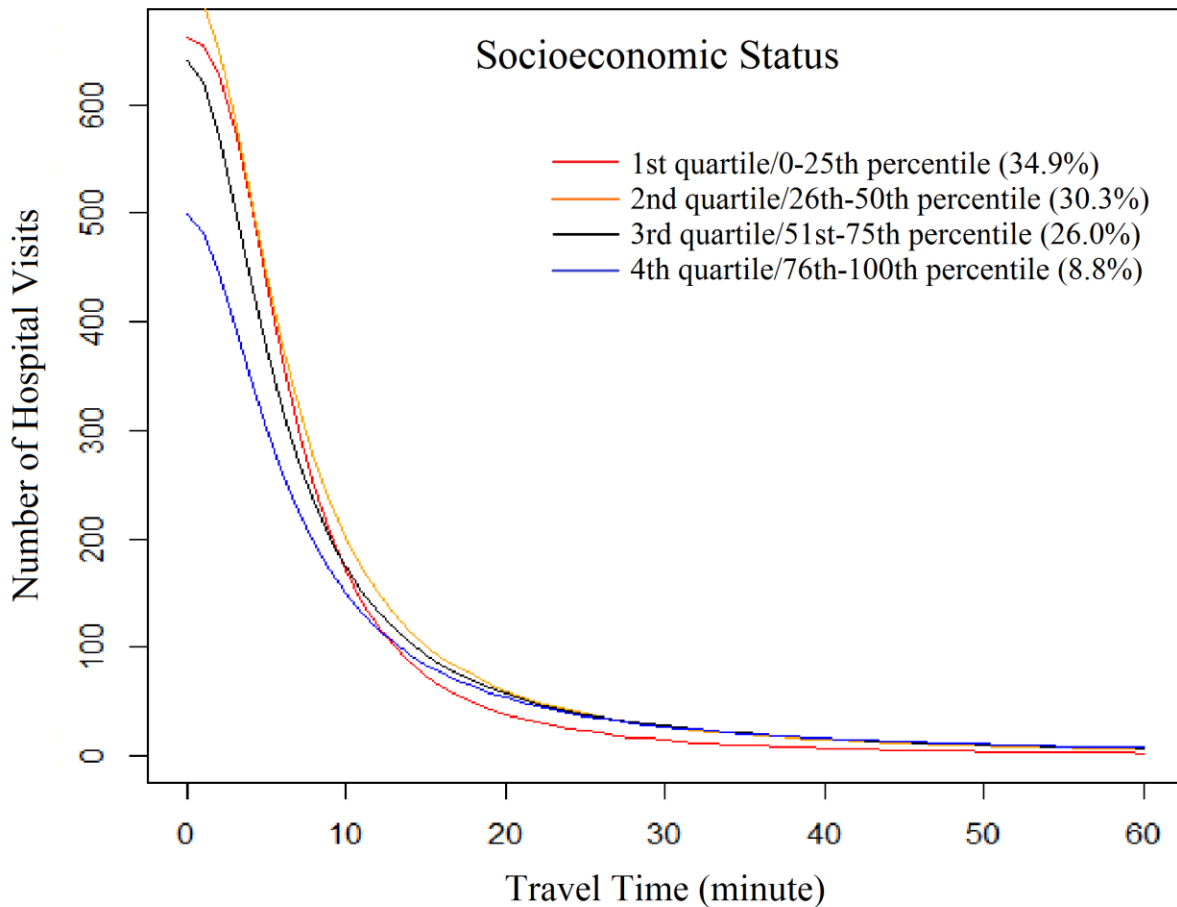


Figure 9. Travel patterns of the patients across socioeconomic subgroups

There are no significant differences found in the travel patterns of patients among four calendar quarters. The distance friction factor during October to December ($\beta = 1.85$) is slightly larger than the other three quarters (1.82-1.84), which might be random or explained by more seasonal residents who spend their winter in Florida while on vacation. The statistical results

here also correlate with the assumption in Chapter 3 that different components of permanent and seasonal residents at different times of the year do not impact the general travel patterns of the patients over a year.

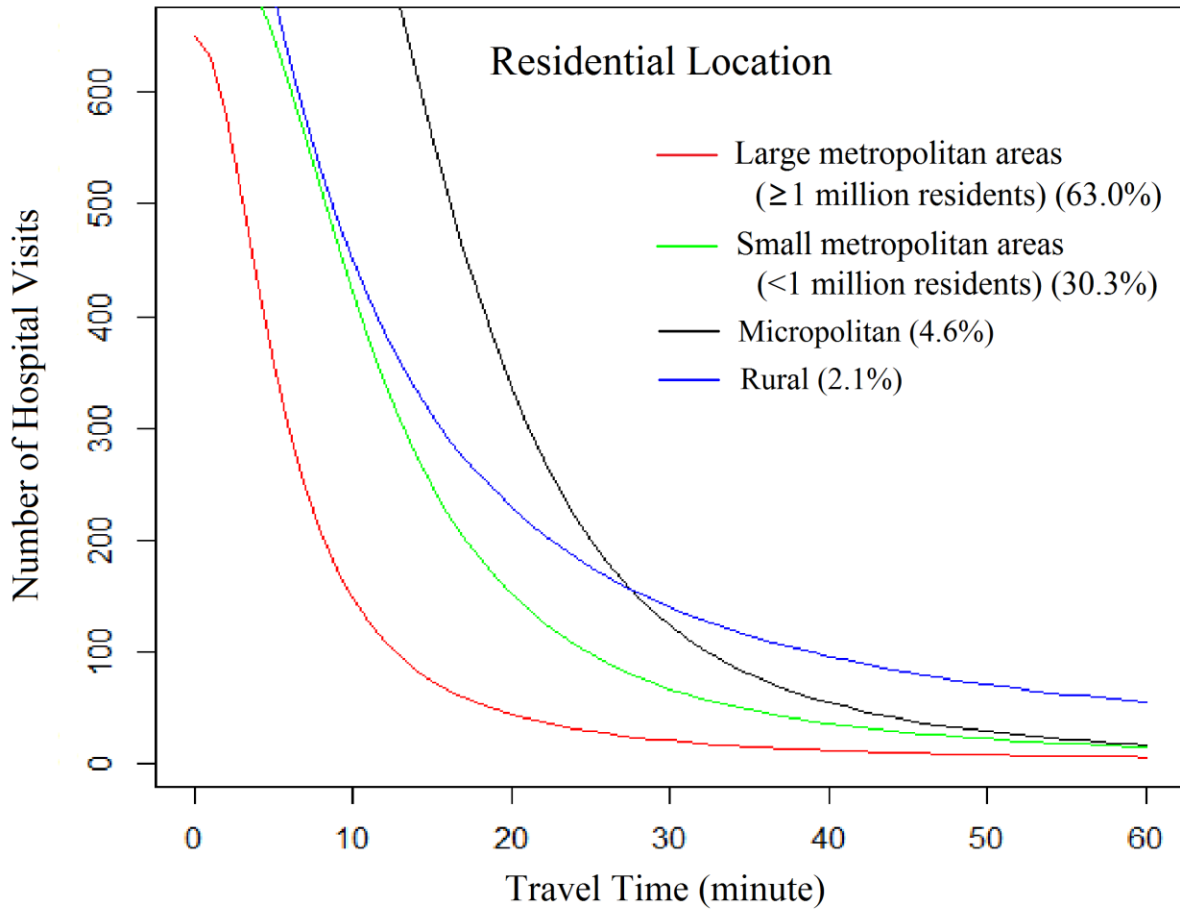


Figure 10. Travel patterns of the patients across socioeconomic subgroups

5.4 Summary

This chapter concludes that the log-logistic function better models the travel patterns of the overall inpatients than the other three commonly used forms (power, exponential, and Gaussian functions), and the outperformance is not simply attributable to its more parameters. The finding is consistent with the study of hospital utilization in Michigan in 2010 (Delamater et al. 2013). In addition to the overall patients, the log-logistic function also better fits the travel pattern of each race/ethnicity, SES, and urbanicity subgroup of patients, which enhances the

understanding of the impacts of these factors on the patients' travel patterns for hospitalization. Three key trends of subpopulations are repeated here: 1) whites spend the longest travel time to hospitals on average and their travel behaviors are least impacted by increasing travel time, followed by Asians, blacks, and Hispanics, but blacks have stronger distance decay effects than Hispanics; 2) patients' average travel time increases and distance decay effects decreases as their SES rises; and 3) patients gradually spend longer travel time to hospitals on average from large metropolitan to rural areas, and rural patients are least affected by increasing travel time, but apart from rural patients, the impacts of increasing travel time on patients' travel patterns gradually become stronger from large metropolitan to micropolitan areas.

Chapter 6 Delineating the HSAs by the Huff Model

The distance decay function for the overall patients derived in Chapter 5 is used in the Huff model, which is used for delineation of the HSAs. Based on the Huff-derived HSAs, linear regression is used to explore how the attributes such as socio-demographic variables are associated with the average travel time within the HSAs. Finally, the Huff-based HSAs are compared with the HSAs derived by the Dartmouth-Swiss method from Chapter 4 in terms of the LI (localization index), number of divisions, number of hospitals within divisions, compactness, and heterogeneity of internal SES and urbanicity.

6.1 The Huff model approach

The Huff model with the embedded distance decay function is used to delineate the HSAs. The probability of being discharged from hospital j to postal zone i , P_{ij} , is calculated as:

$$P_{ij} = T'_{ij} / \sum_{k=1}^n T'_{ik}$$

where T'_{ij} (T'_{ik}) is the expected volume of discharges from hospital j (k) to postal zone i calculated by $T'_{ij} = p_i^\alpha b_j^\sigma f(d_{ij})$ (or $T'_{ik} = p_i^\alpha b_k^\sigma f(d_{ik})$, where α , σ , and f are pre-determined), which represents the theoretical interaction between i and j (k), and n represents the total number of hospitals accessible to the patients in postal zone i . After the best-fit model with an optimal set of parameters is determined, each zip code is assigned to the hospital with the greatest probability of discharging the patients to that zip code ($\max [P_{ik}]$). Nevertheless, the denominator of the Huff model is identical for all the hospitals accessible to each zip code. For example, for a given zip code i , $\sum_{k=1}^n T'_{ik}$ would not change regardless of j . Therefore, each zip code can be simply assigned to the hospital with the largest interaction with that zip code ($\max [T'_{ik}]$).

The zip codes assigned to the same hospitals are merged to form a set of initial HSAs, and for ensuring the geographic continuity of each HSA, the following adjustments are conducted in order:

- 1) If a given initial HSA does not enclose the hospital, with which the initial HSA has the most intensive interaction, this initial HSA will be combined with the HSA, where that most attractive hospital resides. This step is repeated until each combined HSA encloses all hospitals that exert most intensive interactions with their component initial HSAs.
- 2) Any disconnected zip code without hospital(s) enclosed is re-assigned to the neighboring HSA including the hospital most intensively interacting with that zip code.

The above delineation and adjustment process is summarized in a flowchart in Figure 11. After the two-step adjustments, a set of continuous HSAs are produced, with each enclosing the hospital(s) most intensively interacting with it. The resultant HSAs and their attributes, such as population served, bed-to-population ratio, and average travel time, are discussed in next sections. This set of *Huff-based HSAs* is compared to the *2011 overall-derived HSAs* in Section 6.4.

6.2 Discussion on Huff-based HSAs

According to the Huff model approach, the 983 postal zones are allocated to 190 initial HSAs, and ultimately combined into 169 HSAs after the two-step adjustments described in Section 6.1. An obvious inconsistency can be found between HSA and county boundaries, where one county is normally divided into several HSAs, especially for those metropolitan counties (Figure 12).

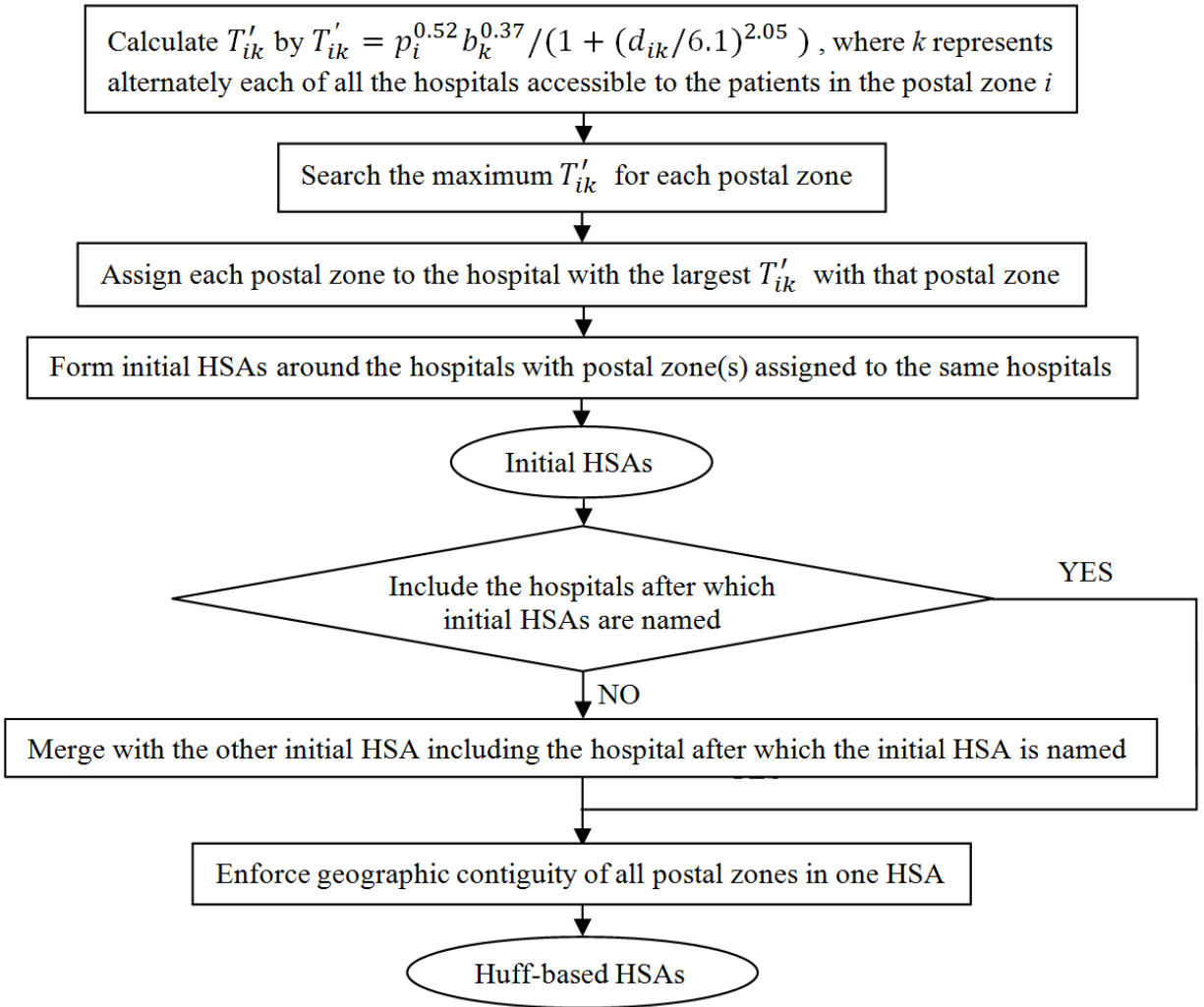


Figure 11. A flowchart of generating the Huff-based HSAs

The population size in a HSA varies from 6,326 (Port St Joe region) to 653,613 (Golden Lakes region), with the largest HSA having population more than 100 times the smallest one (Figure 13). Dividing the population by the total number of beds of all hospitals in a HSA yields the bed-to-population ratio, a primitive measure for disparity in hospital resource allocation. From Figure 14, lower bed-to-population ratios are found in some rural counties, such as Suwannee, Bradford, Glades, and Holmes and Washington County (the east side of Choctawhatchee River), as well as in some micropolitan counties, such as Hendry County, and in some suburbs of metropolitan areas, such as to the north of St. Petersburg (Safety Harbor), the

north of Tampa (Land O' Lakes), the northwest (Apopka) and east of Orlando (Union Park), and the west (Kendale Lakes) and south of Miami (Homestead). Most of the highest bed-to-population ratios are close to the inner city areas, such as Pensacola, Jacksonville, Orlando (Winter Park), the west coast of St. Petersburg (Indian Shores), Fort Lauderdale, south of Miami (Coral Gables), and so forth. Although the Union Park region of Orlando has a low bed-to-population ratio, it is surrounded by three facets of high ratios in the west, south and east.

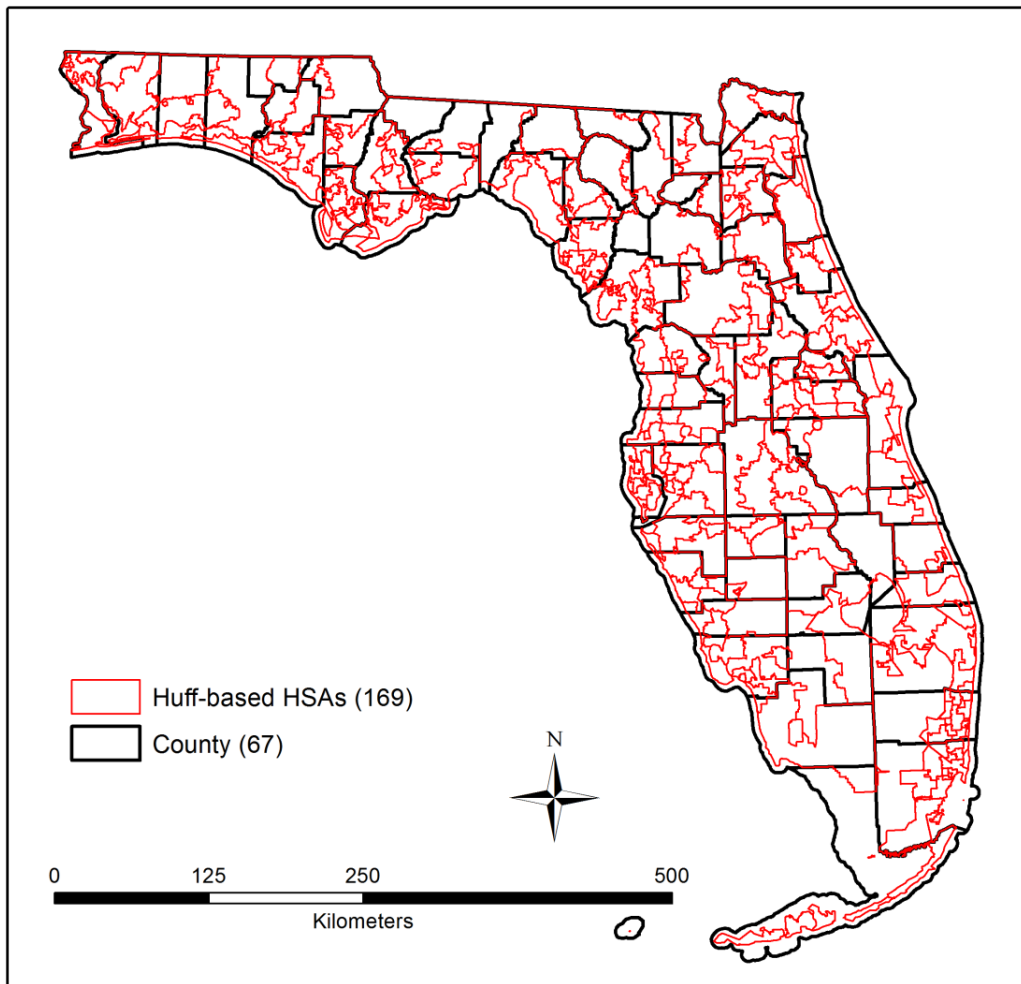


Figure 12. Boundaries of the Huff-based HSAs and counties in Florida

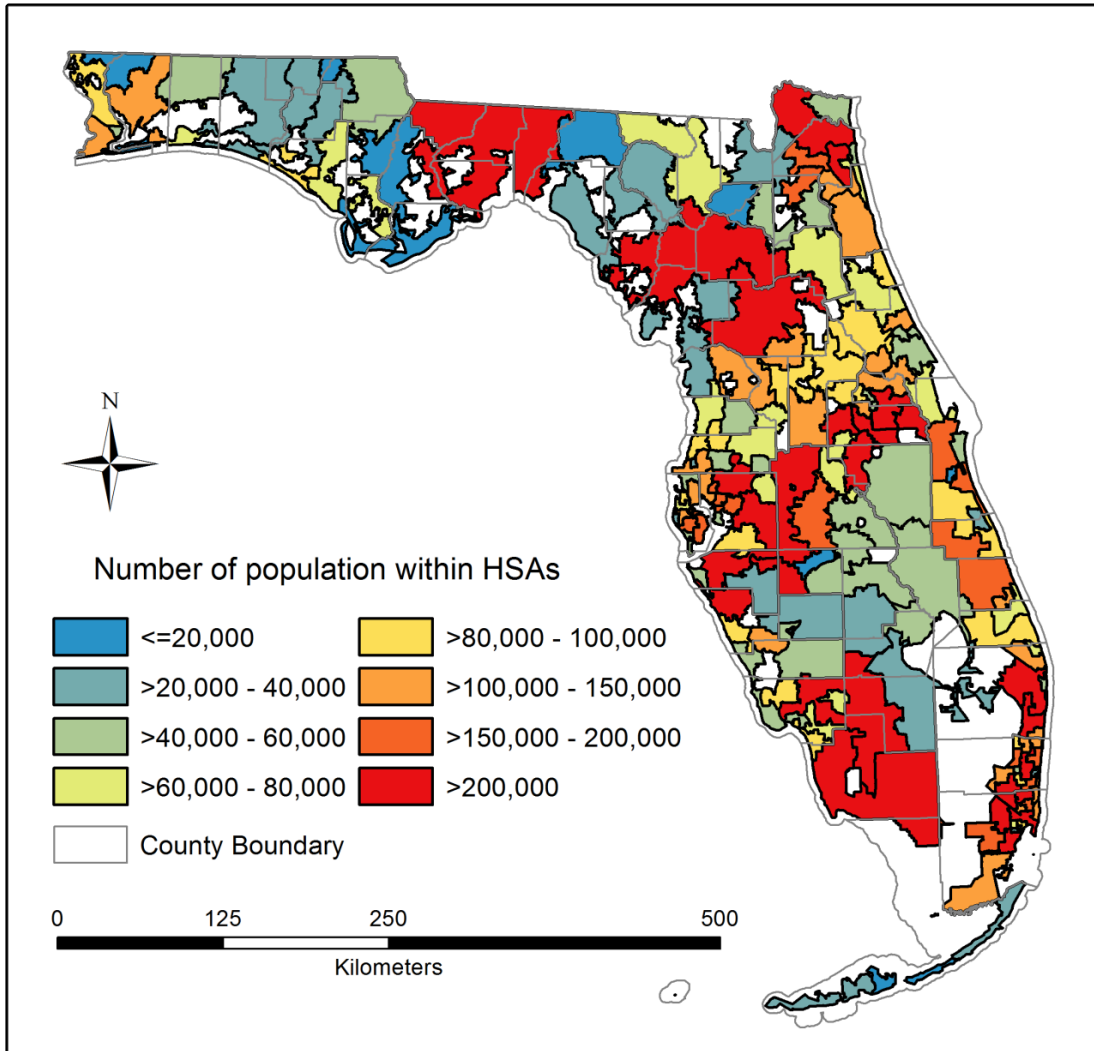


Figure 13. The number of the population within the Huff-based HSAs

The average travel time to hospitals by patients within HSAs ranges from about 6 to 100 minutes (Figure 15). General speaking, the patients in metropolitan areas on average spend less time traveling to hospitals than their counterparts in micropolitan and rural areas. Two of five HSAs with the average travel time beyond 60 minutes are located in rural counties (Franklin and Union County); one is located partly in rural areas (Glades County) and partly in micropolitan (Hendry County); one is in a micropolitan county (Monroe County); and another one is on the Pensacola Beach and Oriole Beach in Pensacola metropolitan area, which is most possibly due to the relatively isolated location. Ten of 14 HSAs with the average travel time less than 10 minutes

are in Miami-Fort Lauderdale-Pompano Beach metropolitan area, three in Tampa-St. Petersburg-Clearwater and one in Orlando-Kissimmee-Sanford metropolitan areas. This uneven distribution, on the other hand, shows that the average travel time to hospitals also varies by metropolitan area. The patients in Orlando-Kissimmee-Sanford metropolitan area on average travel longer than their counterparts in the other three metropolitan areas in Florida. The patients in North Florida, in particular Florida Panhandle, generally spend longer travel time to hospitals than those in Central and South Florida. This spatial variability will be examined in more depth in the next section.

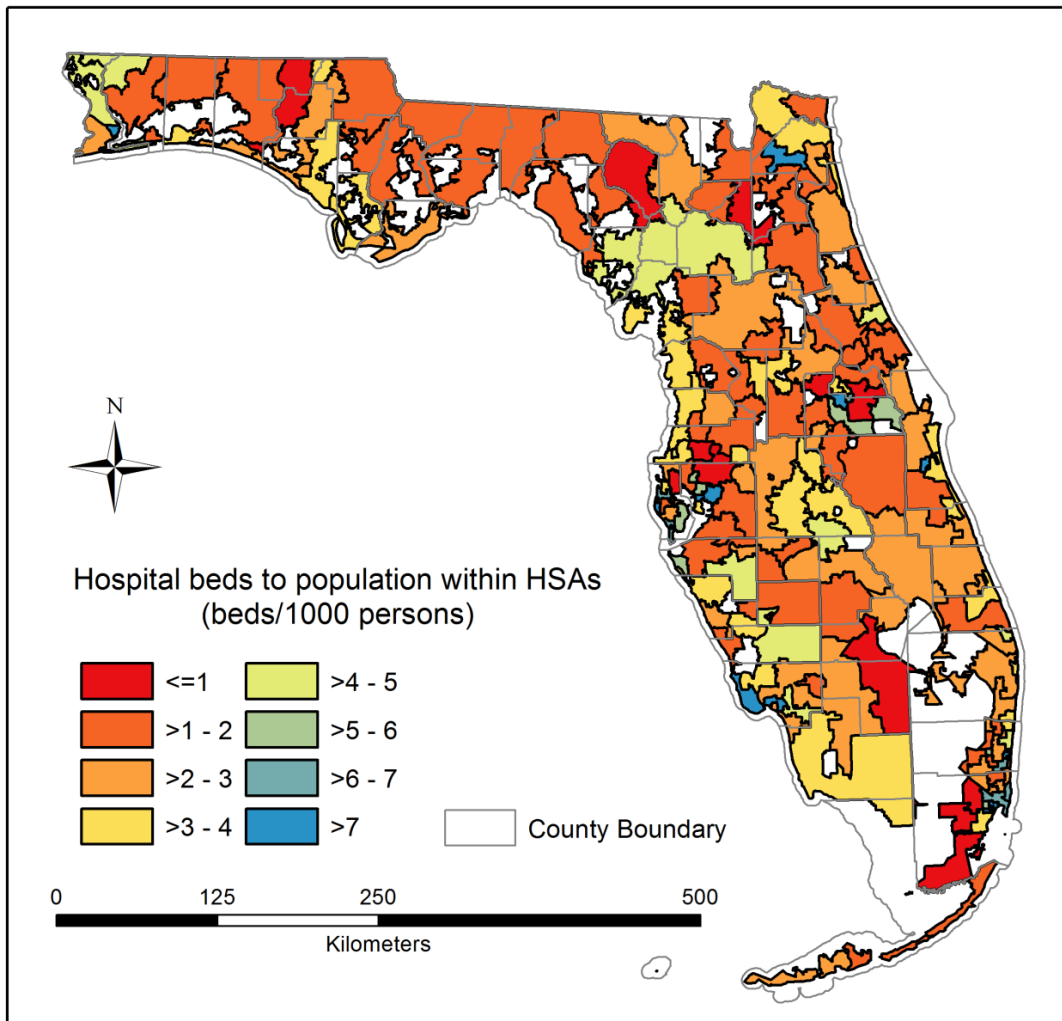


Figure 14. The ratio of hospital beds to population within the Huff-based HSAs

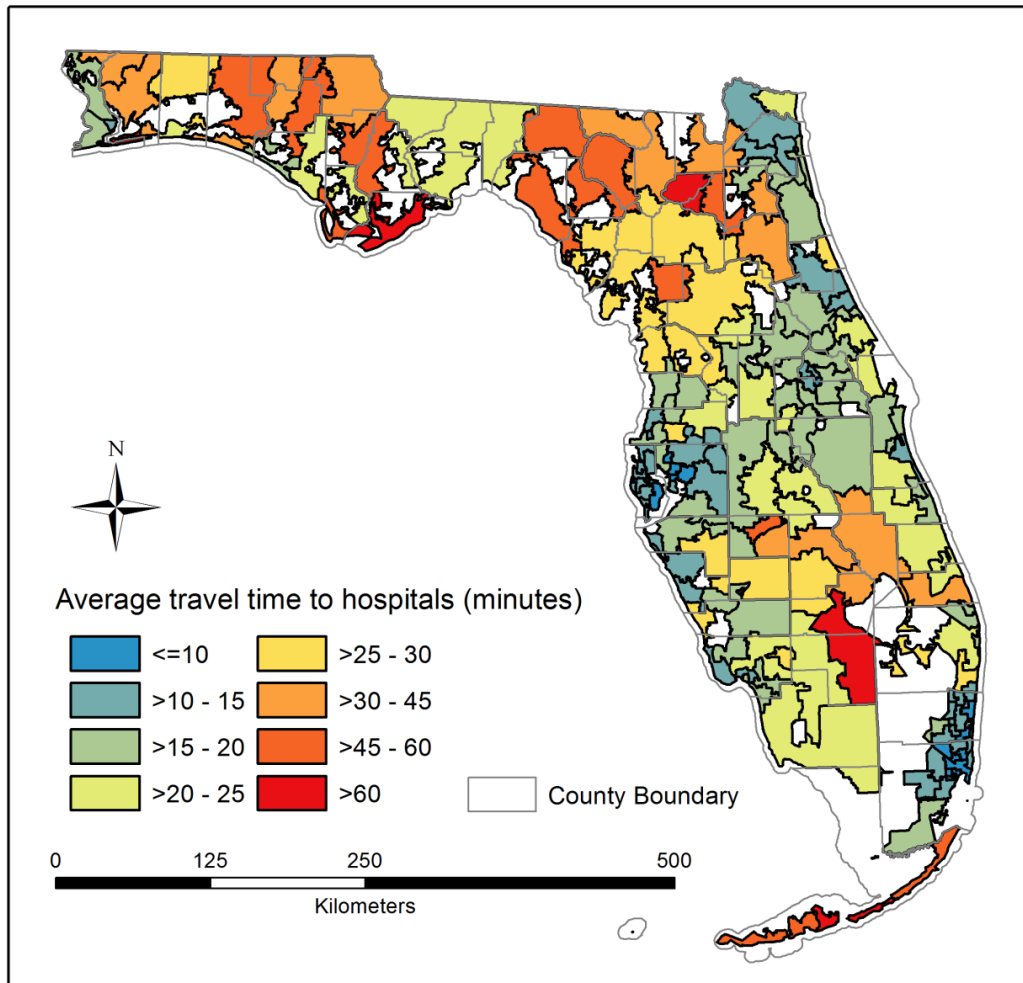


Figure 15. The average travel time of the patients to hospitals within the Huff-based HSAs

6.3 Travel time within the HSAs

As discussed in the previous section, the average travel time by patients varies a great deal across the Huff-based HSAs. One goal of delineating HSAs is to encourage patients to seek care from local hospitals within their HSAs and thus reduce the travel time. This section uses linear regression analysis to explore the underlying factors for this variability. Limited by data availability and measurement feasibility, several variables are proposed. As shown in Table 7, variables such as *area*, *compactness*, *population size*, *bed-to-population ratio*, *number of general and specialized hospitals* within the HSAs are straightforward and can be easily calculated. The remaining variables and their calculation methods are worth further clarification.

Urbanization ratio is calculated as the proportion of urban population (within urbanized areas or urban clusters) over total population in each HSAs, describing its urbanization level (Wang, Wen & Xu 2013). The urban population is aggregated from the block groups with their centroids located in either urbanized areas or urban clusters, defined by the U.S. census. *Poverty rate* is based on the American Community Survey 5-Year Estimates. The *Gibbs-Martin index* (Gibbs & Martin 1962) is used to measure the racial/ethnic heterogeneity in an HSA (Table 7). A value of 0 represents the presence of only one racial/ethnic group, and a value of near 1 represents the maximum heterogeneity by infinite groups (here, the maximum Gibbs-Martin index is smaller than 1 as six racial/ethnic groups are considered). Other factors such as insurance types and other socioeconomic status (SES) variables in addition to poverty rate may also be associated with average travel time of the residents due to their impacts on hospital selection. Due to data limitation, these variables are not considered in this exploratory study.

Table 7. Candidate influential factors for patients' average travel time within the HSAs

Variable	Description	Calculation
area	Area (km ²)	Being automatically calculated in ArcGIS
compact	Compactness	Circumference of equal area circle/Perimeter of area
pop	Population size (million)	Being Aggregated by the zip codes within each HSA
Bratio	Hospital bed-to-population ratio	Hospital beds/Population size
Ghosp	Number of general hospitals	Number of general hospitals within each HSA
Shosp	Number of specialized hospitals	Number of specialized hospitals within each HSA
Uratio	urbanization ratio	Urban population/Population size
poverty	Poverty rate	Population below poverty level/Population size
age	Mean age of patients	Mean age of the patients within the HSAs
HRR	Heterogeneity of races/ethnicities of residents	$1 - \sum_{i=1}^n p_i^2$

p_i is the proportion of the residents in the i th racial/ethnic category within a given HSA, and n is the number of categories within that HSA.

A simple bivariate regression is run between each of the aforementioned variables and average travel time. As shown in Table 8, mean age or racial-ethnic heterogeneity is insignificant ($p = 0.165, 0.09$ and 0.154 , respectively), number of general hospitals or poverty rate is significant at the level of 0.05 , and all other variables are significantly associated with the average travel time at the level of 0.01 . Prior to including all significant factors in one regression, the correlations between independent variables are examined. A high correlation is found between the number of general and specialized hospitals, hence the number of general hospitals is removed from the final regression due to standalone insignificance at the level of 0.05 .

Table 8. Coefficients and significance of the candidate influential factors for patients' average travel time within the HSAs

Variable	Standalone		Stepwise	
	Coefficient	Significance	Coefficient	Significance
area	0.005	<0.001	—	
compact	-40.909	<0.001	-24.494	<0.001
pop	-55.386	<0.001	—	
Bratio	-1.294	<0.001	-37.200	<0.001
Ghosp	-3.569	0.018	removed	
Shosp	-8.736	<0.001	-3.699	0.003
Uratio	-45.884	<0.001	—	
poverty	51.786	0.017	—	
age	-0.283	0.165	—	
HRR	-12.615	0.090	—	

Abbreviations: —, insignificant at the level of 0.05 in the bivariate regressions

The stepwise regression is used to eliminate insignificant variables and examine the joint multivariate effects. After the stepwise process, three variables remain significant ($p < 0.01$) in the final model: HSA compactness ($p < 0.001$), urbanization ratio ($p < 0.001$), and number of specialized hospitals within the HSAs ($p = 0.003$). The *adjusted R*² are 0.533 . In other words, the

combination of the three variables explain more than half of the variation in average travel time across the HSAs. All the three variables are negatively associated with average travel time: (1) more regular and compact HSAs tend to have shorter average travel time; (2) patients in more urbanized areas are more likely to travel shorter than their counterparts in more rural settings; and (3) more specialized hospitals within the HSAs help reduce patients' average travel time.

6.4 Comparison of the HSAs

The 169 Huff-based HSAs are compared with the 78 overall-derived HSAs from Chapter 4 (Section 4.2) in terms of number of divisions, number of hospitals within divisions, compactness, and heterogeneity of internal SES and urbanicity (Table 9). The comparisons of the means and ranges reveal that the Huff-based HSAs on average have (1) fewer hospitals included within each division, (2) more regular and compact shapes, and (3) more homogeneous SES and urbanicity within each division. All are favorable regionalization attributes for the Huff-based HSAs.

Table 9. Comparisons of the mean values of different parameters within the Huff-based HSAs and 2011 overall-derived HSAs (minimum value — maximum value)

	Huff-based HSAs	2011 overall-derived HSAs
Number of divisions	169	78
Number of hospitals within divisions	1.3 (1–7)	2.8 (1–12)
Compactness	0.49 (0.20–0.79)	0.48 (0.16–0.77)
Heterogeneity of internal SES	0.43 (0–0.75)	0.51 (0–0.75)
Heterogeneity of internal urbanicity	0.06 (0–0.59)	0.08 (0–0.61)

Theoretically, the core hospital(s) within a given Huff-based HSA discharge(s) the largest percentage of records to that HSA unit relative to the discharges to other HSAs. Therefore, it does not necessarily mean that the core hospital(s) within an HSA discharge(s) the most records ($\geq 50\%$) to that HSA. For a given HSA, it is possible that the percentages of the records discharged to other units during a specific period, perhaps not largest separately, may add up to

more than 50% and disqualify that HSA for being an independent HSA, according to the definition by the Dartmouth approach based on the LI. Therefore, the hospital discharge records in 2011 are also used for examining, if following the same Dartmouth rules, whether the Huff-based HSAs can lead to a comparable product with the flow-based *2011 overall-derived HSAs*. The LIs are calculated for the Huff-based HSA units based on hospital discharges in 2011. Each HSA with $LI < 0.5$ is merged into the adjacent HSA from which the largest proportion of records occurring outside of its own HSA is discharged, until the LIs of all the combined HSAs are equal or larger than 0.5, also referred to as the *2011 Huff-based HSAs*. Lastly, the LIs of the 2011 Huff-based and overall-derived HSAs are compared by a *t*-test, prior to which a natural log transformation is performed to reduce the influences of the skewed distribution of LI values.

After enforcing $LI \geq 0.5$, the 169 Huff-based HSAs with LIs ranging from 0.01 to 0.95 in 2011, are further aggregated into 76 HSA units with LIs from 0.50 to 0.97. Compared to the 2011 overall-derived HSAs, both sets of HSAs have a mean LI of 0.65. Despite their distinctive configurations (Figure 16), no statistical difference in their LIs is found by both *t*-test (Table 10) and histogram (Figure 17). This indicates that the Huff-based HSAs could be integrated with year specific hospital records to produce a set of adequate HSAs for a given period.

Table 10. Results of *t*-test for the differences of the localization index (LI) values of the 2011 overall-derived and Huff-based HSAs

	Levene's Test		<i>t</i> -test for Equality of Means						
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Lower	95% Upper
Equal variances assumed	0.099	0.754	0.020	152	0.984	0.001	0.029	-0.057	0.058
Equal variances not assumed			0.020	151.338	0.984	0.001	0.029	-0.057	0.058

*95% confidence interval

6.5 Summary

This chapter shows that the Huff model, with a built-in distance decay function capturing patient's travel behavior, serves as a solid foundation to define HSAs. The Huff-based HSAs have multiple advantages over the HSAs derived solely from patients' travel flow data: more regular and compact shapes of the divisions, and more homogeneous SES and urbanicity within the divisions. More than half of the variation in average travel time by patients across the HSAs can be explained by a combination of the HSA compactness, urbanization ratio, and number of specialized hospitals within the HSAs.

Several limitations need to be addressed in future work. First, more studies are needed to evaluate whether the log-logistic function is also the best-fitting distance decay function for the patients' travel patterns in other states and over a certain period of time. Secondly, the stability of the Huff-based HSAs over time needs to be assessed. Finally, some of the Huff-based HSAs have $LI < 0.5$. More in-depth analysis is needed to examine why hospitals in these areas are not enough to attract local patients (e.g., restrictions for admitting Medicare patients, more attractive hospitals in neighboring HSAs).

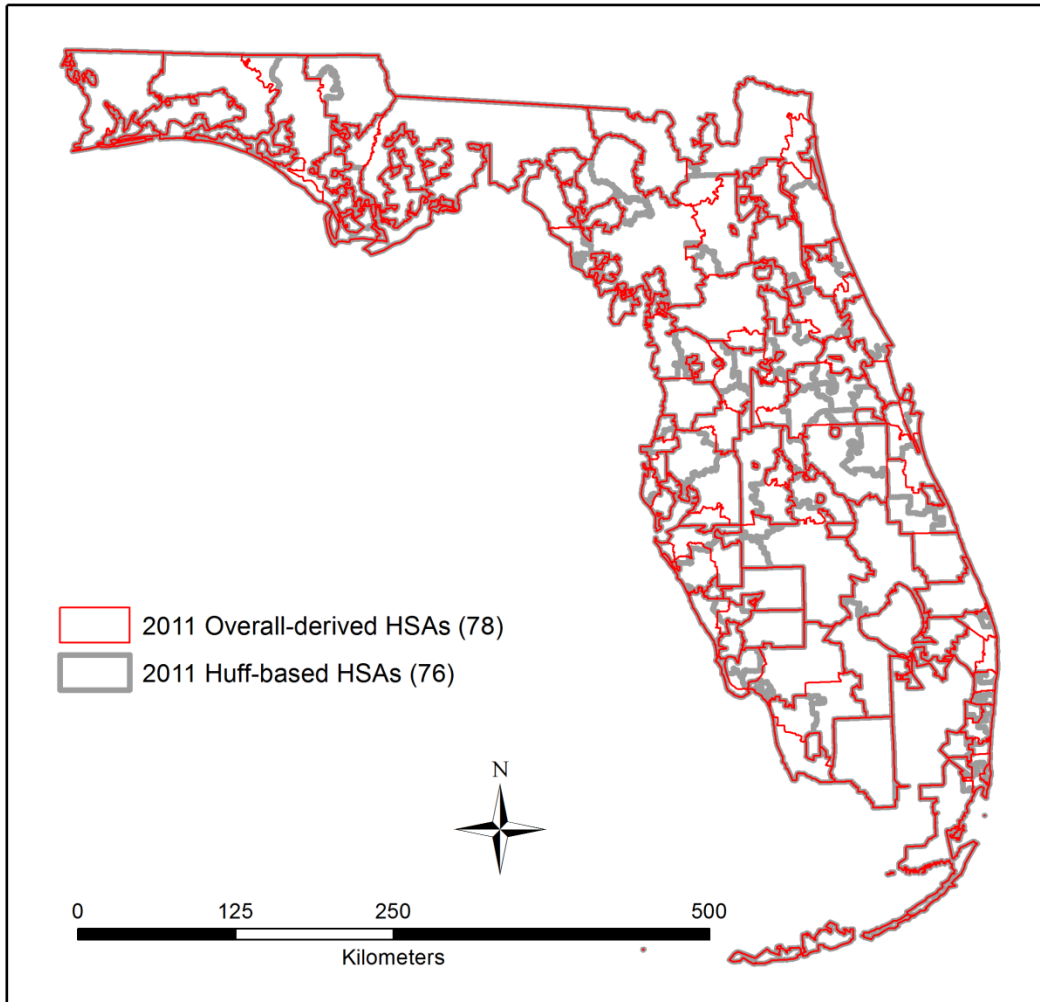


Figure 16. Comparison of the boundaries between the 2011 Huff-based and overall-derived HSAs in Florida

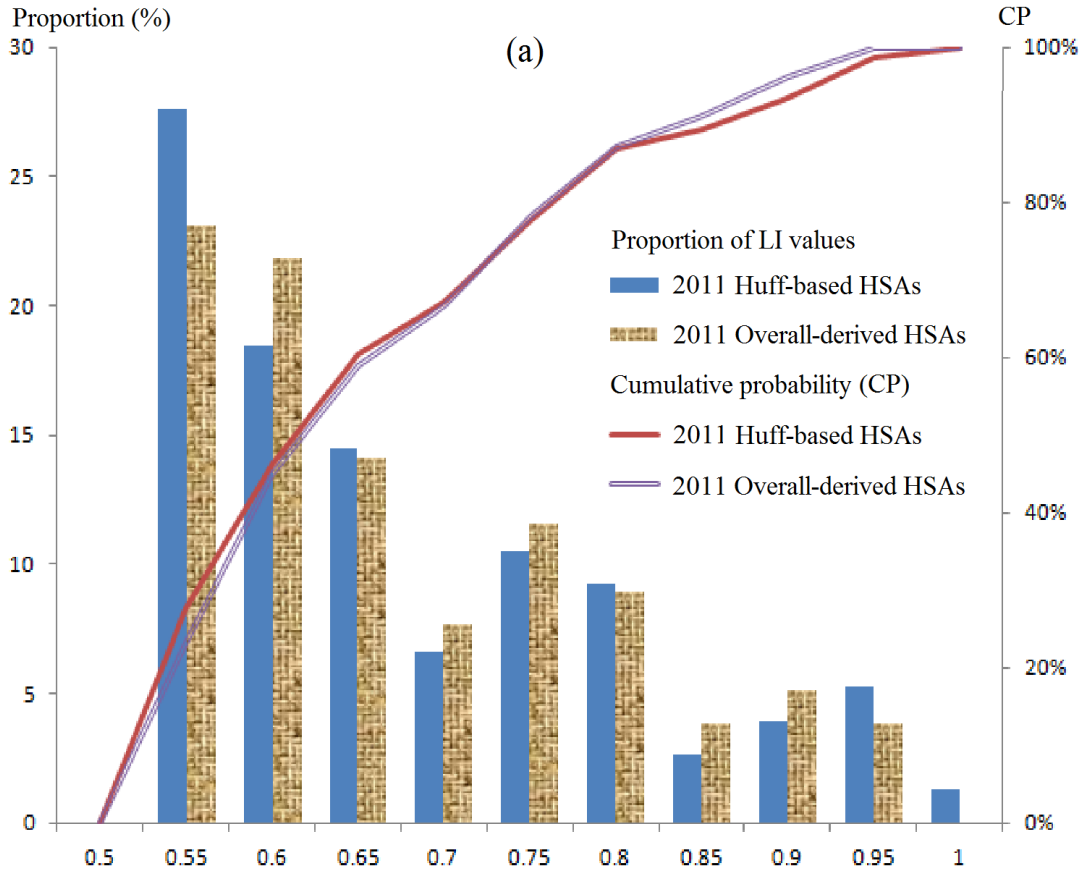


Figure 17. Comparisons of proportion and cumulative probability of the localization index (LI) values between the 2011 Huff-based and overall-derived HSAs in Florida

Chapter 7 Constructing Hierarchical HSAs

The hierarchical HSAs (HSAs-HRRs), in addition to delineating the catchment areas for different levels of hospital services, strengthen the connections between levels of hospitals, coordinate referrals for limited hospital resources, and save patients travel cost and waiting time. However, the value of the most widely used *Dartmouth HSAs-HRRs* at present (Center for Evaluative Clinical Sciences 1999) might be diminished by the outdated and unrepresentative Dartmouth HSAs (Jia, Xierali & Wang 2014). Moreover, the current approach to producing the *Dartmouth HSAs-HRRs* makes it difficult to update the hierarchical system, especially the higher-level HRR boundaries that are based on the configuration of HSAs. This chapter analyzes the travel patterns of patients seeking specialized cares (cardiovascular surgeries and neurosurgeries), based upon which the Huff-based HRRs and hierarchical HSAs-HRRs are delineated.

7.1 Travel patterns of patients for specialized cares

In accordance to the Dartmouth HRRs, the *specialized patients* in this study are defined as the patients receiving at least one surgery on the cardiovascular (cardio) or nervous (neuro) systems in 2011, versus the overall patients. The Clinical Classification Software (CCS) is developed based on the *International Classification of Diseases, 9th Revision, Clinical Modification* (ICD-9-CM), for aggregating patient diagnoses and procedures into a manageable number of clinically meaningful categories. The corresponding attributes in the SID are DXCCS (diagnosis classification) and PRCCS (procedure classification). A total of 377,601 specialized discharge records are extracted from the 2011 SID in Florida, which include 358,487 cardio discharges (PRCCS code between 43 and 63) from 214 hospitals and 22,673 neuro discharges (PRCCS code between 6 and 9) from 181 hospitals, with 33,559 discharges receiving both types

of surgeries during their hospitalization (Agency for Healthcare Research and Quality 2014). Cardiovascular surgical services are available in all 181 hospitals providing neurological services.

The interactions between specialized patients and hospitals, represented by the volume of the specialized discharges from the hospital j to the postal zone i (T_{ij}^S), are formulated as

$$T_{ij}^S = p_i^\alpha b_j^\sigma f(d_{ij})$$

where p_i is total population in postal zone i , b_j is number of beds in hospital j , d_{ij} is travel time from zip code i to hospital j in minutes, $f(d_{ij})$ is a distance decay function (Table 5), and α and σ are the parameters to be estimated.

By aggregating all the specialized discharge records by zip code and hospital, there are totally 979 (out of 983) zip codes containing specialized discharges from 214 hospitals. A total of 209,596 (979×214) estimated patient-to-hospital flows are formed, 13,375 of which include at least one discharge in 2011. Similar to identifying the travel patterns underlying the HSAs in Section 5.2, all the flows including only one specialized discharge are excluded as outliers, and the four functions are used to model the travel patterns of specialized patients based on the remaining flows in R (Table 11). To take into account the stabilization of comparison of the performance among four functions, the *pseudo-R*² and AIC are calculated repeatedly as the minimum number of discharges in one flow increases from two to ten iteratively (Figure 18). Despite proximate *pseudo-R*² and AIC, the log-logistic function slightly outperforms the exponential function due to a higher *pseudo-R*² and lower AIC, hence is used as the best-fitting function for all the flows including at least two discharges. The optimal combination of parameters is $\alpha = 0.36$, $\sigma = 0.35$, $\theta = 5.6$ and $\beta = 1.65$ (Table 11). Therefore, the estimated interactions between postal zone i and hospital j ($T_{ij}^{S'}$) are formulated as:

$$T_{ij}^{s'} = p_i^{0.36} b_j^{0.35} / (1 + (d_{ij}/5.6)^{1.65})$$

where the parameter values are all statistically significant ($p < 2 \times 10^{-16}$).

Table 11. Optimal parameters and assessment of the four candidate functions based on the specialized (cardiovascular and neurological), cardiovascular (Cardio), neurological (Neuro), and overall inpatients (All)

Function		α	σ	θ	β	$pseudo-R^2$	AIC
Power	Specialized	0.32	0.35	-	0.48	0.2217	117,299.0
	Cardio	0.33	0.33	-	0.48	0.2168	112,592.3
	Neuro	0.11	0.25	-	0.24	0.0553	21,640.3
	All	0.45	0.43	-	0.56	0.2788	319,604.2
Exponential	Specialized	0.36	0.35	-	0.12	0.3372	115,639.5
	Cardio	0.36	0.34	-	0.12	0.3331	110,995.4
	Neuro	0.08	0.26	-	0.02	0.0627	21,617.4
	All	0.53	0.38	-	0.14	0.4742	312,474.0
Gaussian	Specialized	0.33	0.33	8.34	-	0.3117	116,029.3
	Cardio	0.34	0.32	8.14	-	0.3087	111,351.3
	Neuro	0.08	0.23	40.99	-	0.0513	21,652.5
	All	0.51	0.35	7.54	-	0.4503	313,479.4
Log-logistic	Specialized	0.36	0.35	5.60	1.65	0.3440	115,534.0
	Cardio	0.36	0.34	5.63	1.69	0.3394	110,902.5
	Neuro	0.09	0.28	19.96	0.79	0.0676	21,604.3
	All	0.52	0.37	6.10	2.05	0.4744	312,467.1

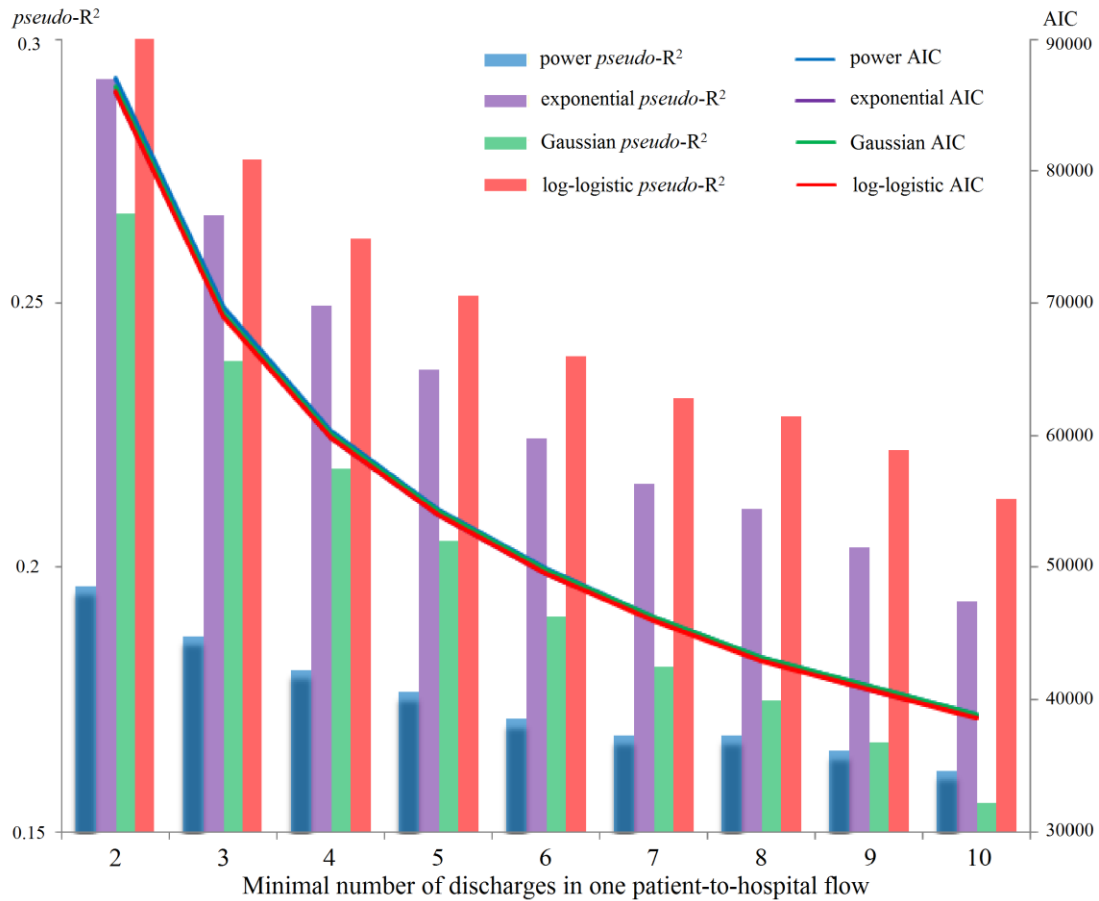


Figure 18. Comparison of the *pseudo-R*² and Akaike information criterion (AIC) of four distance decay functions for specialized patients

In addition, all the four functions are used to model the travel patterns of the cardio and neuro patients respectively, and the log-logistic function is found to outperform the other three functions consistently (Table 11). The neuro patients ($\beta = 0.79$) demonstrate a weaker distance decay effect than cardio patients ($\beta = 1.69$), which reveals that different types of specialized patients may have different degrees of responses in distance (travel time) decay. A longer average travel time by neuro than cardio patients (29.3 versus 19.6 minutes) may also be attributable to the fact that, relative to the number of the hospitals providing cardiovascular surgical procedures, 33 fewer hospitals provided neurosurgery services in Florida in 2011. In this case, according to the central place theory (CPT), cardio services are not as specialized as neuro services due to a larger number of providers. However, a very small *pseudo-R*² (=0.0676 even by

the best-fitting function) implies that the distance decay effect is not strongly present for neuro patients. Incorporating neuro into cardio discharge records decreases the decay factor β from 1.69 to 1.65 without changing the *pseudo-R*² significantly. Therefore, modeling based on the cardio and neuro data combined is a good balance for reflecting the travel patterns of the specialized patients.

The models with the optimal parameters produced based on the overall discharge records (Table 5) are also listed in Table 11 as reference. The distance decay friction factors for cardio and neuro patients, separately or jointly, are lower than those for the overall patients in the power (0.48 versus 0.56), exponential (0.12 versus 0.14), and log-logistic function (1.65 versus 2.05). This confirms that the specialized patients tend to travel longer than general patients for obtaining their corresponding services. Keeping p_i and b_j constant (set both as 1), the best-fitting log-logistic functions for the specialized and overall patients are drawn for comparison (Figure 19). As travel time increases, the number of specialized patients decreases at a lower rate than that of the overall patients.

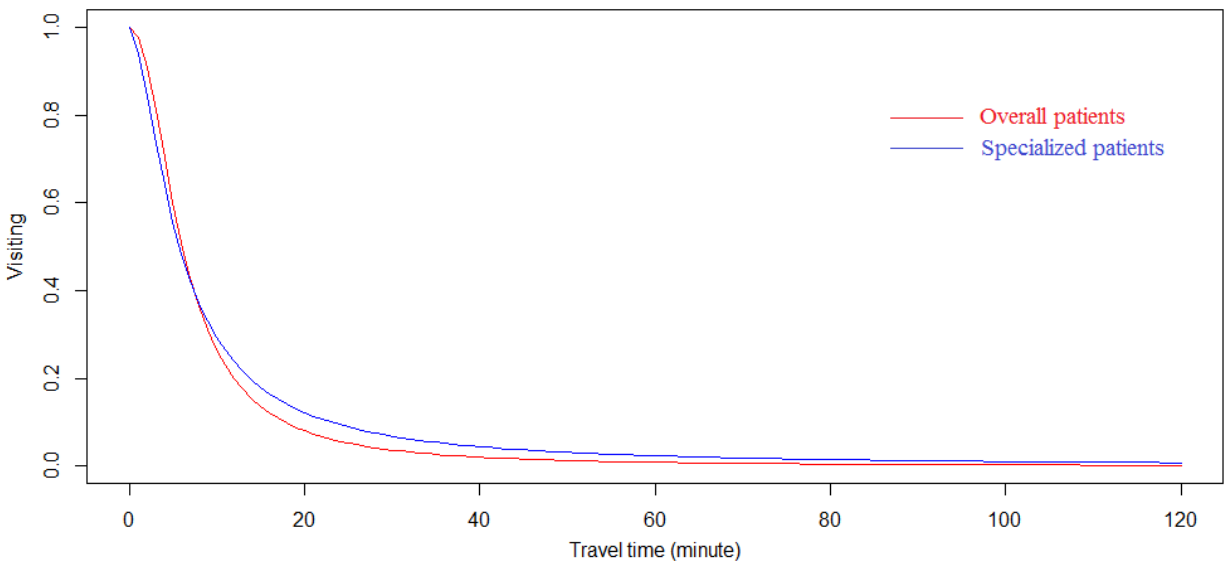


Figure 19. Comparison of the travel patterns of the specialized and overall inpatients by log-logistic function

7.2 Huff-based HRRs and hierarchical HSAs

Figure 20 is a flowchart for the delineation of the HRRs using the Huff model approach and post-delineation adjustments. 181 hospitals providing both cardiovascular and neurological services, also termed *specialized hospitals*, are included in the following analysis. According to the Huff model with the embedded best-fitting function, the expected interactions between each zip code and specialized hospital ($T_{ij}^{S'}$) are computed. Next, each zip code is assigned to the hospital with the greatest interactions with it, and the zip codes assigned to the same HRRs are agglomerated to form 165 initial HRRs. Then, the following adjustments are conducted in order:

- 1) for a given initial HRR that does not enclose the hospital most intensively interacting with that HRR, it would be combined with the HRR including that most attractive hospital. This step is repeated until each combined HRR includes all the hospitals most intensively interacting with each of its component initial HRRs.
- 2) any disconnected zip code without hospital(s) enclosed is re-assigned to the neighboring HRR including the hospital most intensively interacting with that zip code.

After the two-step adjustments, a total of 145 continuous HRRs are produced, with each including the hospital(s) most intensively interacting with it. Population in the preliminary HRRs varies from 13,605 to 672,866 (Figure 21), with the largest HRR almost 50 times the smallest one. A minimum population size of 120,000 for any HRR as required by the Dartmouth HRRs is also enforced in the next step for comparability.

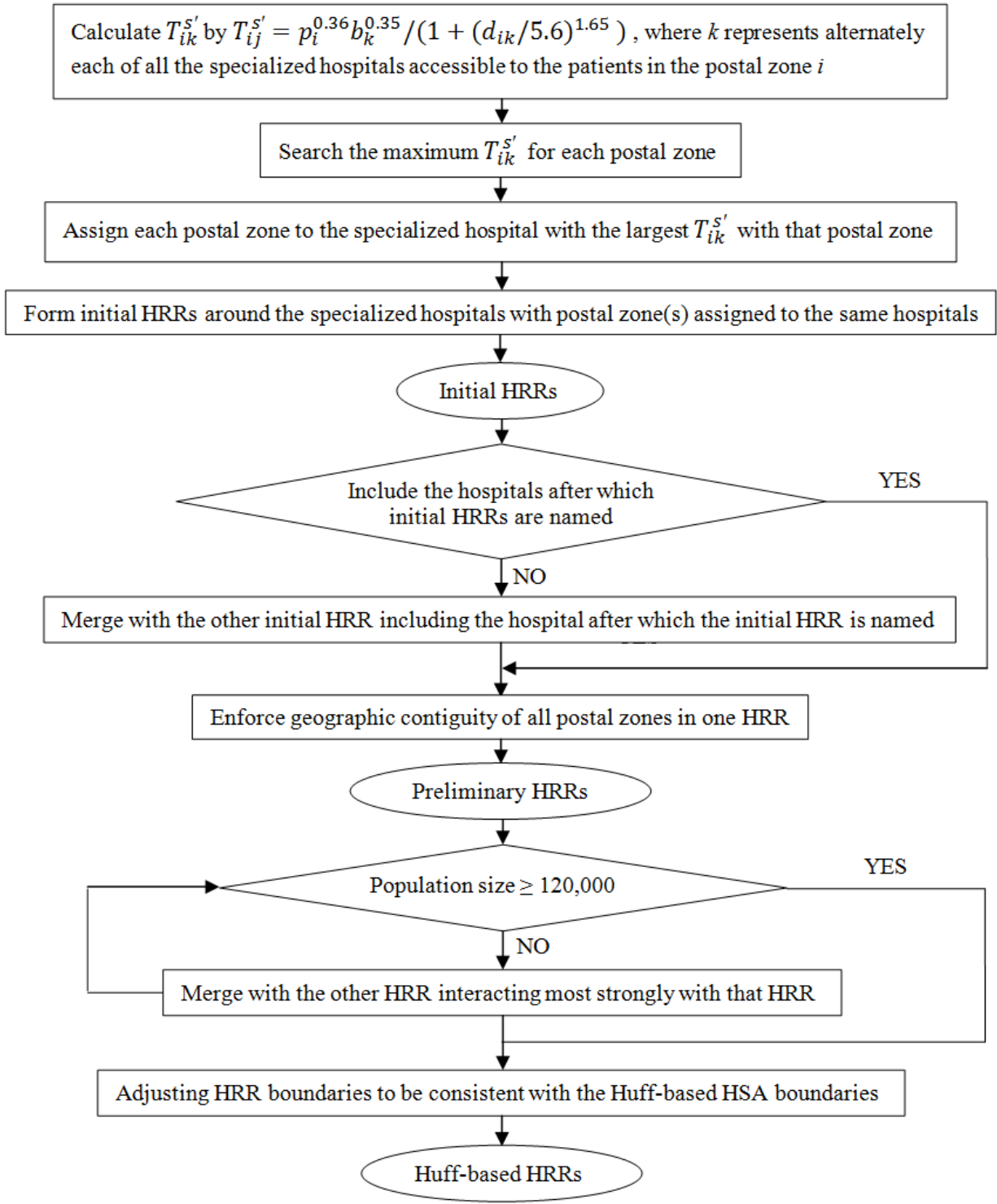


Figure 20. A flowchart of generating the Huff-based HRRs

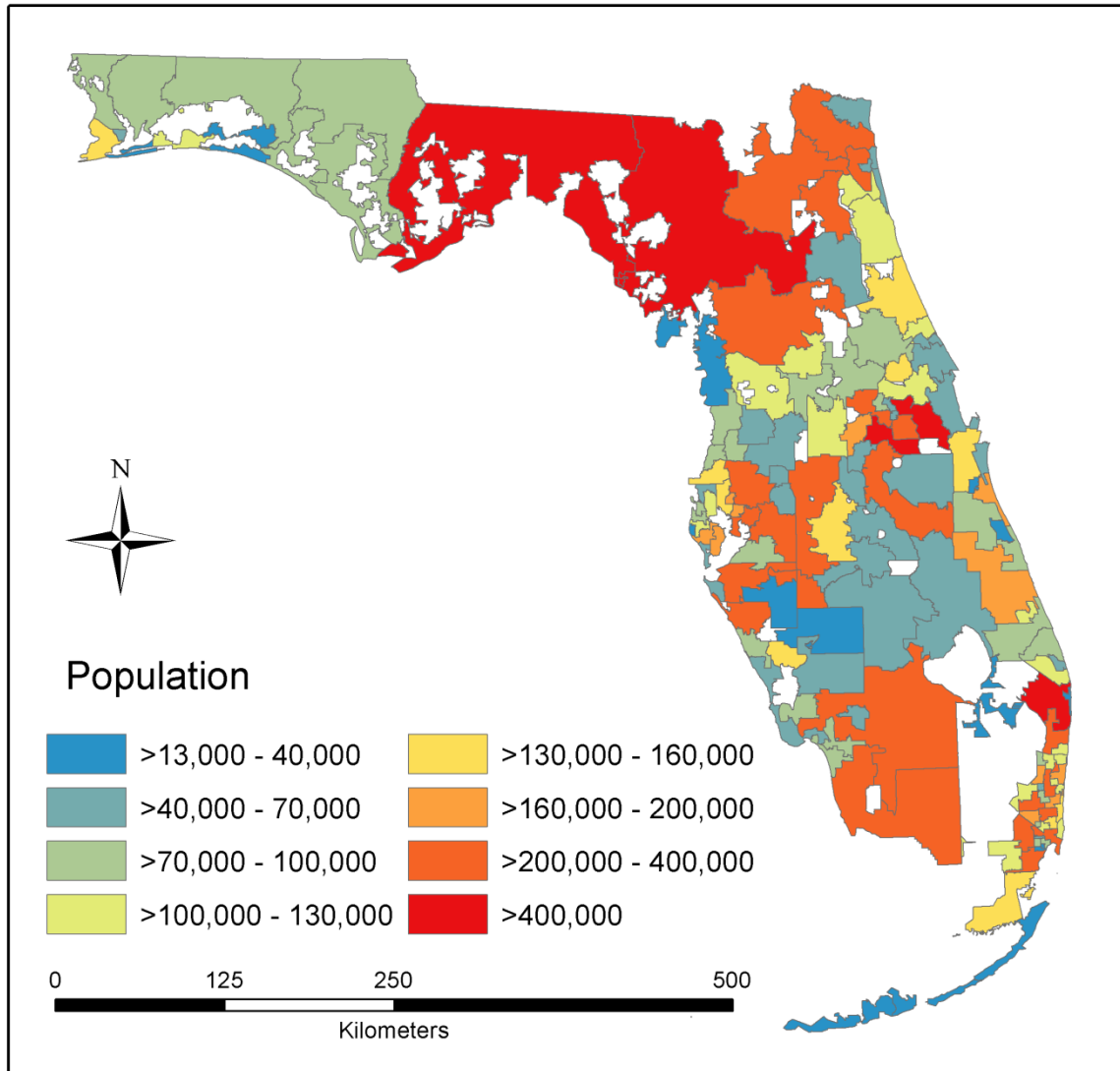


Figure 21. Population within the preliminary HRRs

With multiple zip codes and hospitals included in each HRR, measuring the interactions between HRRs is more complicated than between zip codes and hospitals. Here, the Huff model approach is implemented at the HRR level to model the interactions between HRRs as follows:

$$T_{ij}^R = p_i^\alpha b_j^\sigma f(d_{ij})$$

where T_{ij}^R is the volume of the discharges from all the hospitals in HRR j to all the zip codes in HRR i , p_i is the total population in HRR i , b_j is the total number of beds in all the hospitals in HRR j , α and σ are the parameters describing the effects of the numbers of HRR population and

hospital beds upon interactions, respectively; d_{ij} is the travel time from HRR i to j (unit: minute), $f(d_{ij})$ is a distance decay function described in Section 5.1 (Table 5).

The travel time d_{ij} must be measured between points, so the population-weighted and bed-weighted centroids are generated to represent the locations of the overall population and hospital resources within HRRs, respectively. The population-weighted centroids within HRRs are calculated in a similar approach as discussed in Section 5.1. With all trips by specialized patients aggregated by origin and destination HRRs and converted into travel flows between HRRs, the four parameters (α , σ , β , and θ) are estimated by the non-linear least squares estimators, and the four models are compared by the *pseudo-R*² and AIC (Table 12).

Table 12. Optimal parameters and assessment of the four candidate functions for modeling the interactions between the HRRs (from origin HRRs to destination HRRs)

Function	α	σ	θ	β	<i>pseudo-R</i> ²	AIC
Power	0.32	0.59	-	0.73	0.3203	269,729.7
Exponential	0.42	0.45	-	0.14	0.4940	263,525.4
Gaussian	0.39	0.47	6.39	-	0.4786	264,153.5
Log-logistic	0.40	0.46	6.23	2.18	0.5075	262,957.5

With the largest *pseudo-R*² and smallest AIC, the log-logistic function is used as the best function for fitting all patient-to-hospital flows at the HRR level. The optimal combination of parameters produced is $\alpha = 0.4$, $\sigma = 0.46$, $\theta = 6.23$ and $\beta = 2.18$ (Table 12), so the expected interactions between (origin) HRR i and (destination) HRR j ($T_{ij}^{R'}$) are formulated as:

$$T_{ij}^{R'} = p_i^{0.4} b_j^{0.46} / (1 + (d_{ij}/6.23)^{2.18})$$

where the parameter values are all statistically significant ($p < 2 \times 10^{-16}$).

Upon the completion of the calculation, a list of (destination) HRRs for each (origin) HRR is recorded in ascending order of attractiveness ($T_{ij}^{R'}$). All the HRRs are ranked in

ascending order of the population number within them. From the smallest end, each HRR with a population size < 120,000 is combined with an adjacent HRR based on the two following rules:

- 1) if that HRR has only one neighbor, then it is combined with that neighbor;
- 2) if that HRR has more than one neighbor, then it is combined with its most attractive neighbor.

Following those two rules, the combination continues until the total population of each combined HRR reaches 120,000. At last, for administrative purposes, all HRR boundaries are adjusted to be consistent with the Huff-based HSA boundaries. In other words, if a given HSA is segmented by the HRR boundary, the smaller segment (normally just one or two zip codes) would be re-assigned to the neighboring HRR that contains the larger segment. The resulting HRRs are named the *Huff-based HRRs*.

After adjusting the population size and nesting all the HSAs completely within the HRRs, the 145 preliminary HRRs are ultimately combined into the 72 Huff-based HRRs, in which the 169 HSA units are completely nested (Figure 22). Compared to the 34 overall-derived HRRs in Section 4.3 (Figure 23), the 72 HRRs have more balanced sizes of population ranging from 120,119 to 736,839 (versus from 130,061 to 1,596,791) (Figure 24), with a mean of 258,445 (versus 547,295) and a median of 242,876 (versus 391,782). Similar to the Huff-based HSAs, the Huff-based HRRs are derived by utilizing the travel behavior of specialized patients, which are fairly stable over time. The traditional Dartmouth approach for delineating HRRs relies on detailed travel flows of specialized patients that may not be available for many studies, and thus their applicability is limited.

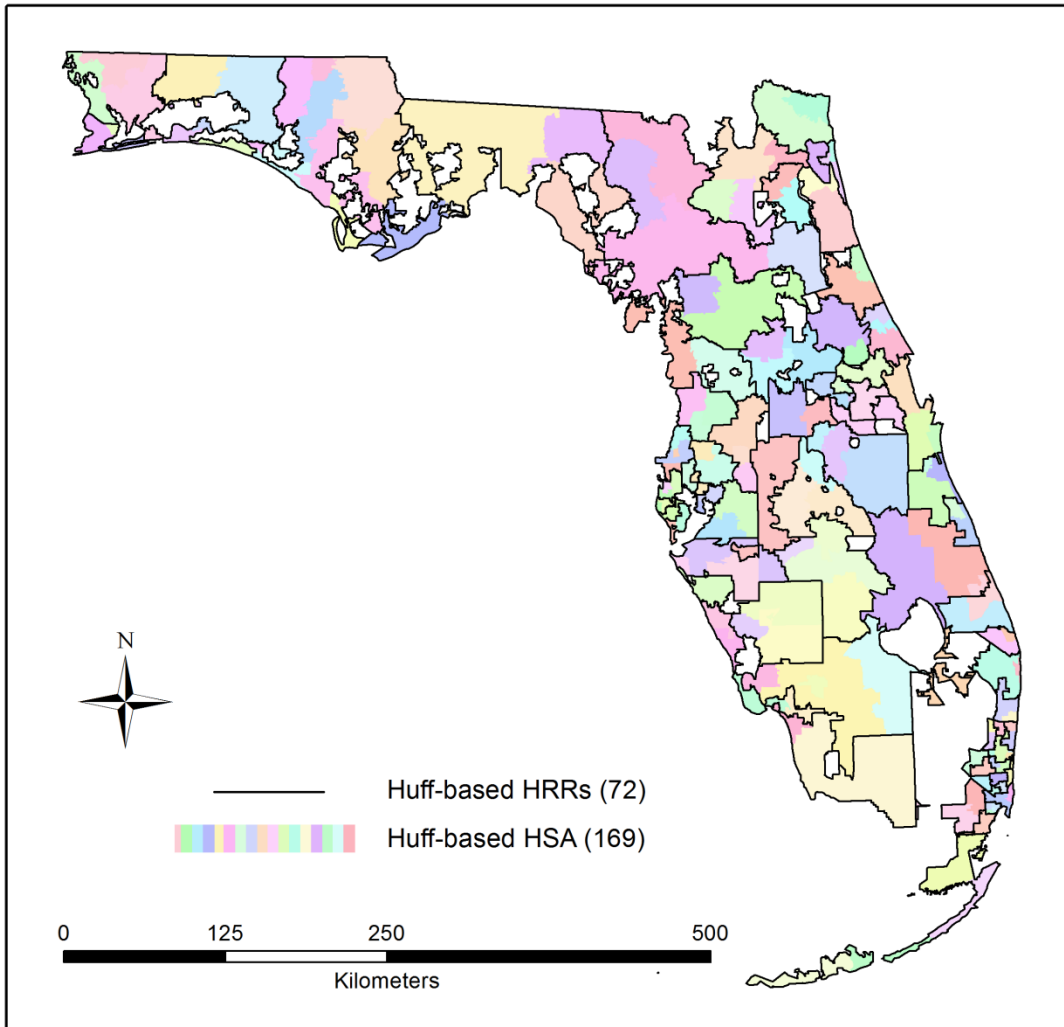


Figure 22. Huff-based hierarchical HSAs-HRRs

7.3 Summary

In this chapter, the travel patterns of patients seeking general and specialized care are compared to identify different distance decay effects, captured by their different best-fitting (log-logistic) functions. However, a major limitation is that only those hospital discharges with procedures on the cardiovascular and nervous systems are used to represent the specialized care patients. The patients with neuro surgeries are few, and also exhibit a very weak distance decay effect. In practice, the records of patients receiving cardio and neuro surgeries are combined, and

the distance decay behavior derived from their travel patterns is used as the foundation for delineating the HRRs.

Three key findings are observed. First, similar to the overall patients, the log-logistic function convincingly outperforms the power, exponential, and Gaussian functions for modeling the travel patterns of the cardiovascular patients. Secondly, according to the distance decay curves, specialized patients are more likely to travel longer for obtaining the corresponding hospital services than general patients. Thirdly, compared to the HSAs, fewer but larger HRRs are produced based on the corresponding travel patterns, and each HRR encloses one or more HSAs after necessary adjustment to form a hierarchical HSA system (HSAs-HRRs).

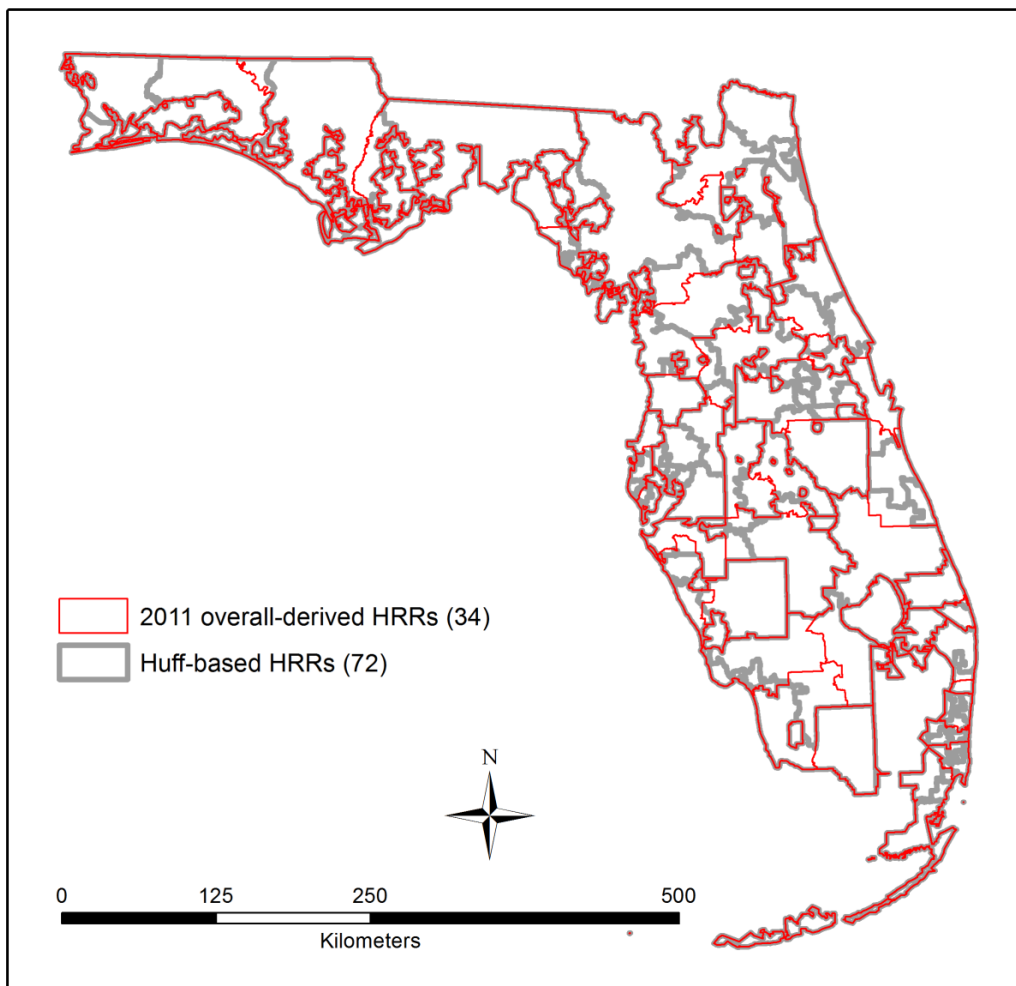


Figure 23. Boundaries of the 2011 overall-derived HRRs and Huff-based HRRs

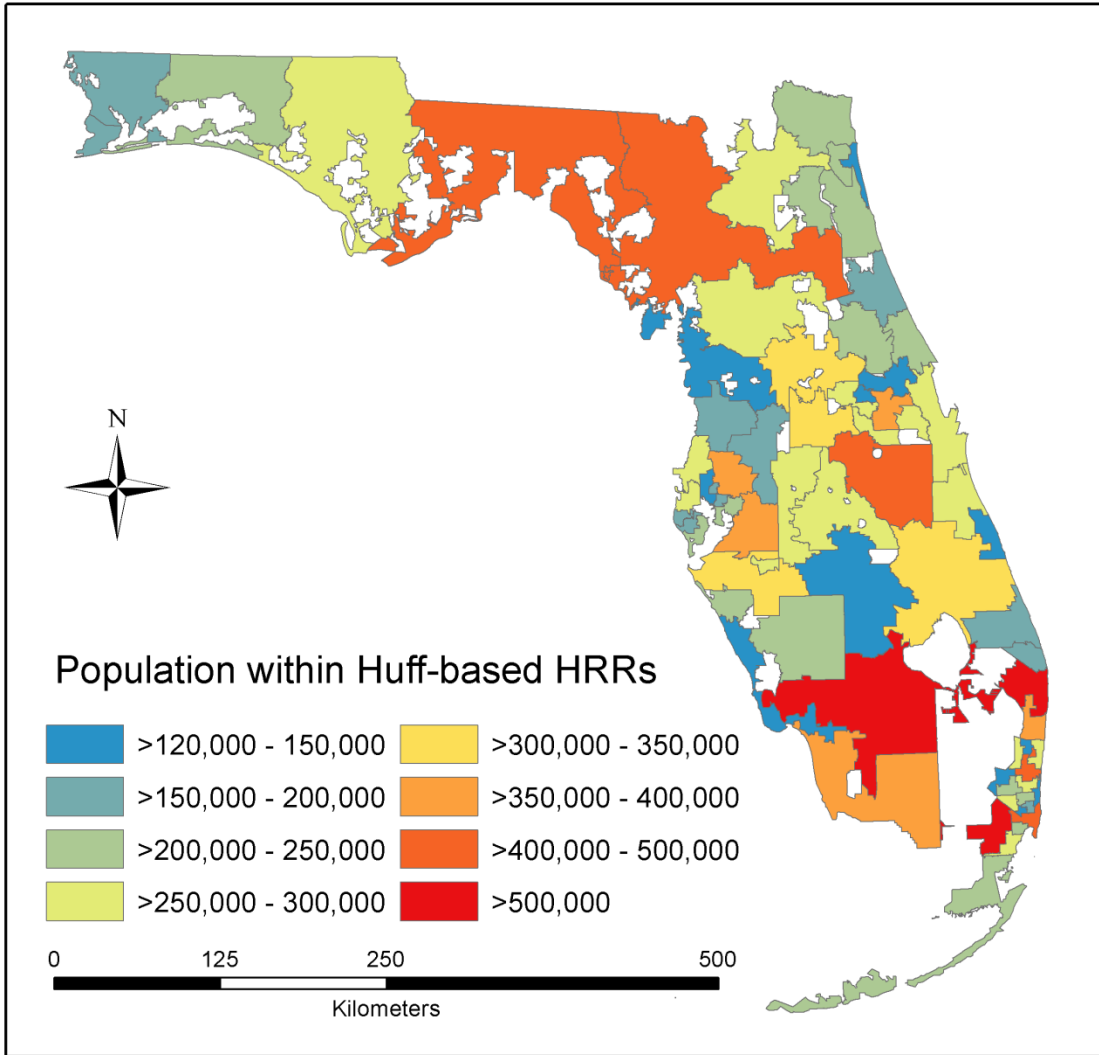


Figure 24. Population within the Huff-based HRRs

Chapter 8 Conclusions

The results of this study challenge the use of Medicare-derived HSAs in studies related to hospital use and health care delivery (Ashton et al. 1999; Fisher et al. 1994; Silverman, Skinner & Fisher, 1999; Wennberg 1999). The boundaries of the Dartmouth Medicare-derived HSAs have significantly shifted in Florida during the past two decades; hence, re-visiting the HSA delineation is necessary to keep them updated. In addition, the Medicare-derived HSAs do not represent the health care seeking behavior of the overall population. It is necessary to update the delineation of HSAs based on more recent data of discharges of all inpatients rather than those of only Medicare inpatients.

Newly derived HSAs offer promise for small area analyses for healthcare market assessment (Gittelsohn & Powe 1995; Wennberg & Gittelsohn 1973). The HSAs enable the comparison of hospitalization practices and surgical procedure rates among different hospitals in various regions (Ashton et al. 1999; Fisher et al. 1994). A detailed review of major methodologies of regionalization suggests that traditional approaches of defining HSAs are inadequate and call for an improved approach. There are some common features between HSAs and other functional areas such as labor market areas (LMAs) and retail trade areas (RTAs). Methods used for defining those functional areas such as the Huff model may also benefit the task of defining HSAs. However, HSAs differ from other functional areas. For example, every day most people commute to work not to hospitals; there are also differences in regularity and urgency between these activities and hospitalization (Huff & McCallum 2008).

It is important to produce a set of HSAs with stable boundaries over time, and when necessary, update the HSA boundaries to reflecting the underlying major changes in a timely manner. A proper way to construct such HSAs is to build them upon patients' travel patterns for

hospitalization. The Huff model has a built-in distance decay function that captures patients' travel behavior and remains fairly stable over time. Therefore, one major advantage of the Huff model approach is the ability of projecting possible adjustment of HSAs when major changes happen to the hospital system in absence of patients' travel data.

This study also utilizes the distance decay functions to capture the specific travel patterns of patients seeking specialized care, and then feeds the results to the Huff model approach for delineating the HRRs. Following the administrative principles ($K=7$) in the Central Place Theory, the HRRs are adjusted to include complete HSAs within them, and form a hierarchical HSA-HRR system.

The GIS technologies facilitate several key tasks in this study: 1) matching the hospital records to HSA boundaries, 2) calculating the Localization Index (LI) within the HSAs, 3) computing the network travel time between patients and hospitals, 4) implementing the Huff model approach to delineating the HSAs and HRRs, and 5) adjusting the HRR boundaries to form the hierarchical HSA-HRR system.

There are several *limitations* for the study. First, with absence of exact home address of patients, their locations are assumed to be the zip code centroids, and therefore the distance measure between patients and hospitals is approximate. Secondly, only the discharges from acute long-term care hospitals and general medical and surgical hospitals are included in the study. This excludes discharge patients from other specialty hospitals, which may affect the analysis of the travel patterns of patient for specialized care, to some extent. In addition, only the number of beds is used to measure the size of a hospital. More characteristics (e.g., quality and level) of the hospitals could be taken into account to construct a more comprehensive index of attractiveness for them. This affects the accuracy in estimating expected interactions between residents and

hospitals by the Huff model. Finally, the travel patterns are modeled based on all the travel flows with at least two discharges, and such a decision is arbitrary. Furthermore, one patient may have multiple hospital visits. Inclusion of multiple trips and exclusion of a single trip, both by a single patient in a zip code area, may introduce bias to the study.

According to the population distribution by age in 2013 (Kaiser Family Foundation 2013), the percentage of the elderly aged 65 and above was 17% in Florida, higher than the national 14% and lowest 11% in Utah and Texas in the contiguous U.S. Despite a relatively small difference in age structure between Florida and the national level, the conclusions of this study should be applied to those younger states with caution. For one thing, the elderly may disproportionately need hospital services, and may not travel as far as young patients due to age or severity of the illness. In addition to some hospital attributes (e.g., quality of services, hospital image, and cost of treatment) that may influence the decision-making of the overall population, the elderly patients' choice of hospitals could also be influenced by health insurance coverage. For example, some types of insurance, such as Medicare Advantage plans, dictate a network in which patients may receive services, making some hospitals unavailable to nearby populations.

This is the first attempt of a comprehensive study of the HSA-HRR system in the U.S. *Future work* can advance in multiple directions. Both cross-sectional and longitudinal hospital discharge data can be analyzed to expand the current work in both temporal and spatial dimensions. Replication of the work in more states may confirm whether the log-logistic function is the best-fitting distance decay function for hospital visits and whether related findings may be generalized to a larger spatial scope. An updated national hierarchical HSA-HRR system conceivably could provide a reliable analysis unit for many health studies. On the other hand,

studies on data of multiple epochs may help use detect possible changes in patients' travel behaviors over time, and find an appropriate time interval for updating the HSAs-HRRs.

Some key definitions and rules for defining the HSAs and HRRs by the Dartmouth approach are used in this study. For example, cardiovascular surgical procedure and neurosurgery are used as the specialized services for defining the HRRs. Both services may not be as specialized as they used to be as they are being provided by an increasing number of hospitals. An appropriate identification of specialized care is important to construct a reliable higher level hospital service areas (i.e., HRRs), and calls for the input of medical professionals and studies of data of larger scale and for longer time.

References

- Agency for Healthcare Research and Quality. (2014). Healthcare Cost and Utilization Project – HCUP, A Federal-State-Industry Partnership in Health Data: Clinical Classifications Software (CCS) 2014. available <https://www.hcup-us.ahrq.gov/toolssoftware/ccs/CCSUsersGuide.pdf> (last accessed on 8-17-2015).
- Arcury, T. A., Gesler, W. M., Preisser, J. S., Sherman, J., Spencer, J., & Perin, J. (2005). The effects of geography and spatial behavior on health care utilization among the residents of a rural region. *Health Serv Res, 40* (1), 135-155.
- Ashton, C. M., Petersen, N. J., Soucek, J., Menke, T. J., Yu, H. J., Pietz, K., Eigenbrodt, M. L., Barbour, G., Kizer, K. W., & Wray, N. P. (1999). Geographic variations in utilization rates in Veterans Affairs hospitals and clinics. *N Engl J Med, 340* (1), 32-39.
- Basu, J., & Cooper, J. (2000). Out-of-area travel from rural and urban counties: a study of ambulatory care sensitive hospitalizations for New York State residents. *J Rural Health, 16* (2), 129-138.
- Basu, J., & Friedman, B. (2007). A re-examination of distance as a proxy for severity of illness and the implications for differences in utilization by race/ethnicity. *Health Econ, 16* (7), 687-701.
- Biello, K. B., Rawlings, J., Carroll-Scott, A., Browne, R., & Ickovics, J. R. (2010). Racial disparities in age at preventable hospitalization among U.S. Adults. *Am J Prev Med, 38* (1), 54-60.
- Brown, P. J., & Hincks, S. (2008). A framework for housing market area delineation: principles and application. *Urban studies, 45* (11), 2225-2247.
- Calvo, A. B., & Marks, D. H. (1973). Location of health care facilities: an analytical approach. *Socio-Economic Planning Sciences, 7* (5), 407-422.
- Carlsson, F., Johansson, M., Petersson, L. O., & Tegsjö, B. (1993). Creating labour market areas and employment zones. In *CERUM report*.
- Center for Evaluative Clinical Sciences. (1999). The Dartmouth Atlas of Health Care in the United States. In J. E. Wennberg (Ed.), Chicago, Illinois.
- Christaller, W. (1933). *Central Places in Southern Germany*.
- Coombes, M. (2000). Defining locality boundaries with synthetic data. *Environment and Planning A, 32* (8), 1499-1518.

- 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* (5), 1038-1052.
- Delamater, P. L., Messina, J. P., Grady, S. C., WinklerPrins, V., & Shortridge, A. M. (2013). Do More Hospital Beds Lead to Higher Hospitalization Rates? A Spatial Examination of Roemer's Law. *PLoS One, 8* (2), e54900.
- Development Core Team. (2011). R: A language and environment for statistical computing. In *Technical report*, Vienna, Austria.
- Dökmeci, V. F. (1973). An optimization model for a hierarchical spatial system. *Journal of Regional Science, 13* (3), 439-451.
- Dökmeci, V. F. (1977). A quantitative model to plan regional health facility systems. *Management Science, 24* (4), 411-419.
- Dramowicz, E. (2005). Retail Trade Area Analysis Using the Huff Model. *Directions Magazine, 2*.
- Escarce, J. J., & Kapur, K. (2009). Do patients bypass rural hospitals?: Determinants of inpatient hospital choice in rural California. *J Health Care Poor Underserved, 20* (3), 625-644.
- Farahani, R. Z., Hekmatfar, M., Fahimnia, B., & Kazemzadeh, N. (2014). Hierarchical facility location problem: Models, classifications, techniques, and applications. *Computers & Industrial Engineering, 68*, 104-117.
- Fisher, E. S., Wennberg, J. E., Stukel, T. A., & Sharp, S. M. (1994). Hospital readmission rates for cohorts of Medicare beneficiaries in Boston and New Haven. *N Engl J Med, 331* (15), 989-995.
- Garnick, D. W., Luft, H. S., Robinson, J. C., & Tetreault, J. (1987). Appropriate measures of hospital market areas. *Health Serv Res, 22* (1), 69-89.
- Ghosh, A., & McLafferty, S. L. (1987). *Location strategies for retail and service firms*. Lexington Books, Lexington, MA.
- Gibbs, J. P., & Martin, W. T. (1962). Urbanization, technology, and the division of labor: International patterns. *American Sociological Review, 667-677*.
- Gittelsohn, A., & Powe, N. R. (1995). Small area variations in health care delivery in Maryland. *Health Serv Res, 30* (2), 295-317.
- Goodchild, M. F., & Lam, N. S. N. (1980). *Areal interpolation: a variant of the traditional spatial problem*, London, Ont.: Department of Geography, University of Western Ontario.

- Goodman, D. C., Fisher, E., Stukel, T. A., & Chang, C. (1997). The distance to community medical care and the likelihood of hospitalization: is closer always better? *Am J Public Health, 87* (7), 1144-1150.
- Goodman, D. C., & Fisher, E. S. (2008). Physician workforce crisis? Wrong diagnosis, wrong prescription. *N Engl J Med, 358* (16), 1658-1661.
- Guagliardo, M. F. (2004). Spatial accessibility of primary care: concepts, methods and challenges. *Int J Health Geogr, 3* (1), 3.
- Guagliardo, M. F., Jablonski, K. A., Joseph, J. G., & Goodman, D. C. (2004). Do pediatric hospitalizations have a unique geography? *BMC Health Serv Res, 4* (1), 2.
- Haynes, K. E., & Fotheringham, A. S. (1984). *Gravity and spatial interaction models* (Vol. 2). Sage Publications, Beverly Hills.
- Healthcare Cost and Utilization Project. (2011). SID Database Documentation. In. Agency for Healthcare Research and Quality, Rockville, MD.
- Hodgson, M. J. (1988). An hierarchical location-allocation model for primary health care delivery in a developing area. *Soc Sci Med, 26* (1), 153-161.
- Huff, D., & McCallum, B. M. (2008). Calibrating the Huff Model Using ArcGIS Business Analyst. In *ESRI White Paper*.
- Huff, D. L. (1964). Defining and estimating a trading area. *The Journal of Marketing, 34*-38.
- Jha, A. K., Li, Z., Orav, E. J., & Epstein, A. M. (2005). Care in U.S. hospitals--the Hospital Quality Alliance program. *N Engl J Med, 353* (3), 265-274.
- Jha, A. K., Orav, E. J., Zheng, J., & Epstein, A. M. (2008). Patients' perception of hospital care in the United States. *N Engl J Med, 359* (18), 1921-1931.
- Jia, P., & Xierali, I. (2015). Disparities in healthcare travel patterns of the inpatients diagnosed with congestive heart failure in Florida. *Preventing Chronic Disease, in press*.
- Jia, P., Xierali, I., & Wang, F. (2014). Evaluating and re-demarcating the Hospital Service Areas in Florida. *Applied Geography, 60*, 248-253.
- Jones, A. P., Haynes, R., Sauerzapf, V., Crawford, S. M., Zhao, H., & Forman, D. (2008). Travel time to hospital and treatment for breast, colon, rectum, lung, ovary and prostate cancer. *Eur J Cancer, 44* (7), 992-999.
- Jones, C. (2002). The definition of housing market areas and strategic planning. *Urban studies, 39* (3), 549-564.
- Joseph, A. E., & Phillips, D. R. (1984). Accessibility and utilization: geographical perspectives on health care delivery. *Sage*.

- Kaiser Family Foundation. (2013). Population Distribution by Age. In S. H. Fact (Ed.).
- Klauss, G., Staub, L., Widmer, M., & Busato, A. (2005). Hospital service areas -- a new tool for health care planning in Switzerland. *BMC Health Serv Res*, *5*, 33.
- Lembcke, P. A. (1952). Measuring the quality of medical care through vital statistics based on hospital service areas; I. Comparative study of appendectomy rates. *Am J Public Health Nations Health*, *42* (3), 276-286.
- Llewelyn-Davies, R., & Macaulay, H. M. C. (1966). Hospital planning and administration. In World Health Organization, Geneva.
- Luo, W., & Qi, Y. (2009). An enhanced two-step floating catchment area (E2SFCA) method for measuring spatial accessibility to primary care physicians. *Health Place*, *15* (4), 1100-1107.
- Luo, W., & Wang, F. (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* (6), 865-884.
- Luo, W., & Whippo, T. (2012). Variable catchment sizes for the two-step floating catchment area (2SFCA) method. *Health Place*, *18* (4), 789-795.
- Medicare.gov. (2014). The Official U.S. Government Site for Medicare.
- Moist, L. M., Bragg-Gresham, J. L., Pisoni, R. L., Saran, R., Akiba, T., Jacobson, S. H., & Port, F. K. (2008). Travel time to dialysis as a predictor of health-related quality of life, adherence, and mortality: the Dialysis Outcomes and Practice Patterns Study (DOPPS). *American Journal of Kidney Diseases*, *51* (4), 641-650.
- Newhouse, J. P., Garber, A. M., Graham, R. P., McCoy, M. A., Mancher, M., & Kibria, A. (2013). *Variation in Health Care Spending: Target Decision Making, Not Geography*. National Academies Press, Institute of Medicine.
- Nicholl, J., West, J., Goodacre, S., & Turner, J. (2007). The relationship between distance to hospital and patient mortality in emergencies: an observational study. *Emerg Med J*, *24* (9), 665-668.
- O'Neill, L. (2004). The effect of insurance status on travel time for rural Medicare patients. *Med Care Res Rev*, *61* (2), 187-202.
- Parr, J. B. (1978). Models of the central place system: a more general approach. *Urban studies*, *15* (1), 35-49.
- Patel, A. R., Fik, T. J., & Thrall, G. I. (2008). Direction sensitive wedge-casting for trade area delineation. *Journal of Real Estate Portfolio Management*, *14* (2), 125-140.

- Pieda, D. T. Z. (2004). Housing market assessment manual. In, Office of the Deputy Prime Minister, London.
- Radley, D. C., & Schoen, C. (2012). Geographic variation in access to care--the relationship with quality. *N Engl J Med*, 367 (1), 3-6.
- Rahman, S. U., & Smith, D. K. (2000). Use of location-allocation models in health service development planning in developing nations. *European Journal of Operational Research*, 123 (3), 437-452.
- Reutstein, D. (1971). Lecture notes from M.I.T.-Harvard Lecture Series on the future of medicine. In M.I.T. (Ed.).
- Şahin, G., & Süral, H. (2007). A review of hierarchical facility location models. *Computers & Operations Research*, 34 (8), 2310-2331.
- Schultz, G. P. (1970). The logic of health care facility planning. *Socio-Economic Planning Sciences*, 4 (3), 383-393.
- Shi, X., Alford-Teaster, J., Onega, T., & Wang, D. (2012). Spatial access and local demand for major cancer care facilities in the United States. *Annals of the Association of American Geographers*, 102 (5), 1125-1134.
- Shortt, N. K., Moore, A., Coombes, M., & Wymer, C. (2005). Defining regions for locality health care planning: a multidimensional approach. *Soc Sci Med*, 60 (12), 2715-2727.
- Shortt, N. K., & Moore, A. J. (2006). Functional localities: an integrated spatial approach towards health care locality definition. *GeoJournal*, 67 (1), 27-40.
- Silverman, E. M., Skinner, J. S., & Fisher, E. S. (1999). The association between for-profit hospital ownership and increased Medicare spending. *N Engl J Med*, 341 (6), 420-426.
- Tobler, W. R. (1970). A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography*, 46, 234-240.
- United Press International, I. (2013). U.S. healthcare: Most expensive, longest waits, most red tape. In.
- Wang, F. (2012). Measurement, Optimization, and Impact of Health Care Accessibility: A Methodological Review. *Ann Assoc Am Geogr*, 102 (5), 1104-1112.
- Wang, F. (2015). *Quantitative Methods and Socioeconomic Applications in GIS* (2nd ed.). Taylor & Francis, Boca Raton, FL.
- Wang, F., Wen, M., & Xu, Y. (2013). Population-Adjusted Street Connectivity, Urbanicity and Risk of Obesity in the U.S. *Appl Geogr*, 41, 1-14.

Wennberg, J., & Gittelsohn, A. (1973). Small area variations in health care delivery. *Science*, 182 (4117), 1102-1108.

Wennberg, J. E. (1999). Understanding geographic variations in health care delivery. *N Engl J Med*, 340 (1), 52-53.

Wilbush, J. (1974). The local hospital: functions, advantages and difficulties. *Can Fam Physician*, 20 (7), 54-56.

Zhang, Y., Baicker, K., & Newhouse, J. P. (2010). Geographic variation in Medicare drug spending. *N Engl J Med*, 363 (5), 405-409.

Vita

Peng Jia is half Mongolian and half Han Chinese, born in Hohhot, Inner Mongolia, China. He earned his Bachelor of Engineering in Environmental Engineering from Nanjing Agricultural University (NJAU, Nanjing, Jiangsu, China) in 2007. In 2010, he earned his Master of Science (M.S.) in Cartography and Geographic Information System (GIS) from the Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences (RADI, CAS), formerly known as the Institute of Remote Sensing Applications (IRSA, CAS).

He has been trained in the Emerging Pathogens Institute (EPI) at the University of Florida (UF) as a medical geographer, and earned his second M.S. in Geography from the UF in 2012. At the same year, he entered the Department of Geography & Anthropology at Louisiana State University (LSU) to pursue his Doctor of Philosophy (Ph.D.) degree in Geography, under the mentor of Fahui Wang. He is the awardee of The 2015 LSU Dissertation Year Fellowship. He also holds the position of the Treasurer of the LSU Chapter (Eta Psi) of Gamma Theta Upsilon Honor Society.

His research focuses on 1) measurement of health behavior and healthcare utilization, 2) obesity and built environment, 3) climate change impacts on zoonotic (e.g., Brucellosis and avian influenza) and vector-borne diseases (e.g., malaria), and 4) global health metrics and evaluation, in particular the Health and Demographic Surveillance System. He specializes in GIS, remote sensing, spatio-temporal analysis, and applied statistics.

Upon completing his Ph.D. study in the summer of 2015, he will join the Department of Epidemiology and Environmental Health at the State University of New York (SUNY) at Buffalo as a postdoctoral associate.