CONSTRUCTING OPTIMAL MEDICAL MANAGEMENT AREAS FOR HEALTH SERVICES RESEARCH

By

Chenxiao Ling

A THESIS

Submitted to
Michigan State University
In partial fulfillment of the requirements
for the degree of

MASTER OF SCIENCE

Geography

2011

ABSTRACT

CONSTRUCTING OPTIMAL MEDICAL MANAGEMENT AREAS
FOR HEALTH SERVICES RESEARCH

By

Chenxiao Ling

This study constructs optimal medical management areas (MMAs) in the State of Michigan for the purpose of visualizing and exploring the spatial patterns of two health indicators: ischemic heart disease (IHD) (ICD-9-CM: 410-414) and diabetes (ICD-9-CM: 250) for the purpose of assessing population-demand for health services. Data on IHD and diabetes are obtained from the Michigan Inpatient Hospital Discharge Database (MIDB) for the year 2008. MMA boundary definitions are optimized using Automated Zone Matching (AZM) methodology software. Optimization is conducted by aggregating the residential ZIP Codes of patients discharged from hospitals with IHD or diabetes, using three constraint parameters: (1) minimum case threshold, to ensure rate stability; (2) maximum shape compactness, to avoid irregular or elongated MMAs; and (3) maximum internal homogeneity, to construct MMAs with populations that demographically similar. The modifiable area unit problem (MAUP) is examined within the context of MMA design and epidemiological scale by evaluating IHD and diabetes in relation to their relevant broader disease groups, diseases of the circulatory system (ICD-9-CM: 390-459) and endocrine, nutritional and metabolic diseases, and immunity disorders (ICD-9-CM: 240-279). Following the optimization of AZM-MMAs area-based proportions, crude rates and age-adjusted rates are calculated to represent the various views of demand for health services. The limitations and benefits of using AZM versus traditional ZIP Code boundary definitions to construct MMAs to assess the demand for inpatient hospital services are discussed for future applications.

This thesis is dedicated to my parents, without whose support and selfless love, the completion of this work would not have been possible.

ACKNOWLEDGEMENTS

I would first like to express my deepest gratitude to my advisor, Dr. Sue C. Grady, for your invaluable guidance, patience and assistance as I learned to conduct this research. Your encouragement inspired confidence in me. It is an honor for me to be one of your students. Special thanks to the distinguished faculty members who served on my committee: Dr. Ashton M. Shortridge and Dr. Joseph P. Messina. Your comments and advices greatly helped with the building of my thesis. I feel fortune to have you on my committee. I would also like to thank Dr. David J. Campbell and Dr. Bruce Wm. Pigozzi, who provided me with comments and opinions on thesis writing.

I gratefully acknowledge the use of the AZTool software, which is copyright Dr. David Martin, Ms. Samantha Cockings and University of Southampton in the UK. Thanks the Michigan Department of Community Health for kindly providing the main database. I also wish to thank Dr. Arika Ligmann-Zielinska for preparing the shapefile for this study. Special thanks to Mr. Jim T. Brown and Mr. Wilson Ndovie, for your technical support to the medical geography lab and my work. Thanks are also due to Dr. Richard E. Groop and Dr. Jiaguo Qi, for bringing me into MSU and this program, and other faculty and staff for assisting and encouraging me in many ways.

I'm so grateful to my mom and dad who provide me with this opportunity to study abroad and conduct this thesis. I'd also like to thank Paul, Kristie, Hui, Ziting, Peter and Katie, for your accompany as lab-mates and friends. Finally, I wish to express my gratitude to all the individuals who helped with the completion of this work.

TABLE OF CONTENTS

LIST OF TABLES

[1] Includes endocrine, nutritional and metabolic diseases, and immunity disorders

LIST OF FIGURES

LIST OF ABBREVIATIONS

AZM	Automated Zoning Matching
AZTool	Automated Zoning Tool
CDC	Centers for Disease Control and Prevention
CI	Confidence Interval
CON	Certificate of Need
EDs	Enumeration Districts
FSA	Facility Sub-Areas
HSA	Hospital System Areas
IAC	Intra-Area Correlation
ICD-9-CM	International Classification of Diseases, Ninth Revision, Clinical Modification
IHD	Ischemic Heart Disease
MAUP	Modifiable Areal Unit Problem
MDCH	Michigan Department of Community Health
MIDB	Michigan Inpatient Hospital Discharge Database
MMAs	Medical Management Areas
P2A	$Perimeter^2/Area$
RSE	Relative Standard Error
SA	Simulated Annealing
WHO	World Health Organization
ZCTA	ZIP Code Tabulation Areas

1. INTRODUCTION

Health service researchers study health care from ‘supply’ and ‘demand’ perspectives. In this thesis, the ‘supply’ perspective refers to the *availability* of inpatient hospital services for the entire population or a subset (i.e., the adequacy of the supply to meet the needs of a population) (Penchansky & Thomas, 1981). Attributes by which to measure inpatient hospital supply include: the potential for capacity or number of beds, the skill set of physicians, nurses, and other health care providers, the services that those providers offer, the length of hospital stay, and the method(s) of payment and reimbursement. Related to supply is *accessibility* or the ability to acquire the available services offered. Common barriers to receiving available services include: low socioeconomic status or poor health insurance coverage, distance in metric or time and hours of operation (Penchansky & Thomas, 1981). Finally, inpatient hospital services may be available and accessible, but not *utilized* by the population. Common barriers to utilization include: fear of diagnosis or treatment, untoward perceptions and prejudices, and language and cultural barriers (Meade et al., 1988). Most health service research focuses on the supply perspective through measurements of availability, accessibility and utilization to understand the health care system.

Important also is the ‘demand’ perspective. In this thesis, ‘demand’ refers to the need of the population for inpatient hospital services. The definition of demand/need has evolved over time. Bradshaw (1972) for example, defined need as: expressed need, comparative need, and normative need. Expressed need refers to “an expression in action of felt need” and comparative need referred to “the comparison of situations with others.” Expressed and comparative needs are subjective definitions. Normative need on the other hand, represents “an experts’ definition or diagnosis”, which could be subjective or objective. Acheson (1978) later defined demand/need as the "relief from the negative states of distress, discomfort, disability, handicap, and the risk of morbidity or mortality." Acheson (1978) argued that "the demand for health care is hard to assess quantitatively, because it is really a reflection of perceived health status." Health care professionals commonly assess patient's perceived health status, to make their diagnoses in practice. Over time, however the definition of demand/need took on a more quantitative meaning, when researcher began to view demand/need in terms of 'incidence' and 'prevalence' of disease. For example, (Bowling, 2002) stated, "it is reasonable that the spatial patterns of disease could reflect the 'demand' for health services, because people with diseases have the 'need' to seek healthcare." (Cromley & McLafferty, 2002) also expressed 'need' as "the prevalence of health conditions that should be addressed by health care services." This thesis therefore, adopts the prevalence of disease as a measure by which to assess demand of inpatient hospital services in future analyses.

In 2010, the Michigan Department of Community Health (MDCH) Division of Health Policy and Access expressed an interest in visualizing the State's critical health indicators using rate-maps to assess the local 'demand' for inpatient hospital services in those areas. Critical health indicators (MDCH, 2009) are defined as high-priority diseases and conditions identified by the State that also require in-depth public health program monitoring-evaluation and health care intervention. In Michigan (MDCH, 2009), heart disease-related deaths and increased diabetes prevalence were among the most important reported diseases/conditions in the State. In response to their request, data on the residential ZIP Codes of patients discharged from hospitals with ischemic heart disease (IHD) (ICD-9-CM: 410-414) and diabetes (ICD-9-CM: 250) were obtained from the Michigan Inpatient Hospital Discharge Database (MIDB) 2008 to map local areas of demand for inpatient hospital services. Ischemic heart disease refers to coronary heart disease (i.e., blood flow and oxygen to the heart muscle is reduced, thereby, increasing the risk of a myocardial infarction (heart attack) (The Merck Manual 16th edition, 1992). Diabetes is an endocrine disorder affecting the pancreas, limiting insulin production, and resulting in elevated blood glucose levels (The Merck Manual 16th edition, 1992). Although IHD-related deaths are decreasing in Michigan (MDCH, 2011), the prevalence of IHD and diabetes morbidity are increasing, in part related to an aging population, but environmental risks and lack of available and accessible high quality health care may also be contributing to this increasing prevalence. This research will focus on mapping the prevalence of demand and through understanding the spatial patterns of diseases, and new causal hypothesis may be generated. For the purpose of this thesis, the boundaries within which the prevalence rates of IHD and diabetes are mapped will be referred to as Medical Management Areas (MMAs).

The purpose of this thesis is to (1) construct optimal MMAs by which to (2) visualize and explore the spatial patterns of IHD and diabetes in the State of Michigan to assist the MDCH staff with evaluating local areas of demand for inpatient hospital services. There are however, known inherent problems with mapping disease rates at the residential ZIP Code level: specifically ZIP Codes with small case (numerator) or population (denominator) numbers will result in unstable rates (Washington State Department of Health, 2010); and the modifiable area unit problem (MAUP), in which there will be changes in disease rates with the modification of geographic scale and/or zone design (Openshaw, 1984). Both of these problems are inherently linked to the epidemiological scale of analysis (i.e., the disease or group(s) of diseases being studied).

To address these problems, residential ZIP Codes of patient’s discharged from hospitals with IHD and diabetes will be aggregated using Automated Zone Matching (AZM) methodology

(AZTool) (Cockings et al., 2011) to achieve an adequate case and population count. AZM-zones will be optimized using three constraint parameters: (1) a minimum case threshold, to ensure rate stability; (2) maximum shape compactness, to avoid constructing irregular or elongated MMAs; and (3) maximum internal homogeneity, to construct MMAs of relative similar demography. The MMA boundary definitions for IHD and diabetes will be validated by running AZM 50 times using the same model-constraint parameters and quantifying the 'optimal' run (i.e., the best output zone design). The modifiable area unit problem (MAUP) will be examined within the context of epidemiological scale by evaluating the spatial patterns of IHD and diabetes in relation to their relevant ICD-9-CM broad disease groups, diseases of circulatory system (ICD-9-CM: 390-459) (herein, referred to as circulatory diseases) and endocrine, nutritional and metabolic diseases, and immunity disorders (ICD-9-CM: 240-279) (herein, referred to as endocrine disorders) in Michigan. Following the optimization of MMAs, area-based IHD and diabetes proportions (cases/100 total hospital discharges) and crude and age-adjusted prevalence rates (cases/1,000 population) will be calculated to visualize and explore AZM-MMA areas of demand in the State. It is recognized that these maps will only comprise those patients who were hospitalized with these diseases and conditions (i.e., chronic, severe or late-stage disease), further justifying the need to construct optimal MMAs to inform health services research.

2. BACKGROUND

2.1 Geographic Health Services Research

During a yellow fever epidemic on New York City in 18th century, Dr. Valentine Seaman depicted a map of yellow fever deaths and tried to link them with what he called "putrid eflluvia" (Seaman et al., 1796) Although his assumption about yellow fever transmission was proved to be incorrect later, his contribution as the first disease spot map to medical geography remains important. The most well known historical disease maps were created by John Snow, when he mapped the residential locations of cholera cases during a cholera epidemic in 1854 in London. John Snow's maps became famous because he was able to show the relationship, i.e., the geographic proximity of drinking water source and cholera incidence. Both of these studies demonstrated the potential use of disease maps to assess demand for inpatient hospital services.

Contemporary health services research appeared in the literature in the 1950s, with an emphasis on health insurance. The first national study of health insurance coverage was undertaken in 1953 by the Health Information Foundation (Institute of Medicine, 1979). Some other landmarks of health service research at that time included a five-year study report of chronic illness by the Commission of Chronic Illness, which highlighted the involvement of economical concepts as an aide in patient screening and monitoring (Institute of Medicine, 1979). In terms of health facility research, the emphasis on progressive patient care (i.e., patients are grouped according to their illness and their need for care) emerged in the late 1950s (Haldeman, 1959). Studies of the 1960s and early 1970s on health services became more diversified in terms of the administration of health service, with research topics including regionalization and rural health care (McNerney & Riedel, 1962), organization of medicine in a sociological context (Freeman et al., 1979), economies of scale in medical practice (Lorant, 1971), and ambulatory medical care (Walker et al., 1964). In terms of health facility research, health care delivery was studied using location theories and various spatial models, e.g., central place theory (Shannon & Dever, 1974) and gravity models. Location theories are primarily concerned with the geographic location of economic activity. For example, central place theory aims to explain the size and location of human settlements in an urban system (Goodall, 1987). The gravity model seeks to simplify the demographic behavior of a large group of people by using the physical "gravity" model. When applied to health care studies, these theories were utilized to study people's behavior to seek health care (utilize), locate health care facilities (available) and identify the locations for future health services. These theories and their applications in health services research laid the framework from within which many algorithms and computer-generated programs were developed and now used today. Hierarchical location analysis primarily provided the theoretical foundation for the medical service referral system, such that with increasing hierarchy each hospital level becomes more specialized (Ghosh & Rushton, 1987). Spatial interaction issues, including the study of patients' travel patterns and the determination of location and size for health service facilities were both developed with the involvement of location-allocation models and algorithms (Ghosh & Rushton, 1987).

The use of Geographic Information System (GIS) in health services research emerged in the 1990s to assess health care need, access, and utilization (McLafferty, 2003). In the past twenty years, an increasing number of researchers published widely on the topic, including health care access (McLafferty & Grady, 2005; Wang et al., 2008), health care disparity and inequality in access (Grady & McLafferty, 2007; McLafferty & Wang, 2009) and health care facility utilization (Bennett et al., 2010; Statler et al., 2011). This thesis will expand on this body of work by utilizing GIS to construct MMAs in Michigan.

2.2 Government Agencies and Health Services Research

The Agency for Healthcare Research and Quality (AHRQ) and the Academy for Health Services Research and Health Policy (AHSR) are federal agencies overlooking the health care system in the United States.

AHRQ defines health services research as: “the examination of how people acquire access to health care, how much care costs, and what happens to patients as a result of receiving health care” (AHRQ, 2002). Within AHRQ the goals of health services research are to (1) identify the most effective ways to organize, manage, finance, and deliver high quality care; and (2) to reduce medical errors; and improve patient safety (AHRQ, 2002). In 2000, the Board of Directors of AHSR, now the Academy for Health Services Research and Health Policy, adopted a new definition for the field of health services research: “the multidisciplinary field of scientific investigation that studies how social factors, financing systems, organizational structures and processes, health technologies, and personal behaviors affect access to health care, the quality and cost of health care, and ultimately our health and well-being” (Lohr & Steinwachs, 2002).

The Certificate of Need (CON) legislation is “a representative of public utility regulation”, which was designed to restrict the expansion of hospitals, particularly in the number of beds and purchase of expensive equipment (CON, 1978). In 1978, the federal government required that all states implement CON programs for cardiac care services, meaning that hospitals had to apply and receive certification prior to implementing a cardiac care program. The intent of implementing such as program was to reduce the costs associated with duplicate investments (Ho et al., 2009). As hospitals merged and the numbers of facilities declined, some states other than Michigan, eliminated their CON programs (Ho, et al., 2009).

The Michigan Department of Community Health (MDCH), Division of Health Policy and

Access is responsible for managing hospitals through the utilization of a CON Commission in Michigan. Hospital regulation is achieved by limiting the number of beds a hospital can use for inpatient services. The MDCH assigns hospitals to Hospital System Areas (HSAs) and Facility Sub-Areas (FSAs) through studying utilization patterns. Hospital beds are then assigned within HSAs and FSAs (Langley et al., 2010). The Michigan implementation of Bed Need Methodology is unique because it is the only state to perform a bed need methodology at the ZIP Code level. This methodology utilizes the annual inpatient records (a.k.a., Michigan Inpatient Data Base, MIDB) collected by the Michigan Health and Hospital Association (MHA). In summary, the methodology calculates a utilization rate for inpatient ZIP Codes to each hospital subarea in the base year, and multiplies this rate by the projected population (five years after base year), within each age group and ZIP Code across the state to obtain a total projected bed need by subarea. This methodology is a 'supply' approach to understanding where hospital beds should be added or reduced. This research will provide information on MMAs for IHD and diabetes to inform bed need in Michigan.

2.3 Important Issues in Constructing Medical Management Areas

In the construction of MMAs using the residential ZIP Code of patients discharged from hospitals, it will be important to address small case numbers and the modifiable areal unit problem (MAUP). If there are too few cases of disease, the rates calculated will be unstable and spatial patterns observed will be inaccurate. If the constructed MMAs are not optimized there will be increased potential for MAUP problems, i.e., less reliable spatial patterns. In this study, the epidemiological scale of analysis will also be considered in the design of MMA construction. For example, studying diseases at a large epidemiological scale, i.e., primary diagnostic groups there will be more case counts but lower specificity and the ability to understand the hospital bed need in Michigan. In contrast, by constructing MMAs using specific diseases, such as IHD and diabetes, there may be low case counts but increased specificity in understanding the types of services and beds that are needed in Michigan. It is therefore important to construct optimal MMAs that have an adequate case count, are minimally sensitive to changes in geographic scale and provide meaningful information on diseases and conditions that can inform future bed need in Michigan. A brief overview of the small numbers problem, the MAUP and epidemiological scale are provided below.

2.3.1. Small Number Problem

The Center for Disease Control and Prevention (CDC, 2002) defines stable rates as those with at least 20 cases in the numerator, which corresponds to a 22% relative standard error (RSE) of the rate (please see the methods section for the equation to calculate the RSE); also referred to as the "rule of twenty" (Indiana State Department of Health, 2005). Likewise, the US Bureau of the Census suggests having at least 50 population in the denominator to define stable rates (2002). Thus, having a minimum of 20 cases in the numerator and 50 population in the denominator is considered a 'stable' rate. In the construction of MMAs it is also important to think about what kind of future analytical analyses may be conducted. For example, in multilevel modeling it will be important to have at least 30 cases per unit (and 30 units) to achieve an adequate sample size to model fixed effects; this is also referred to in the literature as the "30/30 rule" (Kreft, 1996).

2.3.2. Modifiable Areal Unit Problem (MAUP)

The MAUP occurs when statistical results are sensitive to changes in the geographic units of which data are collected (Fotheringham & Wong, 1991). The results may vary with changes in the levels of aggregation (scale effect) and the configuration of the zoning scheme (zone effect). Gehlke and Biehl (1934) first showed that correlation coefficients for variables with absolute values would increase when contiguous areal units were aggregated because the variation in the variable(s) decreased as aggregation increased. The MAUP is also observed with changes in scale for maps of disease rates, ratios and proportions due to changes in case and/or population counts with aggregation *or* de-aggregation. The zoning effect refers to variation in results when the units of analysis are reconfigured, e.g., changes in the configuration of ZIP Codes.

2.3.3. Epidemiological Considerations

The International Statistical Classification of Diseases (ICD) and related health problems is used worldwide to classify diseases and aggregate them into disease groups for diagnostic and epidemiological purposes (WHO, 2011a). The MDCH uses the ICD-9-CM version; however, the ICD-10 version is released and is in use by many US states and counties. The ICD-9-CM disease tabulations are organized into 17 primary diagnostic groups and two supplementary classifications. A list of these groups is provided in Appendix 1 (WHO, 2011b). Within each primary disease group, there is a breakdown of sub-groups and specific diseases. In this study, MMAs are constructed for two sub-groups (ischemic heart disease and diabetes) of two primary diagnostic related groups (diseases of circulatory system and endocrine, nutritional and metabolic diseases, and immunity disorders) to explore the constraint parameters for optimal zone design and their meaning for health services research.

2.4 Automated Zone Matching (AZM) Methodology

This research will adopt the AZM methodology to address the small case-number problem, the MAUP and changes in epidemiological scale in the creation of optimal MMAs. A brief review of the AZM literature is therefore, provided (Table 1).

Table 1. Summary of literature on zone design.

Author(s)	Purpose	Data	Parameters	Conclusions
Martin (2003)	To solve the problem of matching incompatible zonal geographies	2001 UK Census EDs and tracts	Population threshold = 100, Population target = 250, Shape = compact, Homogeneity = tenure Simulated annealing = Yes	AZM offers an automated approach to the matching problem.
Cocking and Martin (2005)	To demonstrate the usefulness of zone designs versus census boundaries; explore the MAUP in traditional versus new zone designs	Townsend score, self-reported LLTI (1991 UK Census) EDs and wards	Population target = from 250 to 4500 in the increments of 250, Population threshold =90% of target value	Purpose-specific automatically designed zoning system is more effective than census boundaries
Haynes et al. (2007)	To identify alternative sets of neighborhood units using zone design and comparing their characteristics	Survey, 1991 UK Census EDs	Shape = increasingly stringent, Adjacency = railways and major roads, Homogeneity = housing type	The administrative neighborhood units were not an improvement
Haynes et al. (2008)	To use multilevel modeling to identify variations between subjective and automated zone design neighborhoods	Survey, 1991 UK Census EDs	Population Target = larger/medium/smaller, Shape = compact (weak), Homogeneity = Townsend/tenure/house type	Subjective neighborhoods did not produce stronger neighborhood effects than computer generated areas.
Riva et al. (2008)	To assess the ability of census tracts as units to measure the active living potential of environments	Canadian Census Dissemination areas (DAs)	Homogeneity = population density, land use mix, and geographic accessibility to proximity services	Census tracts are limited for measuring active living potential

Table 1 (cont'd)

Stafford et al. (2008)	To identify area boundaries using three methods and compare the extent of health inequalities across each area	1999 health survey , 2001 UK Census Wards, output areas	Homogeneity = proportion in rented social housing	Alternative definitions of boundaries have no substantive effect on the estimates of health inequality.
Flowerdew et al. (2008)	To evaluate how the effect of the neighborhood on health differs with different neighborhood definitions	1991 UK Census Wards, EDs	Population threshold = 2651, population target = 8136 (average), Shape = compact (weight 1 or 10), Homogeneity = employment and tenure, Simulated annealing = Yes	Boundaries matter in health studies.
Grady and Enander (2009)	Implement a methodology by which to conduct public health surveillance for low birth weight and infant mortality in Michigan	Michigan's Vital Statistics ZIP Codes	Population threshold = 25, Population target = 20, Shape = compact, Homogeneity = mother's race	AZM proved to be a useful tool to visualize and explore the spatial patterns of low birth weight and infant mortality for public surveillance.
Statistics New Zealand (2009)	To evaluate AZTool for generating robust statistical output zones that fulfill pre-specified optimal characteristics	2006 area units, Meshblocks	Population threshold = Yes, Population target = Yes, Shape = Compact, Homogeneity = population or household size	AZTool has been evaluated to be useful to create new output geographies to replace the current ones.
Grady (2010)	To access the impact of racial residential segregation on low birth weight and preterm birth using optimized neighborhood boundary definitions.	Michigan's Vital Statistics birth registry, Census tracts	Population threshold = 20, Population target = 30, Shape = Compact, Homogeneity = race, rental properties, vacant properties, and the presence or absence of a major road or highway	Optimization of neighborhood boundary definitions is recommended for future studies on impacts of racial segregation.

As Openshaw (1977) points out, the existence of scale and aggregation effects is seen as a fundamental characteristic of spatial data. They cannot be removed without doing possibly irreversible damage. Thus, the only way to minimize these effects is to control the scale and aggregation characteristics of spatially aggregated data. Openshaw then develops a heuristic procedure that turns scale and aggregation problems to one of optimal-zone design. This attempt is made to identify a set of zones, which optimizes an objective function related in some way to the performance of a model subject to whatever constraints may be relevant. The development of this automated zone procedure was regarded as the pioneer work to address the MAUP. This procedure is revised and improved by Openshaw and Rao (1995).

Later on, different software packages incorporate the principles of automated zone design originally conceptualized by Openshaw (1977) came to the world. The relevant packages include SAGE (Haining et al., 1998; Wise et al., 2001), ZDES (Alvanides et al., 2001), A2Z (Daras & Alvanides, 2005), AZM (Martin, 2003) and its newer version AZTool (Cockings, et al., 2011). There are a small but growing number of studies utilizing zone design software to evaluate the spatial patterns of health outcomes and local risk factors. These packages offer different options of parameter input and they all greatly contribute to the development of software used in zone design. Martin (2003) presents his early work with AZM, extending from Openshaw's Automated Zone Procedure (AZP). AZP algorithm is used in order to maximize the match between two zonal geographies to reconcile incompatible zoning systems.

Most of the studies using zone design software focus particularly on MAUP. Cockings and Martin (2005) conducted seminal research using AZM to measure the correlation between self-reported limiting long-term illness (LLTI) and area-level deprivation at the enumeration district (ED) and ward scales in Avon, a former county in the UK. This approach took EDs as building blocks and undertook repeated design of the Avon zoning system at different scales in order to examine both the scale and aggregation aspects of the MAUP in relation to the deprivation-health relationship. LLTI was measured using the Standard Morbidity Ratio (SMR) from the 1991 Census. Areal-level deprivation was measured using the Townsend score involving standardization. The parameters used in this study were population threshold, population target and shape. Population target was set at the values from 250 increasing to 4,500 with step size of 250, covering the range from ED to ward scales. Population threshold was set to 90% of the target value to reduce the variation in acceptable zone sizes and thus aid in the production of alternative zoning systems at predetermined scales. The shape parameter (simple statistics as perimeter^2/area) was minimized. Note that there was no homogeneity constraint specified in the analysis. The best result from the fifty random restarts is used for the analysis. The correlation coefficients of SMR and deprivation for each set of parameters were calculated. The authors concluded that the correlation between variables were markedly affected by the choice of zoning system, and were strongly associated with the scale of aggregation. The authors demonstrated the ability of AZM to explore the influence of pre-defined and alternative zoning systems and recommended the using of zone design tools as potentially important role in environmental and health studies.

A study by Haynes et al. (2007) used A2Z software to compare computer-generated zone design with areal units subjectively defined by local government officers in the city of Bristol, UK. They created seven sets of different parameter input. Three increasingly stringent shape constraints were applied for the first three sets. A railroad and major roads adjacency constraint was used in the fourth set. The fifth attempted to align zone boundaries along ward boundaries. The first five zone design all used deprivation as homogeneity constraint. The sixth used a weak shape constraint, but housing type as a homogeneity constraint; while the seventh also took account of ward boundaries. To create sets of neighborhoods at different scales that would be alternative to the subjectively defined communities, they also created systems of 50, 101 and 150 zones, maximizing the homogeneity of deprivation or housing type in turn, using a weak shape constraint and no population size requirement. They turned out to find that automated zone design at different scales could identify much more homogeneous areas than subjectively design ones, but had less shape compactness. The authors concluded that in the construction of optimal zone designs there needs to be balanced between the use of neighborhood homogeneity constraints and constraints relating to zone shape and boundary alignment.

Following earlier tract, Haynes et al. (2008) used A2Z software to study pre-school children accidents using multilevel modeling to identify variations between alternative sets of subjective and automated zone design neighborhoods in southwest England. They designed 13 different sets of “neighborhoods” for the study area. One consisted of enumeration districts, which are units in 1991 UK Population Census. The other 12 used enumeration districts as building blocks and grouped them into larger units at three spatial scales （N = 100, 201, 307）. Each automated zone set maximized the homogeneity of selected census characteristics of Enumeration Districts (EDs) within zones, subject to a weak shape constraint. Three sets of census measures (the Townsend composite index of material deprivation, housing tenure, and housing type) were used in turn within each spatial scale. Information about accident occurrence and measures of physical activity, total development and conduct difficulties, mothers’ age at delivery, post-natal depression, life events, social support and smoking status were taken as variables. Multilevel modeling was used to identify variations between subjectively defined neighborhoods and computer-generated zones. The risk of accidents to pre-school children, and most of the characteristics of children and mothers associated with accident risk, varied significantly between neighborhoods. Generally, neighborhoods subjectively defined by planners did not produce stronger effects than computer-generated areas.

In contrast, a study conducted by Riva et al. (2008) intended to assess the soundness of census tracts as units of analysis for measuring the active living potential for environments, hypothesized to be associated with walking. K-mean clustering method (performed in SAS program) was used to classify smallest areas into clusters (e.g., types of environments). Through the use of zone design, using the homogeneity constraints population density, land use mix and accessibility to services, the authors identified seven types of environments within which varying levels of active living were possible. Then they compared the census tracts to the designed zones to evaluate the degree of soundness of census tracts as units of analysis. The results showed that the soundness of census tracts for measuring active living potential may be limited.

In the same year, Stafford et al. (2008) used three methods to define area-boundaries and compare the extent of health inequalities across each drawing on data from the London boroughs of Camden and Islington. The first one used administrative boundaries, specifically 2001 census ward boundaries. The second method drew boundaries using physical and man-made features of the environment, i.e., roads, railway lines, canals and areas of parkland. The third method was designed to maximize the socioeconomic homogeneity of residents using ZDES 3b software. Census 2001 output areas were used as the component geography. The proportion in rented social housing was taken as the weighting variable. The software was configured to perform 10 runs in order to select the best performance. The selected optimization criterion ensured that shape-compact zones were produced. The results of the two-level hierarchical models showed that there was a tendency for slightly larger estimated inequalities across areas defined by socioeconomic homogeneity compared with other definitions, but differences between methods were very small. They finally pointed out that, estimates of the extent of variation in health across neighborhoods - neighborhood inequalities in health - were very similar irrespective of the way in which the neighborhood boundaries were defined. Although administrative area boundaries have little theoretical basis for health study, those studies indicated that they have no substantive effect. Based on these findings, they can have greater confidence in the results of the numerous studies which have used administrative boundaries to define the neighborhood.

While a great number of studies indicate that different sets of boundaries don’t have a significant effect on health study outcomes, however, we can still hear the voices that disagree with this statement. Using AZTool, Flowerdew et al. (2008) have used British census Enumeration Districts as building blocks to construct alternative zonal systems using different sets of criteria, and experiment to see if neighborhoods defined in different ways have similar implications for health. Population threshold and population target were determined by the size of ward. The first three sets assigned relative weights 1:1, 10:1, and 1:10 to population: shape, without using any homogeneity constraint. Employment and tenure were used separately as homogeneity constraint in the fourth and fifth set of parameters. They found out that it did matter where the boundary was drawn. The conclusion indicated that the effect of neighborhood conditions should be looked at using several different ways to define neighborhoods, and that the size and composition of the neighborhoods may vary in different parts of a study area.

Statistics New Zealand (Ralphs & Ang, 2009) published a complete report which evaluates

AZTool for generating robust statistical output zones that fulfill pre-specified optimal characteristics such as compactness of shape, minimum population size, standard mean population size, and constrained nesting within larger areas. They find that the new geographies produced by AZTool substantially out-perform the current geographies across almost all of the optimization criteria. The algorithm is stable, and is able to repeatedly generate high-quality solutions in a timely manner. They conclude that the ArcGIS/AZTool toolkit would form the basis of a viable workflow for the automatic production of optimal geographical areas.

Grady and Enander (2009) investigate the spatial patterns of low birthweight and infant mortality in the State of Michigan using AZM methodology and minimum case and population threshold to calculate stable rates and standardized incidence and mortality ratios at the ZIP Code level. Applying AZM with a target population of 25 cases and minimum threshold of 20 cases resulted in the reconstruction of zones with at least 50 births and RSEs of rates 20–22% and below respectively, demonstrating the stability of these new estimates. Other AZM parameters included homogeneity constraints on maternal race and maximum shape compactness of zones to minimize potential confounding. With these model parameters the AZM analysis was conducted by running 50 program restarts with 100 iterations each taking the run (i.e., zone design) with the most compact shape, the strongest internal homogeneity and lastly the best target population statistic. They found that the fifty random restarts were almost equally optimal in terms of detecting the significant areas (i.e., clusters with elevated rates), and thus determined to use the initial random aggregation (IRA) as the zone design for the purpose of public surveillance. The AZM identified areas with elevated low birthweight and infant mortality rates and standardized incidence and mortality ratios. The authors concluded that the

AZM proved to be a useful tool for visualizing and exploring the spatial patterns of low birthweight and infant deaths for public health surveillance.

Following earlier research, Grady (2010) studied the impacts of racial residential segregation on low birth weight using improved neighborhood boundary definition. AZM was used to recombine census tracts into new output zones. A target population of 30 cases and a minimum population threshold of 20 cases were set as parameter constraints. The homogeneity constraints used were race, rental properties, vacant properties and the presence or absence of a major road or highway. The shape constraint which maximized shape compactness was also used in the study. Following the new definition of neighborhood boundary, two-stage hierarchical generalized linear models (the Bernoulli models) were estimated conceptualizing mothers and infants nested within zones of varying levels of racial isolation, racial clusters and poverty to assess the effect of high racial isolation and high racial clusters on intrauterine growth retardation (IUGR) and preterm birth. The results showed that high racial isolation had significant impacts on IUGR, while the odds of preterm birth were higher in racially clustered zones. MAUP effects were not found in the models. Optimization of neighborhood boundary definitions is recommended for future studies on impacts of racial segregation.

Although the conclusions from different studies may vary, the optimized zone design did not perform worse than the subjectively-defined boundaries and could be obtained easily through computation using automated zone design. Since the software packages will enable users to define their own parameters and the outputs are largely dependent on the input parameters, the future research could focus on standardization of parameters and therefore to optimize the zone design outputs.

2.5 Summary of Literature

This review of the literature presented some of the major problems that will be encountered when using the residential ZIP Code of patients discharged with IHD, diabetes, circulatory diseases and endocrine disorders to construct MMAs to assess the 'demand' for inpatient hospital services in Michigan. AZM is presented as a methodology to address these problems.

The review of the AZM literature shows that the field is still in relatively early development in terms of modeling zone designs for health and health services research. For example, very few studies reported on the full set optional of parameters that are available for use in zone design. Previous studies appeared to run the program for a number of times, but some of them simply used one of the zone designs (e.g., the first restart) without validating the optimal one within multiple restarts. This study will advance previous research on AZM zone design by exploring global optima versus local optima used in previous research to further address the MAUP (described in more detail in the Methods Section Step 5). It will also explore different weighting schemes between the three constraint parameters (minimum case threshold versus shape compactness versus intra-area correlation coefficients). Finally, this study will present a statistical approach to evaluate which of 50 restarts (output generated) is the optimal zone design.

Furthermore, most of the studies focus only on the aspect of zone design, i.e., geographic considerations, regardless of epidemiological considerations. For this study, optimal MMAs were constructed based on both larger and smaller epidemiological scales for different purposes. This study will assess zone design in relation to epidemiological scale. Elevated and low areas of the diseases studied will be reported for future research.

2.6 Purposes of Study

The goals of this study are to (1) design optimal Medical Management Areas (MMAs) for IHD and diabetes to (2) visualize and explore the spatial patterns of these disease subgroups to inform future bed need in Michigan. MMAs of these sub-groups will be compared with MMAs of the principal diagnoses to explore geographic and epidemiological scale together.

2.7 Objectives

The objectives of this study include:

1. To demonstrate the need for the use of AZM in the construction of MMAs:

   Hypothesis 1a: Mapping disease rates by ZIP Codes will result in a large proportion of unstable rates.

2. To optimize MMA boundary definitions:

   Hypothesis 2a: The shape constraint parameter will be less important than the internal homogeneity for broad disease groups compared to specific diseases because of the greater variation in population within the broad disease groups;

   Hypothesis 2b: Aggregating on the variables sex, age and race will increase demographic homogeneity within zones.

3. To visualize disease proportions, crude rates and age-adjusted rates of IHD, diabetes, circulatory diseases and endocrine disorders:

   Hypothesis 3a: The spatial patterns of disease groups will vary by method used to map the diseases; new information about the prevalence of these diseases in Michigan will be learned from each method.

3. DATA AND METHODS

3.1 Data

Geographic Data

The ZIP Code boundary file that will be used in this analysis will be obtained from ESRI (2007). ESRI receives these from a private company named Tele Atlas (ESRI, 2011). The 2007 ZIP Code boundaries (n=900) will be used because it provides the boundary that is closest to the year of the MIDB 2008. Due to the requirements of the AZTool, all polygons need to be contiguous; therefore, the topology of the ZIP Code boundary file will be cleaned by removing all isolated islands, overshoots and undershoots, and connecting Michigan's Upper and Lower Peninsulas for statewide analysis.

Health Data

The Michigan Inpatient Hospital Database (MIDB) provided by the Michigan Department of Community Health (MDCH) – Division of Health Policy and Access will be used in this research. The MIDB is a database containing the inpatients' discharge records for all Michigan hospitals and Michigan residents discharged from hospitals outside of Michigan for a calendar year. The MIDBs from 2000 to 2008 were given to researchers in Department of Geography at the Michigan State University by MDCH for research bed need and health care access within the State of Michigan. "Inpatient" refers to a patient who spent as least one night in a hospital. The data that will be used in this analysis from the 2008 discharge records (n=1,174,862) will include patients' demographic information –i.e., sex, age, race/ethnicity, ZIP Code of residence and principal diagnosis at time of discharge (ICD-9 code). Only those records that pertained to the disease groups to be studied will be used in this analysis, specifically: ischemic heart disease

(IHD) (ICD-9-CM: 410-414) (n=58,573), diabetes (ICD-9-CM: 250) (n=17,352), circulatory diseases (ICD-9-CM: 390-459) (n=214,649) and endocrine, nutritional and metabolic diseases, and immunity disorders (ICD-9-CM: 240-279) (n=44,057).

Population Data

Although the 2010 census has already released some primary results, detailed population with age groups at the ZIP Code Tabulation Area (ZCTA) level are not yet available on the census website to calculate area-based crude and age-adjusted rates. Detailed population estimates in 2008 were therefore, obtained from Geolytics, Inc. (http://www.geolytics.com/). These data include the total population and population by age-groups (per below) for each ZIP Code.

3.2 Methods

MMA-zones will be obtained by aggregating ZIP Codes using the optimal zone design generated by AZTool. For the health data, proportions (no. cases/total hospital discharges), crude rates (no. cases/population) and age-adjusted rates will be calculated using the new MMA-zones. Those different views of the health data may be used to inform inpatient hospital services.

Computer Architecture

The data processing and analyses are accomplished on a Sun Microsystems Ultra 20 Workstation, which was running Microsoft Windows XP Professional. This workstation is remotely logged into from a Sun Microsystems Fire V40z Server containing MySQL using

KRDC. The server runs Sun Solaris 10 (x86). Additional software used for this research includes Microsoft Excel 2007, ESRI ArcGIS 9.3, SAS 9.2 and AZTool.

Workflow

Step 1: Database imported into SAS. The complete database MIDB08 is exported from the server using a SELECT query (SELECT * FROM endemic.MIDB08;) in MySQL and stored as a text file on the workstation. No record or field is removed during this process.

Step 2: Clean the databases. The original ESRI ZIP Code boundary file is modified by removing all the slivers and isolated islands and then connecting Upper Michigan and Lower Michigan. Thus, every ZIP Code can find at least one contiguous neighbor. The text file in Step 1 is imported into SAS as a new table. The MIDB data is cleaned to assess any missing data and to identify outliers. Out of state inpatient records are removed (n=18,078). The records containing missing and vague ZIP Codes (e.g., 48400, 48600), are removed as well (n=236). MIDB ZIP Codes that don't have matched records in the ESRI 2007 ZIP Code boundary shapefile are recoded according to visual comparison (n=65) (please refer to Appendix 2 for a list of those ZIP Codes that were removed or recoded).

Step 3: Calculate the number of cases by sex, age and race. The health data was queried from the cleaned SAS table. If IHD is to be studied, the records with the corresponding ICD-9 codes are selected from the cleaned database. Information on sex, age groups (0-14 years old, 15-44 years old, 45-64 years old, 65-74 years old, and 75 years old and older) and race groups (white, black and others) are recoded and aggregated by residential ZIP Code. Additionally, using the same age divisions, the total numbers of patient discharges are also aggregated by ZIP Code to calculate the proportions (no. cases/total number patient discharges) of disease. The proportions were calculated in SAS along with their 95% confidence (upper and lower) intervals (CI) and relative standard errors (RSEs).

The 95% CIs are used to measure the precision around the proportion, i.e., 5% uncertainty that the proportion occurred by chance. Since the major factor determining the length of a confidence interval is the size of the sample, the confidence interval becomes narrower –i.e., the proportion is more valid when the sample size is getting larger. Another way of determining if the proportion is unstable due to small numbers is to calculate the relative standard error.

Below are the formulas used to calculate the RSE and 95% CIs in this thesis (Healthy People 2010 Statistical Note, 2002).

$$RSE = \frac{SE}{rate} \times 100$$

$$SE = \frac{rate}{\sqrt{cases}}$$

$$RSE = \frac{rate}{\sqrt{cases}} \times \frac{1}{rate} \times 100 = \frac{1}{\sqrt{cases}} \times 100$$

```
/* IHD */
data IHD_proportion ; set IHD_age_diss ;
proportion = (cases / total_disch) * 1000 ;
RSE = 100 * (sqrt (1/cases)) ;
if cases >= 100 then
do ;
CIL = proportion - (1.96 * proportion * (RSE/100)) ;
CIU = proportion + (1.96 * proportion * (RSE/100)) ;
end ;
else
if cases < 100 then
do ;

if cases =    1 then L = 0.02532 ; else if cases =    2 then L = 0.12110 ; else if cases =    3 then L =
0.20522 ; else if cases =    4 then L = 0.27247 ; else

etc… up to 100 (please see NCHS 2010)

CIL = proportion * L ;
CIU = proportion * U ;
end;
if cases = '.' then cases = 0 ;
if total_disch = '.' then total_disch = 0 ;
if proportion = '.' then proportion = 0 ;
if RSE = '.' then RSE = 0 ;
if CIL = '.' then CIL = 0 ;
if CIU = '.' then CIU = 0 ;
```

Step 4: Join the health database to the ZIP Code boundary shapefile. In ArcMAP, the health database is joined to ESRI 2007 cleaned shapefile using the key ID = ZIP Code. This shapefile is exported as a new shapefile ready to be imported into AZTool.

Step 5. Checking the instability of IHD and diabetes proportions by ZIP Codes. Of 900 ZIP Codes, 276 (30.7%) had fewer than 30 IHD cases and 578 (64.2%) of ZIP Codes had fewer than 30 cases of diabetes. These numbers demonstrate that using ZIP Codes as the unit of analysis will lead to a high proportion of unreliable rates (i.e., high RSEs of the rates) (Figures 1-2) justifying the need for AZM to construct MMAs in Michigan.

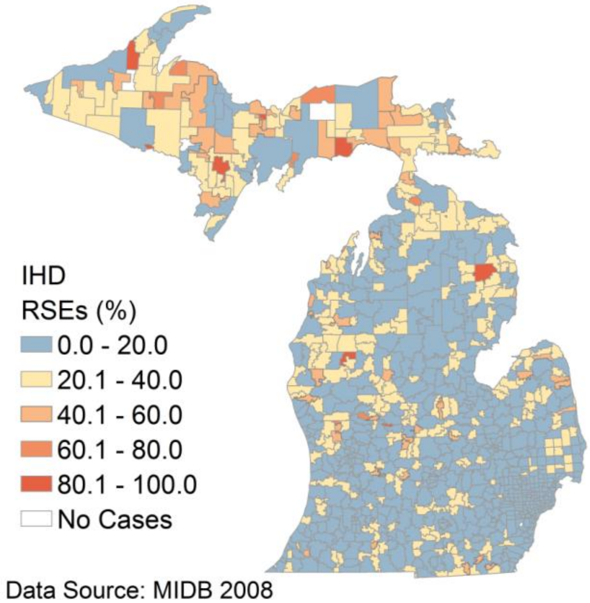

Figure 1. IHD: Relative standard errors (RSEs) of proportions by ZIP Code [1]

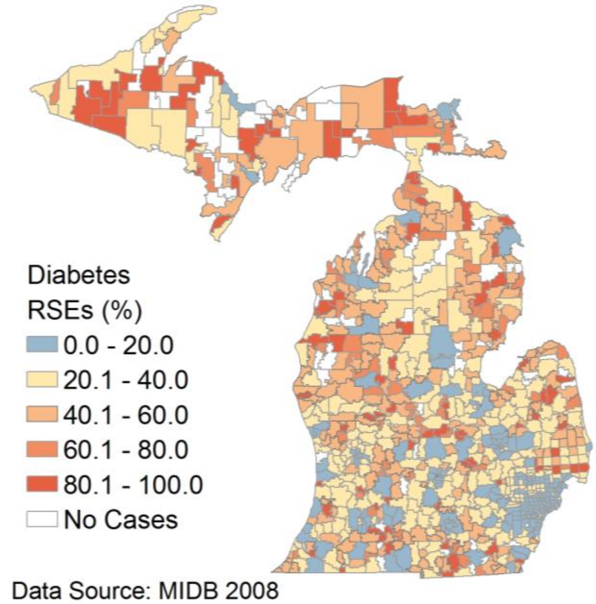

Figure 2. Diabetes: Relative standard errors (RSEs) of proportions by ZIP Code

[1] For interpretation of the references to color in this and all other figures, the reader is referred to the electronic version of this dissertation.

Step 6: Using AZTool to generate MMA-zone assignment files. The cleaned ZIP Code shapefile is input into AZTool – AZTImport is used to generate intersection (.pat) and contiguity (.aat) files. A python code is used to make the contents in every column align to the right in accordance with the requirement of AZTool_Developer. The intersection and contiguity files are input into AZTool_Developer to produce optimal zone assignment using the constraint parameters defined below.

This methodology implements a computationally intensive algorithm that recombines ZIP Codes from the overall dataset into a smaller set of output zones from within which to calculate stable proportions and rates. This is an iterative process, by which one geographic unit is randomly selected and user-defined constraint parameter(s) (below) are evaluated. If the parameters are not met, AZM will search contiguous ZIP Codes until it is achieved, thereafter, aggregating the data and dissolving internal boundaries to create a new zone.

Minimum Population Threshold

The first parameter selected is the minimum population threshold. In this study, the minimum population threshold is set as 30 cases. Thus, all new zones created have at least 30 cases from which to calculate stable rates. The relative weight is set to 100%.

Shape Compactness

The second parameter (perimeter^2/area, P2A) selected is the shape compactness, defined as:

$$\sum \frac{1}{4\pi} \frac{q_k^2}{A_k}$$

Where q_k is the perimeter of zone k and A_k is its area. The automated zone design function is to minimize the perimeter squared divided by area, which maximizes shape compactness of zones. In this study it is desirable to have compact zones in order to minimize the potential for local variability across the zone. The relative weight is set to 50%.

Internal Homogeneity

The third parameter selected is the homogeneity constraint. The homogeneity constraint promotes homogeneity within zones and heterogeneity between zones. This study uses inpatients' sex, age and race as homogeneity constraint. Although the theoretical maximum score for internal homogeneity is 1.0, this will not be found in census areas, and any value above about 0.05 implies a reasonable degree of homogeneity (Martin & Cockings, 2001).

Consider a grouping variable, with k=1,2,…, K categories. This study has K=3 categories (sex, age, race). For each category, e.g. sex measure of homogeneity can be obtained using the intra-area correlation (IAC). This is calculated as:

$$\delta_k = \frac{\frac{1}{M-1}\sum_{g=1}^{M} N_g \left(P_{kg} - P_k\right)^2}{\left(\overline{N^*} - 1\right) P_K \left(1 - P_K\right)} - \frac{1}{\left(\overline{N^*} - 1\right)}$$

Where, $\overline{N}*$ is the mean case size of the ZIP Codes, with an adjustment (Tranmer & Steel, 1998) to take into account variation in the case size of the units. In practice, $\overline{N}*$ is close to $N*$. N_g is the number of cases of a ZIP Code, P_K is the overall proportion of the cases in category K, and P_{kg} is the proportion in category k in ZIP Code g.

After each variable is calculated, the overall IAC is calculated as:

$$\delta = \frac{1}{K-1}\sum_{k=1}^{K}\left(1 - P_k\right)\delta_k$$

The overall relative weight is set to 100% for the sub-groups of disease and 150% for the primary diagnostic groups. Circulatory diseases and endocrine disorders generally have more variability in patients within ZIP Codes than IHD and diabetes because of the epidemiological scale of analysis. Within internal homogeneity each constraint parameter is given equal weight.

Simulated/Synthetic Annealing (SA)

AZTool will be run in SA mode, swaps which cause a deterioration in the overall solution will be allowed in the first half of the iteration cycle (50 iteration cycles). This enables the program to search for the global-optima by escaping from the local-optima. The size of initial margin is set to 1 (default) in this study.

Region

The last constraint parameter is that the new zones will be constrained within existing Hospital Service Areas (HSAs). Because ZIP Codes are not perfectly nested within HSAs, a ZIP Code is considered to belong to an HSA if the centroid of ZIP Code falls within the HSA. Constraining zones within HSAs was preferred because these are health services administrative units from within which, hospital services are managed. From an administrative perspective it was therefore, important to create new zones that respected these boundaries.

Optimization

With these parameters, the AZM analysis is conducted by running 100 times with 20 iterations taking the run with the minimum threshold, the most compact shape and the maximum homogeneity as the optimal AZM-MMA boundary definition. All the input route and parameters are described in an .xml file. A batch mode is utilized to execute a single .xml file for 50 times for each disease/disease groups. Each output assigns a tract ID (repeatable) for a ZIP Code, as well as generates a statistical report containing the information about zone number, zone size, P2A score, IAC score, etc. Of the 50 outputs, the one with the lowest P2A and highest IAC is chosen to be the optimal (MMA) zone design. With the same parameters, AZTool is conducted by running 50 program restarts with 100 runs and 20 iterations, each of which takes the most compact shape and maximum IAC. The optimal output of the program restarts is identified using Wilcoxon signed-rank test (a paired difference test) to select the lowest

P2A and highest IAC score. Because the P2A score is almost 1 million times larger than the IAC score, they are not comparable on different orders of magnitude. To rank the outputs in a direct way will result in simply ranking the P2A score. Thus, the IAC score is magnified by multiplying the ratio of the average of P2A scores to the average of IAC scores. The modified IAC score will then be ranked along with the P2A score.

Step 7: Dissolve the ZIP Code boundaries. Join the zone assignment file and second table generated in step 3 on the boundary shapefile. Dissolve the boundary using new TractID and keep the variables of summation of cases, total discharge and total population by age groups. The attribute table should contain the information with all age groups. Bring the new MMA-Zone table back into SAS to calculate proportions and rates. Herein only the IHD example is provided.

Step 8: Calculate the proportion of IHD. The total number of hospital discharges by MMA-Zone is used in the denominator to calculate the proportions.

$$\text{Proportion} = \frac{\sum case_i}{\sum total\_disch\arg es_i} \times 100$$

Step 9: Calculate the rates of IHD and diabetes. The total population is used as the denominator to calculate the crude rate.

$$\text{Crude rate} = \frac{\sum case_i}{\sum pop_i} \times 100$$

Age-adjustment is a statistical process applied to rates of disease, death, injuries or other health outcomes which allows communities with different age structures to be compared (NYSDOH, 1999). A large number of diseases occur at different rates in different age groups. Most chronic diseases, for example heart diseases, occur more often among older people. Some other outcomes, such as injuries, occur more often among younger people. Thus, an area with more older people tend to show higher rates of chronic diseases, and similarly an area with more younger people tend show higher rates of injuries or other youth- susceptible diseases. The high rates don't necessarily indicate that population within areas are less healthy. In order to minimize the effect brought by different age distribution within an area (e.g., zone), the standard population from a higher political unit (e.g., state) should be used to adjust proportion and rate. In this study, age standardization (direct method) is used to adjust the proportion and rate using the statewide population as standard population.

Age-adjusted rate = $\sum_{i=1}^{5} \frac{\frac{age_i}{pop_i} \times pop_i}{\sum pop_i} \times 100$, i=1, 2, …, 5

Step 10: Visualize the proportion and rate maps. Join the proportion, crude and age-adjusted rates (tables) by MMA-Zone to the dissolved zone-shapefiles and create choropleth maps.

The steps above were repeated for diabetes, the circulatory diseases and endocrine conditions.

4. RESULTS

4.1 Descriptive Statistics

Table 2 shows the descriptive statistics for IHD and diabetes in Michigan. In 2008, there were 1,174,862 hospitalized patients in Michigan. Of these, 58,573 (5.0%) were patients discharged with IHD and 17,352 (1.5%) with diabetes. Of patients discharged with IHD 60.1% were male compared with 51.6% for diabetes. Almost all patients discharged with IHD were over 44 years of age; however, a substantial proportion of patients with diabetes were less than 44 years of age. Of patients discharged with diabetes, 33.9% were Blacks compared with 13.0% for IHD.

Table 2. Patient discharge characteristics for Ischemic Heart Disease (IHD) and diabetes in Michigan, 2008.

		IHD		Diabetes	
		No.	%	No.	%
Sex[1]					
	Male	35,200	60.1	8,951	51.6
	Female	23,367	39.9	8,401	48.4
Age					
	0-14	3	<1.0	906	5.2
	15-44	2,812	4.8	5,389	31.1
	45-64	22,972	39.2	6,224	35.9
	65-74	14,566	24.9	2,215	12.8
	75+	18,220	31.1	2,618	15.1
Race					
	White	50,141	85.6	11,069	63.8
	Black	7,624	13.0	5,874	33.9
	Others	808	1.4	409	2.4
Total		58,573	100.0	17,352	100.0

[1] Data variable missing = 6

4.2 Automated zone matching (AZM) methodology output

The optimal zone design in this study met the requirements of a minimum case threshold of

30, maximum shape compactness (P2A) and maximum internal homogeneity (IAC) (patient's sex, age and race). Based on the statistical results from these constraint parameters the optimal zone design was selected from 50 program restarts (20 runs with 100 iterations in each restart). Tables 3-6 display the statistical results from the constraint parameters (P2A and IAC) for each of the 50 restarts. Thus, most optimal restart was that with the combination of a low P2A and high IAC scores. However, the lowest P2A and highest IAC scores do not necessarily occur in the same 'restart'. Therefore, to select the optimal zone design the Wilcoxon signed rank test was used to prioritize the 50 restarts. The Wilcoxon signed rank test is a method used to compare two repeated measurements on a single sample. In this study it was used to rank the P2A and IAC output from the 50 restarts.

The direct method of Wilcoxon signed rank simply ranks the difference between the P2A and IAC scores from the lowest to the highest in order to get a relatively small P2A and IAC combination score (4th column). However, the P2A score was over 100,000 times larger than the IAC score, and therefore the variation of IAC score ($\pm$0.1) was negligible compared to the variation of P2A score ($\pm$400~500). Thus, this method appeared to rank only the P2A score.

In order to make the IAC score more comparable to the P2A score, the IAC score was magnified by multiplying the ratio of the average of P2A scores to the average of IAC scores. The new IAC score (5th column) was then used to conduct the Wilcoxon signed rank along with the raw P2A score. The final signed rank is shown in the last column of each table.

Comparing the results from the two ranking methods the 'top restarts' are identical for IHD, diabetes, circulatory diseases and endocrine disorders, although the other restarts are not the same. Thus this top restart (zone design) was used to construct the MMAs. Figures 3-6 show the scatter charts of P2A and IAC scores. The marked pair of P2A and IAC in each table is the one that was chosen.

For the 50 restarts for IHD, the P2A score ranged from 12159.985 to 13097.956 and the raw IAC score ranged from 0.101 to 0.104. There was no certain order or obvious pattern for the 50 restarts. For this disease the optimal run was the run 50. In order to validate that the 51st or later runs would not offer a better solution AZM was run for an additional 10 restarts. A better solution was not found within the additional 10 restarts (results not shown).

Table 3. Ischemic Heart Disease (IHD) results from AZM constraint parameter 50 restarts.

# Restart	P2A Score	IAC Score	direct rank	IAC Score*	sign	x_i-y_i	$\|x_i$-$y_i\|$	rank	signed rank
1	12706.433	0.102	32	12581.069	1	125.364	125.364	25	25
2	12578.864	0.103	18	12704.413	-1	-125.549	125.549	26	-26
3	12587.606	0.101	19	12457.725	1	129.881	129.881	28	28
4	12665.61	0.103	27	12704.413	-1	-38.803	38.803	7	-7
5	12650.174	0.102	25	12581.069	1	69.105	69.105	17	17
6	12311.717	0.102	4	12581.069	-1	-269.352	269.352	44	-44
7	12659.281	0.103	26	12704.413	-1	-45.132	45.132	9	-9
8	12530.21	0.102	17	12581.069	-1	-50.859	50.859	11	-11
9	12479.754	0.102	14	12581.069	-1	-101.315	101.315	21	-21
10	12702.271	0.102	31	12581.069	1	121.202	121.202	23	23
.									
.									
40	12406.211	0.102	8	12581.069	-1	-174.858	174.858	35	-35
41	12422.611	0.101	11	12457.725	-1	-35.114	35.114	5	-5
42	12516.299	0.102	16	12581.069	-1	-64.770	64.770	15	-15
43	12606.672	0.103	21	12704.413	-1	-97.741	97.741	20	-20
44	12667.231	0.103	28	12704.413	-1	-37.182	37.182	6	-6
45	12445.146	0.101	13	12457.725	-1	-12.579	12.579	1	-1
46	12731.338	0.102	34	12581.069	1	150.269	150.269	32	32
47	12632.239	0.103	22	12704.413	-1	-72.174	72.174	18	-18
48	12829.676	0.103	42	12704.413	1	125.263	125.263	24	24
49	12412.041	0.102	10	12581.069	-1	-169.028	169.028	34	-34
50	12159.985	0.103	1	12704.413	-1	-544.428	544.428	50	-50

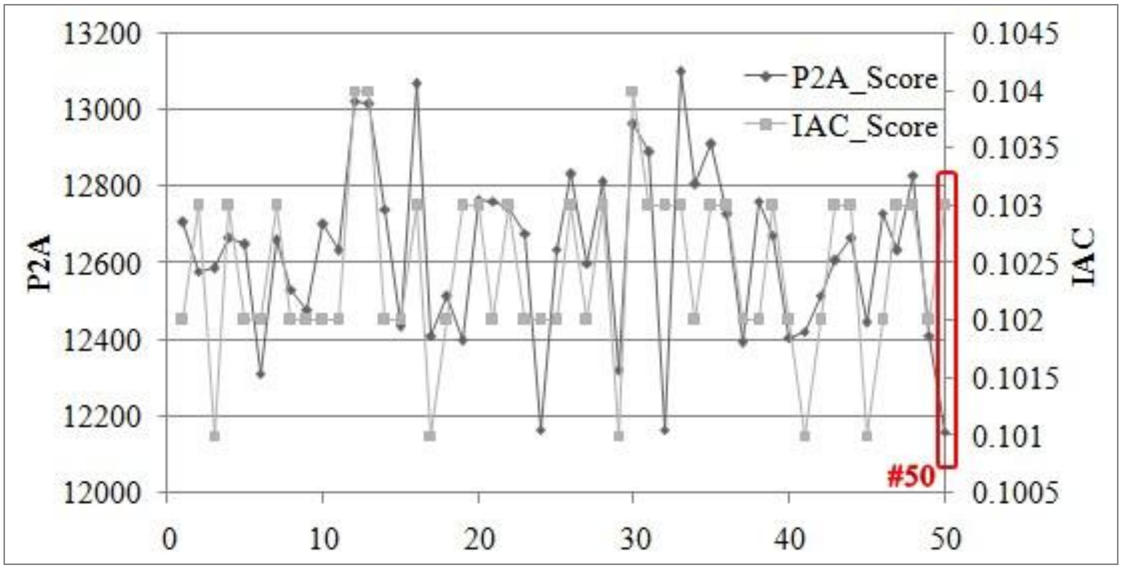

Figure 3. Ischemic Heart Disease (IHD) results from AZM constraint parameter 50 restarts.

For circulatory diseases the 50 restarts showed the P2A scores ranging from 12290.7 to 13180.417 and the IAC scores ranged from 0.105 to 0.107. The P2A score was slightly higher than that of IHD but the IAC score was also slightly higher than that of IHD. This finding suggests that the internal homogeneity of MMAs for circulatory diseases is better than IHD, but at the cost of shape compactness. These findings were the result of the different weighting schemes (150% versus 100%). Again there was no certain order or obvious pattern for the 50 restarts. For circulatory diseases the best run occurred at the 44th restart.

Table 4. Circulatory diseases results from AZM constraint parameter 50 restarts.

# Restart	P2A Score	IAC Score	direct rank	IAC Score*	sign	x_i-y_i	$\|x_i$-$y_i\|$	rank	signed rank
1	12698.642	0.106	11	12830.206	-1	-131.564	131.564	32	-32
2	12814.638	0.106	19	12830.206	-1	-15.568	15.568	8	-8
3	12817.251	0.106	21	12830.206	-1	-12.955	12.955	6	-6
4	13180.417	0.107	50	12951.245	1	229.172	229.172	43	43
5	12833.441	0.106	24	12830.206	1	3.235	3.235	2	2
6	12882.058	0.106	31	12830.206	1	51.852	51.852	19	19
7	12587.874	0.106	7	12830.206	-1	-242.332	242.332	44	-44
8	12653.838	0.105	9	12709.166	-1	-55.328	55.328	20	-20
9	13142.169	0.107	49	12951.245	1	190.924	190.924	41	41
10	12886.871	0.105	32	12709.166	1	177.705	177.705	38	38
.									
.									
40	12852.792	0.106	27	12830.206	1	22.586	22.586	11	11
41	12510.993	0.106	3	12830.206	-1	-319.213	319.213	48	-48
42	12586.063	0.105	6	12709.166	-1	-123.103	123.103	29	-29
43	13113.917	0.106	47	12830.206	1	283.711	283.711	46	46
44	12290.7	0.105	1	12709.166	-1	-418.466	418.466	50	-50
45	12465.152	0.106	2	12830.206	-1	-365.054	365.054	49	-49
46	12939.848	0.106	37	12830.206	1	109.642	109.642	28	28
47	12874.901	0.106	29	12830.206	1	44.695	44.695	16	16
48	12825.617	0.106	23	12830.206	-1	-4.589	4.589	3	-3
49	12997.182	0.107	42	12951.245	1	45.937	45.937	17	17
50	12994.496	0.107	41	12951.245	1	43.251	43.251	15	15

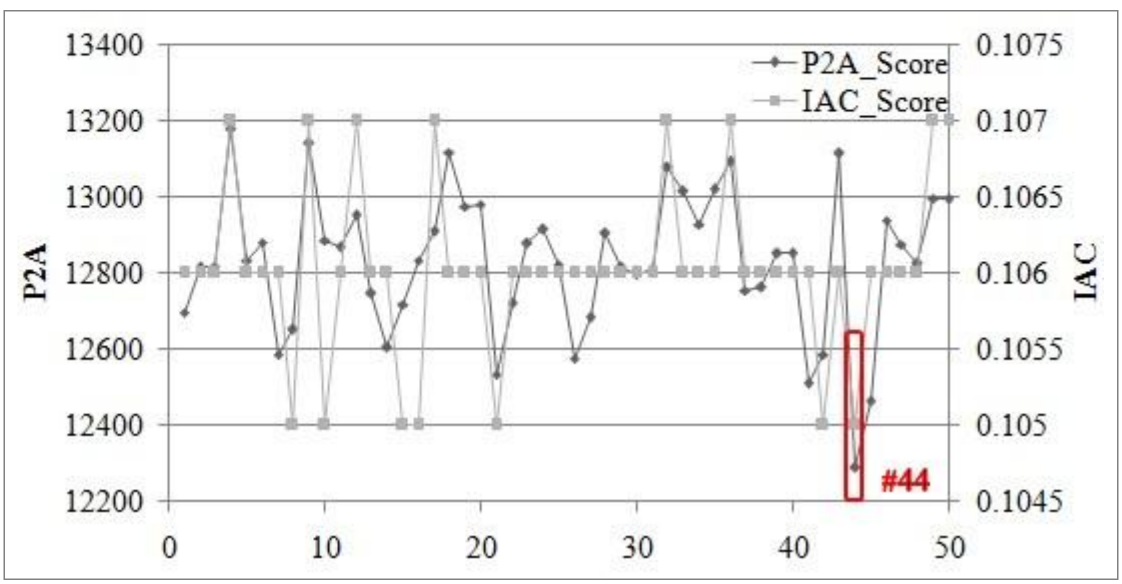

Figure 4. Circulatory diseases results from AZM Constraint Parameters 50 restarts.

For diabetes the P2A scores ranged from 7723.661 to 8382.684 and the IAC scores ranged from 0.113 to 0.115. Both P2A and IAC scores performed better than those of IHD, indicating that MMAs for diabetes has an overall more compact shape and higher levels of internal homogeneity. The best run was the 42nd restart.

Table 5. Diabetes results for AZM constraint parameter 50 restarts.

# Restart	P2A Score	IAC Score	direct rank	IAC Score*	sign	x_i-y_i	$\|x_i$-$y_i\|$	rank	signed rank
1	8199.467	0.114	41	8081.797	1	117.670	117.670	36	36
2	7958.955	0.114	9	8081.797	-1	-122.842	122.842	37	-37
3	8229.87	0.114	45	8081.797	1	148.073	148.073	44	44
4	8073.524	0.113	22	8010.904	1	62.620	62.620	26	26
5	8201.623	0.115	42	8152.690	1	48.933	48.933	19	19
6	8115.091	0.115	28	8152.690	-1	-37.599	37.599	14	-14
7	8010.526	0.113	12	8010.904	-1	-0.378	0.378	1	-1
8	8122.558	0.114	32	8081.797	1	40.761	40.761	17	17
9	8103.236	0.115	25	8152.690	-1	-49.454	49.454	20	-20
.									
40	8143.834	0.115	36	8152.690	-1	-8.856	8.856	3	-3
41	8127.077	0.114	33	8081.797	1	45.280	45.280	18	18
42	7723.661	0.114	1	8081.797	-1	-358.136	358.136	50	-50
43	8054.704	0.114	20	8081.797	-1	-27.093	27.093	8	-8
44	8101.63	0.114	24	8081.797	1	19.833	19.833	7	7
45	7779.17	0.114	2	8081.797	-1	-302.627	302.627	49	-49
46	8226.145	0.114	44	8081.797	1	144.348	144.348	42	42
47	7986.754	0.114	10	8081.797	-1	-95.043	95.043	31	-31
48	8184.887	0.114	40	8081.797	1	103.090	103.090	34	34
49	8029.364	0.113	15	8010.904	1	18.460	18.460	6	6
50	8024.535	0.114	13	8081.797	-1	-57.262	57.262	22	-22

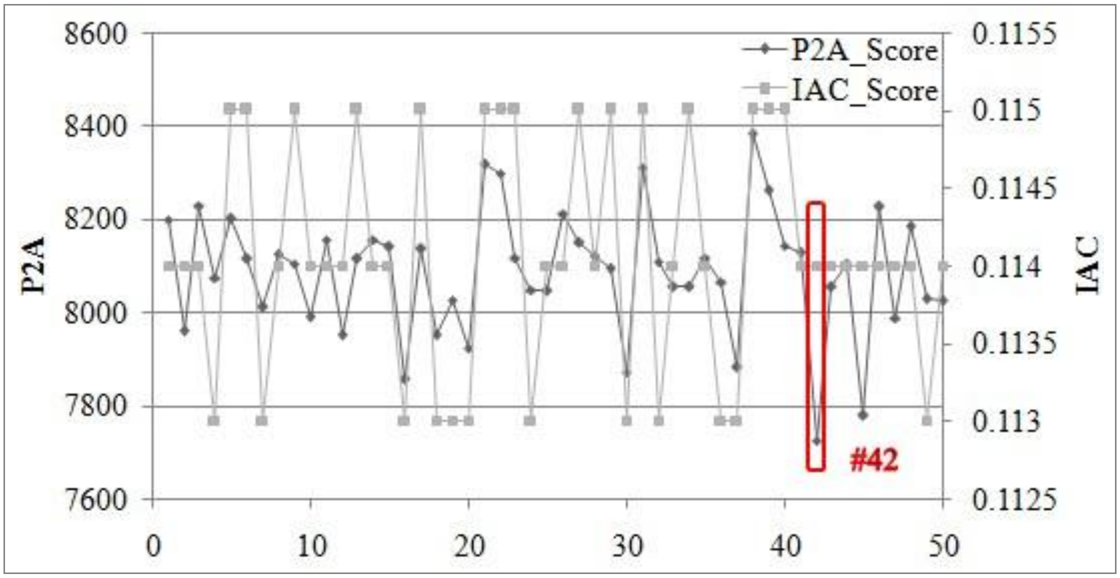

Figure 5. Diabetes results for AZM constraint parameter 50 restarts.

Finally for endocrine disorders the P2A scores ranged from 11205.813 to 12094.776 and the IAC scores ranged from 0.105 to 0.108. The shape compactness is better than for IHD and circulatory diseases, but not as good as diabetes. The internal homogeneity of endocrine disorders (0.105-0.108) is similar to circulatory diseases (0.105-0.107), which is between IHD (0.101-0.104) and diabetes (0.113-0.115). The 45th restart was the best.

Table 6. Endocrine disorders results for AZM constraint parameter 50 restarts.

# Restart	P2A Score	IAC Score	direct rank	IAC Score*	sign	x_i-y_i	$\|x_i$-$y_i\|$	rank	signed rank
1	11978.361	0.108	28	11818.119	1	160.242	160.242	30	30
2	11708.393	0.106	32	11599.265	1	109.128	109.128	20	20
3	11372.357	0.106	6	11599.265	-1	-226.908	226.908	40	-40
4	11772.844	0.107	37	11708.692	1	64.152	64.152	10	10
5	11544.485	0.107	17	11708.692	-1	-164.207	164.207	31	-31
6	11775.188	0.107	38	11708.692	1	66.496	66.496	11	11
7	11811.23	0.107	39	11708.692	1	102.538	102.538	18	18
8	11330.497	0.106	3	11599.265	-1	-268.768	268.768	46	-46
9	11478.534	0.106	13	11599.265	-1	-120.731	120.731	23	-23
10	11449.538	0.106	11	11599.265	-1	-149.727	149.727	28	-28
.									
.									
42	11563.882	0.106	21	11599.265	-1	-35.383	35.383	5	-5
43	11418.243	0.107	9	11708.692	-1	-290.449	290.449	47	-47
44	11836.446	0.107	41	11708.692	1	127.754	127.754	24	24
45	11205.813	0.106	1	11599.265	-1	-393.452	393.452	50	-50
46	11528.659	0.106	16	11599.265	-1	-70.606	70.606	12	-12
47	11376.88	0.106	7	11599.265	-1	-222.385	222.385	39	-39
48	11900.819	0.107	47	11708.692	1	192.127	192.127	38	38
49	11731.725	0.106	34	11599.265	1	132.460	132.460	26	26
50	11372.179	0.106	5	11599.265	-1	-227.086	227.086	41	-41

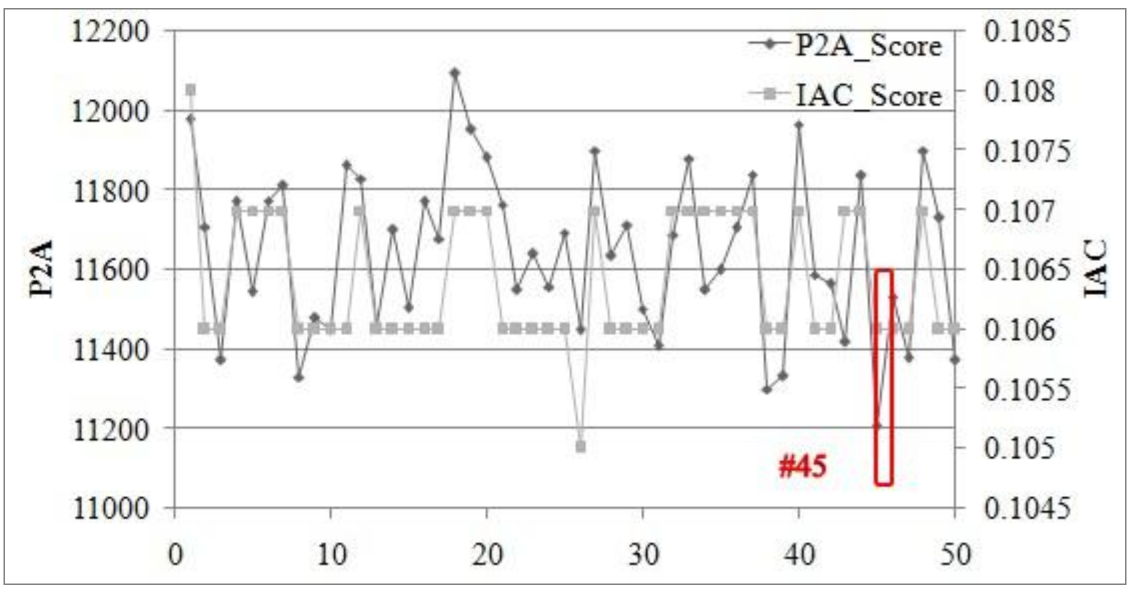

Figure 6. Endocrine disorders results for AZM constraint parameter 50 restarts.

4.3 Medical Management Areas (Proportions)

Following automated zone design the number of ZIP Codes was reduced from 900 to 274 zones for IHD (Figure 7), 310 zones for circulatory diseases (Figure 8), 183 zones for diabetes (Figure 9) and 274 zones for endocrine disorders (Figure 10). Their proportions were calculated within the optimal zone designs. Each zone had at least 30 cases and therefore met the case threshold for stable proportions.

Figure 7 and Table 7 shows the MMAs of crude proportions for IHD. The spatial patterns show MMAs with high proportions (11.8, 95% CI 9.3, 14.7) of patients discharged from hospitals in Au Gres and the northwestern portion of the state (hospital system area) (HSA=6) with IHD. High proportions of discharges are also observed near the southern portion of the state (9.3, 95% CI 8.3, 10.4) (HSA=2) bordering with Indiana and Ohio and in western Michigan (10.6, 95% CI 7.6, 14.3) (HSA=4).

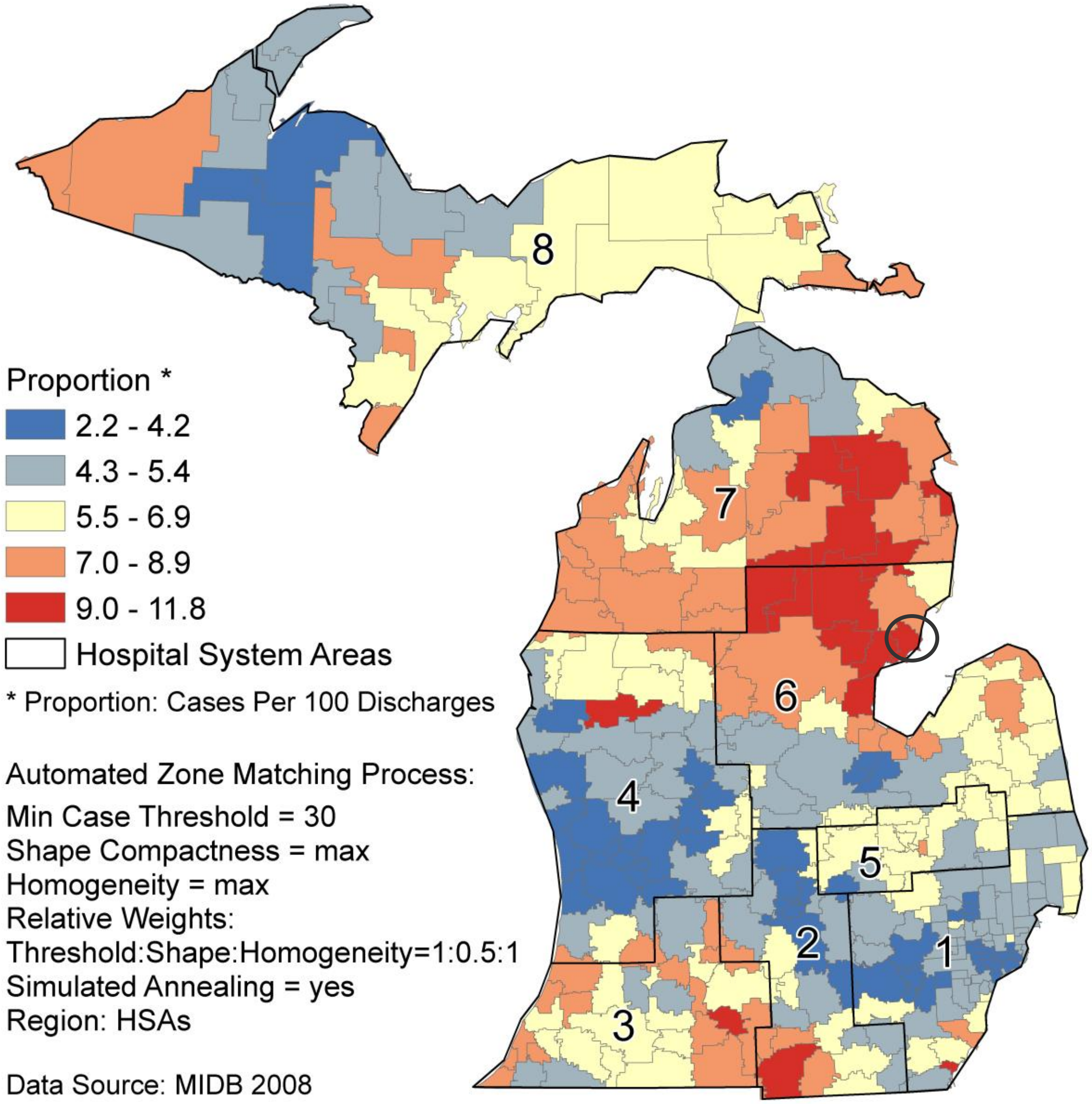

Figure 7. Ischemic heart disease: proportions by MMA (n=302), Michigan 2008.
(statewide proportion: 5.0 per 100 discharges)

Table 7. MMAs with high[1] proportions of patients discharged with IHD, Michigan, 2008.

# MMA	Cases	total	crude proportion[2]	95% CI[3]	RSE[4]
316	77	654	11.8	9.3, 14.7	11.4
627	57	504	11.3	8.6, 14.7	13.2
42	75	679	11.0	8.7, 13.8	11.5
557	39	365	10.7	7.6, 14.6	16.0
210	42	398	10.6	7.6, 14.3	15.4
412	136	1,293	10.5	8.8, 12.3	8.6
170	167	1,598	10.5	8.9, 12	7.7
36	87	857	10.2	8.1, 12.5	10.7
535	268	2,713	9.9	8.7, 11.1	6.1
468	112	1,166	9.6	7.8, 11.4	9.4
393	159	1,658	9.6	8.1, 11.1	7.9
52	297	3,183	9.3	8.3, 10.4	5.8
381	57	613	9.3	7, 12	13.2
151	36	388	9.3	6.5, 12.8	16.7
573	289	3,150	9.2	8.1, 10.2	5.9
302	83	909	9.1	7.3, 11.3	11.0
18	38	426	8.9	6.3, 12.2	16.2

[1] High crude proportions are marked as red areas in the maps.

[2] Proportion: cases per 100 discharges

[3] 95% CI: 95% Confidence Interval

[4] RSE: Relative Standard Error

Figure 8 shows the MMAs of crude proportions of patients discharged with circulatory diseases. These spatial patterns are somehow similar to that of IHD. The northeast and northwestern sides of Lower Michigan (HSA=7), MMAs near the southern border (HSA=2) and some scattered areas of the Upper Peninsula. The highest crude proportion was 30.5 per 100 discharges (95% CI 23.6, 38.7) in Black River and Spruce (Table 8). Those large urban areas, such as Detroit, Lansing, Grand Rapids and Marquette, tended to have low proportions of hospital discharges for circulatory diseases, which conforms to the pattern of IHD.

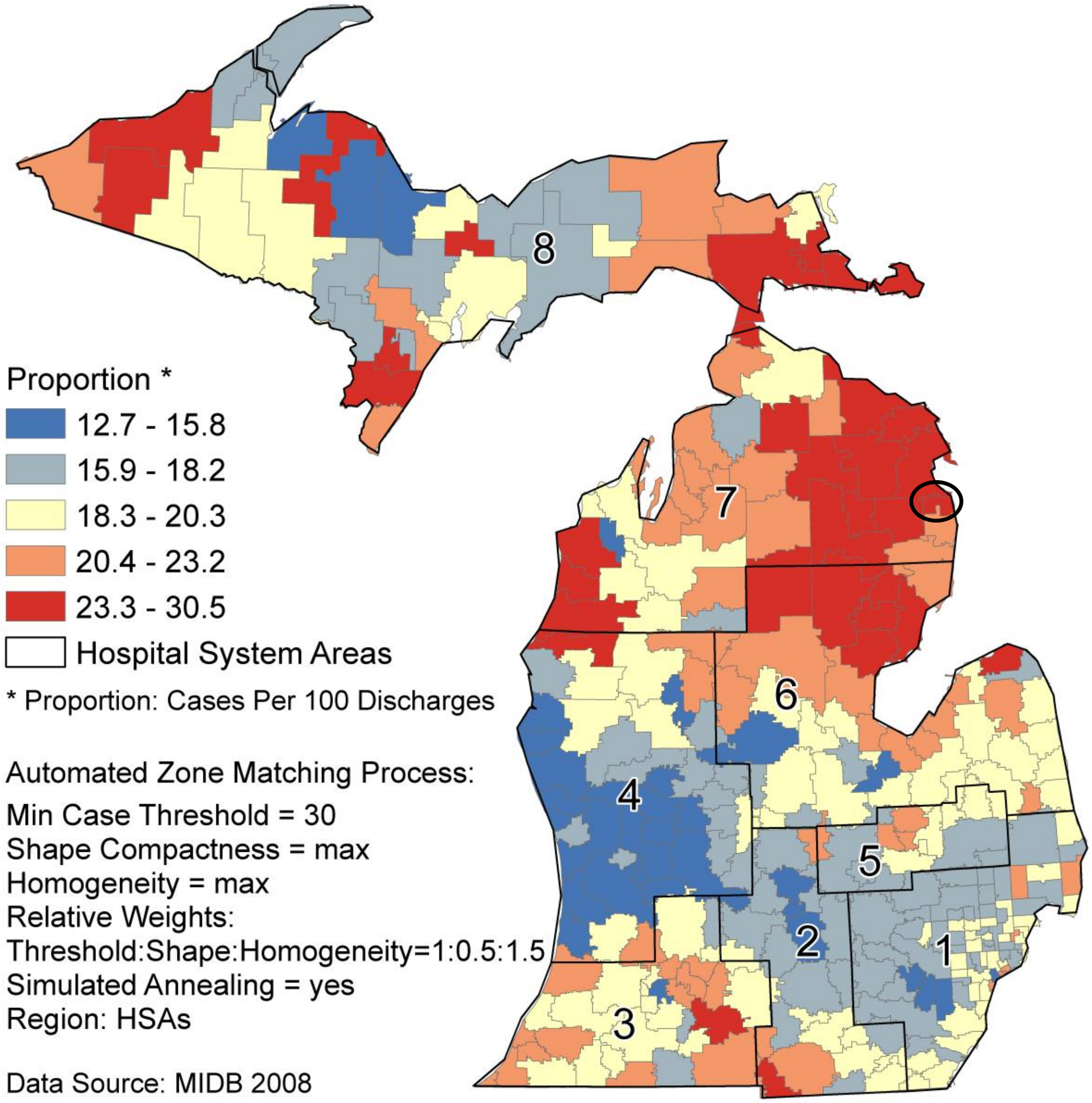

Figure 8. Circulatory diseases: proportions by MMA (n=310), Michigan 2008.
(statewide proportion: 18.3 per 100 discharges)

Table 8. MMAs with high[1] proportions of patients discharged with circulatory diseases, Michigan, 2008.

# MMA	Cases	total	crude proportion[2]	95% CI[3]	RSE[4]
781	67	220	30.5	23.6, 38.7	12.2
28	206	679	30.3	26.2, 34.5	7.0
466	98	327	30.0	24.3, 36.5	10.1
484	61	219	27.9	21.3, 35.8	12.8
533	96	354	27.1	22, 33.1	10.2
753	71	262	27.1	21.2, 34.2	11.9
763	166	613	27.1	23, 31.2	7.8
463	117	433	27.0	22.1, 31.9	9.2
500	644	2,473	26.0	24, 28.1	3.9
774	267	1,026	26.0	22.9, 29.1	6.1
386	188	728	25.8	22.1, 29.5	7.3
495	370	1,437	25.7	23.1, 28.4	5.2
559	410	1,598	25.7	23.2, 28.1	4.9
811	42	164	25.6	18.5, 34.6	15.4
713	1210	4,748	25.5	24, 26.9	2.9
343	424	1,671	25.4	23, 27.8	4.9
426	144	576	25.0	20.9, 29.1	8.3
694	58	233	24.9	18.9, 32.2	13.1
593	173	695	24.9	21.2, 28.6	7.6
262	64	258	24.8	19.1, 31.7	12.5
618	36	146	24.7	17.3, 34.1	16.7
354	416	1,692	24.6	22.2, 26.9	4.9
37	549	2,241	24.5	22.4, 26.5	4.3
166	906	3,728	24.3	22.7, 25.9	3.3
585	248	1,021	24.3	21.3, 27.3	6.4
652	171	704	24.3	20.6, 27.9	7.6
379	61	252	24.2	18.5, 31.1	12.8
165	100	416	24.0	19.3, 28.8	10.0
624	150	625	24.0	20.2, 27.8	8.2
512	387	1,616	23.9	21.6, 26.3	5.1
48	143	598	23.9	20, 27.8	8.4
660	238	999	23.8	20.8, 26.9	6.5
802	101	425	23.8	19.1, 28.4	10.0

[1] High crude proportions are marked as red areas in the maps.

[2] Proportion: cases per 100 discharges

[3] 95% CI: 95% Confidence Interval

[4] RSE: Relative Standard Error

Figure 9 shows the MMAs of crude proportions of patients discharged with diabetes. Different from previous maps of MMAs for IHD and circulatory diseases, the MMAs with high and low proportions of diabetes were more dispersed. Those large cities, such as Detroit, Lansing, Grand Rapids Kalamazoo, Saginaw, Muskegon and Flint, tended to have high proportions of patients discharged with diabetes, which was opposite to the pattern of IHD. The highest crude proportion was 3.4 per 100 discharges (95% CI 2.9, 3.9) in Detroit (Table 9). Suburban and rural areas were more likely to have lower proportions of patient's discharged with diabetes that urban areas.

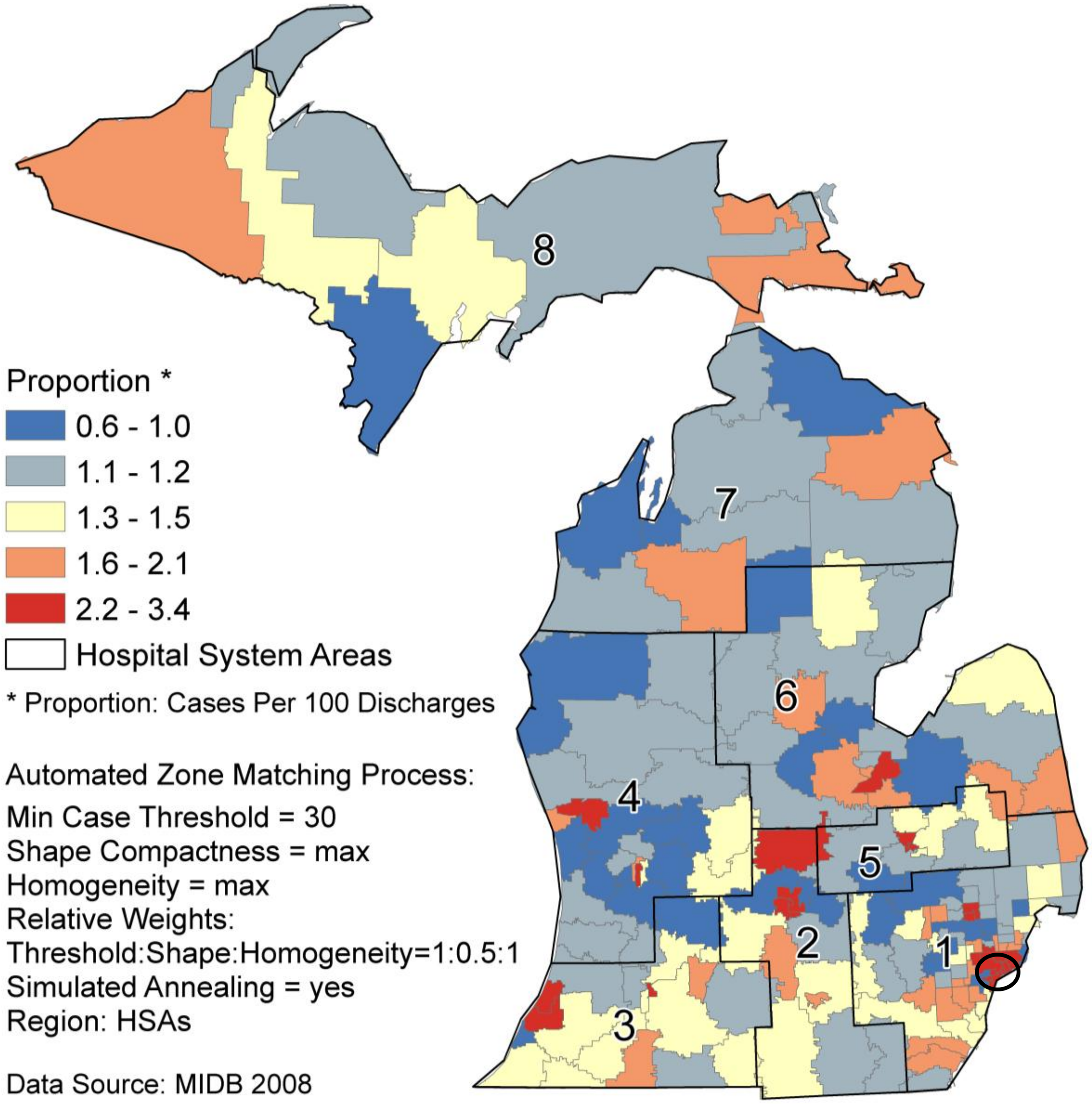

Figure 9. Diabetes: proportions by MMA (n=183), Michigan 2008.

(statewide proportion: 1.5 per 100 discharges)

Table 9. MMAs with high[1] crude proportions of patients discharged with diabetes, Michigan, 2008.

# MMA	cases	total	crude proportion[2]	95% CI[3]	RSE[4]
343	194	5,712	3.4	2.9, 3.9	7.2
250	421	14,410	2.9	2.6, 3.2	4.9
323	175	6,379	2.7	2.3, 3.1	7.6
199	968	36,453	2.7	2.5, 2.8	3.2
344	214	8,078	2.6	2.3, 3	6.8
40	208	7,870	2.6	2.3, 3	6.9
182	669	25,883	2.6	2.4, 2.8	3.9
154	103	3,988	2.6	2.1, 3.1	9.9
10	280	10,965	2.6	2.3, 2.9	6.0
241	277	10,899	2.5	2.2, 2.8	6.0
2	112	4,448	2.5	2.1, 3	9.4
228	36	1,442	2.5	1.7, 3.5	16.7
298	163	6,779	2.4	2, 2.8	7.8
297	40	1,668	2.4	1.7, 3.3	15.8
252	114	4,812	2.4	1.9, 2.8	9.4
169	75	3,224	2.3	1.8, 2.9	11.5
331	128	5,568	2.3	1.9, 2.7	8.8
167	246	10,932	2.3	2, 2.5	6.4
108	123	5,470	2.2	1.9, 2.6	9.0
235	111	4,957	2.2	1.8, 2.7	9.5
43	212	9,528	2.2	1.9, 2.5	6.9
311	246	11,060	2.2	1.9, 2.5	6.4
332	103	4,718	2.2	1.8, 2.6	9.9
270	127	5,835	2.2	1.8, 2.6	8.9
306	154	7,228	2.1	1.8, 2.5	8.1
341	118	5,581	2.1	1.7, 2.5	9.2
334	124	6,033	2.1	1.7, 2.4	9.0

[1] High crude proportions are marked as red areas in the maps.
[2] Proportion: cases per 100 discharges
[3] 95% CI: 95% Confidence Interval
[4] RSE: Relative Standard Error

Figure 10 is the map of crude proportions of patients discharged with endocrine disorders. The spatial pattern is generally similar to diabetes, which seldom has large clusters. Urban areas like Detroit, Lansing, Muskegon, Kalamazoo and Midland, as well as some other areas show MMAs with elevated proportions of hospital discharges with endocrine disorders in the Upper Peninsula and Lower Michigan. The highest crude proportion of discharges with endocrine disorders was 6.16 per 100 discharges (95% CI 4.27, 8.61) in Munith (Table 10).

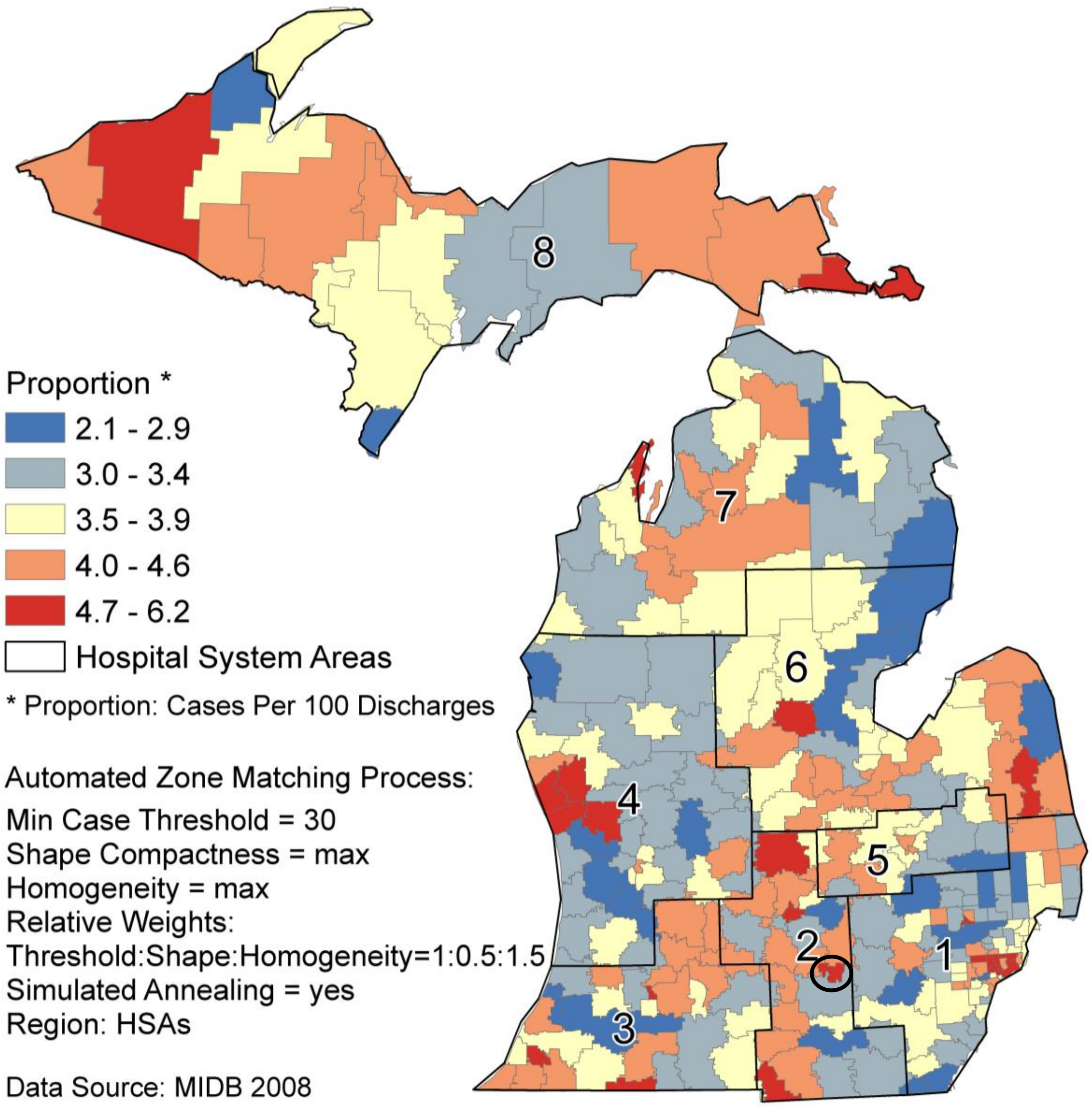

Figure 10. Endocrine disorders: proportions by MMA (n=274), Michigan 2008. (statewide proportion: 3.8 per 100 discharges)

Table 10. MMAs with high[1] crude proportions of patients discharged with endocrine disorders, Michigan, 2008.

# MMA	Cases	total	crude proportion[2]	95% CI[3]	RSE[4]
29	34	552	6.2	4.3, 8.6	17.1
365	30	519	5.8	3.9, 8.3	18.3
107	67	1,186	5.6	4.4, 7.2	12.2
453	1,319	24,355	5.4	5.1, 5.7	2.8
499	90	1,668	5.4	4.3, 6.6	10.5
369	570	10,578	5.4	4.9, 5.8	4.2
8	456	8,633	5.3	4.8, 5.8	4.7
186	56	1,068	5.2	4, 6.8	13.4
508	627	12,195	5.1	4.7, 5.5	4.0
391	620	12,086	5.1	4.7, 5.5	4.0
518	786	15,545	5.1	4.7, 5.4	3.6
539	44	878	5.0	3.6, 6.7	15.1
23	35	704	5.0	3.5, 6.9	16.9
143	114	2,300	5.0	4, 5.9	9.4
424	218	4,448	4.9	4.3, 5.6	6.8
416	166	3,431	4.8	4.1, 5.6	7.8
417	481	9,986	4.8	4.4, 5.2	4.6
316	653	13,565	4.8	4.4, 5.2	3.9
410	130	2,705	4.8	4, 5.6	8.8
129	50	1,052	4.8	3.5, 6.3	14.1
510	41	868	4.7	3.4, 6.4	15.6
229	31	659	4.7	3.2, 6.7	18.0
310	60	1,283	4.7	3.6, 6	12.9
80	69	1,480	4.7	3.6, 5.9	12.0
169	373	8,078	4.6	4.1, 5.1	5.2

[1] High crude proportions are marked as red areas in the maps.
[2] Proportion: cases per 100 discharges
[3] 95% CI: 95% Confidence Interval
[4] RSE: Relative Standard Error

4.4 Medical Management Areas (Crude and Age-Adjusted Rates)

Figure 11 shows the MMAs of crude rates of patients discharged with IHD in the population using the optimal zone design. The spatial patterns showed that high rates of IHD in the population did not scatter evenly throughout the state. The northeast and southern portions of Lower Michigan (HSA=2,6,7) showed elevated crude rates of discharges with IHD. The highest crude rate was 24.2 per 1,000 population (95% CI 19.1, 30.3) and the age-adjusted rate was 17.2 in Au Gres (Table 11).

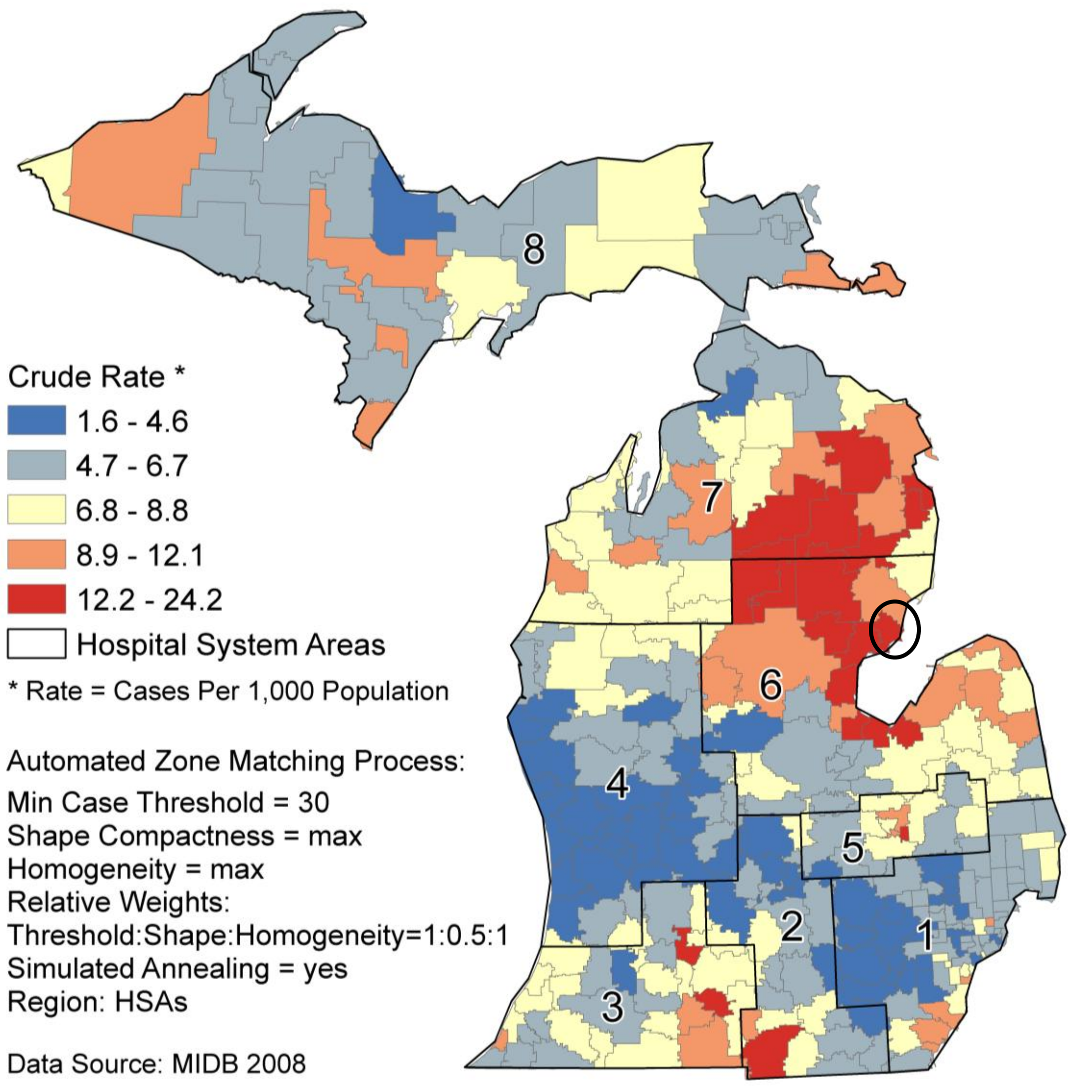

Figure 11. Ischemic heart disease: crude rates by MMA (n=302), Michigan 2008. (statewide rate: 5.8 per 1,000 population)

Table 11. MMAs with high[1] crude rates of patients discharged with ischemic heart disease, Michigan, 2008.

# MMA	cases	population	crude rate[2]	95% CI[3]	RSE[4]	age-adjusted rate[2]
316	77	3,180	24.2	19.1, 30.3	11.4	17.2
170	167	9,888	16.9	14.3, 19.5	7.7	10.8
151	36	2,307	15.6	10.9, 21.6	16.7	8.5
573	289	19,283	15.0	13.3, 16.7	5.9	10.1
468	112	7,535	14.9	12.1, 17.6	9.4	10.9
412	136	9,331	14.6	12.1, 17	8.6	13.5
535	268	18,437	14.5	12.8, 16.3	6.1	11.8
55	38	2,685	14.2	10, 19.4	16.2	8.6
557	39	2,844	13.7	9.8, 18.7	16.0	11.1
336	469	34,425	13.6	12.4, 14.9	4.6	12.9
36	87	6,411	13.6	10.9, 16.7	10.7	10.7
302	83	6,124	13.6	10.8, 16.8	11.0	10.5
182	502	37,945	13.2	12.1, 14.4	4.5	12.3
354	180	13,811	13.0	11.1, 14.9	7.5	10.3
393	159	12,252	13.0	11, 15	7.9	11.7
627	57	4,560	12.5	9.5, 16.2	13.2	11.4
150	39	3,130	12.5	8.9, 17	16.0	10.9
52	297	23,928	12.4	11, 13.8	5.8	12.2
202	177	14,323	12.4	10.5, 14.2	7.5	11.6
585	516	41,988	12.3	11.2, 13.3	4.4	10.3
70	50	4,144	12.1	9, 15.9	14.1	7.8

[1] High crude proportions are marked as red areas in the maps.
[2] Rate: cases per 1,000 population
[3] 95% CI: 95% Confidence Interval
[4] RSE: Relative Standard Error

Figure 12 shows the MMAs of crude rates of patients discharged with circulatory diseases using the optimal zone design. The spatial pattern still showed MMAs with high proportions of patient's discharged with IHD throughout the state. A large cluster of high proportion of discharges with circulatory diseases was located in northeast of Lower Michigan (HSA=6,7) and two large clusters of low proportion were located in southwest and southeast of Lower Michigan.

The highest crude rate was 54.7 per 1,000 population (95% CI 44.4, 66.7) and its age-adjusted rate was 34.15 in Comins and Fairview (Table 12).

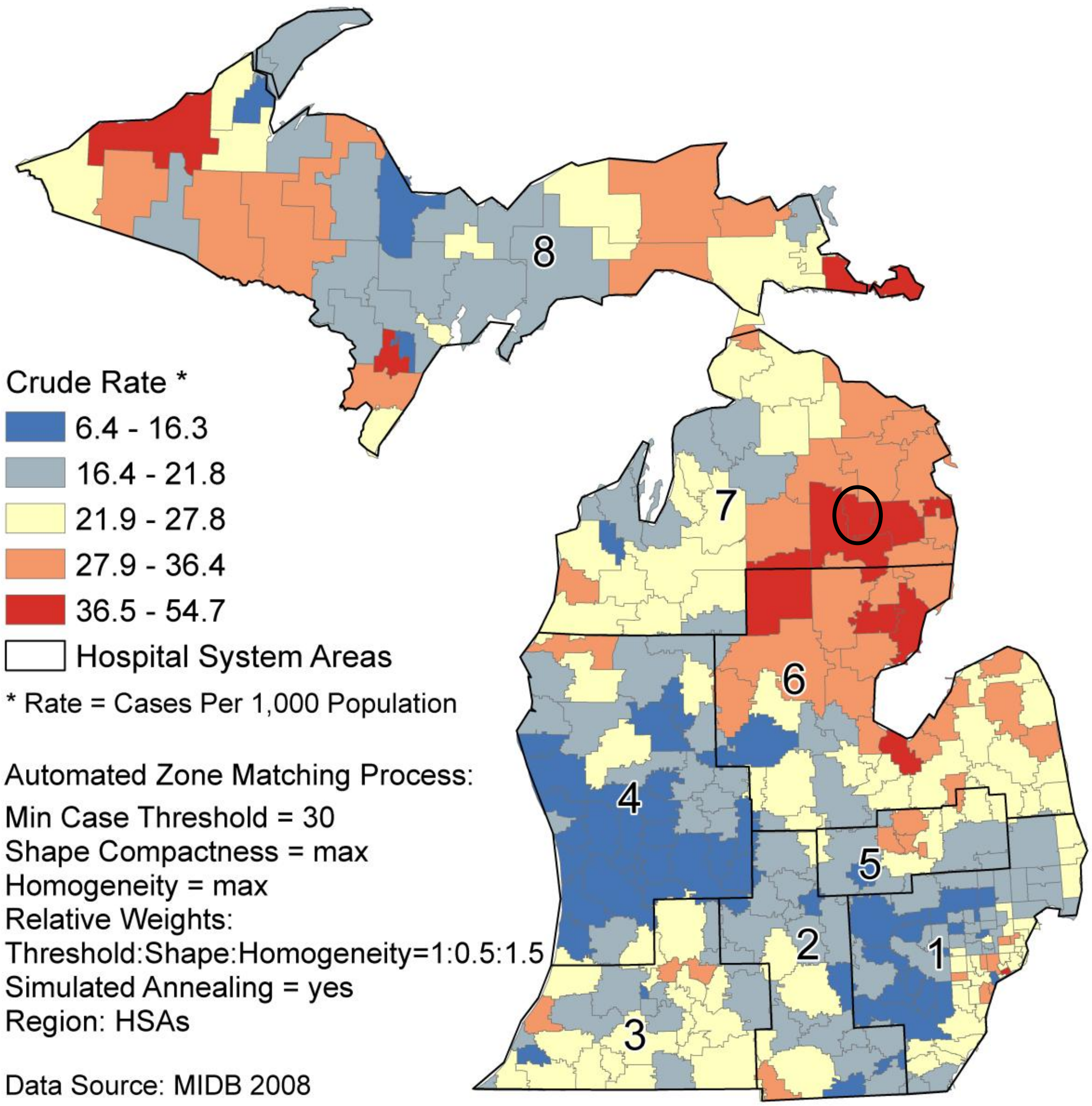

Figure 12. Circulatory diseases: crude rates by MMA (n=310), Michigan 2008.
(statewide rate: 21.2 per 1,000 population)

Table 12. MMAs with high[1] crude rates of patients discharged with circulatory diseases, Michigan, 2008.

# MMA	cases	population	crude rate[2]	95% CI[3]	RSE[4]	age-adjusted rate[2]
466	98	1,790	54.7	44.4, 66.7	10.1	34.1
774	267	5,886	45.4	39.9, 50.8	6.1	34.8
152	1,558	34,638	45.0	42.7, 47.2	2.5	43.9
343	424	9,777	43.4	39.2, 47.5	4.9	29.7
618	36	848	42.5	29.7, 58.8	16.7	23.2
713	1,210	29,171	41.5	39.1, 43.8	2.9	26.5
753	71	1,747	40.6	31.7, 51.3	11.9	30.4
781	67	1,663	40.3	31.2, 51.2	12.2	24.7
533	96	2,383	40.3	32.6, 49.2	10.2	20.2
593	173	4,385	39.5	33.6, 45.3	7.6	25.1
559	410	10,511	39.0	35.2, 42.8	4.9	25.5
671	1,587	41,622	38.1	36.3, 40	2.5	34.6
694	58	1,593	36.4	27.6, 47.1	13.1	24.2

[1] High crude proportions are marked as red areas in the maps.
[2] Rate: cases per 1,000 population
[3] 95% CI: 95% Confidence Interval
[4] RSE: Relative Standard Error

Figure 13 shows the age-adjusted rates of MMA-IHD. IHD age-adjusted rates were still generally high in the northwest portion of the state and other rural areas. The highest age-adjusted rate was 17.2 per 1,000 population in Flint (MMA 316). MMAs that had high age-adjusted rates also included Kingsley and Buckley (MMA 452: 10.7 per 1,000 population) and Monroe (MMA 388: 10.1 per 1,000 population). In contrast, some other areas seem not to have very high rates in this map, such as Glennie, compared to the map of crude rates.

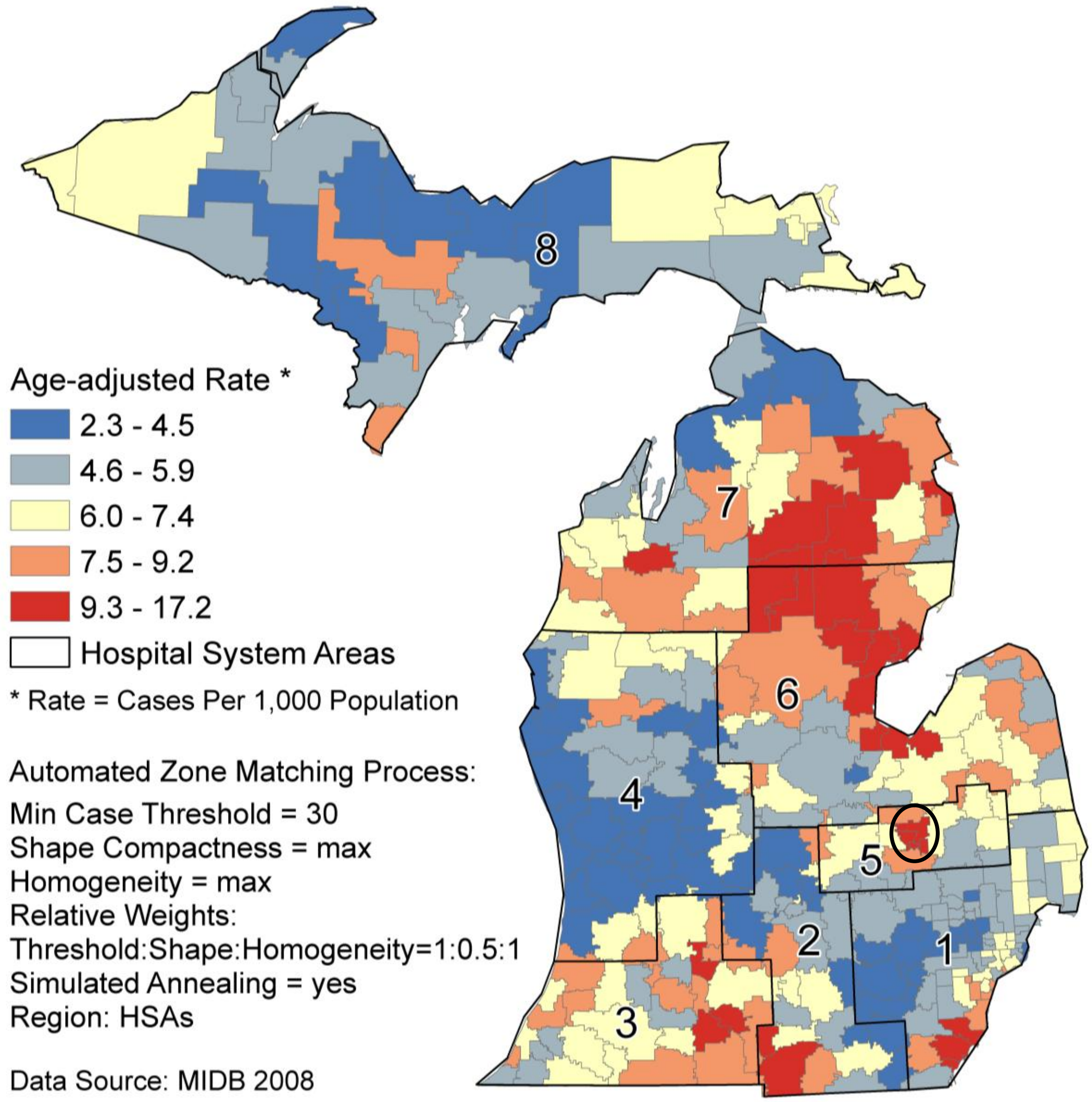

Figure 13. Ischemic heart disease: age-adjusted rates by MMA (n=302), Michigan 2008.

Figure 14 shows the age-adjusted rates of patient's discharged with circulatory diseases. The spatial pattern of discharges with circulatory diseases was somehow similar to that of IHD. Compared to the map of crude rates, areas with high and low age-adjusted rates were even more scattered. Areas with high age-adjusted rates generally scattered in the Lower Michigan. The highest age-adjusted rate was 43.9 per 1,000 population in Detroit (MMA 152). Some other urban areas, such as Battle Creek (MMA 337: 33.5 per 1,000 population) and Flint (MMA 654: 39.5 per 1,000 population), showed elevated age-adjusted rates as well.

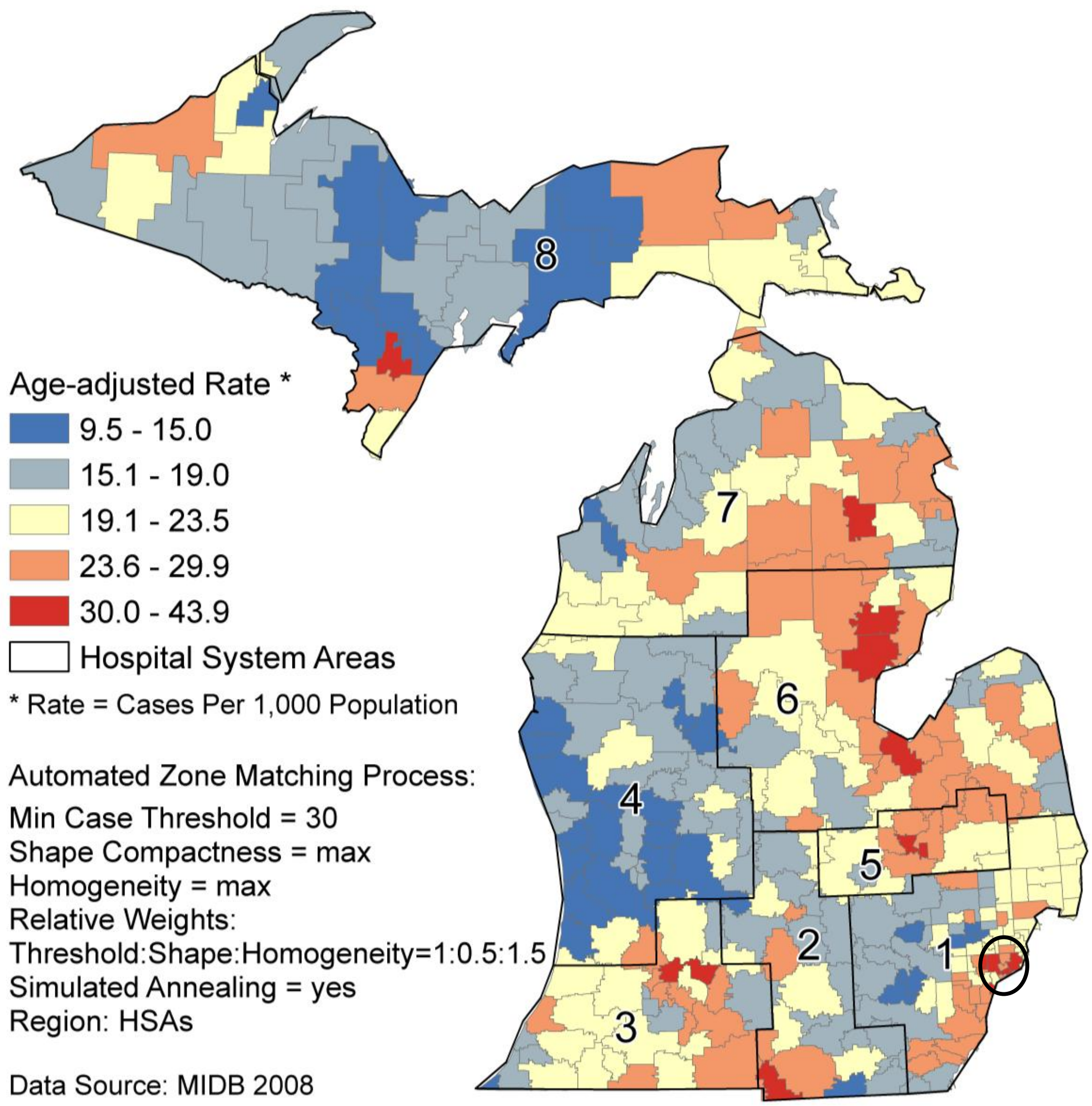

Figure 14. Circulatory diseases: age-adjusted rates by MMA (n=310), Michigan 2008.

Figure 15 shows the crude rates of discharges with diabetes using the optimal zone design. The spatial pattern of diabetes was more evenly distributed throughout the state than IHD. The highest crude rate was 6.31 per 1,000 population (95% CI 5.42, 7.2) and the age-adjusted rate was 6.46 in Detroit (Table 13). There are two more high areas in Detroit too. The other obviously elevated crude rate (in the middle of the map) was 4.62 per 1,000 (95% CI 4, 5.24) and its age-adjusted rate was 5.12. Most of the state seems to have medium to low rates.

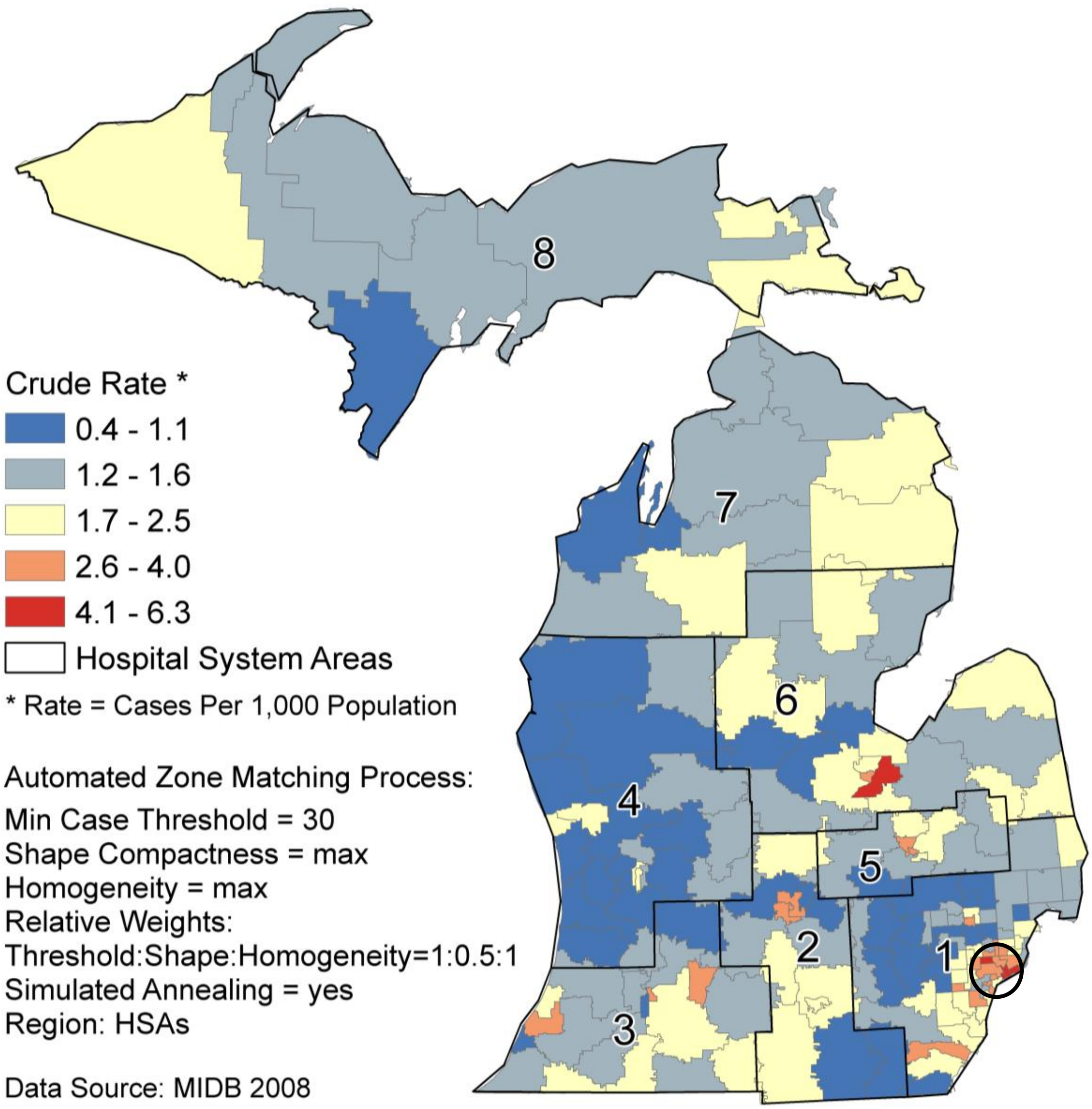

Figure 15. Diabetes: crude rates by MMA (n=183), Michigan 2008.
(statewide rate: 1.7 per 1,000 population)

Table 13. MMAs with high[1] crude rates of patients discharged with diabetes, Michigan, 2008.

# MMA	cases	population	crude rate[2]	95% CI[3]	RSE[4]	age-adjusted rate[2]
343	194	30,733	6.3	5.4, 7.2	7.2	6.5
154	103	20,273	5.1	4.1, 6.1	9.9	5.0
250	421	86,598	4.9	4.4, 5.3	4.9	4.9
344	214	46,337	4.6	4, 5.2	6.8	5.1
40	208	46,155	4.5	3.9, 5.1	6.9	4.7
10	280	70,276	4.0	3.5, 4.5	6.0	4.5

[1] High crude proportions are marked as red areas in the maps.
[2] Rate: cases per 1,000 population
[3] 95% CI: 95% Confidence Interval
[4] RSE: Relative Standard Error

Figure 16 shows the crude rates of discharges with endocrine disorders. The spatial pattern of endocrine conditions were scattered throughout the state. The highest crude rate was 9.54 per 1,000 population (95% CI 9.02, 10.05) and the age-adjusted rate was 10.14 still in Detroit (Table 13). There are a couple more high areas all over the state as well, including Saginaw, Lansing, Flint and west and east corners of the Upper Michigan. The suburban and rural areas are more likely to have low crude rates.

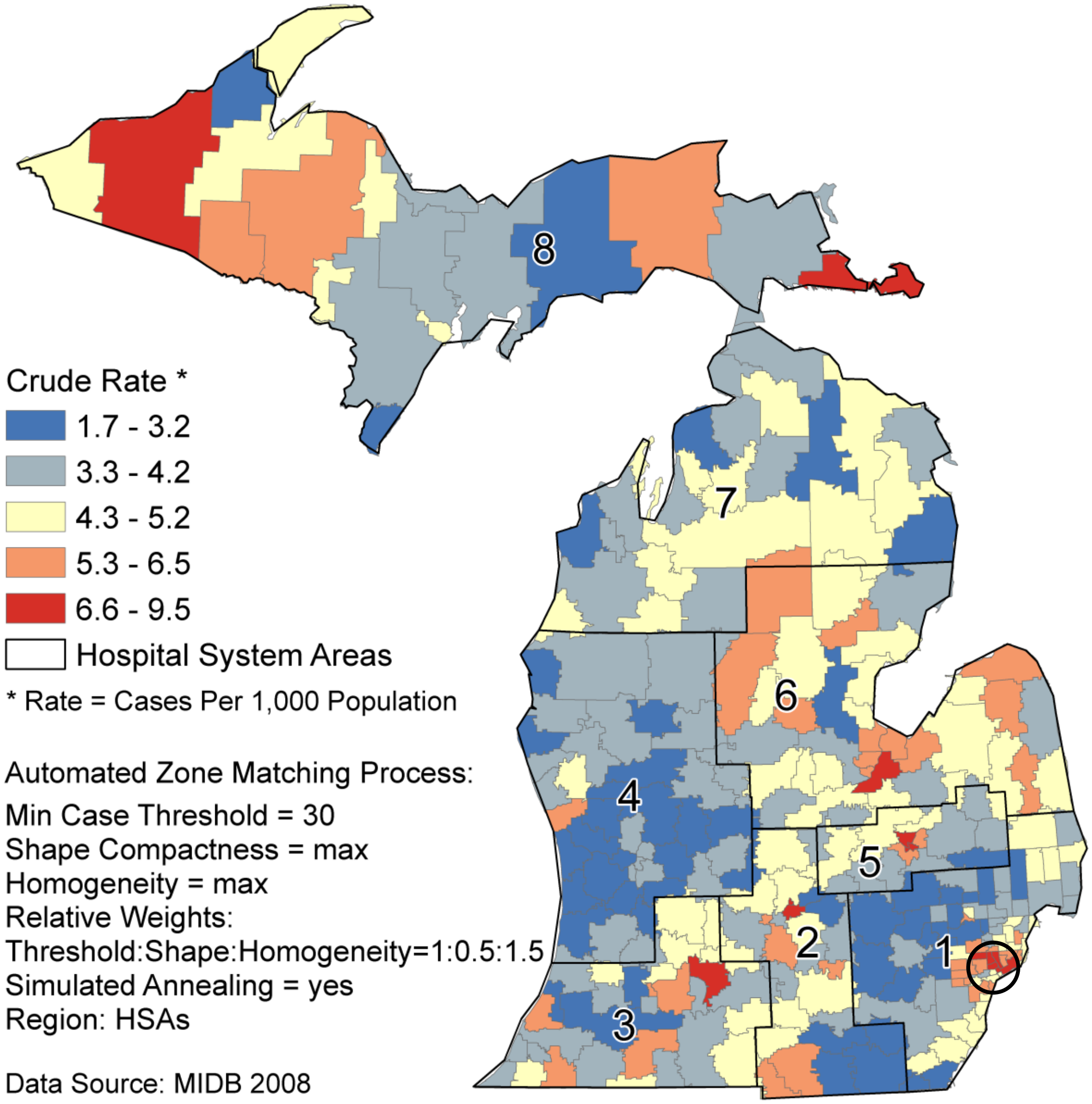

Figure 16. Endocrine disorders: crude rates by MMA (n=274), Michigan 2008.
(statewide rate: 4.4 per 1,000 population)

Table 14. MMAs with high[1] crude rates of patients discharged with endocrine disorders, Michigan, 2008.

# MMA	Cases	population	crude rate[2]	95% CI[3]	RSE[4]	age-adjusted rate[2]
453	1,319	138,325	9.5	9, 10.1	2.8	10.1
499	90	10,868	8.3	6.7, 10.2	10.5	9.1
169	373	46,337	8.0	7.2, 8.9	5.2	8.9
369	570	72,281	7.9	7.2, 8.5	4.2	8.2
518	786	99,998	7.9	7.3, 8.4	3.6	8.5
508	627	80,461	7.8	7.2, 8.4	4.0	8.1
365	30	3,908	7.7	5.2, 11	18.3	6.0
8	456	60,885	7.5	6.8, 8.2	4.7	8.2
107	67	8,966	7.5	5.8, 9.5	12.2	5.8
417	481	64,926	7.4	6.7, 8.1	4.6	8.1
487	215	29,113	7.4	6.4, 8.4	6.8	7.7
535	390	55,949	7.0	6.3, 7.7	5.1	6.9
30	487	70,276	6.9	6.3, 7.5	4.5	8.0
23	35	5,367	6.5	4.5, 9.1	16.9	6.6

[1] High crude proportions are marked as red areas in the maps.
[2] Rate: cases per 1,000 population
[3] 95% CI: 95% Confidence Interval
[4] RSE: Relative Standard Error

The map of age-adjusted rates for diabetes is shown in Figure 17. Besides Saginaw, which also had an elevated crude rate, Flint, Pontiac, Lansing, Kalamazoo and Detroit all had high age-adjusted rates. The highest age-adjusted rate was 6.5 per 1,000 population in Detroit (MMA 343). Direct age-standardization generally makes the rates higher than crude ones.

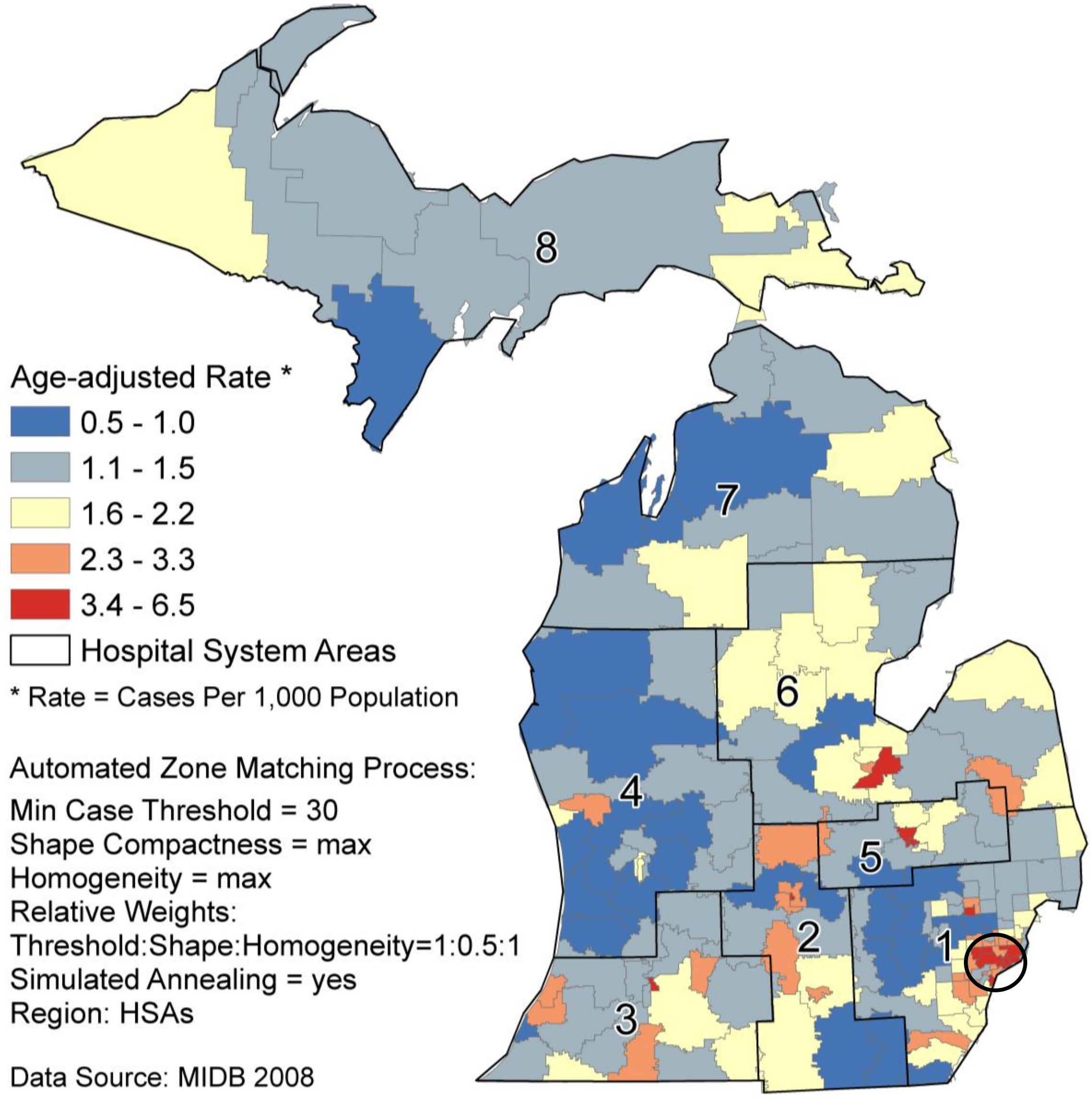

Figure 17. Diabetes: age-adjusted rates by MMA (n=183), Michigan 2008.

The map of age-adjusted rates for endocrine disorders is shown in Figure 18. This map is very similar to the map of crude rates, except rates become higher after age-standardization. The highest age-adjusted rate was 10.1 per 1,000 population still in Detroit (MMA 453). Those areas that popped up in the map of crude rates seem not to be apparent in this map, such as Battle Creek and two other areas in the Upper Michigan. Rural and suburban areas still tend to have lower age-adjusted rates than urban areas.

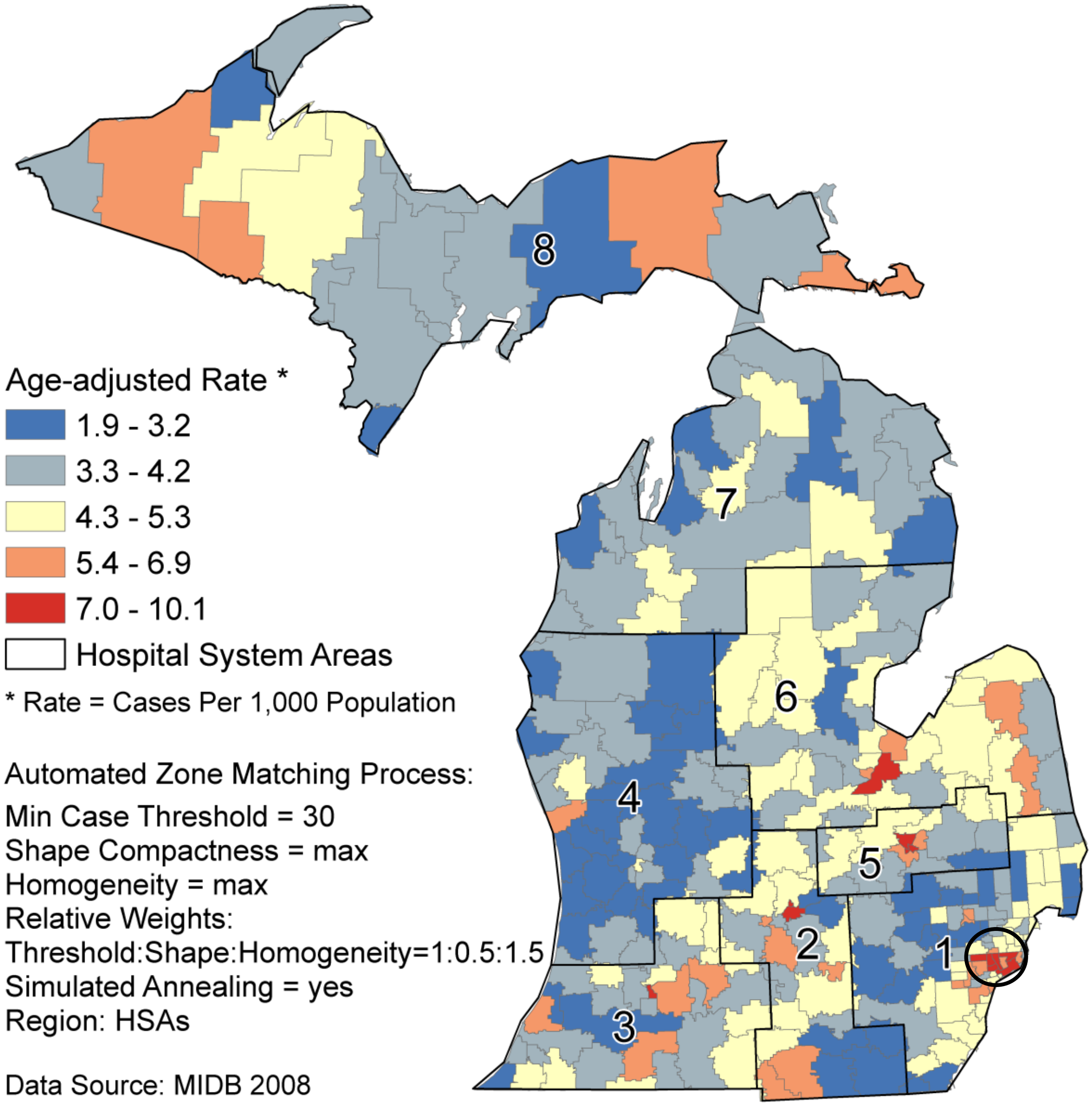

Figure 18. Endocrine disorders: age-adjusted rates by MMA (n=274), Michigan 2008.

4.5 Alternative Zone Designs

Following the implementation of the AZM constraint parameters listed in the Methods Section it was observed that zones constrained within the HSAs may also be prone to the MAUP thus, reducing the quality of the overall zone design via simulated annealing by achieving regional optima. To examine the differences in these possible effects, three different ways to use regions were implemented one by one. The first model conformed to the previous zone design, i.e., using all the HSAs as regions and respect the region boundary. The second model disenabled the region option, i.e., didn't respect any region boundary. The third model ran the program within each HSA, i.e. took HSAs one by one as a region, and merged the eight HSA outputs. All the other constraint parameters (minimum case threshold, maximum shape compactness and maximum internal homogeneity), weighting scheme and simulated annealing option, remained the same as in previous zone designs. To target the specific diseases that are representative, IHD and diabetes were used in this part. Due to limited time, the AZTool program only ran each method once (i.e., one restart) and compare the outputs of the three zone designs. The zone designs and statistical outputs are shown in Table 15.

The second concern while constructing the optimal MMAs is whether simulated annealing improved the overall zone design. Since the zone design improvements by using simulated annealing is still controversial (Ralphs & Ang, 2009), this study attempted to further understand the impact of simulated annealing with and without regional constraints. The AZTool program was operated again without enabling simulated annealing to supplement earlier results. The zone designs and statistical outputs are shown in Table 16.

Table 15. Statistical outputs of zone designs using alternative methods with simulated annealing.

Region	IHD			Diabetes		
	Zones No.	IAC score	P2A score	Zones No.	IAC score	P2A score
Statewide with HSA	307	0.102	12706.400	192	0.114	8199.470
Statewide without HSA	310	0.101	12156.900	191	0.114	8002.900
With each HSA	296	0.105	12221.575	184	0.117	7568.432
HSA 1	79	0.114	2387.320	74	0.121	2311.410
HSA 2	28	0.033	1286.960	15	0.083	669.343
HSA 3	34	0.033	1545.980	20	0.055	940.288
HSA 4	45	0.045	1952.860	23	0.063	1038.820
HSA 5	14	0.127	696.123	10	0.096	456.856
HSA 6	41	0.083	1597.520	22	0.113	922.021
HSA 7	33	0.015	1645.530	12	0.047	767.508
HSA 8	22	0.044	1109.280	8	0.031	462.185

Table 16. Statistical outputs of zone designs using alternative methods without simulated annealing.

Region	IHD			Diabetes		
	Zones No.	IAC score	P2A score	Zones No.	IAC score	P2A score
Statewide with HSA	304	0.105	12153.600	191	0.117	7892.021
Statewide without HSA	300	0.105	11692.740	190	0.117	7684.949
With Each HSA	294	0.105	12122.198	187	0.120	7907.426
HSA 1	77	0.114	2319.089	77	0.127	2457.142
HSA 2	27	0.032	1175.123	14	0.085	726.182
HSA 3	31	0.032	1288.262	19	0.054	893.580
HSA 4	45	0.048	2035.757	26	0.068	1255.703
HSA 5	12	0.121	554.482	9	0.095	432.163
HSA 6	43	0.085	1769.000	22	0.118	934.050
HSA 7	36	0.015	1859.815	12	0.043	712.513
HSA 8	23	0.041	1120.669	8	0.028	496.092

The first model conformed to the previous zone design, i.e., using all the HSAs as regions and respect the region boundary. The second model disenabled the region option, i.e., didn't respect any region boundary. The third model ran the program within each HSA, i.e. took HSAs one by one as a region, and merged the eight HSA outputs.

The results from these analyses showed that the differences between the first and second models (statewide with/without HSAs) in terms of the number of zones were not substantially different (n=307 versus 310); however, the numbers of zones using each HSA as the region constraint (the sum of number of zones in each HSA – third model) was slightly less (n=296). The IAC was similar across all three models demonstrating that the zones had similar levels of demographic homogeneity regardless of the different number of zones and whether or not the zones where constrained within HSAs. These results indicate that the number and size of zones across the three alternative methods are generally similar.

Since the P2A and IAC scores vary greatly by HSA, only the overall zone design which merged the eight HSAs is compared to the other two statewide methods. The overall internal homogeneity generated by merging HSAs appear to be better than the other two methods using statewide geographic units for both IHD and diabetes. Again, the differences of internal homogeneity between the two methods using statewide geographic units with or without HSA are not substantial. In contrast to internal homogeneity, it's too early to conclude anything from the shape compactness scores. The unclear results of shape compactness possibly result from their lighter relative weights compared to case threshold and maximum homogeneity. The shape constraint parameter might be sacrificed to obtain better results of the other two. Two statewide maps showing the crude rates and selected RSE values with and without using HSA (SA enabled) are shown in Figure 19 and Figure 20.

Comparing the output results of enabling and disabling simulated annealing, it is interesting to observe that the three zone design models using simulated annealing do not appear to perform better than the three zone design models without simulated annealing as evidenced by the similar number of zones and slightly lower IAC and higher P2A scores. Thus simulated annealing does not seem to be an important parameter in the construction of new zones. This finding was similar to that reported in research conducted in New Zealand (Ralphs & Ang, 2009). Future research should continue to explore the use of simulated annealing in AZM research.

Another approach to examining how constraining new zones impacts the overall study results is to assess the RSE of rates on each side of regional boundaries. Statewide maps showing the crude rates and selected RSE values for the three models are shown in Figure 21 and Figure 22. Note that these maps use the optimal zone design from the first random start, rather than the ‘best’ from 50 random restarts. Future research should also use multiple restarts to prioritize the zone design and comparing the effects of regional constraints and simulated annealing based on the zone construction.

The assessment of RSE’s on each side of HSA boundaries show that all rates are stable (i.e., RSEs < 20%) and do not demonstrate substantial differences, suggesting that regional constraints do not impact the local rate results.

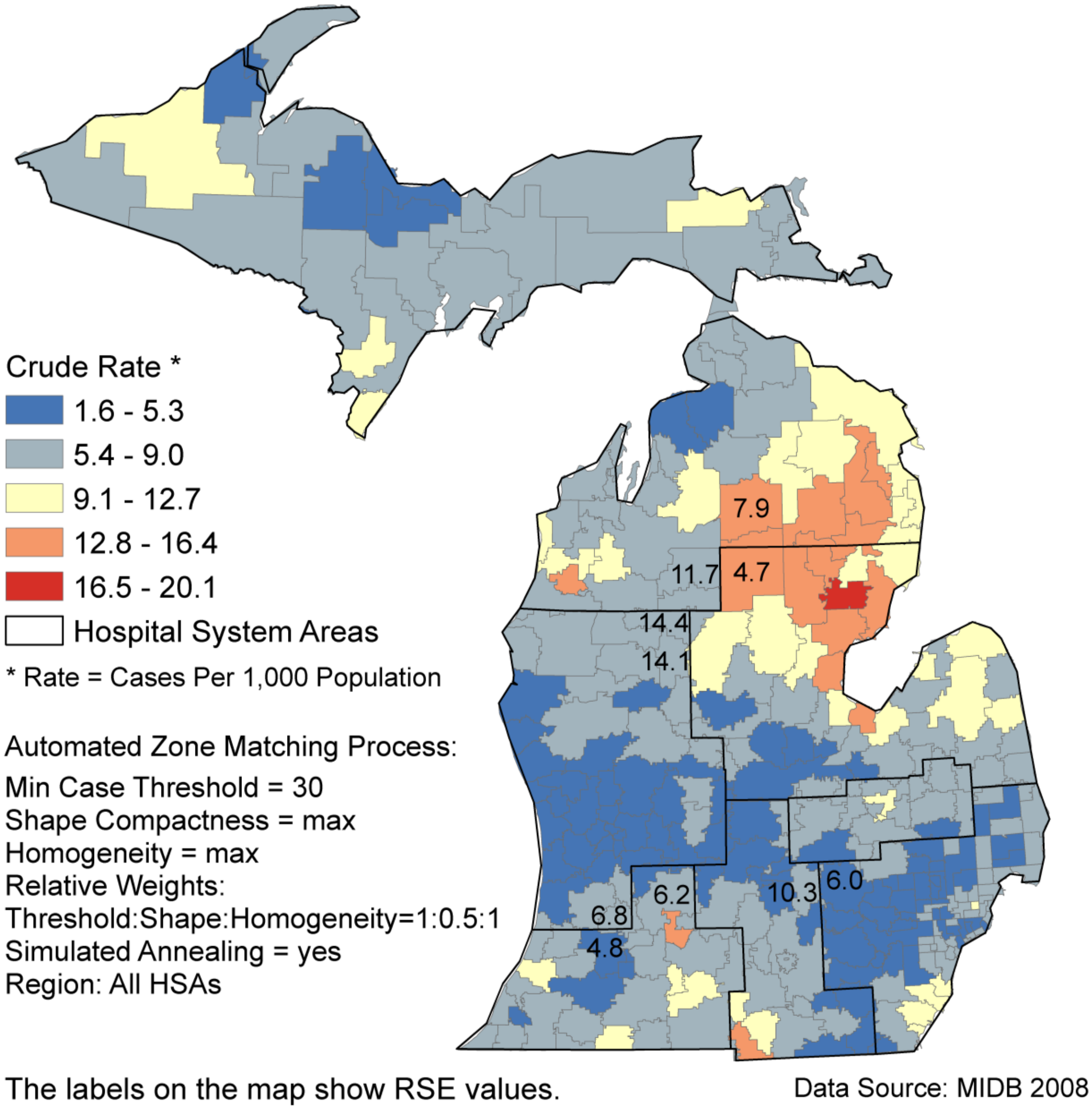

Figure 19. Ischemic heart disease: crude rates by MMA (n=307) and their RSE using all HSAs (SA enabled), Michigan 2008.

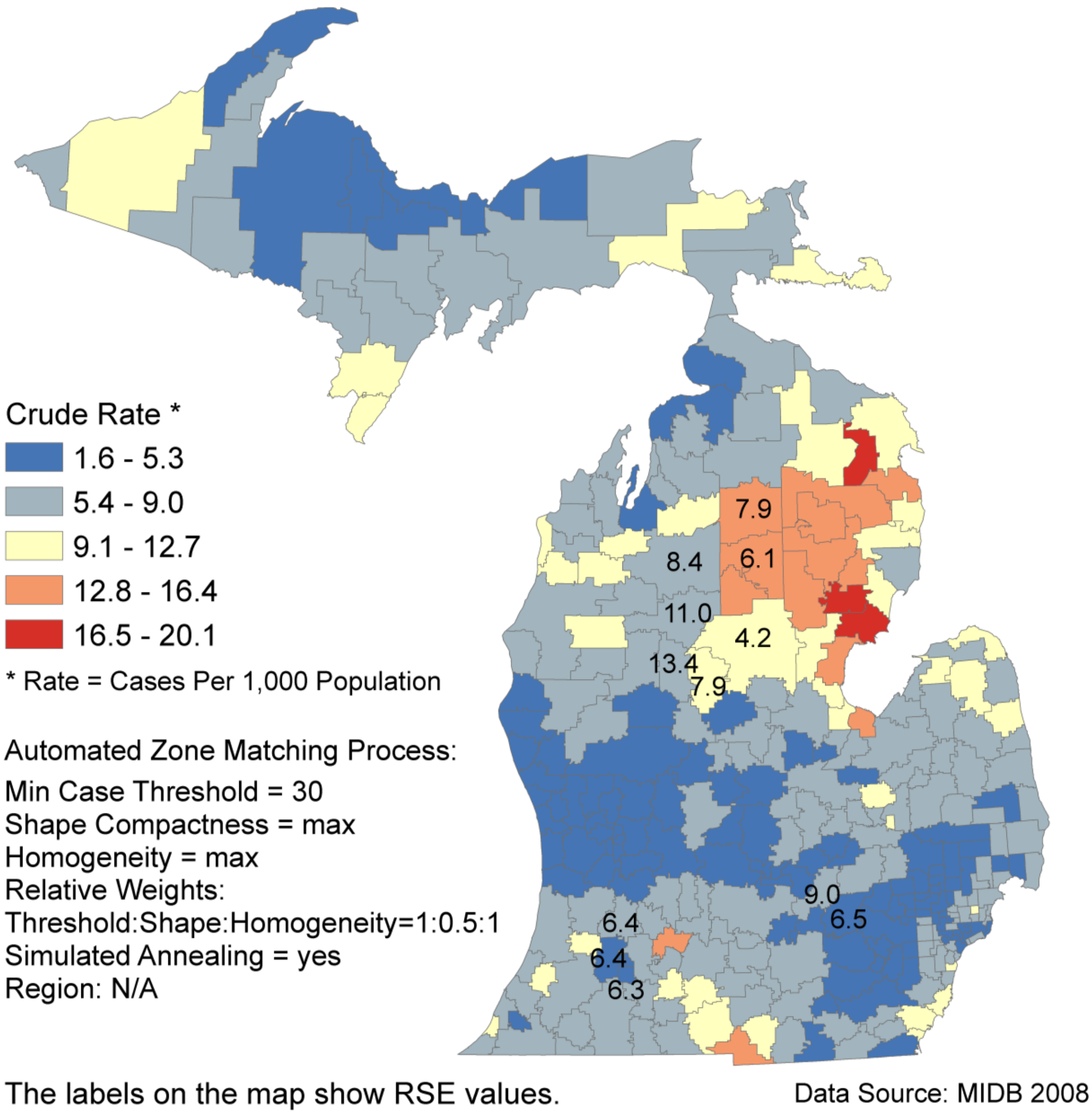

Figure 20. Ischemic heart disease: crude rates by MMA (n=310) without using HSAs (SA enabled), Michigan 2008.

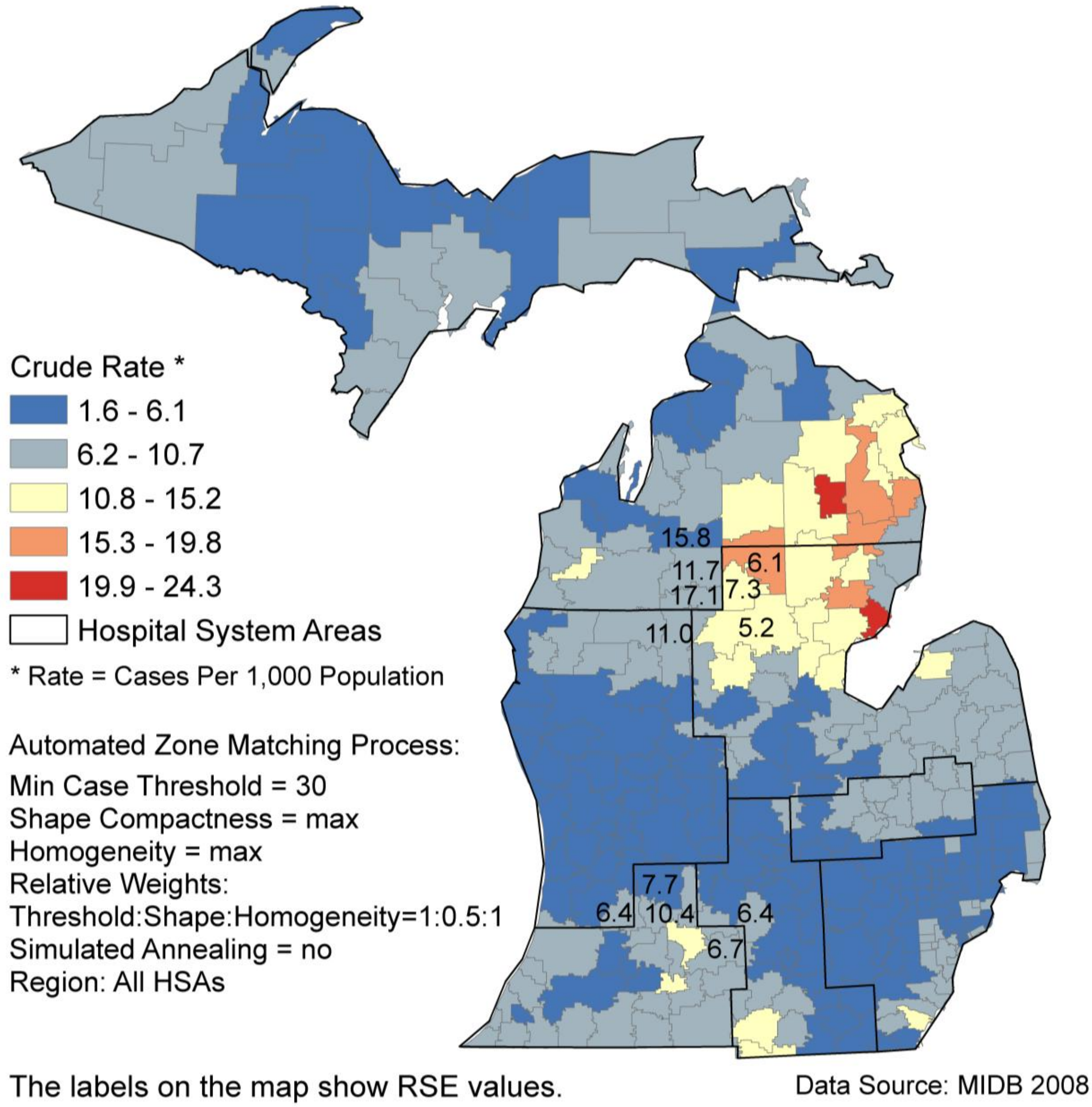

Figure 21. Ischemic heart disease: crude rates by MMA (n=304) using all HSAs (SA disabled), Michigan 2008.

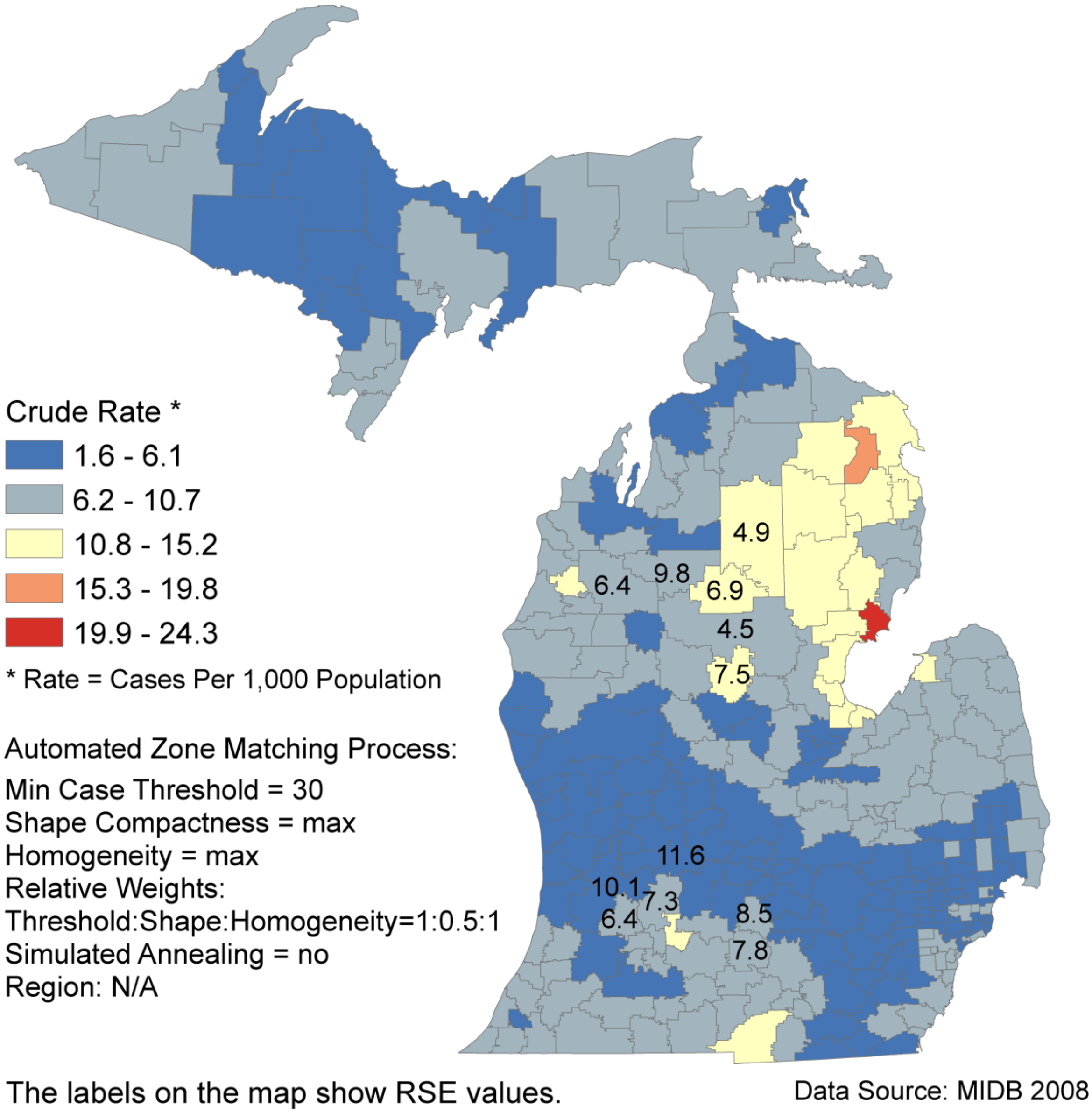

Figure 22. Ischemic heart disease: crude rates by MMA (n=300) without using HSAs (SA disabled), Michigan 2008.

5. DISCUSSION

The goals of this study were to (1) design optimal Medical Management Areas (MMAs) for IHD and diabetes to (2) visualize and explore the spatial patterns of these disease subgroups to inform future bed need in Michigan. MMAs of these sub-groups were compared with MMAs of the principal diagnoses to explore geographic and epidemiological scale together.

The objectives of this study included and the results from the analyses are provided below.

<u>Objective 1</u>: To demonstrate the need for the use of AZM in the construction of MMAs.

Hypothesis 1a: Mapping disease rates by ZIP Codes will result in a large proportion of unstable rates.

Much health services research focuses on the 'supply' perspective. This thesis provided a detailed methodology by which demand could also be measured to inform inpatient bed need in Michigan. The methodology was justified after constructing rate maps of the diseases being studied using ZIP Codes and identifying large areas across the state with unstable rates defined by the RSE > 20%. By creating the RSE maps, this study finds hypothesis 1a to be true and further illustrates that it is inappropriate to map disease rates using ZIP Codes in Michigan. Therefore, AZM, as a program to provide new zones that ensure the stability of rates, is needed to construct optimal MMAs to show the demand of health services.

<u>Objective 2</u>: To optimize MMA boundary definitions.

Hypothesis 2a: The shape constraint parameter will be less important than the internal homogeneity for broad disease groups compared to specific diseases because of the greater

variation in population within the broad disease groups. Since when a couple of ZIP Codes are aggregated into a zone, the population is becoming larger which results in more homogeneity.

Table 17 is a summary of the AZM outputs for IHD, circulatory diseases, diabetes and endocrine disorders. Note that generally the best restart occurred between 40 and 50 restarts. These best restarts met the constraint parameter requirements of the minimum case threshold, maximizing the shape compactness and maximizing the internal homogeneity. This study advanced earlier research by exploring different weighting schemes for the constraint parameters and finally using the combination of minimum threshold (100%), shape compactness (50%), IAC (sex, age, race) for IHD and diabetes (100%) and circulatory diseases and endocrine disorders (150%). In other words, the IAC was modeled according to the epidemiological scale of analysis (i.e., twice as important as the shape constraint for IHD and diabetes and three times as important for circulatory disease and endocrine disorders). The IAC was considered more important than the shape constraint in order to ensure greater demographic-homogeneity within zones, which conceptually should be helpful when assessing bed need in areas. At a lower epidemiological scale, a lower IAC weight was needed because the patients discharged with IHD or diabetes were expected to be more similar than patients discharges with diagnoses within the primary diagnostic related groups, circulatory diseases and endocrine disorders. Interestingly, all sets of MMAs constructed had a similar number of zones in the their output demonstrating the success of the modeling because if the IAC had been weighted similar to shape compactness there would have been fewer and more heterogeneous zones which would have been less informative to the inpatient bed need in Michigan. The reason that IHD has a lower IAC score is probably due to the great variety of etiology within IHD. More specific disease such as diabetes has higher level of internal homogeneity. Thus the internal homogeneity of population discharged with certain diseases is dependent on the specificity of diseases. *Objective 2* of optimizing MMA boundary has been completed and thus *hypothesis 2a* stating that the shape constraint parameter will be less important than the IAC for broad disease groups compared to specific diseases is found to be partly true.

Table 17. Summary of 50 restarts for diseases and disease groups.

	Best restart	Cases No.	Zones No.	P2A Score	IAC Score	ZIP Code IAC Score
IHD	50	58,573	302	12160-13098	0.101-0.104	0.109
Circulatory diseases	44	214,649	310	12291-13180	0.105-0.107	0.112
Diabetes	42	17,352	183	7724-8383	0.113-0.115	0.127
Endocrine disorders	45	44,057	274	11206-12095	0.105-0.108	0.113

Objective 2: To optimize MMA boundary definitions.

Hypothesis 2b: Aggregating on the variables sex, age and race will increase demographic homogeneity within zones.

The results showed that as the zones get larger, they become more similar and therefore the differences between zones become smaller, i.e., less heterogeneous. That explains why the ZIP Code level homogeneity is better than the MMA level. Therefore, hypothesis 2b is found to be true. This study also advanced the AZM approach by applying the Wilcoxon signed-rank method to prioritize the top 'restart' out of the 50 restarts conducted. While the Wilcoxon signed rank may not be the only test available to prioritize the restarts it applied nicely to this study and is recommended for use in future AMZ modeling because of it is relatively easy to automate and implement. Other related approaches to prioritize the 50 restarts could be explored and compared in the future.

Region

The use of HSA boundary as region constraint parameter doesn't appear to have adverse effect on the overall zone design because the statewide zone outputs with and without HSA boundary have similar numbers of zones, P2A scores and IAC scores. That greatly reduces that possibility that using HSAs as regions will result in sub-optimal zone design. Thus in the main section, there is no significant effect to use region constraint.

Simulated Annealing (SA)

Finally, this study used simulated annealing (SA) in the construction of MMAs, which is not documented in previous studies. Theoretically, SA helps to achieve the better global-optima, which would be an ideal solution to reduce bias associated with the MAUP. However, only one research (Ralphs & Ang, 2009) had done the experiment to compare the results between running AZM in a SA mode or not and the conclusion is that SA didn't improve the performance of AZM significantly in terms of the population target, and also at the cost of shape compactness. This study also tested the SA effect and found that running AZM in a SA mode didn't produce better zone designs. Since there isn't any final conclusion yet, this could be a promising topic for future research.

<u>Objective 3</u>: To visualize disease proportions, crude rates and age-adjusted rates of IHD, diabetes, circulatory diseases and endocrine disorders:

Hypothesis 3a: The spatial patterns of disease groups will vary by method used to map the diseases; new information about the prevalence of these diseases in Michigan will be learned from each method.

Area-based proportions of patient discharges, crude rates and age-adjusted rates of patient discharges in the population were calculated using the new MMAs and their spatial patterns were visualized. Some of the interesting findings and interpretation of those findings is presented below.

Ischemic Heart Disease

Generally, MMAs that showed elevated proportions of discharged patients represented demand for inpatient hospital services in Michigan. If more people were discharged from hospitals with IHD in 2008, it could be assumed that people would still have that demand in the near future. The medical information of proportion is especially useful for local health department and hospitals, because they would know that the amount and quality of health facilities and equipment should meet the demand of patients with IHD.

On the map showing proportions of patients discharged with IHD, HSA 6 and 7 showed elevated proportions. The highest proportion was also located in HSA 6. The general pattern of crude map is similar to that of proportion map, indicating that patients attended hospitals were fairly representative of the total population. Since the areas with elevated rates and proportions were mostly rural areas where people were less rich, people with IHD should be taken better care of by providing availability and accessibility of health services to them. However, the true reasons that those areas showed elevated rates remain unknown. For example, doctors might refer quite a lot of patients to those areas where high-quality health care services were offered in local hospitals and clinics. Thus it might be too early to conclude that people were less healthy over in those areas.

From the legend of maps showing age-adjusted rates and crude rates, it is found that age-adjustment reduced the overall rates significantly. The highest rates reduced from 24.2 to 17.2 per 1,000 population, indicating the elder people could largely contribute to the high rates. However, even after age-adjustment, the rate maps still showed that HSA 6 and 7 and other scattered MMAs had elevated age-adjusted rates. This illustrates that more elder people is not the only factor causing the high crude rates, and other factors (e.g., more doctors' referral, more health insurance coverage, high socioeconomic states) could also contribute to the high rates. Future research should continue to explore the factors that are associated with the high rates.

The circulatory diseases always show similar patterns as IHD's, no matter proportion maps or rate maps. Since IHD is the main cause of heart diseases and even circulatory diseases, it's not surprising to see similar patterns. As a broad disease group, circulatory diseases show the demand of health services from people with more various characteristics.

Diabetes

The proportion maps of patients discharged with diabetes and endocrine disorders are less similar than IHD and circulatory diseases, although of all people discharged with endocrine disorders, the cases of diabetes were more than one third in 2008. Thus, if the spatial pattern of diabetes is the solely one to be studied, using its broader group (i.e., endocrine disorders) to obtain more cases will not be a wise idea. But one thing in common for these two maps is, the high proportions of patient discharges are mostly located in urban areas, indicating that people with endocrine disorders sought more health services or their physicians were more likely to admit them to hospitals than in rural areas.

The age-adjustment doesn't reduce the rate of patients discharged with diabetes in urban areas. These findings suggest that the population of diabetic patients is relatively similar to the state's population. It is worth noting that the highest crude rate and the highest age-adjusted rate were located in the same metropolitan area – Detroit. More elderly people might not be the reason of high rates in Detroit. Thus, it will be interesting to explore the causes of high rates over there.

Another important point is, diabetes is usually less severe and harmful than IHD, and thus there may be a substantial number of outpatients throughout the state. Moreover, people may not be fully aware of their state of diabetes. Because diabetes is an all-age disease, young people may not have health insurance coverage, which impedes their ability to access health care. People living in rural areas may not be able to see a doctor because hospitals are hardly accessible or affordable. Those factors should be taken into consideration while make decisions. Future research should target on the causes of high rates or low rates to explore the actual reasons of certain spatial patterns.

Limitations

Though AZM offered an optimal zone design for health service study, some limitations are still inevitable. First, using simulated annealing, the aggregation of contiguous ZIP Codes is random until there is not random ZIP Code left and then selection criteria begins. Although the number of zones and IAC/P2A coefficients were similar for each restart, the ways that ZIP Codes were aggregated were completely different. AZM cannot produce two set of zones with exactly the same design within a finite number of running. As a result, the disease patterns might be slightly different in each restart. But the optimal zone design could always be selected meeting the criteria of strongest IAC and smallest P2A coefficients to reduce the uncertainty. Future study should continue to evaluate the global versus local optima (with the local optima the optimal selection occurs first increasing the likelihood that the boundary definitions will appear similar) and simulated annealing. While previous studies (Flowerdew, et al., 2008; Martin, 2003; Riva, et al., 2008) largely tested the zone effect by conducting analysis at different scales, a further step may also consider the scale effect.

Second, this study didn't test how different weights would have influences on AZM zone output. While a set of parameters was recommended in this study, weights of parameters were determined more on an empirical basis. It is highly possible that changing the weights could result in different zone design outputs and different number of zones. David Martin also suggested that great attention should be paid while determining the weights (Atkinson & Martin, 2000). Further study could focus on choosing the optimal weight for each parameter based on knowledge of how it may relate to the disease in question.

Third, for the maximum homogeneity parameter, only sex, age and race were used because of the data limitation. However, the choices of parameters should depend on the specific need. However, as implemented, only a limited number of variables (i.e., a maximum of 12 variables) could be taken into homogeneity parameter.

Another limitation in the mapping process is that we used multiple data sources. MIDB collected data from patients who usually reported their USPS ZIP Codes. Population estimates from the private company Geolytics were originally derived from census 2000 which used ZCTAs. The boundary used in this study is from ESRI 2007 who appeared to use USPS ZIP Codes. As is known to people, the USPS ZIP Codes and ZCTAs do not align very well. Although recoding has been performed, some of the boundaries are still inaccurate. It might lead to the misinterpretation if population was misallocated due to approximate matching of geographic units. Future research should perform analysis and make maps based on the same geographic boundary.

Finally, natural break was the classification method used in the choropleth maps because a good distribution of data is preferable. However, a different classification scheme (e.g., quantiles) may have a substantial visual impact on the perceived spatial patterns of the diseases. Since quantile was proven to be suitable for showing health data on choropleth maps (Brewer & Pickle, 2002) because it keeps the same number of units in each class, it should be explored in future research with the knowledge that interesting data outliers may be smoothed over. Besides quantile, equal interval also provides the possibility to compare different maps.

Future Research

In this analysis health service demand was assessed by constructing Medical Management Areas to visualize and explore the proportions, crude rates and age-adjusted rates of IHD and diabetes. Similarities and differences between these measures were analyzed. Future research could explore underlying mechanisms associated with high or low level of patient discharges to further inform health service researchers and also future epidemiological research. Sociologists and public health researchers may further examine the disparities of health service demand between urban and rural areas, and between younger and elder people in the state of Michigan. These methods may also be applied in other states to assess hospital demand in those areas.

6. CONCLUSIONS AND RECOMMENDATIONS

Ischemic heart disease (IHD) and diabetes, as two of the critical health indicators in the state of Michigan, receive substantial attention from the Michigan Department of Community Health (MDCH). MDCH is interested to know the spatial patterns of the critical health indicators for the purpose of health service regulation from demand perspective. This study illustrates that the spatial patterns of IHD and diabetes derived from administrative units (i.e., ZIP Codes), can result in a substantial number of units with high RSEs and therefore unstable rates throughout the state when the numbers of cases within units are small. One of the methods to solve this small number problem is to aggregate ZIP Codes into larger zones to obtain more cases within zones. In this study, AZM dissolves the boundaries of ZIP Codes with less than 30 cases, and constructs new MMAs with maximum shape compactness and maximum internal homogeneity. The corresponding broader groups of IHD and diabetes in ICD-9-CM (circulatory diseases and endocrine disorders) are also utilized to construct MMAs for the comparison with IHD and diabetes. Choropleth maps serve as the media to represent the spatial patterns of those diseases and disease groups. Proportions (cases per 100 discharges), area-based crude rates (cases per 1,000 population) and age-adjusted rates are shown as different health views. MMAs with high and low prevalence are informative to health care regulators and are useful to assist decision-making and health service management. AZM is therefore an efficient tool that could be used in health service research to explore health care demand.

While AZM is useful in constructing MMAs for health service studies, future researchers should be aware that the construction of zone designs are a relatively subjective process. The output of this program is significantly influenced by the purposes of study, data availability and the input parameters. Researchers should therefore, be careful with the selection of input parameters in the optimization of new zones. With these issues considered, AZM is still recommended for future health service and geographic-epidemiological research.

APPENDICES

Appendix 1. Diseases and Injuries Tabular Index.

1. INFECTIOUS AND PARASITIC DISEASES (001-139)

2. NEOPLASMS (140-239)

3. ENDOCRINE, NUTRITIONAL AND METABOLIC DISEASES, AND IMMUNITY DISORDERS (240-279)

4. DISEASES OF THE BLOOD AND BLOOD-FORMING ORGANS (280-289)

5. MENTAL DISORDERS (290-319)

6. DISEASES OF THE NERVOUS SYSTEM AND SENSE ORGANS (320-389)

7. DISEASES OF THE CIRCULATORY SYSTEM (390-459)

8. DISEASES OF THE RESPIRATORY SYSTEM (460-519)

9. DISEASES OF THE DIGESTIVE SYSTEM (520-579)

10. DISEASES OF THE GENITOURINARY SYSTEM (580-629)

11. COMPLICATIONS OF PREGNANCY, CHILDBIRTH, AND THE PUERPERIUM (630-679)

12. DISEASES OF THE SKIN AND SUBCUTANEOUS TISSUE (680-709)

13. DISEASES OF THE MUSCULOSKELETAL SYSTEM AND CONNECTIVE TISSUE (710-739)

14. CONGENITAL ANOMALIES (740-759)

15. CERTAIN CONDITIONS ORIGINATING IN THE PERINATAL PERIOD (760-779)

16. SYMPTOMS, SIGNS, AND ILL-DEFINED CONDITIONS (780-799)

17. INJURY AND POISONING (800-999)

SUPPLEMENTARY CLASSIFICATION OF FACTORS INFLUENCING HEALTH STATUS AND CONTACT WITH HEALTH SERVICES (V01-V89)

SUPPLEMENTARY CLASSIFICATION OF EXTERNAL CAUSES OF INJURY AND POISONING (E800-E999)

Appendix 2. List of incompatible ZIP Codes.

Table 18. List of incompatible ZIP Codes.

Incompatible ZIP Codes	Operation
48028	removed
48033	divided into 48033, 48034
48138	removed
48139	recode to 48189
48143	recode to 48169
48168	divided into 48167, 48168
48190	recode to 48191
48193	divided into 48192, 48193
48243	removed
48434	recode to 48465
48437	recode to 48458
48476	recode to 48429
48620	removed
48627	recode to 48653
48630	recode to 48629
48633	recode to 48632
48638	divided into 48603, 48638
48710	divided into 48706, 48710
48724	recode to 48604
48758	removed
48824	divided into 48823, 48824
48852	recode to 48829
48853	recode to 48879
48874	recode to 48871
48894	recode to 48835
48896	recode to 48858
49027	recode to 49056
49104	divided into 49103, 49104
49115	recode to 49125
49119	recode to 49128
49282	recode to 49233
49289	recode to 49279
49312	recode to 49309
49335	recode to 49348
49406	recode to 49453
49434	recode to 49423
49458	recode to 49402
49519	divided into 49509, 49519
49534	divided into 49534, 49544
49611	recode to 49659
49626	recode to 49660
49627	recode to 49648

Table 18 (Cont'd)

49628	recode to 49635
49634	recode to 49660
49666	recode to 49649
49674	removed
49722	recode to 49770
49748	recode to 49728
49757	removed
49764	removed
49775	removed
49782	removed
49791	recode to 49721
49793	recode to 49780
49796	recode to 49713
49805	recode to 49950
49808	recode to 49814
49852	removed
49863	removed
49864	removed
49871	recode to 49866
49872	recode to 49878
49901	recode to 49950
49915	recode to 49935
49917	recode to 49913
49918	recode to 49913
49922	recode to 49945
49929	recode to 49953
49934	recode to 49930
49955	recode to 49916
49959	recode to 49911
49660	recode to 49953
49961	recode to 49967
49963	recode to 49905
49971	recode to 49953

REFERENCES

REFERENCES

Acheson, R. M. (1978). The definition and identification of need for health care. *Journal of Epidemiology and Community Health, 32*(1), 10.

AHRQ. (2002). Helping the Nation With Health Services Research Fact Sheet Agency for Healthcare Research and Quality. from http://www.ahrq.gov/news/focus/scenarios.htm

Alvanides, S., Openshaw, S., & Macgill, J. (2001). Zone design as a spatial analysis tool. In N. J. Tate & P. M. Atkinson (Eds.), *Modelling Scale in Geographical Information Science* (pp. 141-157). New York: Wiley.

Atkinson, P. M., & Martin, D. (2000). *GIS and Geocomputation*: CRC.

Bennett, K. M., Scarborough, J. E., & Vaslef, S. (2010). Outcomes and health care resource utilization in super-elderly trauma patients. *The Journal of surgical research, 163*, 127-131.

Bowling, A. (2002). *Research methods in health*: Open University Press Filadelfia.

Bradshaw, J. (1972). A Taxonomy of Social Need *Problems and Progress in Medical Care: Essays on Current Research*. Oxford: Oxford University Press for Nuffield Provincial Hospitals Trust.

Brewer, C. A., & Pickle, L. (2002). Evaluation of methods for classifying epidemiological data on choropleth maps in series. *Annals of the Association of American Geographers, 92*(4), 662-681.

Bureau, C. (2002). Source and accuracy of the data for the March 2001 current population survey microdata file, 2001. *National Vital Statistics Reports, 56*(698).

CDC. (2002). Healthy People 2010 Statistical Note: Healthy People 2010 criteria for data suppression: US Dept of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics.

Cockings, S., Harfoot, A., Martin, D., & Hornby, D. (2011). *Maintaining existing zoning systems using automated zone design techniques: methods for creating the 2011 census output geographies for England and Wales*. Environment and Planning A.

Cockings, S., & Martin, D. (2005). Zone design for environment and health studies using pre-aggregated data. *Social science & medicine (1982), 60*, 2729-2742.

CON. (1978). Chapter IV-Certificate of Need. *Medical care, 16*(10), 21-26.

Cromley, E. K., & McLafferty, S. (2002). *GIS and public health*: The Guilford Press.

Daras, K., & Alvanides, S. (2005). Zone design in public health policy. In M. Campagna (Ed.), *GIS for sustainable development* (pp. 247-167): CRC.

ESRI. (2011). Redistribution rights data, from http://help.arcgis.com/en/arcgisdesktop/10.0/help/index.html#//001z00000003000000.htm

Flowerdew, R., Manley, D. J., & Sabel, C. E. (2008). Neighbourhood effects on health: does it matter where you draw the boundaries? *Social science & medicine (1982), 66*, 1241-1255.

Fotheringham, A. S., & Wong, D. W. S. (1991). The modifiable areal unit problem in multivariate statistical analysis. *Environment and Planning A, 23*(7), 1025-1044.

Freeman, H. E., Levine, S., & Reeder, L. G. (1979). *Handbook of medical sociology*: Prentice-Hall, Inc., Englewood Cliffs, New Jersey

Gehlke, C. E., & Biel, K. (1934). Certain effects of grouping upon the size of the correlation coefficient in census tract material. *Journal of the American Statistical Association, 29*(185A), 169-170.

Ghosh, A., & Rushton, G. (1987). *Spatial analysis and location-allocation models*. New York: Van Nostrand Reinhold.

Goodall, B. (1987). *The Penguin dictionary of human geography*: Penguin Books.

Grady, S. C. (2010). Racial residential segregation impacts on low birth weight using improved neighborhood boundary definitions. *Spatial and Spatio-temporal Epidemiology, 1*(4), 239-249.

Grady, S. C., & Enander, H. (2009). Geographic analysis of low birthweight and infant mortality in Michigan using automated zoning methodology. *International Journal of Health Geographics, 8*.

Grady, S. C., & McLafferty, S. (2007). Segregation, nativity, and health: Reproductive health inequalities for immigrant and native-born black women in new york city. [Article]. *Urban Geography, 28*(4), 377-397.

Haining, R., Wise, S., & Ma, J. (1998). Exploratory Spatial Data Analysis. *Journal of the Royal Statistical Society: Series D (The Statistician), 47*(3), 457-469. doi: 10.1111/1467-9884.00147

Haldeman, J. C. (1959). Progressive patient care: A challenge to hospitals and health agencies. *Public Health Reports, 74*(5), 405.

Haynes, R., Daras, K., Reading, R., & Jones, A. (2007). Modifiable neighbourhood units, zone design and residents' perceptions. *Health & place, 13*, 812-825.

Haynes, R., Jones, A. P., Reading, R., Daras, K., & Emond, A. (2008). Neighbourhood variations in child accidents and related child and maternal characteristics: does area definition make a difference? *Health & place, 14*, 693-701.

Healthy People 2010 Statistical Note. (2002). Healthy people 2010 criteria for data suppression: Centers for Disease Control and Prevention, Department of Health and Human Services, National Center for Health Statistics.

Ho, V., Ku-Goto, M.-H., & Jollis, J. G. (2009). Certificate of Need (CON) for cardiac care: controversy over the contributions of CON. *Health services research, 44*, 483-500.

Indiana State Department of Health. (2005). Data Users Guide - Rates, Small Numbers, Percents, Etc. , from http://www.in.gov/isdh/23986.htm

Institute of Medicine. (1979). *Health Services Research: Report of a Study*. Washington DC: The National Academies Press.

Kreft, I. G. G. (1996). *Are multilevel techniques necessary? An overview, including simulation studies*. Unpublished manuscript, California State University, Los Angeles.

Langley, S. A., Fuller, S. P., Messina, J. P., Shortridge, A. M., & Grady, S. C. (2010). A methodology for projecting hospital bed need: a Michigan case study. *Source Code Biol Med, 5*, 4.

Lohr, K. N., & Steinwachs, D. M. (2002). Health services research: an evolving definition of the field. *Health services research, 37,* 7-9.

Lorant, J. H. (1971). Empirical Studies in Health Economics. *JAMA: The Journal of the American Medical Association, 215*(2), 301.

Martin, D. (2003). Extending the automated zoning procedure to reconcile incompatible zoning systems. *International Journal of Geographical Information Science, 17*, 181-196.

Martin, D., & Cockings, S. (2001). AZM Online Help

McLafferty, S., & Grady, S. (2005). Immigration and geographic access to prenatal clinics in Brooklyn, NY: A geographic information systems analysis. *American Journal of Public Health, 95*(4), 638-640. doi: 10.2105/ajph.2003.033985

McLafferty, S., & Wang, F. H. (2009). Rural Reversal? Rural-Urban Disparities in Late-stage Cancer Risk in Illinois. *Cancer, 115*(12), 2755-2764. doi: 10.1002/cncr.24306

McLafferty, S. L. (2003). GIS and health care. *Annual review of public health, 24,* 25-42.

McNerney, W. J., & Riedel, D. C. (1962). Regionalization and rural health care. An experiment in three communities. *Regionalization and rural health care. An experiment in three communities.*

MDCH. (2009). *2009 edition of the Michigan Critical Health Indicators report*. Retrieved from http://www.michigan.gov/documents/mdch/Critical_Health_Indicators_2007_198949_7.pdf.

MDCH. (2011). Comparison of Michigan Critical Health Indicators Report & Healthy People 2010 Targets.

Meade, M., Florin, J., & Gesler, W. M. (1988). *Medical geography*: Guilford Publications.

NYSDOH. (1999). Age-Adjusted Rates - Statistics Teaching Tools, from http://www.health.state.ny.us/diseases/chronic/ageadj.htm

Openshaw, S. (1977). A geographical solution to scale and aggregation problems in region-building, partitioning and spatial modelling. *Transactions of the institute of british geographers, 2*(4), 459-472.

Openshaw, S. (1984). *The modifiable areal unit problem*: Geo Books.

Openshaw, S., & Rao, L. (1995). Algorithms for reengineering 1991 Census geography. *Environment and Planning A, 27*, 425-425.

Penchansky, R., & Thomas, J. W. (1981). The Concept of Access - Definition and Relationship to Consumer Satisfaction. *Medical care, 19*(2), 127-140.

Ralphs, M., & Ang, L. (2009). Optimised geographies for data reporting: zone design tools for Census output geographies *Statistics New Zealand Working Paper No 09-01*. Wellington: Statistics New Zealand.

Riva, M., Apparicio, P., Gauvin, L., & Brodeur, J. M. (2008). Establishing the soundness of administrative spatial units for operationalising the active living potential of residential environments: an exemplar for designing optimal zones. *International Journal of Health Geographics, 7*, 43.

Seaman, V., Hopkins, G. F., & Webb, J. D. (1796). *An Account of the Epidemic Yellow Fever, as it Appeared in the City of New-York in the Year 1795: Containing, Besides Its History, &c., the Most Probable Means of Preventing Its Return, and of Avoiding It, in Case it Should Again Become Epidemic*: Printed by Hopkins, Webb & Co. no. 40, Pine-Street.

Shannon, G. W., & Dever, G. E. A. (1974). *Health care delivery: Spatial perspectives*: McGraw-Hill.

Stafford, M., Duke-Williams, O., & Shelton, N. (2008). Small area inequalities in health: are we underestimating them? *Social science & medicine (1982), 67*, 891-899.

Statler, K. D., Dong, L., Nielsen, D. M., & Bratton, S. L. (2011). Pediatric stroke: clinical characteristics, acute care utilization patterns, and mortality. *Child's nervous system : ChNS : official journal of the International Society for Pediatric Neurosurgery, 27*, 565-573.

The Merck Manual 16th edition. (1992). *The Merck manual of diagnosis and therapy.*: Merck & Co., Inc.

Tranmer, M., & Steel, D. G. (1998). Using census data to investigate the causes of the ecological fallacy. *Environment and Planning A, 30*(5), 817-831.

Walker, J. E. C., Murawski, B. J., & Thorn, G. W. (1964). An experimental program in ambulatory medical care. *New England Journal of Medicine, 271*(2), 63-68.

Wang, F. H., McLafferty, S., Escamilla, V., & Luo, L. (2008). Late-stage breast cancer diagnosis and health care access in illinois. *Professional Geographer, 60*(1), 54-69.

Washington State Department of Health. (2010). Guidelines for Using and Developing Rates for Public Health Assessment, from http://www.doh.wa.gov/data/Guidelines/Rateguide.htm#unstablerates

WHO. (2011a). International Classification of Diseases (ICD), from http://www.who.int/classifications/icd/en/

WHO. (2011b). International Classification of Diseases (ICD), ICD-10 2nd Edition Volume 2 Instruction Manual, from http://www.who.int/classifications/icd/en/

Wise, S., Haining, R., & Ma, J. S. (2001). Providing spatial statistical data analysis functionality for the GIS user: the SAGE project. [Article]. *International Journal of Geographical Information Science, 15*(3), 239-254. doi: 10.1080/13658810151072877