This is to certify that the
thesis entitled

CHILDHOOD LEAD POISONING IN MICHIGAN:
SPATIAL ANALYSES OF THE DISTRIBUTION OF AND
FACTORS RELATING TO COMMUNITY ELEVATED BLOOD
LEAD LEVELS

presented by

ERIC SANDBERG

has been accepted towards fulfillment
of the requirements for the

M.S. degree in Geography

Major Professor's Signature

14 Nov 08

Date

CHILDHOOD LEAD POISONING IN MICHIGAN: SPATIAL ANALYSES OF THE DISTRIBUTION OF AND FACTORS RELATING TO COMMUNITY ELEVATED BLOOD LEAD LEVELS

By

Eric Allen Sandberg

A THESIS

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

MASTER OF SCIENCE

Geography

2008

ABSTRACT

CHILDHOOD LEAD POISONING IN MICHIGAN:
SPATIAL ANALYSES OF THE DISTRIBUTION OF AND FACTORS RELATING TO COMMUNITY ELEVATED BLOOD LEAD LEVELS

By

Eric Allen Sandberg

Lead poisoning, defined by the Centers for Disease Control as equal-to or greater-than ten micrograms per deciliter of blood, afflicts children in Michigan at a higher rate than the national average. The primary, though not exclusive, source of exposure is lead-based paint in households that dates to before the 1978 ban on this product. Since lead exposure causes permanent neural damage and is difficult to extract from the body, primary prevention by removing the hazards is the only solution to this problem. This thesis uses point-based clustering and regression techniques to examine the spatial patterns and characteristics of childhood blood lead levels in Michigan. The Michigan Lead Database results of blood lead tests from 1998 to 2005 are employed for this objective. Only children insured by Medicaid, a majority of the database and typically at higher risk of lead poisoning, are included in this thesis. Results indicate that the inner city children in Michigan suffer the greatest from lead exposure. Regression analysis reveals that older housing within an area is the best predictor of mean blood lead levels. Spatial techniques used in this thesis have the potential to greatly enhance primary prevention efforts.

This thesis is dedicated to my family

ACKNOWLEDGEMENTS

I wish to express my thanks to Dr. Joseph Messina, who has mentored and advised me throughout my time in graduate school and welcomed me warmly to Michigan and MSU.

I also wish to express my gratitude towards Dr. Sue Grady, who has warmly advised and encouraged me during the past two years, and Dr. Stan Kaplowitz, who has provided invaluable help, advice, and encouragement in pursuing research into lead poisoning.

I wish to recognize and thank Dr. Ashton Shortridge, who provided much of the instruction and technical assistance in running the methods of this thesis.

I wish to acknowledge Mark Finn, Ivan Ramirez, Annalie Campos, and Lindsey Campbell for their assistance.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Images in this thesis are presented in color.

LIST OF ABBREVIATIONS

µg/dL: Micrograms per Deciliter

BLL: Blood Lead Level

CDC: Centers for Disease Control and Prevention

CLPPP: Childhood Lead Poisoning Prevention Program

EPA: Environmental Protection Agency

FDA: Food and Drug Administration

GAM: Geographic Analysis Machine

GIS: Geographic Information Systems

GWR: Geographically Weighted Regression

HSA: Health Systems Agencies

LBPPPA: Lead-Based Paint Poisoning Prevention Act

LIA: Lead Industries Association

MCD: Minor Civil Division

MCGI: Michigan Center for Geographic Information

MDCH: Michigan Department of Community Health

OLS: Ordinary Least Squares Regression

PCA: Principal Components Analysis

TEL: Tetraethyl Lead

1 Introduction

1.1 Introduction

Lead has adversely affected humans for thousands of years (Bellinger and Schwartz 1997). Though the harmful effects of lead were recognized in antiquity, it has continued to be used in many manufactured items. Recent events such as the lead paint found in Chinese-manufactured toys emphasize the risk which still exists from products found on store shelves (Barboza 2007). But the greatest hazards from lead are from the vestiges of an earlier time period when lead was commonly used in house paint and gasoline. Many people still suffer needlessly from the effects of lead particle inhalation or ingestion within their homes and neighborhoods. Children suffer the most because of the small size of their bodies, and their behaviors put them at greater risk (Centers for Disease Control and Prevention 2005a). The children who are insured by Medicaid, a government-funded health care coverage program for low-income individuals and families, are known to typically have higher blood lead levels than the general population (Kemper et al. 2005a). Thus all children on Medicaid are required by law to be tested by two years of age, and others are encouraged to be tested during a health visit (Michigan Department of Community Health 2001). Michigan is sixth in the nation for percentage of children with elevated blood lead levels (Task Force to Eliminate Childhood Lead Poisoning 2004). Indications are that the distribution of children with high blood lead levels (BLL) in Michigan is not random, but is associated with historical patterns of development and current place-based socio-demographic and economic characteristics (Frost 2004). This research focuses on exploring the spatial distribution of BLL in

children in the State of Michigan (Figure 1), emphasizing the patterns observed and the common socio-demographic and economic characteristics associated with them.

Figure 1: Reference map of Michigan

Michigan children have historically had higher BLL than the national average stemming from a variety of risk factors. Heavy industrialization throughout the late 19th

and early 20th century caused atmospheric lead deposition in the state from the combustion of coal and leaded gasoline from cars (Yohn et al. 2004). In many urban areas in Michigan and throughout the United States, soil depositions from leaded gasoline (1929-1986) created a large persistent reservoir of lead (Mielke 1999). This input is frequently coupled with lead house paint, both interior and exterior. Though lead paint was banned from use in residential homes in 1978, an estimated 64 million homes in the United States still contain layers of lead-based paint (Jacobs et al. 2002). Children living in states with older housing are at greater risk of lead poisoning because lead paint chips are often in or around the outside of the house. The chips and dust of lead can amass in areas of the house, accessible for children to inhale. According to the U.S. Census Bureau, nearly three-fourths of Michigan houses were built during or before the 1970s (US Census Bureau 2001). While many substantial sources of lead such as leaded paint and gasoline are no longer in production, used lead is environmentally stable and continues to be a hazard to which Michigan children could be exposed.

With the threat to children of lead firmly established, Governor Jennifer Granholm (2002 – present) recently created a task force to lead "a statewide effort to successfully address the goal of the elimination of childhood lead poisoning in Michigan by 2010" (Task Force to Eliminate Childhood Lead Poisoning 2004). In 1997, regulations were put into place that required Michigan laboratories to report the results of all blood lead tests to the Michigan Department of Community Health (MDCH), replacing the voluntary reporting set up in 1992 (Michigan Department of Community Health 2005a). Within MDCH, the Childhood Lead Poisoning Prevention Program (CLPPP) coordinates lead-related activities. The results are received by CLPPP,

reviewed for data entry errors, and put into the statewide child lead database. CLPPP then relays results of children with elevated BLL to the local health departments, so they can target homes and neighborhoods for environmental remediation.

Since Michigan's push for the elimination of lead poisoning began, there have been positive developments. The percentage of children in Michigan with elevated BLL (>= 10 μg/dL) decreased from 9.7% (n = 7,100 out of 73,643 tests) of those tested in 1998 to 2.3% (n = 3,137 out of 132,913 tests) of children tested in 2005, possibly indicating CLPPP methods have been successful (Michigan Department of Community Health 2005a). New legislation passed by the Michigan Legislature in 2004 sanctions testing of more children within the state, including ensuring follow-up tests for children with elevated BLL results and faster reporting by labs to CLPPP.

Unfortunately, progress has begun to stall on some fronts. Recent budget challenges within Michigan have put state funds for lead poisoning prevention in jeopardy (Lam 2007). The result is that less money will be available to local health departments for environmental testing and removal (remediation) of environmental lead sources. A recent survey of health officers from local health departments throughout Michigan found that 74% of the respondents reported that lead poisoning was not adequately addressed in their health district (Kemper, Uren, and Hudson 2007). At the same time that funding for lead programs is being cut, new medical and epidemiological research has found that children with BLL lower than the 10 μg/dL cutoff point considered elevated by the Centers for Disease Control and Prevention (CDC) suffer damaging effects (Lanphear et al. 2005b; Finkelstein, Markowitz, and Rosen 1998; Canfield et al. 2003). These studies have shown that effects of lead exposure, such as IQ

loss, can actually occur at a faster rate below the current CDC threshold (Canfield et al. 2003).

The geographic aspects of lead poisoning have received more attention in recent years in community health because of advances in computing technologies such as Geographic Information Systems (GIS), geocomputation, and spatial statistics (Cromley and McLafferty 2002). Analyses of the geographic distribution of lead poisoning are useful for finding "hot spots" where clusters of children with elevated blood lead levels reside and for creating models for where lead exposure is likely higher based on socio-demographic and housing variables (Griffith et al. 1998). The overall population hazard from lead has dropped due to the metal being largely taken out of industrial use and exposure has become more concentrated in older areas. As this drop has occurred, disparities between areas of high and low incidence of lead poisoning have developed (Lanphear 2005a). This divergence can be observed in geographic variations in neighborhood characteristics as well as public health intervention (Bailey, Sargent, and Blake 1998).

1.1.1 Purpose of Study

The purpose of this study is to use the Michigan statewide yearly database of lead test results in children from year to year to explore spatial patterns and processes over time and to measure the extent to which geographic variation in BLL can be explained by US Census socio-demographic variables. This will be accomplished using spatial statistics, spatial clustering techniques, and geographic regression modeling. Building on previous research on the geographic dimensions of lead exposure, this research explores

spatio-temporal variations in lead test results in Michigan. The main questions that this study aims to address are:

Are there spatial clusters of elevated BLL in Michigan? At what spatial scales do these patterns manifest?

Are socio-demographic and economic variables in the US Census able to predict and explain the geographic variation in elevated blood lead levels in Michigan children?

Can a model based on US Census socio-demographic and economic variables accurately predict the spatial distribution of elevated BLL in Michigan over time?

This thesis is organized into four chapters. The remainder of Chapter 1 provides a review of relevant literature and the research hypothesis. Chapter 2 describes data and methods used in investigating these research questions. The results from these analyses and a discussion of their implications are presented in Chapter 3. Finally, Chapter 4 concludes with recommendations for policy and programmatic changes and suggestions of future research.

1.2 Literature Review

1.2.1 Lead Uses and Consequent Problems

Lead is a bluish-gray metal that occurs naturally within the Earth's crust (Centers for Disease Control and Prevention 2005a). There are several elemental properties that make it of use to humans. Lead is very dense, able to be shaped easily, and resistant to corrosion (United States Geological Survey 2007). It is soft enough that it can be rolled into a sheet and shaped into rods and pipes (Hunter 1969). Lead has a very low melting

point, allowing it to be softened in a temperatures as low as a campfire (Angier 2007). Because of these qualities, lead has been distributed widely throughout the environment through extensive human use. Lead does not break down naturally, a fact which separates it from many other environmental contaminants (Kitman 2000).

Archaeological evidence of human use of lead dates back thousands of years. A lead figurine in the British Museum has been dated to 5,800 yrs ago in the Neolithic Period (Clarkson 1995). Lead was also found in Bronze Age pottery and was extensively mined by the Ancient Greeks and Romans (Brill and Wampler 1967; Weiss, Shotyk, and Kempf 1999). Roman use included making lead pipe for plumbing and as a preservative in wine, inducing high lead levels among the Roman aristocracy and suspicion among modern researchers that lead might have played a role in the decline of the empire (Nriagu 1983; Waldron 1973). Evidence of lead's durability is found in excavated 2,000 year old perfectly preserved Roman water pipes (Hunter 1969).

Though lead was continuously used in pre-industrial societies, studies conducted in various environmental archives such as peat bogs and glaciers confirm that lead production and use in the environment exponentially increased after the industrial revolution (Weiss, Shotyk, and Kempf 1999). Lead has been used in many products such as batteries, water pipes, ammunition, ceramic glazes, roofing, and lead sheet for lining buildings. But the two applications that caused the most damage to American children were lead paints and in leaded gasoline (Centers for Disease Control and Prevention 2005a).

Leaded gasoline was developed to reduce engine knock. The solution settled on in the 1920s by the automotive industry was Tetraethyl lead (TEL), selected over several

safer alternatives such as ethanol (Kitman 2000). TEL improved engine performance and was an effective anti-knocking agent, which led to it being called "a gift from God" by an industry executive (Nriagu 1990). Despite early warning signs such as refinery worker deaths, the industries involved in the production and use of leaded gasoline continued to resist any efforts by the public health community for a ban and worked to fund its own research (Kovarik 2005). Leaded gasoline is documented as the source of nearly all the lead found in the environment (Hernberg 2000).

Lead historically has been used in paints because of its anti-corrosive properties. Two lead compounds, white and red lead, were commonly used in paints through the 20th century. While red lead was used primarily in painting of ships, white lead paint was used in households because it was resistant to water and prevented mildew (Hunter 1969). Lead was considered a valuable addition to paint, making the cost of house paint rise with the amount of lead added into the mixture (Beam 2007). The paint industry as well as the Lead Industries Association (LIA), a lead industry trade group, heavily marketed lead paint (Markowitz and Rosner 2000). Advertisements appeared in popular periodicals touting the durability of leaded paint. The industry also created a mascot of the Dutch Boy, a young boy who appeared in many advertisements encouraging children to use lead paint (Markowitz and Rosner 2002).

There are several ways lead can enter a child's body once it is in the local environment. Lead has a sweet taste, which makes young children (under two years of age) especially vulnerable to lead around the home because children have a tendency to put objects in their mouth, a condition known as pica (Gaston 1972). Also in the home, lead paint can chip, and the dust can accumulate in areas of the house such as

windowsills, carpet, and other accessible places (Lanphear et al. 1998d). Inhalation of lead paint dust by children also can occur when the old paint layers are sanded during home renovation (Lanphear 2005a). Another pathway by which children may be exposed to lead is through the soil around the child's residence. Left over lead from the leaded gasoline era has been found to have accumulated in areas of high traffic congestion (Tong 1990). Children who play in such environments often get lead particles on their hands which can easily be transferred to the mouth and ingested (Mielke 1999). Thus oral ingestion and inhalation are the two main routes by which children are exposed.

Lead is able to disrupt many essential nervous system functions at a cellular level, particularly affecting the developing bodies of children (Garza et al. 2005). Lead is a potent neurotoxin that has been established as a poison for centuries (Lidsky and Schneider 2003). It has been suggested that the root of the neurotoxicity goes far back in the evolution of living cells and lead's role as a non-essential metal. Lead levels in modern humans are estimated to be 50-200 times higher than in estimated blood lead levels before human lead usage following the industrial revolution (Flegal and Smith 1992). Tests on animals have shown similar negative effects of exposure which show up in humans (Finkelstein, Markowitz, and Rosen 1998). Once lead is inside the human system, it is able to mimic the role of other essential metals for cell function like calcium (Clarkson 1995). No known life forms rely on lead for survival (Angier 2007).

Once inside the body, lead effects on children are serious and long-term even at very low levels. Lead exposure is typically measured in micrograms of lead per deciliter (μg/dL) of blood. The current threshold for what is considered lead poisoning by the CDC is 10 μg/dL. This is equivalent to a teaspoon of lead in a swimming pool 100 feet

by 40 feet and five feet deep (Richardson 2005). At clinical levels of lead exposure, generally above 60-70 μg/dL, a child will begin to show outward signs that poisoning has occurred. These include loss of the ability to coordinate muscular movement, convulsions, anemia, stupor, colic, coma, and possibly death (Agency for Toxic Substances & Disease Registry 2007). Such high levels of lead were once quite common in the United States, but since the gradual phasing out of leaded paint and gasoline, lead exposure usually occurs at a sub-clinical level where testing is needed to confirm poisoning. Sub-clinical effects of lead exposure include decreased impulse transmission through the nervous system, reduced cell and nerve function, loss of IQ points, and decreased hearing and growth (Bellinger and Bellinger 2006). Follow-up studies of children with high blood lead levels as toddlers have found links with loss of IQ points once the child enters school (Chen et al. 2005). There has been recent interest in studying the effects of lead exposure below the CDC threshold 10 μg/dL for lead poisoning (Canfield et al. 2003; Finkelstein, Markowitz, and Rosen 1998; Lanphear et al. 2005b; Needleman and Bellinger 1991a). Research has shown children with blood lead levels within this lower range (<10μg/dL) experience adverse effects. Needleman and Bellinger (1991a) summarized the research and found a strong link for loss of IQ points at lower levels. Finkelstein, Markowitz, and Rosen (1998) studied the effects of lead on the central nervous system and found that any amount of lead within the body was hazardous. Canfield et al. (2003) found that IQ loss occurred more rapidly at BLL concentrations below the CDC threshold than at higher concentrations. Lanphear et al. (2005b) confirmed this finding by surveying IQ test scores and BLL levels. Their research found an inverse relationship between IQ and BLL with the steepest drop under the 10 μg/dL.

This development has led to greater concern among public health officials for the safety of children who have been exposed but have a blood lead level under the CDC threshold, as well as initiated calls for the threshold to be lowered (Gilbert and Weiss 2005).

Treatment for lead exposure is time consuming and often cannot undo the damage already caused (Silbergeld 1997). Because lead is absorbed into the body at a cellular level, it is very difficult to extract. Chelation therapy is a process where a chelating agent is added to the body which binds with lead, making it inert and speeding up bodily excretion (Ettinger 1999). It is has been licensed by the Food and Drug Administration (FDA) to be used when the child's blood lead level is above 45 μg/dL (Dietrich et al. 2004). The process can take many treatments as BLL often rebounds following initial dosage. Chelation therapy has come under scrutiny because of its ineffectiveness of preventing neurological damage (Rosen and Mushak 2001). Medical professionals increasingly stress that the only effective way of treating lead exposure is primary prevention of lead hazards within the children's environment.

1.2.2 Research, Industry, and Public Policy

Through the lens of hindsight, many early warnings of the danger of lead were missed or ignored (Figure 2). A few observers in Roman times made the connection between ship builders and lead poisoning, but modern discovery of the etiologic connections between lead and various symptoms of poisoning dates to the 19th century (Hernberg 2000). Early studies of the effects of lead examined factory workers who were exposed to massive amounts of lead dust (Tong, Schirnding, and Prapamontol 2000). The first study of the source of lead in children was conducted by an Australian doctor, J.

Lockhart Gibson, who identified lead paint as the source of exposure (Gibson 1904). News of the Australian results reached American researchers when mentioned within a medical textbook in 1907 and Gibson's call for lead paint to be banned from places near children in 1911 (Markowitz and Rosner 2002). Very soon, articles about lead began to appear in the American academic journals. Early research came from John Hopkins Hospital in Baltimore, where in 1917 physician Kenneth Blackfan described the horrible condition of children suffering from clinical lead poisoning and called for measures to keep children from lead paint (Fee 1990). Mounting pressure began to build around the world for lead to be banned from house paint.

During the first few decades of the 20th century, an assortment of countries banned lead from household interior paint. France, Belgium, and Austria were the first to ban indoor lead paint in 1909, followed by bans in Tunisia and Greece as well as a resolution supporting outlawing lead paint by the League of Nations in 1922 (Chisolm 2001). By 1927, Great Britain, Australia, Czechoslovakia, Sweden, Belgium and Poland had followed suit (Richardson 2005). But the United States would not take this step for another 50 years.

The creation of the Lead Industries Association (LIA) trade group in 1928 had a profound effect on US policy relating to lead products. The group was able to successfully lobby for the industry and stifle any attempt at regulation of lead paint. At the same time, the health community was debating TEL gasoline. The lead gasoline industry turned to Robert Kehoe, a researcher out of the University of Cincinnati, for scientific aid to support their case. Kehoe is widely recognized as the originator of a paradigm still used by industry today, that burden of proof for proving a product

hazardous enough for removal lies with health experts and not industry (Nriagu 1998). In Kehoe, the industry found their spokesman scientist who would point to lead being a natural element within the human body (Needleman 1998). For most of the middle part of the 20th century, the only research funding for studying lead came from industry, and most of those funds went to Kehoe. His research on behalf of the makers of TEL and his primacy in lead research helped keep regulation at bay (Kitman 2000). At a 1925 conference commissioned by the surgeon general to debate regulations on TEL, Kehoe successfully defended its use against other health advocates who called for a ban. With no formidable opposition, the lead industry began to advertise heavily. LIA began to intensely promote white lead paint in residential homes, producing pamphlets for children, buying ad space in popular magazines, and having representatives travel around the country promoting its use to a variety of state and local governments. This promotion of lead by LIA included advocating its use in some Michigan public school districts (Markowitz and Rosner 2002).

The tide began to turn against the lead industry in the 1940s. A rash of lead related sickness and deaths during the Great Depression made the issue harder for the medical community to ignore. As blood lead testing became more widely available, medical consensus grew on the harm of lead, and the chorus of criticism put the lead industry increasingly on the defensive. Randolph Byers and Elizabeth Lord published a study in 1943 where they followed children who had been poisoned by lead in early childhood, finding nearly all experienced behavioral problems and struggled in school (Chisolm 2001). *Time* magazine picked up the story and brought it to a national audience (Markowitz and Rosner 2000). Many other stories about lead poisoning began to appear

in magazines and on television news over the next decade (Markowitz and Rosner 2002). However, while the paint industry voluntarily reduced lead content in its paints in the mid-1940s, it did not remove lead completely from house paint. As environmental awareness grew during the 1960s, public tolerance of industrial contamination waned.

In 1970 there were no federal regulations regarding lead paint, and only four states and ten cities in the United States had bans on the indoor use of paint (Hernberg 2000). Early legislation in the United States was meant to respond to lead poisoning rather than prevent it. Congress passed the first federal legislation against lead paint in 1971, a half-century after many other developed nations. Known as the Lead-Based Paint Poisoning Prevention Act (LBPPPA), the measure prohibited lead-based paint (defined as more than 1% lead by weight) in residential structures built by the federal government, set the lead poisoning threshold at 60 μg/dl, and set abatement standards (Department of Housing and Urban Development 2004). The newly created Environmental Protection Agency (EPA) followed in 1973 with the first regulations of leaded gasoline, beginning a gradual phase-out that lasted until 1986. In 1975 model year, automobile manufacturers began building vehicles which had a new emission control system including a catalytic converter, which required unleaded gasoline (Environmental Protection Agency 1996). The final major policy regulations came in 1977, when the US Consumer Product Safety Commission ruled that residential house paint could not contain more the 0.06% lead by dry weight (Bellinger and Bellinger 2006). With the regulations of the 1970s, major sources of childhood lead poisoning were no longer being manufactured, though the vestiges of earlier usage remained a threat.

Effects of the new legislation were immediate and striking. In the National Health and Nutrition Examination Survey (NHANES II) conducted by the CDC, average BLL of people surveys dropped from 16 μg/dl to 9 μg/dl between 1976 and 1980 (Needleman 2004). But the same survey estimated that 700,000 children likely had elevated blood lead levels (30μg/dL at this time), leading to a continued push by the public health community for more funds (Rabin 1989). In the research community, the priority began to shift from demonstrating the harm of lead to targeting the source of elevated blood lead levels in communities. The new population-based studies began to look at what locales were at risk in order to aid the removal of hazards and the prevention of exposure before it occurs.

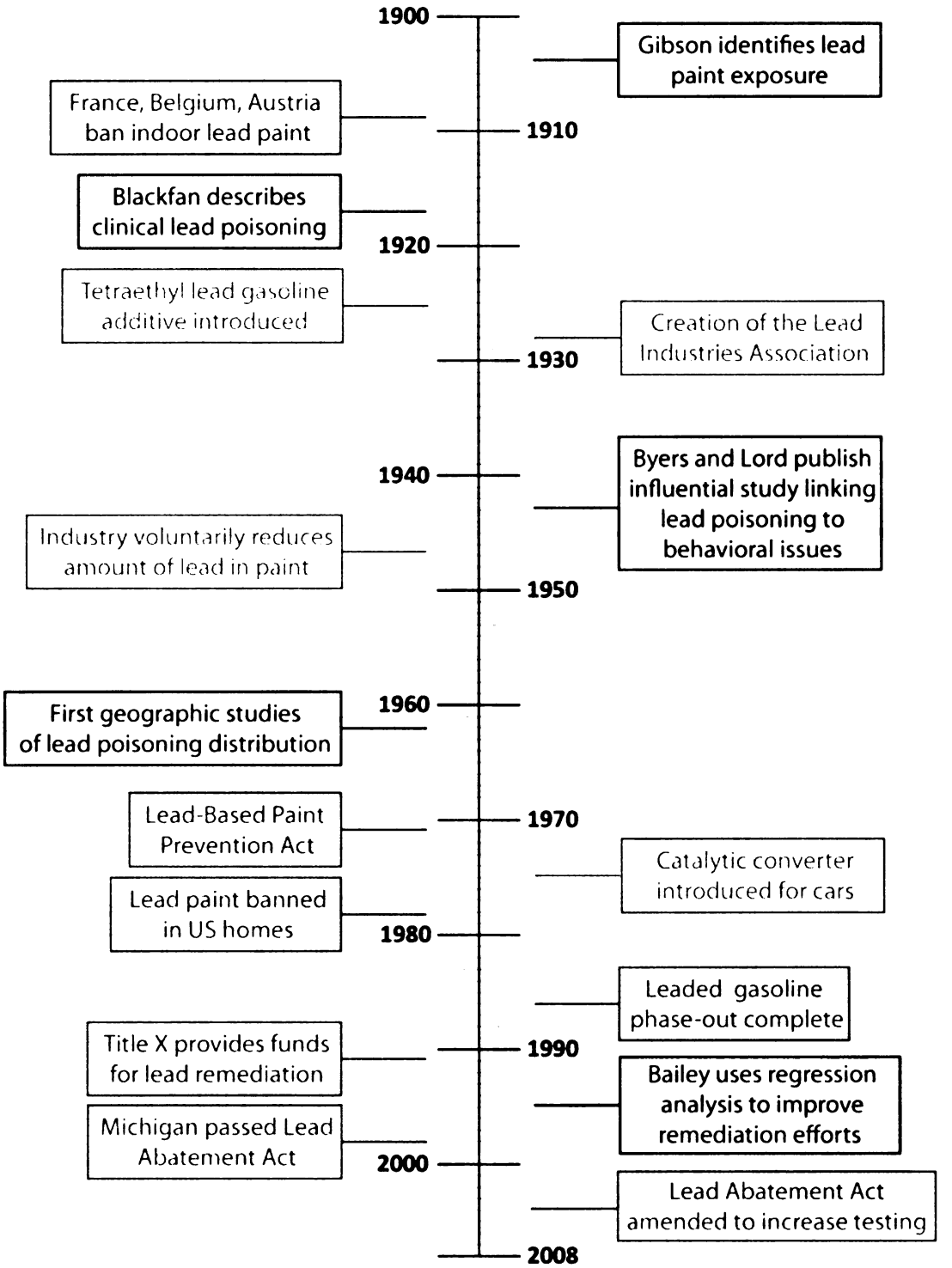


Figure 2: Timeline of events relating to lead poisoning. Legislation is marked in blue, business and industry marked in orange, and research is marked in green.

In the early 1990s, legislation was passed at the federal level to provide funding for primary prevention of lead poisoning. Coupled with the lowering of the elevated BLL threshold to 10 μg/dL in 1991, the passage of Title X of the Housing and Community Development Act of 1992 made federal funding available for remediation programs and broadened the official definition of a lead-based hazard. Remediation of lead involves removal of all lead paint dust, removal of lead-based paint, removal of lead-contaminated topsoil, and replacing painted fixtures (Environmental Protection Agency 2001). It has to be carried out by a state-certified contractor. The bill made grants available for state and local governments to reduce lead paint in private sector housing. It required that housing sold by the federal government be lead-free, extended the LBPPPA to all housing, and ensured disclosure of the danger to residents (Richardson 2005). Title X marked a change in policy from treating specific cases to prevention of lead poisoning before it occurs. Lead-based hazards were extended from just paint chips to dust within the house and bare soil on the property (Department of Housing and Urban Development 1993). Individual states were now expected to draft abatement plans or risk loss of federal funding.

The threat of funding shortfall prompted the Michigan Legislature to pass the Lead Abatement Act in 1998. This provided local health departments throughout Michigan with funds to conduct blood tests on children and remediate the child's environment if necessary. A screening plan (Appendix 1) was developed to cover children thought to be at risk is based on the CDC recommendations (Michigan Department of Community Health 2007). Universal screening is now recommended for zip codes in Michigan where 27% of housing was built before 1950 (national average),

12% incidence of lead poisoning among children 12 to 36 months of age in 2000, or high percentages of pre-1950 housing and children living in poverty. Zip codes that are deemed high-risk by those standards are shown in figure 3. If a child is not in one of these zip codes but is insured by Medicaid, a blood lead test is required and paid for by the federal government (Kemper and Clark 2005c). Though follow-up screening is required for children who have BLL above the 10μg/dL limit, this mandate is not followed nearly half the time (Kemper et al. 2005b). Finally, if the child is not insured by Medicaid and does not live in a high risk zip code, MDCH recommends that the parents or guardians be given a questionnaire to determine if a blood lead should be given. The questions ask if the child lives in or visits a building built before 1950, has a sibling or playmate with lead poisoning, lives around an adult who works with lead, is subject to cultural practices or remedies containing lead, or is included in a special population group that may had suffered previous exposure such as a foreign adoptee. A yes answer to any of these questions prompts a blood lead test (Michigan Department of Community Health 2007).

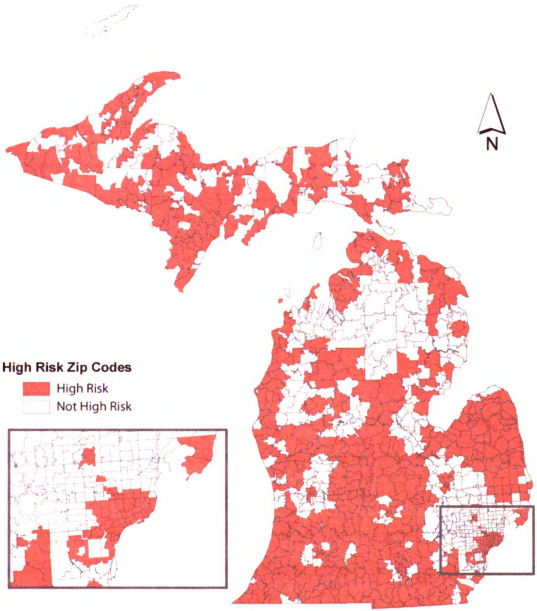


Figure 3: Map of zip codes deemed "high risk" by CDC standards

Following press reports on lead poisoning in 2003, the Michigan Legislature amended the Lead Abatement Act in 2004 to increase testing of vulnerable children (Centers for Disease Control and Prevention 2005b). The Lead Task Force appointed by the governor crafted a plan to rid Michigan of lead poisoning by eliminating lead hazards

in housing, expanding testing, assuring capacity to serve kids who need medical help, and securing funding (Task Force to Eliminate Childhood Lead Poisoning 2004).

1.2.3 Geographic Studies of Lead

Research in how lead exposure varies by geographic location began in the 1960s. The geography of lead poisoning was a component of the wider research into clinical lead poisoning (Gaston 1972). Many studies were based in large cities where the residence of children who were treated in a hospital was plotted on a city map. For example, Jacobziner and Raybin (1962) investigated cases of lead poisoning reported by New York City hospitals. Analysis was restricted to disease mapping, where locations of the residences of lead poisoned children were plotted on a map. The authors found a spatial pattern of children with elevated BLL, uncovering a "lead belt" through the low income, largely minority neighborhoods which was attributed to substandard housing with lead-based paint (Jacobziner and Raybin 1962). Other studies based their spatial analysis on blood lead samples collected throughout study areas, such as the cities of Chicago and Philadelphia (Gaston 1972). Disease maps of the samples confirmed that lead poisoning (above 60 μg/dL at the time) generally afflicted lower income neighborhoods that often contained older housing and politically dispossessed citizens. The spatial patterns found by these community samples were later confirmed through larger statewide population surveys and screening programs (Griffith et al. 1998).

Larger population-based studies at county, state, and national levels that looked at using population variables to focus primary prevention strategies were completed in the 1980s and 1990s. The NHANES II survey from 1976-1980 conducted the first

population-wide study of children with lead poisoning (Bailey et al. 1994). Results showed that the problem was the worst in urban areas, and African-American children suffered more exposure to lead than others (Mahaffey et al. 1982). Children under the age of six were found to have the highest mean BLL. Unlike adults where men had higher average BLL, the child's sex was found to not be predictor of lead exposure (Mahaffey et al. 1982). While statewide screening programs generally came after Title X, several studies looked at lead poisoning in cities that had programs. Daniel (1990) found that while BLL in New York City was declining overall, the older urban areas were more likely to have housing with layers of lead paint than housing outside the city. African-Americans accounted for nearly two-thirds of lead poisoning cases, and children between six months and two years old were found to be at the highest risk (Daniel et al. 1990). Guthe et al. (1992) used GIS to examine at the spatial pattern of blood lead test results compared to major roadways and industrial sites in Newark, New Jersey. The lack of conclusive links between these sites and the occurrences of elevated BLL caused the authors to call for additional research (Guthe et al. 1992). Since these studies revealed the same patterns with the same population markers, research into the spatial distribution of lead poisoning turned to using regression analyses to discover areas where exposure was more likely.

To better target screening programs that proliferated after the passage of Title X, researchers studying the geography of lead poisoning turned to regression models based on enumerative unit variables (Table 1). An early example was Bailey et al. (1994), who looked at lead poisoning in children in Massachusetts at the minor civil division scale. Though the research was criticized because the state screening program at the time used a

surrogate marker rather than the actual blood lead level, the paper did indicate that many population risk factors that had been identified earlier indeed helped explain the distribution of lead poisoning throughout Massachusetts. Several common indicators of community lead risk were found to explain the geographic variation of lead poisoning in the state including percentage of African-Americans, percentage of housing units built before 1940, and percentage of households headed by a female (Bailey et al. 1994). Bailey also looked at the role of an area's industrial heritage in lead poisoning by creating a dummy variable for minor civil divisions that bordered the industry-heavy Merrimack River and found that adjacency to this waterway was statistically significant in predicting elevated BLL.

The next regression model for lead poisoning that appeared in the literature was Sargent et al. (1995), who also looked at lead poisoning in Massachusetts. Many of the same variables were observed to affect geographic variation of lead poisoning as Bailey (1994), this time at a community level (Sargent et al. 1995). In each case, impoverished communities had greater difficulty with childhood lead poisoning. Similar to the Bailey model, this regression did suffer from the fact that Massachusetts used a surrogate marker for BLL. Two years later, both authors were involved in creating a model for lead exposure, this time at the census tract level in Providence, Rhode Island (Sargent et al. 1997). While many of the same poverty and racial characteristics were found to predict geographic variations as the earlier models, additional variables were used which were found to have a significant effect. One such factor was the percentage of recent immigrants to the United States (< 5 years). The authors speculate that the lack of

understanding of the dangers of lead paint and the language barrier might have placed immigrants at greater risk for lead exposure (Sargent et al. 1997).

The first regression model for lead poisoning that considered the spatial component was Griffith et al. (1998). The study looked at Syracuse, New York with three US Census scales: blocks, block groups, and tracts. New variables found to explain geographic variation of BLL were average household value and average rent. Griffith also used buffering analysis around major roadways and found the BLL of children living next to roadways to be similar to the rest of the study population, which indicated that leaded gasoline did not contribute to elevated BLL. But the main contribution of the study was the combination of regression analysis with spatial analysis. Griffith found that incorporating space into the regression analysis through the use of a spatial autoregressive model helped further explain the geographic variance. Elevated BLL in Syracuse was found to cluster at every scale (block group, tract, and zip code) tested, which led the authors conclude that community childhood lead exposure cannot be understood completely without accounting for the geographic dimension (Griffith et al. 1998).

Author	*Study Site*	*Spatial Scale*	*Method*	*Dep. Variable*
Bailey (1994)	Massechusetts	Minor Civil Division	Poisson Regression	Count > 25 mg/dL
Sargent (1995)	Massechusetts	Minor Civil Division	Logistic Regression	Cases / Tests
Sargent (1997)	Providence, RI	Census Tract	Linear Regression	% > 10 mg/dL
Griffith (1998)	Syracuse, NY	Census Block, Blk Group,Tract	Spatial Regression	Number of Cases
Lanphear (1998)	Rochester, NY	Block Group	Logistic Regression	% > 10 mg/dL
Talbot (1998)	New York State	Zip Code	Linear Regression	Ln(% > 10 mg/dL)
Litaker (2000)	19 Ohio Counties	Census Tract	Logistic Regression	12% of more > 10 mg/dL
Miranda (2000)	6 NC Counties	Tax Parcel	Linear Regression	Ln(BLL)
Haley (2004)	New York State	Zip Code	Linear, Spatial Error	Ln(% > 10 mg/dL)
Kaplowitz (n/a)	Michigan	Individual, Blk Group	Linear	Ln(BLL)

Table 1: Summary of previous geographic studies of lead poisoning

Several other local scale studies in the literature have produced interesting results. Lanphear et al (1998b) studied childhood BLL at the census block group level in Rochester, New York. While their regression model did not use any new variables, they tested the model against individual data collected by a testing clinic in a local area. Results showed the block group level data in the community predicted elevated BLL as well as the individual level data (Lanphear et al. 1998b). Litaker et al (2000) used a risk score based on housing, ethnicity, education, and housing rental for their regression model of 19 Ohio counties. They found that their model predicted the spatial distribution of elevated BLL better than the CDC guidelines, which are the same as the screening plan by MDCH (Litaker et al. 2000). The study by Miranda (2002) is the only lead regression

model organized at the parcel level. Though not practical for a statewide study, the authors used tax parcel data for six counties in North Carolina to estimate the areas most in need of primary prevention. The finer scale of the analysis allowed a residence-by-residence analysis based on the year each structure was built (Miranda, Dolinoy, and Overstreet 2002). While the study worked at a microscale for the counties surveyed, the difficulty of gathering household data on other variables did not allow the authors to look at many other socio-economic factors.

The largest population-based geographic elevated BLL study was done in New York State (Haley and Talbot 2004; Talbot, Forand, and Haley 1998). Authors of the study used zip code level variables to predict areas in the state where the percentage of children with elevated BLL would be higher. A linear regression model and a spatial error regression model were used throughout the entire state. Perhaps the most interesting result in the research was that the same variables of percentage housing built before 1940, percentage high school graduates, and percentage African-American births were the best predictors of childhood BLL in both New York City as well as the rest of the state (Talbot, Forand, and Haley 1998). Generally, lower levels of BLL found in New York City are attributed to the fact the lead paint was banned by the local government in residential areas within the city two decades earlier than the federal ban, though the result still surprised the authors. Conclusions of the study were that when working with a large study area, variables that explain BLL variance at finer scales might not persist. For example, population density was noted to not have an effect at the statewide level, unlike earlier localized studies (Haley and Talbot 2004).

The faculty of the Sociology Department at Michigan State University has studied common factors of BLL in Michigan. A detailed survey was used to sample around 4,200 children throughout Michigan to determine significant indicators of elevated BLL(Frost 2004). Children who lived in urban, low-income areas were sampled. The variables found to significantly predict BLL in a child were water through lead pipes, siblings with elevated BLL, adults in the house with elevated BLL, the child is African-American, and household income below $20,000. The data were later used to create a predictive model based on census variables (Kaplowitz, Perlstadt, and Post 2007). As the first study to use a continuous dependent variable for BLL, the authors found that Medicaid status, race of the child, and ethnic character of the neighborhood were strong predictors of BLL. Other interesting finds included that exposure risk was higher with pre-1940 housing than the housing built between 1940 and 1950 (Kaplowitz, Perlstadt, and Post 2007).

Author		Independent Variable	+/-	P - Value
Bailey (1994)		Log (Number of children screened)	+	<0.001
		Percentage African-American	+	0.004
		Percentage Female-Headed Households	+	0.003
		Percentage Houses built before 1940	+	<0.001
Sargent (1995)		Median Per Capita Income	-	<0.001
		Percentage African-American	+	<0.001
		Percentage Houses built before 1950	+	<0.001
		Screening Rate	+	<0.001
		Poverty Scale	+	0.007
Sargent (1997)		Percentage Screened	+	0.01
		Percentage Houses built before 1950	+	<0.001
		Natural Log (Number of Vacant Houses)	+	<0.001
		Percentage Recent Immigrants (< 5 years)	+	0.003
Griffith (1998)	Tract	Population Density	+	undisclosed
		Average House Value	-	undisclosed
		Percentage Under 18 years old	+	undisclosed
	Block Group	Population Density	+	undisclosed
		Average House Value	-	undisclosed
		Percentage African-American	+	undisclosed
	Block	Percentage African-American	+	undisclosed
		Average House Value	-	undisclosed
		Percentage Under 18 years old	+	undisclosed
		Percentage Hispanic	+	undisclosed
		Percentage Renter Occupied Housing	+	undisclosed
Lanphear (1998)		City Residence	+	<0.001
		Percentage Screened	+	<0.001
		African-American Population	+	<0.001
		Percentage Houses built before 1950	+	<0.001
		Population Density	+	<0.001
		Low House Value	+	<0.001
		High Poverty	+	<0.001
		Low High School Graduation Rates	+	0.004
		Low Owner Occupied Housing	+	0.012

Table 2: Regression results from earlier studies. Columns are author, independent variable, whether the coefficient is positive or negative, and the p-value

Author		Independent Variable	+/-	P - Value
Talbot (1998)		Percentage African-American births	+	<0.001
		Percentage High School Graduates	-	<0.001
		Percentage Houses built before 1940	+	<0.001
Litaker (2000)		Percentage living in rural areas	-	0.005
		Percentage African-American	+	<0.001
		Percentage Houses built before 1950	+	<0.001
		Percentage Under 6 years old	+	<0.001
		Percentage Male Under 6 years old	+	0.001
		Percentage without High School Diploma	+	<0.001
		Percentage below 150% poverty line	+	<0.001
		Percentage Housing Renters	+	<0.001
		Percentage Female Headed Households	+	<0.001
Miranda (2002)		Residence Year of Construction	-	<0.001
		Median Income	-	<0.001
		Percentage African-American	+	0.001
Haley (2004)	New York City	Percentage Houses built before 1940	+	<0.001
		Percentage without High School Diploma	+	0.02
		Percentage African-American	+	<0.001
	New York State	Percentage Houses built before 1940	+	<0.001
		Percentage without High School Diploma	+	<0.001
		Percentage African-American	+	<0.001
Kaplowitz (unpublished)		Percentage below 185% poverty line	+	<0.001
		Percentage African-American	+	<0.001
		Percentage Latino	+	<0.001
		Percentage without High School Diploma	+	<0.001
		Percentage Houses built before 1950	+	<0.001

Table 3: Continuation of Table 2 showing regression results from earlier studies

Previous geographic studies of lead exposure have shown the usefulness of using regression models (Tables 2 and 3). While many similar variables have been shown to be predictive of childhood BLL, the geographic element of lead poisoning has proved to be important. Factors such as population density have influence at certain spatial scales, but not others.

1.2.4 Theoretical Basis and Hypothesis

Medical geography is a research field which draws upon concepts from a range of disciplines (Meade and Earickson 2000). While interest in how disease varies through space goes back centuries, the organization of medical geography as an academic field dates to the middle of the 20th century (Akhtar 1982). The work of Jacques May in the 1950s introduced the ecology of disease where human behavior-based factors determined the limitations of disease incidence (Meade 1977). The disease ecology approach resulted in a shift from studying disease itself, a process rooted in germ theory, to studying the environment where the disease grows and occurs (Akhtar 1982). Disease became to be viewed as a interrelationship of factors occurring at a certain time and space (Jones and Moon 1987). Disease agents are constrained by the typical environments where they can survive, creating a characteristic spatial distribution, also called landscape epidemiology (Mayer 1986). Disease mapping became a valuable tool for the study of the pattern of disease, although without an underlying process theory (Mayer 1982).

The human ecology model came to medical geography from the biological sciences by way of sociology (Honari 1999). According to Meade and Earickson (2000), human ecology refers to the "patterns of human interaction with the physical environment, including not only behavior but genetic adaptation and physiological reaction to environmental stimuli." Human ecology is a holistic model, concerned with interactions at all scales (Honari 1999). The human-ecology triangle (Figure 4) was created to show that human health is based on the interactions between individual or population characteristics, behavior, and habitat (Meade and Earickson 2000). Population

is concerned with the individual or groups of individuals with common characteristics, looking at how factors such as age, gender, and genetics affect human health. Behavior refers to the observable aspect of culture, which manifests itself in conditions humans create through alteration of the landscape, customs and social norms, and utilization of resources (Meade 1977). Habitat is the environment, both natural and human constructed, in which a person lives as well as the social environment that controls the structure of the person's surroundings (Meade and Earickson 2000). The study of elevated BLL in children that utilizes the human ecology perspective is important because of the clear relationship between children and their behavior in their local environment. The concern among many researchers is not so much with lead itself, but with the environment where it is prevalent and the children who are at risk of exposure. The state of a child's health as related to lead exposure depends on factors related to all three vertices of the triangle, meaning each should be considered.

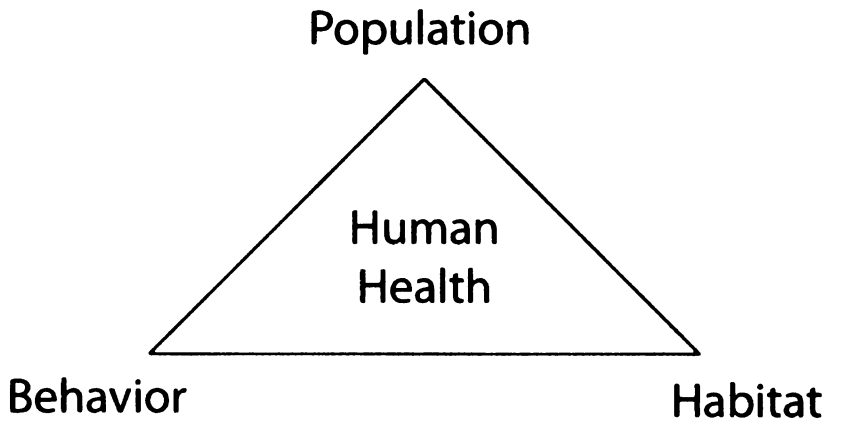


Figure 4: The human ecology triangle

The behavioral aspect of the human ecology triangle for lead has been the most influential due to the preventable nature of lead exposure. Lead poisoning is a disease that is entirely produced by human use of resources. The decision to use lead as an

additive to paint and gasoline for most of the 20th century is the driving reason behind the problem today. Political indifference to the seriousness of lead poisoning also contributed greatly to the prevalence of lead in the American environment. In terms of a spatial lead study, human behavior comes into play in several ways. The first is through the marginalization of impoverished areas, which are known to be the areas of highest lead exposure risk (Pirkle et al. 1998). The expense of remediation and the historically lukewarm response from the public sector has left lower income areas without a correcting mechanism for eradicating the lead in their environment (Rabin 2008). Studies of lead exposure have shown that the effect of human behavior does not always come from industrial or political decision-making (Bailey, Sargent, and Blake 1998). Local efforts to screen children for lead in the bloodstream have an effect on BLL, as well as the educational attainment levels in the community. Individual behavior of both the parent and child influence lead exposure as well. Parents who are employed where lead is present can unknowingly bring it home on their clothes (Frost 2004). Other parental behaviors which affect childhood lead exposure are remodeling an older house with lead paint, using foreign-made products such as cosmetics which might contain lead, and not complying with lead paint removal regulations. The main behavior of children that puts them at risk is pica, the compulsive need to ingest non-food substances (Gaston 1972).

The child's environment, or habitat, affects lead exposure. It figures prominently in the human ecology model for a variety of diseases, but is not a large factor in childhood lead exposure. Pre-industrial levels of lead were much lower than today, indicating lead posed virtually no risk before human's began altering the environment

(Kovarik 2005). Current background concentrations in the soil have been found to be highest near industrialized areas (Murray, Rogers, and Kaufman 2004). Still, it is from the child's human-constructed environment where children live that poses the highest risk of lead exposure. A young child's world is much more constrained than an adult, meaning that more often than not the trigger for lead exposure lies within the house. Lead products lie in older housing stock, dating from years of leaded paint and lead water pipes, and they generally make housing age among the best predictors of child BLL (Pirkle et al. 1998). Other habitat features include the settlement patterns of towns and cities. Michigan cities tend to be decentralized, leading to greater use of cars (Vojnovic et al. 2006). This long-term trend could create lead reservoirs near major roadways that were heavily trafficked during the leaded gasoline era (Hunter 1976).

The human ecology model also considers the social environment in which the child is living. Social environment in the human ecology triangle refers to the "groups, relations, and societies which people live (Meade and Earickson 2000)." Recent immigrants to the United States demonstrate an example of how the social environment around a child could affect BLL. Often, the communities live in substandard housing, do not speak English, are unaware of the dangers of lead, or have residents in the country illegally who cannot come forward for testing (Centers for Disease Control and Prevention 2005b).

Individual level factors are an important part of the human ecology model, but generally are not that important in lead exposure studies. Because lead toxicity is harmful to everyone, typical population factors such as genetics do not make a difference. The ethnic makeup of a neighborhood does predict the elevated BLL, but this is not due

to any physical factor which falls under the population vertices of the human ecology triangle. Researchers also have looked at disparity in BLL between the two genders and uncovered no significant difference in BLL between male and female children (Mahaffey et al. 1982). Age and race are normally the only individual factor that has an effect (Goyer 1993). Typically the peak age for childhood BLL has been found to be about two years of age (Lanphear et al. 2005b).

With knowledge of previous research and the background of the human ecology triangle, this thesis will attempt to answer the questions posed earlier by developing a geographically based regression model. The goal is to create a useful model that illuminates the spatial character of elevated BLL in Michigan and provides a tool for use in primary prevention. From past research, I hypothesize that:

1. Clusters of elevated BLL exist in Michigan. These clusters are within older urban neighborhoods. Similar to Griffith et al (1998), these patterns will manifest at several spatial scales.
2. Variables associated with older housing, lower income, lack of education, and recent immigration to the US will best predict the spatial distribution of BLL. The predictive power of each variable will also vary by place throughout the state and at different geographic scales.
3. The model will work across time ranges due to the underlying socio-economic factors causing the same distribution of BLL every year.

2 Data and Methods

2.1 Data

Lead in the environment remains a hazard for Michigan children. The only viable solution is to prevent exposure at the source (Rosen and Mushak 2001). Primary prevention remains a key strategy for eliminating lead in the human environment (Centers for Disease Control and Prevention 2005b). This thesis divides the geographic study of blood lead levels (BLL) into two phases, the identification of the patterns of affected children and an examination of the socio-economic correlates. Two datasets were used for the geographic study of BLL within the state of Michigan. The primary dataset used is the Michigan Lead Database, created and maintained by Michigan Department of Community Health (MDCH), which contains information and BLL results of each child under the age of six who took a blood lead test. To make sense of the spatial patterns of BLL observed in the lead database, data tables containing possible independent variables were downloaded from the United States Census Summary Files for the 2000 Census. These two sources were used to create both the geocoded BLL test results point dataset and the statewide areal units.

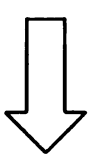
2.1.1 Michigan Lead Database

Since 1997, all laboratories that conduct lead tests within Michigan have been required to report all results to MDCH (Michigan Department of Community Health 1998). These results were originally sent by the labs as paper copies of the Blood Lead Analysis Report, but 2004 legislation now requires electronic reporting (Kemper et al. 2005a). Blood lead analysis reports filed by the testing labs are reviewed for

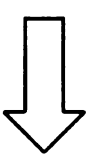
completeness, entered into the database, and run through quality control checks to find any data entry errors (Michigan Department of Community Health 1998). A 2002 internal study that tested the registry's ability to link to other state-maintained datasets such as the Medicaid enrollment files found it to be over 99% accurate (Kemper et al. 2005a). Once the test information is entered into the database, MDCH notifies the child's health care provider and local public health organization of the results (Michigan Department of Community Health 2006). In the case of children with elevated BLL, a local environmental investigation may follow to determine the source of exposure

MDCH Database

Child ID	Address	Birth Date	Race	Insurance	Testing Date	Test Type	BLL
000001	431 I St	3/8/2003	White	Self-Pay	6/3/2004	Capillary	2
000002	682 I St	4/24/2003	White	Medicaid	6/6/2004	Venous	10
000002	682 I St	4/24/2003	White	Medicaid	9/17/2004	Venous	4

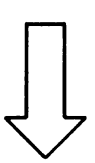


Duplicate tests removed (highest BLL kept)
Addresses Geocoded

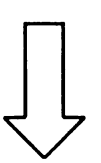


MSU Database

Child ID	Address	Birth Date	Race	Insurance	Testing Date	Test Type	BLL
000001	431 I St	3/8/2003	White	Self-Pay	6/3/2004	Capillary	2
000002	682 I St	4/24/2003	White	Medicaid	6/6/2004	Venous	10



Non-Medicaid Children Removed



Thesis Database

Child ID	Address	Birth Date	Race	Insurance	Testing Date	Test Type	BLL
000002	682 I St	4/24/2003	White	Medicaid	6/6/2004	Venous	10

Table 4: Example highlighting the changes between the original BLL database and the database used in this thesis

The MDCH database contains information about each lead test from 1998 to 2005 and personal information for the examined child. The microgram per deciliter result of the child's blood lead test is recorded as an integer value, with 1 being the lowest number. Also included is whether the test was a capillary or venous test. Capillary tests, also known as finger stick, draw only a small amount of blood (under 100 μL) and are cheaper to administer than the venous test (Parsons, Reilly, and Esernio-Jenssen 1997). General consensus holds that the venous test is more accurate and less susceptible to contamination, so any child who has a high blood lead result on a capillary test is given a venous test to confirm elevated BLL (Michigan Department of Community Health 2007). For this reason, venous tests are the preferred method for investigators (Dignam et al. 2004).

In addition to the information on the actual test, the registry contains some personal information about the child. Age of the child and date of the blood test are included, which allow the data to be separated by year and age. The race of the child is recorded as well as whether or not the child is covered by Medicaid. The test is required for all children covered by Medicaid, so such children constitute a majority of the registry. Finally, the testing labs record the address of the child's residence.

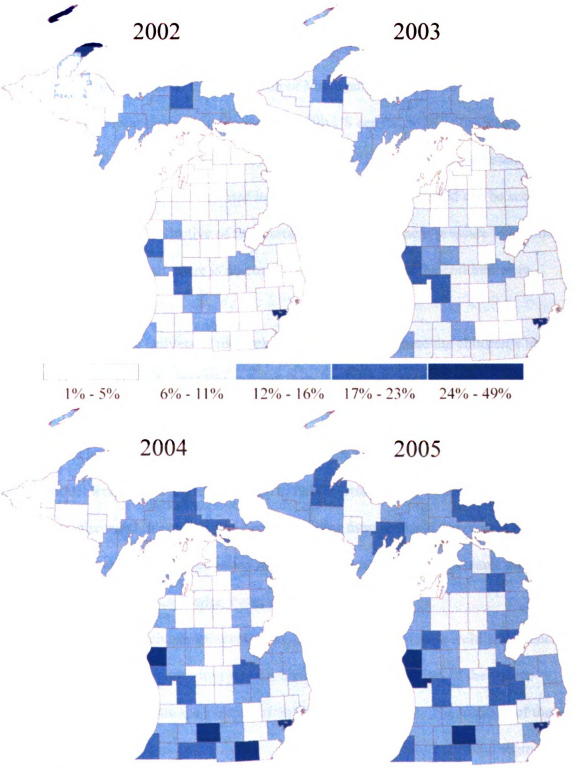


Figure 5: Percentage of children under six years of age tested for lead. All test results for Michigan counties and Detroit included.

Certain assumptions must be made when relying on data acquired from another source rather than collected first hand. Beside the question of data entry and locational accuracy, what proportions of the population of Michigan children were tested remains a concern. In every year since the release of the 2000 US census, MDCH has listed the percentage of children within each county and the city of Detroit who were tested during that year (Figure 5). A general increase in the number of children tested can be seen across the state. This is reflective of the increased state government pressure to eliminate elevated BLL. But overall, there is no county where over 50% of the children were tested.

Michigan State University researchers were able to examine the children's test results in this database. A grant was secured from the Centers for Disease Control for the MSU team to work with the MDCH blood lead test results (Kaplowitz, Perlstadt, and Post 2007). The researchers used the test data to create a regression model with a mix and individual from the database and group variables from the US census. Some test results were discarded in order to avoid complications from multiple samples of the same child. For children who had been tested more than once, the highest test result was kept and the others removed (Kaplowitz, Perlstadt, and Post 2007).

The MSU research team found the geographic location of each child's residence through geocoding. The geocoding process uses a GIS vector data set of the streets within Michigan to estimate the location of each child's residence. The location of the address point is determined by two factors. One is the location along the road segment, estimated by using the address range of the segment as a guide to find the address point location. Another factor is perpendicular offsetting the address point from the road

segment for an accurate estimate of the actual residence site. The process is subject to error but is a commonly used method for GIS-based spatial analysis in health geography (Zandbergen and Green 2007).

Roughly two-thirds of the children in the MSU database were on Medicaid (Kaplowitz, Perlstadt, and Post 2007). This number is much higher than the proportion of children statewide on Medicaid. Because of the concerns over the sampling protocol, it was decided that this thesis would focus exclusively on children covered by Medicaid. Children who are on Medicaid are three times as likely to have elevated BLL as children who are not enrolled (Kemper and Clark 2005c). Since two-thirds of the MSU database is children on Medicaid, these children are more likely to represent the population on Medicaid than the entire MSU database represents the general population. The percentage of Michigan children who are enrolled in Medicaid is around 33% (American Academy of Pediatrics 2003).

With approval from the MSU Human Research Protection Program (IRB # 07-362), the MDCH blood lead database was made available for this thesis. The database was imported into Microsoft Access in order to view descriptive statistics on the children who have been tested. Summary statistics of this database are in figure 6. The number of children tested steadily increased through the years in the registry. There is an especially large rise in the number of tests between 2003 and 2004 after the state government made remediation of lead poisoning a higher priority (Task Force to Eliminate Childhood Lead Poisoning 2004). Another trend is the steady decline in both the mean BLL level in the registry and the percentage of the children whose BLL was elevated (above 10 μg/dL). This decline would likely signal the effectiveness of the primary prevention programs and

remediation, but could also be a product of the increased number of tests. According to Kemper (2005a), the number of children tested likely increased due to requirements by daycare enrollment or early education programs. This might explain why the age of children tested is older than what the CDC recommends.

The donut graphs show that there has been little change in characteristics of the children tested between 1998 and 2005. Children on Medicaid are required to get tested for lead before the age of two or between three to five years of age if not previously tested (Kemper et al. 2005a). Testing under the age of two is generally preferred because children around the age of two tend to show the highest BLL (Ozden et al. 2004). In this dataset, there does not seem to be a preference of testing for children under the age of two. This could be further confirmation that many tests occur later when the child enters educational programs.

The second donut graph shows the proportion of children in the dataset who received a venous test as opposed to a capillary (stick) test. The majority of tests in this dataset, between 60 and 70 percent depending on the year, are venous blood tests. This is encouraging for this research because the venous test is less affected by contamination of the sample (Kemper, Bordley, and Downs 1998).

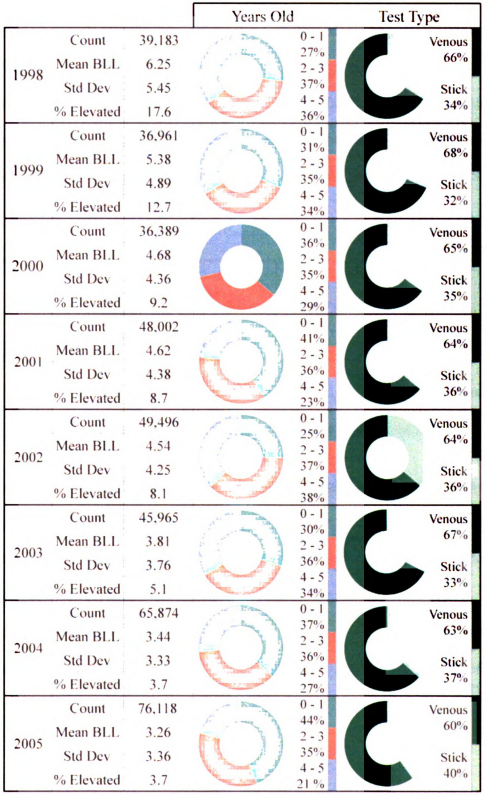


Figure 6: Descriptive statistics of the thesis lead database. Note that elevated means above 10 μg/dL and numbers are for Medicaid insured children.

The process of moving the database to a GIS data format began with importing the MSU database into Microsoft Access (Figure 7). After non-Medicaid children were removed, the new thesis database was divided into eight dBASE (.dbf) files containing the test results for each year. The .dbf format was chosen because of the ease of moving the tables into the GIS program ArcMap. The .dbf files were brought into ArcGIS in order to geocode them.

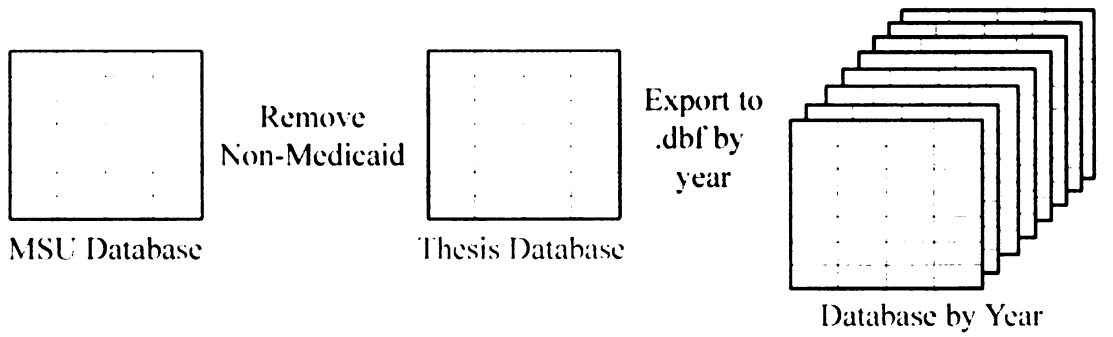


Figure 7: Migration of MSU database to GIS-utilizable .dbf format

A vector data set of Michigan based on the Michigan GeoRef projection was downloaded from MCGI (www.michigan.gov/cgi). The GeoRef projection is preferred when working with Michigan data because it accurately projects the entire state rather than dividing it into sections (Michigan Department of Natural Resources 2001). Latitude and longitude coordinates were used to locate the child's address (Figure 8). The result was eight point-based vector data sets representing each year with all of the database information included.

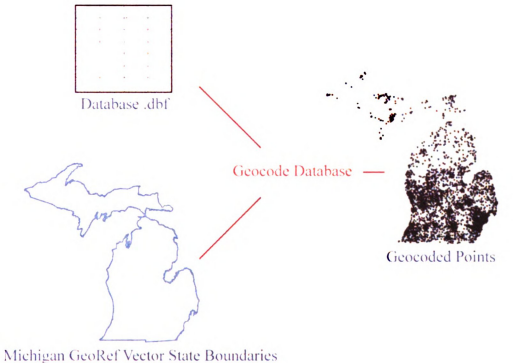


Figure 8: The geographic coordinates were geocoded to a point vector data set through use of the MCGI state boundary vector data set

2.1.2 United States Census

To supply the socio-demographic and economic variables for the regression portion of this thesis, ASCII text data files from the 2000 US census were obtained. Each summary file is available for download from the US census web site (www.census.gov). The various tables can be linked to a variety of geographic divisions through the logical record number. For this thesis, the regression analysis is limited to the geographic levels used in previous spatial BLL studies. This includes census tract, five digit zip code, and minor civil divisions.

The finest scale geographic unit in which the Census Bureau aggregates data for public use is the census block. A block is an areal unit contained within the surrounding

streets or a water body, similar to a city block (US Census Bureau 2000). Census blocks are generally not used in medical geography because they include only raw population counts, not socio-economic variables. Summary File3 is not aggregated by the US Census Bureau because of the small number of census long-form sample respondents within a block. But census blocks provide the basis for every larger geographic unit.

The block group is a cluster of contiguous census blocks. The first digit in the three-digit census block number indicates block groups. Participation by a local statistical committee is taken into account when forming block groups. Each block group is contained entirely within a census tract. A census tract is a statistical subdivision containing between 600 to 3,000 housing units that are delineated by a local committee of data users (US Census Bureau 2000). Census tracts boundaries follow permanent geographic features such as streets, railroads, rivers, and canals. Tract boundaries are geographically contained within individual counties and are designed to be as homogenous as possible with respect to the characteristics of the population within them (US Census Bureau 2000). The tract is a common unit of analysis in medical geography and was used in this thesis.

The final two geographic units of analysis, five digit zip codes and minor civil divisions, are based on federal and local government divisions. Zip codes are service areas created by the United States Postal Service. The Census Bureau aggregated to this unit of analysis for the first time in 2000. This is an important unit of analysis in BLL research because it is often used in testing standards of the CDC and subsequently MDCH. Unlike any other spatial unit, the definition of minor civil divisions (MCD) varies from state to state. In Michigan, MCD refers to townships and incorporated cities

(US Census Bureau 2000). MCD are often preferred as a unit of analysis that the size of each enumerative unit remains fairly constant across the entire state. This is the case in Michigan, where most townships are 36 square mile units created by the Public Land Survey System.

Previous research has identified important variables for the prediction of elevated BLL in children (Bailey, Sargent, and Blake 1998; Talbot, Forand, and Haley 1998; Kaplowitz, Perlstadt, and Post 2007; Griffith et al. 1998; Haley and Talbot 2004; Lanphear et al. 1998b; Litaker et al. 2000; Miranda, Dolinoy, and Overstreet 2002; Sargent et al. 1997; Sargent et al. 1995). The matrices containing significant independent variables noted in tables2 and 3 were downloaded from the census website into Microsoft Access. From there, an identifier called the log record number was used to link the census data with the desired geographic unit. The output table was exported into a .dbf file and joined in ArcMap to census-based vector data sets that were downloaded from MCGI (Figure 9).

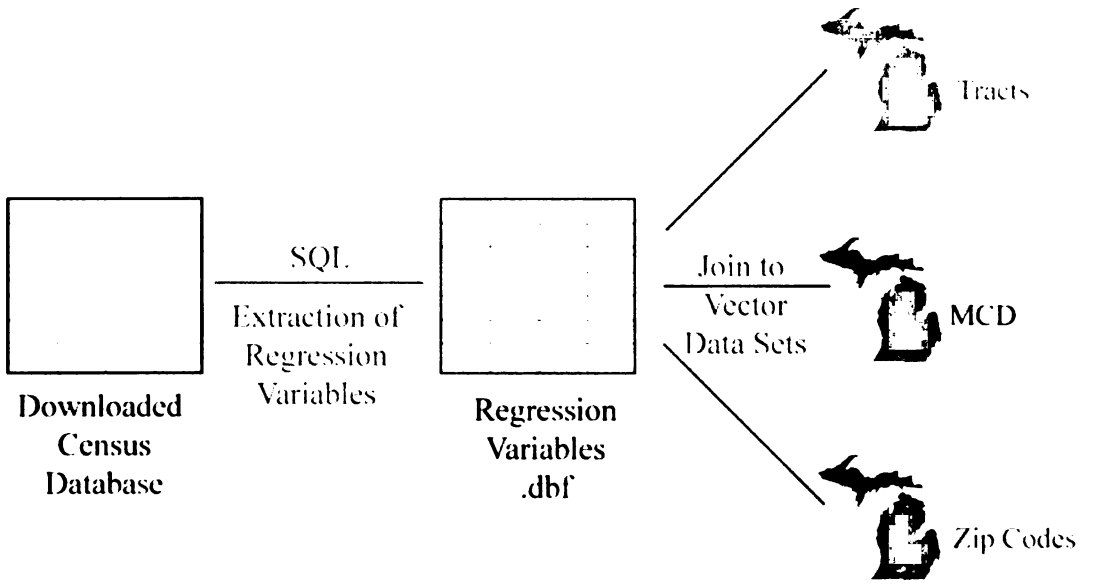


Figure 9: Schemata of the transfer of census variables to vector data sets

2.2 Methods

2.2.1 Clustering

Each child's geocoded address was used to find areas where higher BLL values cluster. Clustering techniques typically involve the division of the point dataset into cases of disease and control cases representing the population at large. With elevated BLL, the thresholds of lead representing a case of disease are vague and the current level of 10 μg/dL has been the designation only since 1991 (Sargent et al. 1995). Disease-clustering techniques seek to study point patterns in order to find areas where the likelihood of disease occurrence is greater than would be expected by chance. A variety of methods are available to study point patterns of disease. This thesis employed three methods, each of which revealed characteristics of clusters. The Cuzick-Edwards statistic reveals the occurrence and size of the clusters, the difference of K-function finds the distance between elevated lead clusters compared to the background population, and the Geographic Analysis Machine creates a visualization of the point pattern (Waller and Gotway 2004; Wheeler 2007; Dockerty, Sharples, and Borman 1999; Dolk et al. 1998; Openshaw et al. 1988).

This thesis sought to test the clustering of "cases" of lead poisoning at several levels of μg/dL. The control points were children with a BLL test result are 1 μg/dL, the lowest value in the database. These children represent a majority of the results and provide a background population representing the spatial distribution of children on Medicaid within the state. Several aspects of lead clustering were investigated, such as the number of cases near each other, distances at which cases cluster, and where these

clusters tend to occur. The linkage between these methods is that they are a different display of the underlying pattern. The neighbor method and the distance method are both expressing the same pattern in a different way. Underlying each is the notion that when controlling for how the population is spread, are the cases of elevated BLL more likely to be near each other. The two methods express this nearness in different ways. The neighbor method says are these cases likely to be neighbors compared to the background population, while the distance method analyzes whether these cases are closer to each other in distance compared to the background population. The link between the two clustering significance tests and the mapping the clusters is not perfect. Questions can arise as to whether any clusters that appear in the neighbor and distance methods are displayed in the map. But mapping is necessary to give clustering analysis any practical purpose. Without knowing the location of clusters of elevated BLL, the exercise of testing for clustering is academic. The distance based clustering tests sketch a rough outline of how large the diameter of the cluster is. More often than not, clear clusters present in the test methods show up at roughly the same size on the maps.

The decision was made to look at possible clustering by individual year rather than aggregating all or several years results together. There were two main reasons for this decision. The first was to see if patterns of clustering or changes in the size of the clusters changed over time. Differences between different years could reflect possible effects of on the ground efforts for testing programs and remediation. The second reason was a matter of computing time. The software required to perform the clustering analysis cannot support a distance matrix of test results for all eight years in many parts of the state.

The tens of thousands of data points for each year in the blood lead database required that the Michigan study area be subdivided into sections for the clustering analysis. This was carried out for a couple reasons. The first was computer processing time. The amount of data points created distance matrices too large to process in a timely manner or at all. Another is the difference in scale between a cluster in an urban area and a cluster in a rural area. In more urban areas, data points are close together, often within a few yards of each other. The rural areas of the state could have several miles between data points within the database.

The state was divided up initially by Health Systems Agencies (HSA). These were areas defined in the 1970s for health care planning in Michigan (Finn 2007). The boundaries followed county lines and divided the state into eight zones. Two of these zones were too large to run the GAM analysis with the hardware available, so they were divided into two. The Upper Peninsula HSA was divided into two pieces, an East and West, based on a gap in the location of test results. The Bay HSA was divided into two pieces based on the Shiawassee/Saginaw Rivers. Because the HSAs in southern Michigan were too large for the number of data points within them, the large urban areas were selected out by the Federal Aid Urban Boundary and analyzed separately. The federal urban aid areas selected were Detroit, Flint, Saginaw/Bay City, Lansing, Battle Creek, Grand Rapids, and Kalamazoo. The Detroit study region still had too many data points for analysis, and was divided into North and South Detroit based on the Wayne County border with Oakland and Macomb counties. In all, the state was divided into 19 sections (Figure 10) each of which, with the exception of South Detroit, had between

1,000 and 4,000 data points. The South Detroit study area had a yearly data point value typically 18,000 to 24,000.

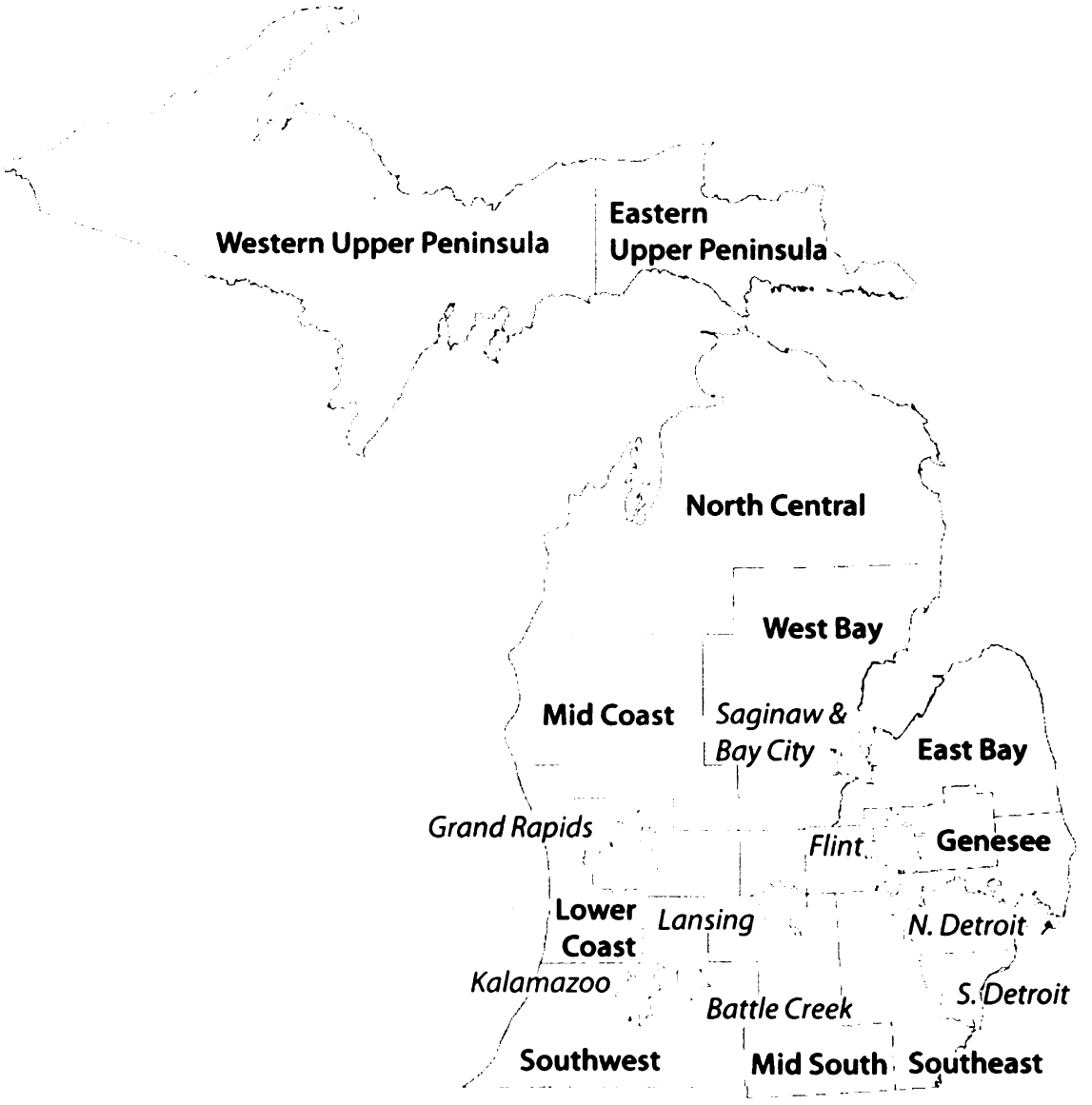


Figure 10: Study areas identified for the clustering techniques. Areas based on HSA boundaries are outlined with black and labeled in bold, while areas based on urban boundaries are outlined in blue and labeled in italics

Nearest neighbor statistics look at where disease cases are located in relation to other nearby cases as well as the general population. In terms of this thesis, the nearest neighbor for each child is the nearest other child in the database. This is determined by

radial distance between the two residences. A popular statistic called the Cuzick-Edwards k-nearest neighbor statistic uses nearest neighbor statistics to estimate the vicinity of disease cases to each other (Waller and Gotway 2004). The basic premise of the statistic is to count every instance where the nearest neighbor to a case is another case. The case-case count can be expanded to several nearest residences. The k-nearest neighbors equation is written as:

$$T_k = \sum_i \sum_j m_i m_j a_{ij}$$

Equation 1: Cuzick-Edwards test statistic

where k is the number of nearest neighbors allowed for each case, m_i is the child in question, m_j is the every other child, and a_{ij} is an indicator variable equal to one when i and j are k nearest neighbors (Waller and Gotway 2004). If i and j are cases, then m_i and m_j equals one. All three variables have to equal one to add to the final result. An example is shown below in figure 11where there are four instances where the nearest neighbor to a case was another case.

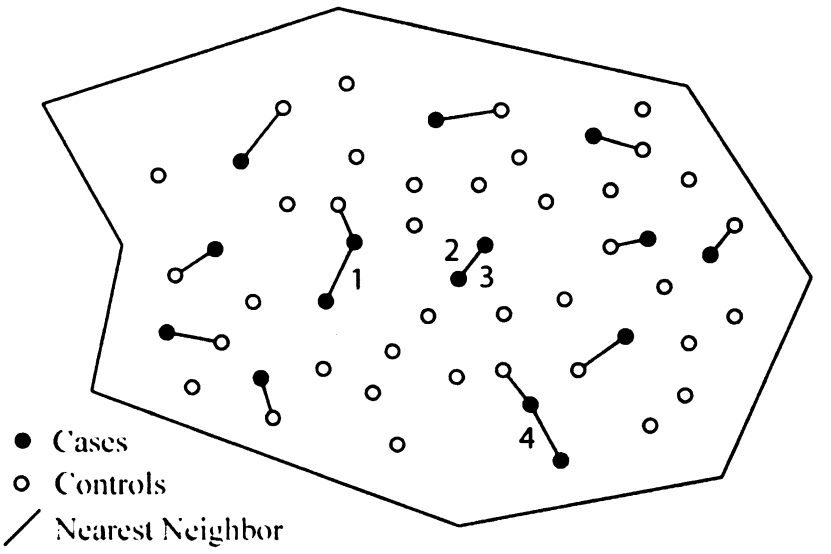


Figure 11: Example of Cuzick-Edwards statistic based on one nearest neighbor

A random labeling hypothesis can be used to test the significance of the k-nearest neighbor result (Wheeler 2007; Waller and Gotway 2004). Each child's residence is randomly labeled as a case or control in the same proportion as the actual data. The results of the random simulations form a normal distribution of test statistics and where the rank of the actual test result falls permits the calculation of a p-value. Many k values of nearest neighbors are used to find if clusters occur in small (one or two neighbors) or large groups (ten or above). The Bonferroni adjustment p-value is used to test clustering across all k values by multiplying the number of tests by the minimum p-value (Wheeler 2007).

The Cuzick-Edwards statistic has been used for both environmental and animal-borne diseases. Dockerty (1999) used the statistic to study clustering of childhood leukemia and lymphoma in New Zealand. The results showed no significant clustering in any age group or nearest neighbor value (Dockerty, Sharples, and Borman 1999). Wheeler (2007), who studied childhood leukemia in Ohio, looked at possible clustering

of leukemia cases versus the background child population in the state. He found no significant clustering at any level of k, meaning that there is no evidence that childhood leukemia cases are geographically dependent (Wheeler 2007).

The software program Clusterseer™ was used to conduct the Cuzick-Edwards statistic tests. Clusterseer is a computer package designed to study spatial and temporal clusters of disease (Wheeler 2007). Case/control boundaries of 5, 10, and 25 μg/dL were tested. The statistic was calculated for k values of 1 through 20. To determine if the Cuzick-Edwards statistics were significant, 999 Monte Carlo simulations were run.

The main drawback of nearest-neighbor statistics is that they do not take distance into account. The nearest neighbor to an event may be far away and therefore less likely to be related. The difference of K-functions seeks to find at what distances cases of disease cluster (Waller and Gotway 2004). The statistic is based on Ripley's K, a common point pattern analysis tool. The Ripley's K function is often used in health studies to find spatial dependence between individual points at different spatial scales. The basic formula for Ripley's K is:

$$\widehat{K}(h) = \frac{R}{n^2} \sum_{i=1}^{n} \sum_{j=1, i \neq j}^{n} \frac{I_h(d_{ij})}{w_{ij}}$$

Equation 2: Equation for Ripley's K

where R is the region of interest with *n* number of cases. On the right side of the equation, d_{ij} is the distance between point *i* and the surrounding point j and I_h is an indicator variable equal to 1 if *j* is within distance h of *i*, otherwise it equals zero

(McKnight 2006). W_{ij} refers to the proportion of the circle around point i which falls within the study area (Waller and Gotway 2004). Ripley's K works by placing a series of concentric circles of increasing radii around each disease event and counting events within that circle. If the number of disease events within the circle is greater than what would be expected based on the number of total events and the size of the study area, that spatial scale is considered clustered. An example of the Ripley's K can be seen in figure 12.

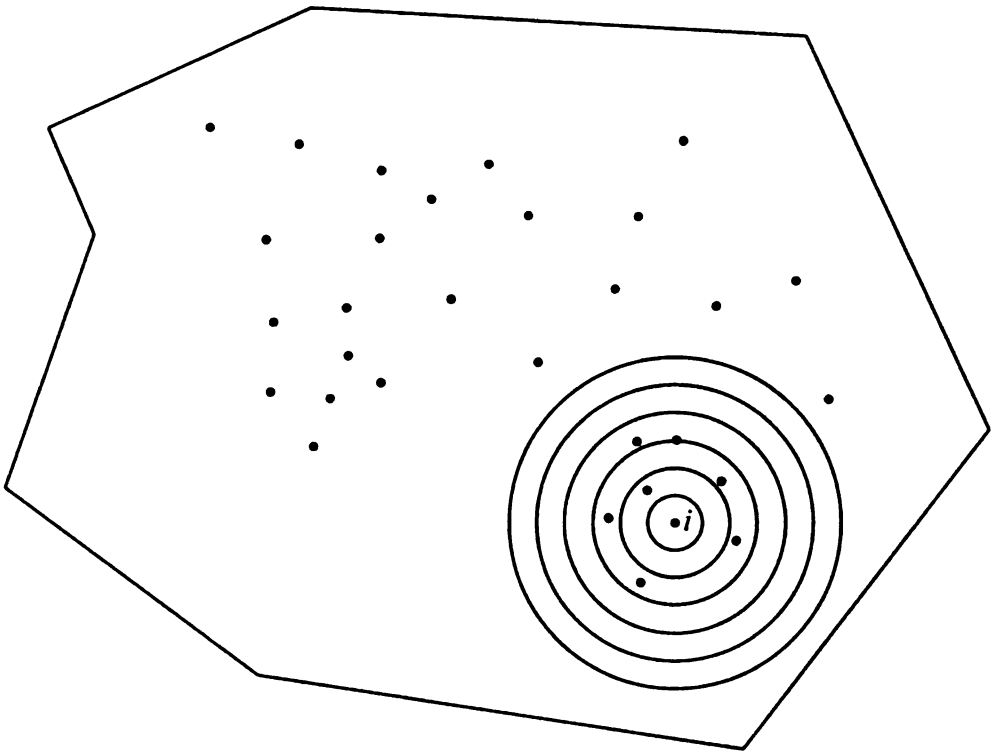


Figure 12: Ripley's K function with circles of distance h around event *i*. Clustering of events are present within four circles around event *i*.

The Ripley's K results are typically compared on a graph with complete spatial randomness patterns in order to find significant clustering or inhibition at different spatial scales. With the childhood BLL data, it is not assumed that the underlying distribution of children is spatially random because a majority of the population of Michigan lives in

metropolitan areas. The clusters of urban settlements within the state make the Ripley's K comparison against spatial randomness useless. Therefore, the distribution of elevated BLL cases must be compared against the background pattern of settlement within Michigan in order to tell if the results are noteworthy. The difference of K-function takes care of this by taking the difference between the K results of the primary pattern of cases and the secondary pattern of controls.

$$KD(h) = K_{cases}(h) - K_{controls}(h)$$

Equation 3: Difference of K

The control pattern is assumed to represent the underlying population from which the cases of disease are picked. The difference of K functions can reveal spatial scales where disease cases tend to cluster more than the population from which they are drawn. If the difference between the two K-functions is zero, the cases of disease are random within the background population. With a positive difference between the K-functions, the cases are clustered together at that spatial scale, while a negative difference indicates dispersion of the cases. A random labeling simulation can be used to test for significance (Waller and Gotway 2004). Each point within the dataset is randomly assigned as a case or control based on the proportion of each label in the original dataset. The simulation results form a normal distribution at each distance, which can be used to create an envelope of results. The true difference of K results can be compared to this envelope to determine significance.

Difference of K analyses has been used in geographical studies in both the human health and veterinary fields. Dolk et al (1998) used the difference of K function to look at congenital diseases related to pesticide use. Difference of K functions showed a lack

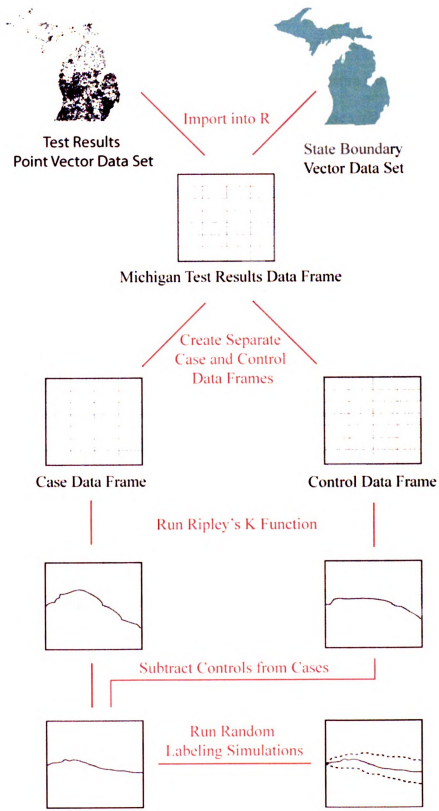
of localized clustering in cases, leading the authors to conclude that there is little geographic variation (Dolk et al. 1998). Another study that looked at biologically similar cancers in dogs and humans in Michigan showed a strong dependence between dog and human cancer, indicating that for certain types of cancer one may be used as a proxy for the other (O'Brien et al. 2000). Foley (2001) also looked at dogs and the spatial distribution of a certain tick-borne disease. Results showed that the dogs with the disease where significantly more spatially clustered than the dog population at large (Foley, Foley, and Madigan 2001). Finally, Prince et al (2001) studied a liver disease with unknown environmental risks using the difference of K method. A high amount of clustering was found at nearly all distances, leading the researchers to conclude that there was a strong link between the disease and local environmental conditions (Prince et al. 2001).

The difference of K functions analysis was performed in R, which is "an integrated suite of software facilities for data manipulation, calculation, and graphical display (Venables and Smith 2008)." This software is open source, command line-based, and utilizes the S computer language. Individual library packages can be uploaded into the program in order to provide statistical functions within the R framework. Three packages were used: splancs, spatstat, and maptools. Splancs and spatstat are packages designed for spatial point pattern analysis, and maptools is a package for working with geographical data and can handle the importation of vector data sets.

Using the maptools package, each yearly lead test results point data set was imported into R. A vector data set representing the state boundary was also imported. The points data are then converted into a data frame to create separate point features for

the cases and controls. Similar to the Cuzick-Edwards test, the case control thresholds of 5, 10, and 25 μg/dL were used. Once the case and control point features were created, the Ripley's K values were computed on each feature using the khat function in the package splancs. The distances specified for the concentric circles ranged from 0.5 kilometers to 10 kilometers, with increments of half a kilometer. These distances were selected with a mind to strike a balance between urban and rural study areas. The output of this function is a graph showing how the Ripley's K value changes with distance. For each year and case/control threshold, the control K values were subtracted from the case K values. Finally, to test for the significance of the difference of K values, the splancs function Kenv.label was used to generate difference of K values from random labeling simulations. The final result was a simulation envelope of the maximum and minimum simulation produced K values for comparison with the actual difference of K (Figure 13).

Figure 13: Method for obtaining difference of K values for each year at case/control thresholds of 5, 10, and 25 µg/dL.

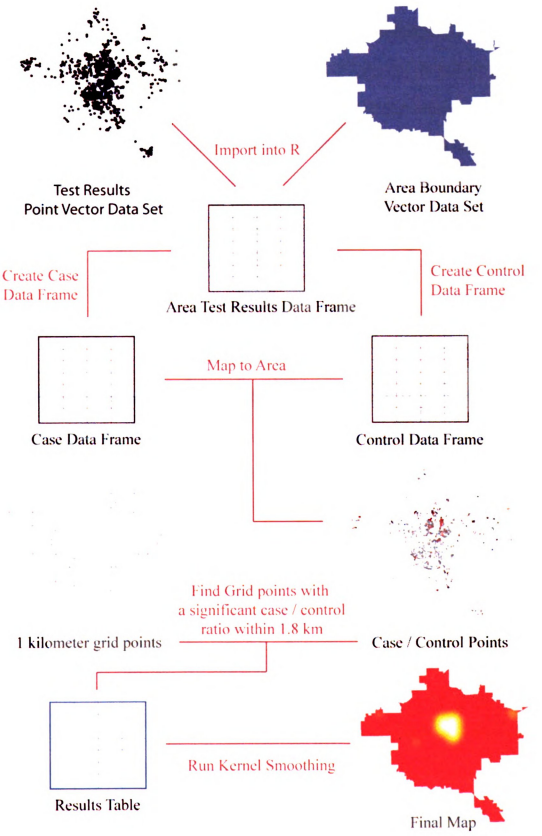

Test Results
Point Vector Data Set
Import into R
State Boundary
Vector Data Set
Michigan Test Results Data Frame
Create Separate
Case and Control
Data Frames
Case Data Frame
Control Data Frame
Run Ripley's K Function
Subtract Controls from Cases
Run Random
Labeling Simulations

Geographic Analysis Machine (GAM) is a technique created by Stan Openshaw at the University of Leeds in 1987 to study childhood leukemia clusters (Openshaw et al. 1988). It is a computationally expensive, but well used, exploratory analysis technique. The method begins with overlaying down a fine mesh grid over an entire study area. Each mesh point of the grid is the center point of a series of concentric circles that overlap each other (Openshaw et al. 1988). The GAM algorithm counts the number of cases and controls within the circle and determines significance either through a random labeling simulation or a Poisson distribution (Waller and Gotway 2004). In a random labeling simulation, if the observed value of disease counts within the circle is higher than the results from random labeling, the circle is drawn on a map. The Poisson test involves using the percentage of cases to total points as the mean of the distribution. The probability of observing the number of observed cases in each circle is calculated, and circles above a significance threshold are retained for the map. The final map usually features many overlapping circles of varying sizes. To make the pattern easier to interpret, a kernel-smoothing technique can be used. The final result of this process is a map showing hotspots within the study region. These hotspots look like large, brightly colored blotches that define the area where cases of lead poisoning occur at a significantly higher rate than the background population. The usefulness of this method is that by converting the point pattern into an area-based hotspot map, the pattern of elevated BLL can be cataloged and interpreted with easier comparison to the geographic unit based maps in regression analysis.

As with the difference of K function, GAM was run in R (Figure 14). The analysis was accomplished with the R library "splancs," which contains a tool for spatial

point pattern analysis. First, the geocoded locations and Michigan boundary files were imported into R. For each case-control threshold, the background rate used is the local ratio of cases to controls across all years. To find clusters of cases, a grid of points1kilometer apart within the Michigan border was created. The distance between the grid points and the geocoded address of each child were calculated with a Euclidean distance function and placed in a distance matrix. If the percentage of cases to controls within 1.8 kilometers of a grid point was less than the 5% chance from randomness predicted by the Poisson distribution, the grid point was marked as having a significantly amount of cases. For better visibility of the resulting pattern, a kernel-smoothing process was used to create the final maps.

Figure 14: Method in R for creating GAM maps.


Import into R
Test Results
Point Vector Data Set
Area Boundary
Vector Data Set
Create Case
Data Frame
Area Test Results Data Frame
Create Control
Data Frame
Map to Area
Case Data Frame
Control Data Frame
Find Grid points with
a significant case / control
ratio within 1.8 km
1 kilometer grid points
Case / Control Points
Run Kernel Smoothing
Results Table
Final Map

2.2.2 Geographically Weighted Regression

Regression models are commonly used in medical geography in order to find explanations for the spatial patterns of disease (Nakaya et al. 2005). Global linear regression models such as Ordinary Least Squares (OLS) are popular for their ability to offer insight into the variations in the data. The basic model is:

$$Y = \beta_0 + \sum_{k=1}^{P} \beta_k X_k + \theta$$

Equation 4: OLS regression model

where Y is the dependent variable, X_k are the independent variables, B_k are the regression coefficients, and θ is the error term (Huang and Leung 2002). The regression coefficients are calculated in matrix form:

$$\beta^* = (X^T X)^{-1} X^T Y$$

where

$$X = \begin{bmatrix} 1 & X_{11} & \cdots & X_{1P} \\ 1 & X_{21} & \cdots & X_{2P} \\ \vdots & \vdots & \cdots & \vdots \\ 1 & X_{n1} & \cdots & X_{nP} \end{bmatrix},$$

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}, \beta^* = \begin{pmatrix} \beta_0^* \\ \beta_1^* \\ \vdots \\ \beta_P^* \end{pmatrix}$$

Equation 5: Matrix calculation of the OLS coefficients

The X matrix is composed of the independent variable values as well as a column of 1 values to stand in for the intercept (O'Sullivan and Unwin 2003). X^T matrix is transposed from the X matrix. The Y matrix is made up of the values of the dependent variable.

While the OLS method is extremely popular, researchers interested in the geographic dimension of regression analysis have been looking into other options. The main problem with OLS regression is that spatial homogeneity (i.e. variable coefficients are constant across space) is assumed to be valid. This runs counter to much research within the social sciences which observes that most social processes are not stationary (Fotheringham, Brunsdon, and Charlton 2002). In global regression models, space can

only be explored through the residuals of each observation, but the variable within the model responsible for the error remains unclear. The spatial pattern of the residuals can reveal spatial autocorrelation, meaning the errors are not independent and the model systematically fails across space.

With the static nature of global regression illustrated, new methods have been devised to bring geographic location into regression modeling. Some methods, such as spatial lag or spatial error models, keep the global framework and bring geography into the equation as another independent variable. A new method that is becoming increasingly popular is Geographically Weighted Regression (GWR). The roots of GWR lie in the growing field of local spatial statistics (Fotheringham, Brunsdon, and Charlton 2002). It is based on the idea that each location is unique, and different processes occur in different areas (Shearmur et al. 2007). GWR breaks down global regression so the changes in model coefficients and predictive power can be analyzed for each geographic unit. Coefficients for each location are estimated by a weighted least squares regression equation (Leung, Mei, and Zhang 2000). The basic equation is:

$$Y_i = \beta_0(u_i, v_i) + \sum_{k=1}^{P} \beta_k(u_i, v_i) X_{ik} + \theta_i$$

Equation 6: Geographically Weighted Regression model

where *i* is the geographic unit and u_i and v_i are the coordinates. The matrix calculation of GWR is similar to OLS except that a diagonal weight matrix is included.

$$\beta^*_{(i)} = \left(X^T W_{(i)} X\right)^{-1} X^T W_{(i)} Y$$

where

$$W_{(i)} = \begin{bmatrix} W_{i1} & 0 & \cdots & 0 \\ 0 & W_{i2} & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & W_{iN} \end{bmatrix}$$

Equation 7: Matrix calculation of GWR coefficients for location *i*

The diagonal matrix gives weights to each other location as they relate to location *i*. GWR has several different weighting functions, all of which are based on the geographic axiom that nearby locations exert more influence than distant locations (Fotheringham, Brunsdon, and Charlton 2002). The most commonly used weighting function is fixed distance and based on a Gaussian curve:

$$W_{ij} = e^{\left(-\beta d_{ij}^2\right)}$$

Equation 8: Fixed weighting scheme based on Gaussian curve

where d_{ij} is the distance from location *i* to location j and β is the bandwidth of the Gaussian curve (Huang and Leung 2002). For polygon features, the distance is measured between the centroids of the area features. This weighting scheme has the same fixed bandwidth for each observation point *i*. As the bandwidth increases, the weights of a location at any distance decreases. The choice of bandwidth can be arbitrary, but a

common method of selecting the bandwidth is to minimize the residual sum of squares for all data points:

$$\sum_{i=1}^{N} [Y_i - Y^*_{\neq i}(\beta)]^2$$

Equation 9: Sum of squares method to determine the bandwidth

where $Y^*(\beta)$ is the fitted value of Y when the bandwidth β is used. The bandwidth that produces the lowest sum of squares is used in the GWR weighting function. The location *i* is not included in the function because it will overpower all other observations if the bandwidth is small, the estimates will fluctuate wildly and be of little value (Fotheringham, Brunsdon, and Charlton 2002). GWR can use an adaptive bandwidth, where the size of the bandwidth of the Gaussian weighting curve at point *i* depends in part on the density of data points within the nearby area. This method is useful is study regions where the density of data points varies across space (Fotheringham, Brunsdon, and Charlton 2002). This thesis chose to use the fixed bandwidth exclusively after the final results showed no difference between the two.

The biggest advantage of GWR is that it can model spatial non-stationarity, which is important when using a large and diverse study area such as the entire state of Michigan (Shearmur et al. 2007). Localized parameters allow visualization of how well each variable and the whole model work across space. Another advantage of GWR is that the results can be visualized through the use of GIS. Unlike the parameters of OLS regression that focus on similarity throughout the study, the results of GWR can only be

easily understood through the use of maps (Fotheringham, Brunsdon, and Charlton 2002). GWR is less prone, though not immune, to spatial autocorrelation in the residuals.

Leung et al (2000) developed a test statistic, similar to the F-test, which reveals if the GWR model works better than the global model. It uses the F-distribution to compare the residual sum of squares from the local GWR model to the global OLS model. The formula is:

$$F_1 = \frac{RSS_g/\delta_1}{RSS_o/(n-P-1)}$$

Equation 10: Leung test statistic

where RSS_g is the residual sum of squares for the geographically weighted regression model, δ_1 is the degrees of freedom in the GWR model, RSSo is the residual sum of squares in the OLS model, and (n – p – 1) is the degrees of freedom in the OLS model.

Ten US Census variables selected from tables 2 and 3 were used to create a GWR model to explain the variation in elevated BLL. Each variable used had been identified as a predictor of lead poisoning in a previous study:

1. Percentage pre-1940 housing - This variable is a measure of housing units within a geographic area that were built before 1940. It has been used before because housing built in that time period would certainly have originally had lead paint(Haley and Talbot 2004).
2. Percentage African-American – The number of African-American residents within a geographic unit has often been used as a predictor because minority

communities have historically suffered from lead poisoning to the greatest extent (Griffith et al. 1998).

3. Percentage Latino – Similar to African-Americans, Latino residents have been found to suffer from excess lead poisoning (Lanphear et al. 1998b).
4. Percentage recent immigrants – Immigrants to the United States may suffer from lead poisoning due to exposure in their country of origin or from imported products or cultural practices (Sargent et al. 1997).
5. Percentage under six years of age – If there is a greater pool of children available, the chance of childhood lead exposure increases.
6. Percentage of rental housing – Children who live in rental housing are often at higher risk of lead poisoning due to lack of disclosure and neglect from the landlord.
7. Percentage of houses headed by a female – Single parent households are often an indicator of lower socio-economic status, thought to be a leading indicator of lead poisoning (Sargent et al. 1995).
8. Percentage vacant housing – Areas with many housing units lying vacant are thought to show signs of age and neglect (Bailey et al. 1994).
9. Percentage of residents without a high school diploma – Education attainment is thought to be significant because it is an indicator of socio-economic status (Talbot, Forand, and Haley 1998).
10. Percentage below 185% of the poverty line – Lower income is believed to correlate with lead poisoning and 185% of the poverty line covers residents in

poverty as well as those in danger of falling into poverty (Kaplowitz, Perlstadt, and Post 2007).

The first step involved taking the point datasets of the children's addresses and aggregating them to the same enumeration units as the census variables. This process began by using the intersect tool in ArcGIS to code each child's location with the appropriate census tract, MCD, and zip code of their residence. Once all of the children's test results were coded, dbf files were exported into Microsoft Access. An SQL query was used to compile the dependent variable, mean BLL, for each census unit. The query for each year exported as a dbf file back into ArcGIS and joined to the census vector data sets to create the final enumeration units to run the analysis.

The three vector data sets containing the census data and aggregated lead data were imported into R. The function "lm", or linear model, was used to create global regression models and eliminate variables in each areal unit that were not significant. Once the significant ($\alpha = 0.05$) variables for each US census level were established, the resulting model was run on each individual year to study possible changes over time. For the GWR portion of the thesis, the R library "spgwr" was used. A Gaussian weighting scheme was used for weighting all other location values with relation to each location *i*, with the bandwidth calculated for each census unit by reducing the sum of squares. The results were exported out of R as a text file and joined with ArcMap vector data sets for visualization.

3 Results

3.1 Clustering Results

The purpose of testing for clustering of disease is to determine if pockets of cases are spatially arranged in a manner that would not have occurred from random chance. Clustering analysis in this thesis used three different techniques. The first was the Cuzick-Edwards statistic. This approach looked at the size of clusters through the relationship of cases to other nearby blood test addresses. The second technique was the difference of K method. It functioned by finding the Ripley's K value for cases of elevated BLL in a study area as well as the Ripley's K value for the background or control child population. The difference of K value is the result of subtracting the K value from the control population from the K value of the cases of elevated BLL. The final method is the Geographic Analysis Machine (GAM). This is a visualization tool used to find "hotspots" where cases of disease cluster.

Due to the size of Michigan and the enormous amount of test data, the state was divided into19 study areas for the cluster analysis. Rural areas were represented by the Hospital Service Areas (HSA). Two of these districts had to be divided into 2 pieces because the land area was too large for the GAM analysis. The Bay HSA was divided into East and West along the Saginaw/Shiawassee Rivers, while the Upper Peninsula HSA was divided along border between Luce/Mackinac and Alger/Schoolcraft Counties. One urban area was broken into two study areas in order to cut down on processing time. The Detroit Federal Urban Aid Boundary was divided in two different study regions along the Wayne County border with Oakland and Macomb Counties.

The results of the clustering analyses followed a similar pattern across different study areas. With the Cuzick-Edwards tests, the 5 µg/dL level often exhibited clustering. This was particularly true in the urban areas, but often extended to less populated parts of the state. The 10 µg/dL cutoff exhibited more variability across the state. In the larger urban areas, a high amount of clustering among cases was present. This persisted through all years in the lead database. In smaller cities, clustering of cases of 10 µg/dL and above were smaller and more common in the earlier years covered by the study. In more rural areas of the state, the low number of cases resulted in clustering being much less common. At the 25 µg/dL case level, only the large urban areas showed any signs of clustering. Other study areas typically did not have enough cases at the 25 µg/dL level.

The difference of K results generally agreed with the Cuzick-Edwards findings. In interpreting difference of K graphs, clustering is noted when the K values at any distance are above the simulation envelope of random labeling test results. At the 5 µg/dL level, in urban areas the K value rises above the simulation envelopes immediately and remains above for the entire 10 kilometer distance tested. In smaller midsized city study areas, the K values sometimes drop back down to zero at greater distances due to the edge effects caused by the small study area size. In the larger HSA study areas, results are mixed depending on if there is a central city within the study area. Clustering is only present at the 25 µg/dL level in the largest cities.

The GAM maps were used in this thesis to determine the spatial location of clusters of elevated BLL cases. Rather than being a significance test of clustering, GAM is a visualization technique that finds hotspots of likely clustering. In urban areas with many test cases, GAM provided good results of where the hot spots of elevated BLL

cases were located. GAM worked fairly well is areas where there were strong clusters consistently through time. This method did not work as well in the rural areas. Since significance values were locally based, one elevated BLL case could be considered a cluster in a rural area because of the lack of cases overall.

This section of results covering clustering techniques is presented by individual study area. Key points and diagrams are shown. Tables are used to display the Cuzick-Edwards results. Years that have a significant ($\alpha = 0.05$) Bonferroni p-value for all k levels are highlighted in orange. The numbers under each k value is the Cuzick-Edwards value, or the amount of neighbor connections at that level. Cuzick-Edwards test statistics that are significantly higher than the previous k level values are highlighted in orange. For the difference of K and GAM analysis, figures of individual years were chosen which best represented the overall pattern in the study area. The code used to create the graphs and maps is available in Appendices 2 and 3. In this section, the 5 µg/dL threshold refers to the tests where 5 µg/dL was the cutoff between cases of elevated BLL and the control population of unaffected children. This phrasing is repeated for 10 and 25 µg/dL.

3.1.1 South Detroit

The region of South Detroit in this thesis represents the Detroit Federal Urban Aid Boundary area south of the northern boundary of Wayne County (Figure 15). This area includes the cities of Detroit, Dearborn, Grosse Pointe, and others in Wayne County. It is the most heavily populated area of the state and seems to have the most robust testing for lead in children. The number of blood tests performed in this region, 15 to 20 thousand each year, was at least three times higher than any other part of the state.

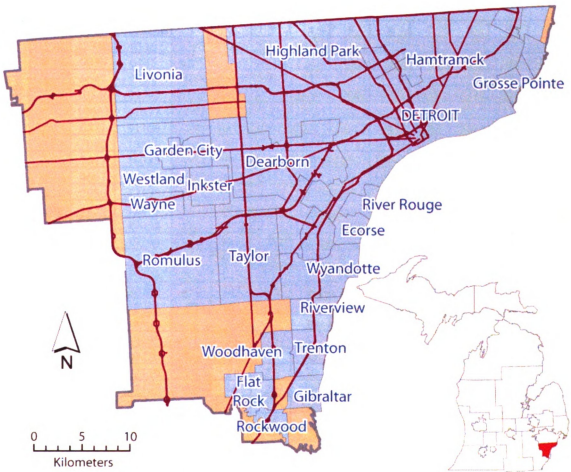


Figure 15: Map of the South Detroit study region

The Cuzick-Edwards results reveal high levels of clustering across all years and threshold levels (Table 5). At the 5 and 10 µg/dL threshold levels, Monte Carlo tests reveal that total number of case-case nearest neighbors to be highly significant for every k value. South Detroit was also the only area of the state that had a large amount of children with BLL at or above 25 µg/dL. The South Detroit study area is the only region of the state where the Bonferroni p-value, an indication of clustering across all k values, is significant at all of years in the database for the 25 µg/dL threshold.

				K									
	Year	Cases	Controls	1	2	3	4	5	6	7	8	9	10
5 Threshold	1998	12350	6136	9255	18226	27241	36169	45135	54038	62995	72023	81059	90058
	1999	9036	6145	6268	12228	18043	23921	29861	35648	41476	47311	53069	58906
	2000	6647	7706	3917	7583	11228	14758	18401	21950	25592	29209	32771	36330
	2001	10346	11592	6189	12084	17836	23490	29042	34791	40552	46334	52053	57798
	2002	10663	11677	6597	12825	18887	24849	30881	36979	43003	48999	55038	61089
	2003	6571	11166	3596	6810	10005	13133	16361	19479	22429	25514	28537	31736
	2004	7080	13866	3643	6995	10218	13425	16723	19972	23252	26483	29712	33019
	2005	7615	16385	3677	6996	10230	13404	16523	19661	22799	25957	29119	32354
10 Threshold	1998	4667	13819	1973	3608	5242	6893	8487	10107	11714	13277	14865	16486
	1999	2986	12195	1123	1983	2796	3606	4424	5198	6005	6750	7514	8264
	2000	1854	12499	547	972	1350	1738	2149	2533	2935	3288	3676	4117
	2001	2880	19058	884	1573	2184	2768	3348	3933	4521	5141	5743	6333
	2002	2941	19399	956	1666	2281	2901	3547	4178	4796	5421	6028	6664
	2003	1586	16151	417	696	937	1162	1411	1625	1841	2051	2289	2542
	2004	1523	19423	355	613	807	1024	1256	1497	1728	1939	2160	2361
	2005	1791	22209	450	735	971	1212	1471	1741	1986	2239	2477	2701
25 Threshold	1998	391	18095	70	105	138	160	177	198	210	227	243	257
	1999	231	14950	20	29	32	36	46	46	51	58	63	66
	2000	158	14195	21	34	39	47	54	60	65	70	74	81
	2001	212	21726	20	33	37	45	48	54	57	62	63	66
	2002	221	22119	18	21	25	30	34	38	41	44	48	51
	2003	165	17572	13	16	20	20	24	28	32	40	49	53
	2004	142	20804	11	17	17	21	26	31	32	34	37	37
	2005	142	23858	13	21	24	24	27	35	39	40	40	41

Table 5: Cuzick-Edwards results for South Detroit

The difference of K graphs for the South Detroit region show a very high degree of spatial clustering of elevated BLL cases. The K values for each threshold level continue to rise even as the distance increases. This is unlike any other region of the state, and would seem to confirm that the spatial clusters of elevated BLL are quite large. Because the K values fall well above the simulation envelopes created from random labeling tests, the degree of clustering is significant. This can be seen in figure 16. The second graph in figure 16 shows the difference of K values rise as high as 18 times as high as the upper bound of the simulation envelope. There is no other study region where the difference of K values rise immediately and continue to rise all the way to ten

kilometers. Since this occurs at all threshold levels, it is safe to say that this study region has the largest cluster of lead poisoning victims in the state.

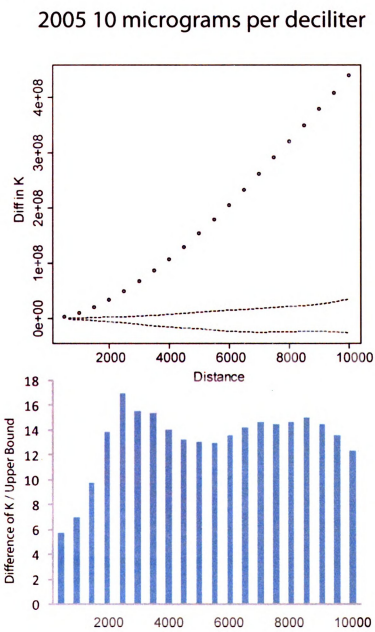


Figure 16: The 2005 South Detroit difference of K graph for the 10 µg/dL threshold

The GAM analysis reveals the spatial location of the clusters of elevated BLL to be squarely within the city of Detroit. The level of intensity of the hotspots fades in later years of the database, but generally falls within the same areas of the city. Figure 17

reveals the two main hotspots that showed up at all threshold levels. These two regions are located to the east and the west of the downtown Detroit area. The western hotspot extends towards the boundary with Dearborn and the eastern hotspot occupies the eastern part of the city of Detroit.

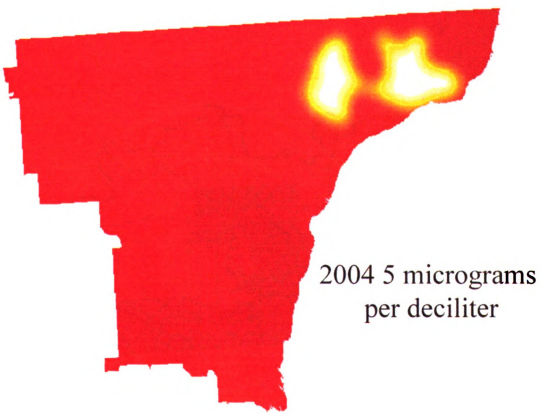


Figure 17: The 2004 GAM map of South Detroit for the 5 μg/dL threshold

3.1.2 North Detroit

North Detroit covers the area of the Detroit Federal Urban Aid Boundary area that falls within Oakland or Macomb Counties (Figure 18). The region contains many suburbs of Detroit and covers a mostly developed landscape. This includes cities such as Pontiac, Warren, St. Clair Shores, Novi, and others. The Detroit Federal Urban Aid

Boundary was divided along the county line due to the large differences in the number of test results between North Detroit and South Detroit. North Detroit has far fewer test results, 2 to 7 thousand per year, than South Detroit.

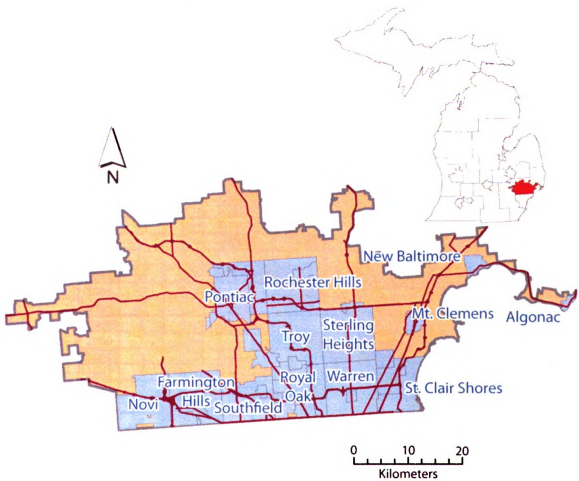


Figure 18: Map of the North Detroit study region

The Cuzick-Edwards results for North Detroit reveal a strong clustering pattern at lower threshold levels and very little clustering at higher threshold levels. At the5 µg/dL threshold level, the total case-case neighbors run far ahead of the number expected at every level of neighborhood. This pattern is consistent across all years (Table 6). There is overall clustering at 10 µg/dL threshold, but the clusters grow very slowly after the k = 3 level. This suggests that the clusters of cases within North Detroit are smaller than

what was seen in South Detroit. At the very high 25 µg/dL threshold, the low number of cases makes it difficult to find any consistency between the years. These very high cases do seem to be near each other, but it does not always constitute a cluster.

				K									
Threshold	Year	Cases	Controls	1	2	3	4	5	6	7	8	9	10
5	1998	376	1430	132	272	388	512	632	740	860	962	1089	1200
	1999	337	1889	100	192	271	335	400	475	534	621	699	792
	2000	363	2405	121	205	298	355	428	488	558	632	698	775
	2001	367	2881	84	148	212	270	340	391	434	482	549	611
	2002	415	3156	101	169	248	312	396	467	537	614	677	743
	2003	374	3900	57	105	158	211	269	329	374	422	466	504
	2004	600	6046	117	201	269	344	421	500	566	640	724	806
	2005	658	6864	147	267	360	458	557	652	748	839	926	1022
10	1998	70	1736	13	23	33	36	45	51	65	70	76	81
	1999	59	2167	9	14	16	19	20	23	25	28	31	33
	2000	58	2710	11	15	22	24	26	28	30	31	34	37
	2001	43	3205	8	12	13	15	18	20	24	26	29	30
	2002	52	3519	7	7	10	12	13	15	16	17	19	21
	2003	41	4233	1	1	3	3	4	4	6	6	7	9
	2004	59	6587	6	10	11	12	13	15	15	17	17	20
	2005	52	7470	0	1	4	4	5	5	7	7	8	9
25	1998	12	1794	6	6	6	6	6	6	6	6	6	6
	1999	8	2218	2	2	2	2	2	2	2	2	2	2
	2000	7	2761	0	0	1	1	2	2	2	2	2	2
	2001	3	3245	0	0	0	1	2	2	2	2	2	2
	2002	3	3568	0	0	0	0	0	0	0	0	0	0
	2003	2	4272	0	0	0	0	0	0	0	0	0	0
	2004	8	6638	0	0	0	0	0	0	0	0	0	0
	2005	4	7518	0	0	0	0	0	0	0	0	0	0

Table 6: Cuzick-Edwards results for North Detroit

The difference of K graphs confirms the clustering within the North Detroit region. The 5 µg/dL threshold shows the rise of the difference of K being well above the simulation envelope. At around five kilometers, the K values begin to drop off, a signal that cases are no longer being added as quickly as controls. This drop occurs in every yearly difference of K graph, and can be seen in figure 19. While the difference of K values peak at five kilometers, the second graph indicates that the fastest growth occurs less than two kilometers. At two kilometers in figure 19, the difference of K values are 9

times as high as the upper bound of the simulation envelope. The 10 µg/dL threshold patterns rise immediately and then fall below the envelope, revealing fairly small clusters. The 25 µg/dL threshold shows no degree of clustering.

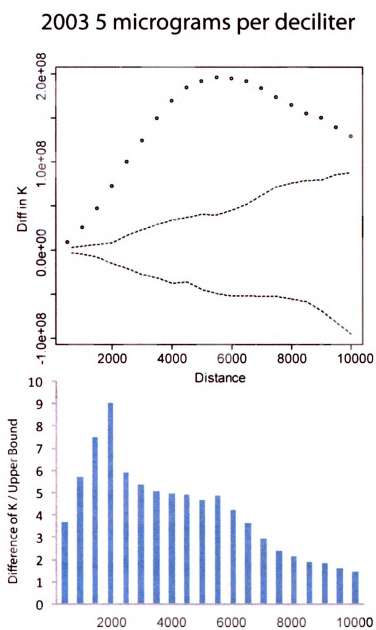


Figure 19: The 2003 North Detroit difference of K graph for the 5 µg/dL threshold

The GAM analysis of North Detroit suggests that Pontiac has the largest cluster of high BLL test results in the region. The city has visible clustering in every year for both

5 and 10 µg/dL thresholds. A secondary area of high BLL clustering is the area which borders the city of Detroit. This includes Warren, Royal Oak, and Southfield. Both of these hotspots are visible in figure 20. Unlike Pontiac, the secondary cluster near the city of Detroit disappears over time, possibly due to increased testing rates. At the very high 25 µg/dL threshold, Pontiac is the only area which consistently shows any hotspots, but the other tests make this seem like these are not very significant.

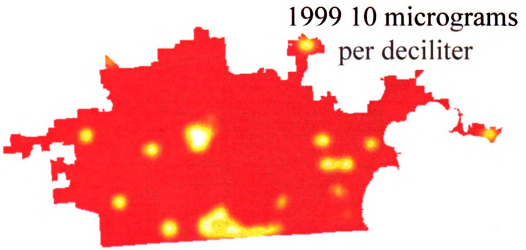


Figure 20: The 1999 GAM map of North Detroit for the 10 µg/dL threshold

3.1.3 Southeast Michigan

The Southeast Michigan region includes all of the Southeast HSA which does not fall within the Detroit urban boundary (Figure 21). While this region is mostly rural, it does have several cities mixed in with surrounding rural areas. The two Detroit study areas do take a large bite out the original HSA, but the vast gulf in the number of tests between the study areas make it reasonable to keep them separate. The three main cities of the Southeast region are Ann Arbor, Monroe, and Port Huron. For every year between

1998 and 2003, the number of blood tests is under 2,000. The number of tests doubles to around 3,500 in 2004 and increases again to nearly 4,000 in 2005.

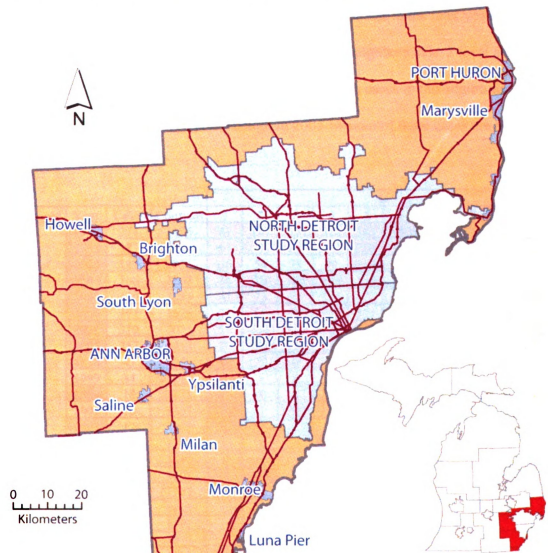


Figure 21: Map of the Southeast Michigan study region

The Cuzick Edwards results for this region show clustering through all years at the 5 μg/dL threshold (Table 7). The Bonferroni p value confirms there is clustering across all k values, but Monte Carlo analysis reveals that the clustering is strongest at k values of 5 or less. Still, many years have fairly large clusters at the 5 μg/dL threshold.

At the 10 µg/dL threshold, the clusters are smaller. The number of case-case neighbors is high at the k=1 level, indicating small pockets of elevated BLL within the region. The clustering is stronger in the earlier years, but is less prominent in the later years of the database with the exception of 2005 where there are 10 neighbors at k=2 level among the 31 cases. At the 25 µg/dL threshold, there are not enough cases in this region for a cluster analysis in nearly every year, though in 2005 two out of three cases are nearest neighbors.

				K									
Threshold	Year	Cases	Controls	1	2	3	4	5	6	7	8	9	10
5	**1998**	274	1125	81	149	226	310	383	465	534	599	668	733
	1999	183	961	67	107	162	214	266	308	366	402	456	506
	2000	142	1051	35	54	75	90	115	137	162	187	213	236
	2001	166	1207	36	70	100	131	158	181	205	225	246	272
	2002	150	1385	42	73	97	113	131	146	164	183	201	229
	2003	171	1626	22	51	78	102	121	142	165	190	215	232
	2004	310	3051	59	103	150	190	246	299	351	398	435	492
	2005	352	3453	70	132	193	249	301	354	407	467	509	567
10	**1998**	43	1356	12	20	27	35	37	41	46	46	47	47
	1999	34	1110	4	5	6	9	11	12	15	18	22	24
	2000	18	1175	7	8	8	8	8	9	10	11	11	11
	2001	16	1357	2	2	4	4	4	6	7	7	8	8
	2002	12	1523	2	2	2	2	2	2	2	2	2	2
	2003	16	1781	0	1	2	4	4	4	4	4	4	4
	2004	25	3336	0	0	0	0	1	1	1	1	1	1
	2005	31	3774	7	10	10	10	10	10	10	10	11	11
25	**1998**	2	1397	0	0	0	0	0	0	0	0	0	0
	1999	2	1142	0	0	0	0	0	0	0	0	0	0
	2000	1	1192										
	2001	1	1372										
	2002	1	1534										
	2003	1	1796										
	2004	2	3359	0	0	0	0	0	0	0	0	0	0
	2005	3	3802	2	2	2	2	2	2	2	2	2	2

Table 7: Cuzick-Edwards results for Southeast Michigan

The difference of K graphs for Southeast Michigan show that where clustering exists, it is small. Depending on year, the difference of K result may be above the upper bound of the simulation envelope at shorter distances, but the results fall back down as

the distance grows. Often the K values hug the upper bounds of the simulation envelopes like in figure 22. There is a quick rise in difference of K values, as high as 2.5 to 3 times above the upper bound of the simulation envelope, fall back down in the envelope by four kilometers. The initial jump is visible in the 5 μg/dL threshold graphs, but less so in the 10 μg/dL threshold graphs. Since the simulation envelopes change with every simulation, this low of a degree of separation means that clustering cannot be confirmed. The fact that clustering is obvious in the Cuzick-Edwards tests but not the difference of K could be a sign that it is confined to a small area that is picked up more easily by neighborhood measures than distance measures.

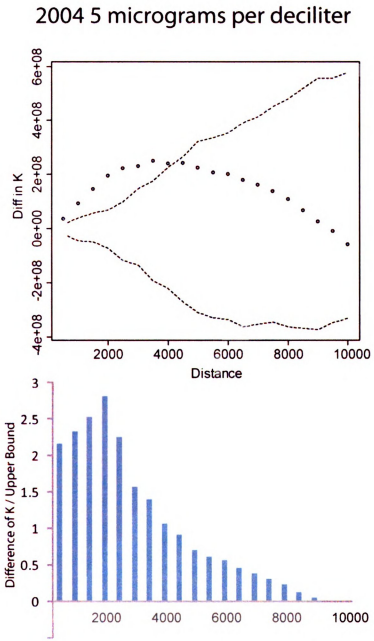


Figure 22: The 2004Southeast Michigan difference of K graph for the 5μg/dL threshold

The results of the GAM analysis show the small pockets of clusters. At the 5 μg/dL threshold, there are a large number of very small hotspots whose placement varies year to year. While it is difficult to pin down the location, Monroe County in the south has very high number of tiny clusters. Both Port Huron and Monroe are visible hot spots

across all years. Ann Arbor is a hotspot only in 1998 (Figure 23). This distinction is apparent at the 10 μg/dL threshold as well, where Ann Arbor quickly disappears as the years progress. Monroe also disappears in later years, while Port Huron remains a hot spot.

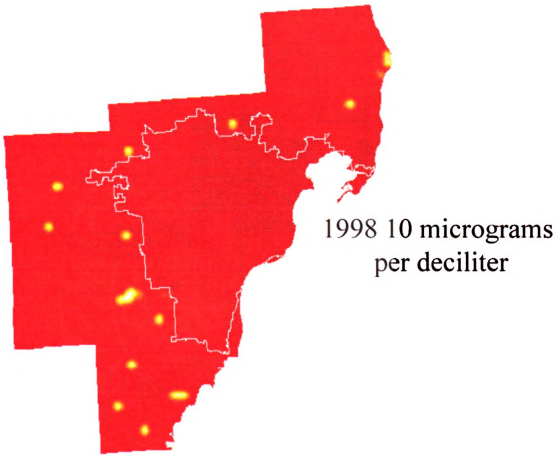


Figure 23: The 1998 GAM map of Southeast Michigan for the 10 μg/dL threshold

3.1.4 Flint

The Flint region covers the Flint Federal Urban Aid Boundary (Figure 24). It covers the city of Flint as well as surrounding cities such as Burton, Grand Blanc, and

Fenton. The region is mostly urban and developed. The number of blood tests with the Flint study area rises from under 1,000 in 1998 to over 4,000 in the year 2005.

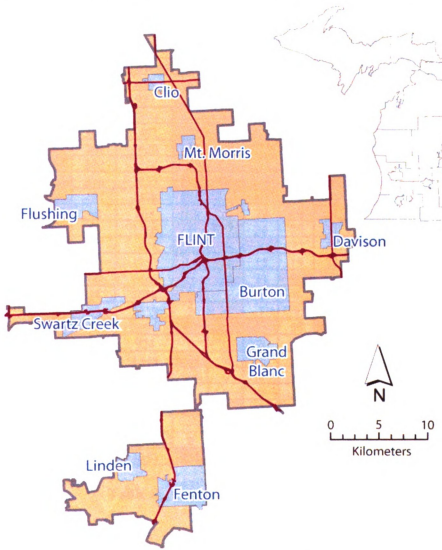


Figure 24: Map of the Flint study region

The Cuzick-Edwards results for Flint show strong clustering at both the 5 and 10 µg/dL thresholds (Table 8). For the 5 µg/dL threshold, this significance remains high even as the number of neighbors grows, indicating the larger cluster of cases. The 10 µg/dL threshold displays significant test statistic values at smaller k values, indicating tight clusters of cases. The 10 µg/dL threshold clustering is higher than similar sized

cities within Michigan, which could indicate the severity of elevated BLL in Flint. A couple years even have a significant Bonferroni p-value for the 25 µg/dL threshold due to two cases being nearest neighbors at the k = 1 level.

				K									
Threshold	Year	Cases	Controls	1	2	3	4	5	6	7	8	9	10
5	1998	298	630	141	259	390	512	625	745	877	1013	1135	1263
	1999	375	977	149	288	430	568	712	856	997	1147	1287	1431
	2000	400	1290	154	269	383	503	615	732	849	970	1100	1223
	2001	309	1492	102	183	260	335	415	490	557	630	691	759
	2002	306	1620	119	210	279	359	434	518	591	659	737	818
	2003	272	1844	70	123	175	232	282	332	392	452	505	564
	2004	278	3023	66	111	153	185	237	273	314	357	385	419
	2005	487	3578	126	255	378	488	589	683	791	894	1003	1098
10	1998	60	868	12	21	28	39	52	61	72	81	90	96
	1999	66	1286	13	22	26	29	31	33	38	51	55	59
	2000	62	1628	15	23	30	32	37	39	42	46	48	53
	2001	49	1752	11	18	18	19	22	24	24	24	26	28
	2002	60	1866	16	28	31	37	46	54	57	60	62	70
	2003	28	2088	2	2	2	4	5	8	8	8	10	11
	2004	42	3259	2	6	7	9	9	12	15	16	19	20
	2005	74	3991	17	21	31	35	42	45	47	54	55	59
25	1998	4	924	0	0	0	0	0	0	0	0	0	0
	1999	6	1346	2	2	2	2	2	2	2	2	2	2
	2000	3	1687	0	0	0	0	0	0	0	0	0	0
	2001	6	1795	0	0	0	0	0	0	0	0	0	0
	2002	7	1919	1	2	2	2	2	2	2	2	3	3
	2003	0	2116										
	2004	3	3298	0	0	0	0	0	0	0	0	0	0
	2005	8	4057	2	2	2	2	2	2	2	2	2	2

Table 8: Cuzick-Edwards results for Flint

Results from the difference of K test confirmed the presence of significant spatial clusters at the 5 and 10 µg/dL thresholds. Each level has K values above the upper bound of the simulation envelope. At the 5 µg/dL threshold, the K values rise immediately and stay above the upper bound for the entire ten kilometer distance. They do fall at large distances, but this is could be due to edge effects of the study area. With the 10µg/dL threshold, the K values rise quickly before falling below the upper bound of the simulation envelope around a distance of six or seven kilometers, as illustrated by figure

25. The K values reach a height of about 2.5 times the upper bound around four kilometers, indicating significant clustering. The 25 μg/dL threshold numbers do not indicate any significant clustering in any year.

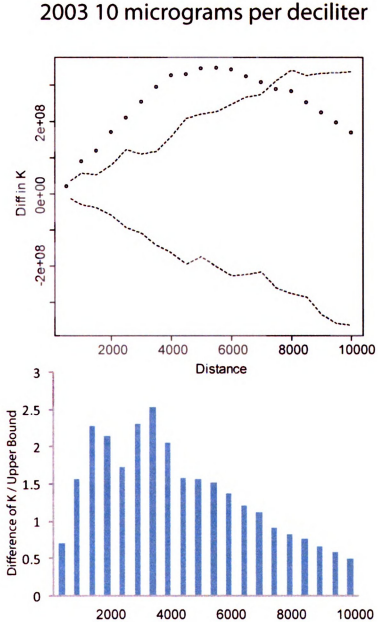


Figure 25: The 2003 Flint difference of K graph for the 10 μg/dL threshold

GAM results for the Flint study area show the clustering of elevated BLL is contained almost exclusively within the city of Flint. The worst areas in all threshold

levels tend to be the neighborhoods to the northwest of downtown and north of the Flint River (Figure 26). While the shape of the hotspot varies year to year, at each threshold level it is centered in these Northwest Flint neighborhoods. This area is likely the source of the elevated BLL clustering seen in the other tests.

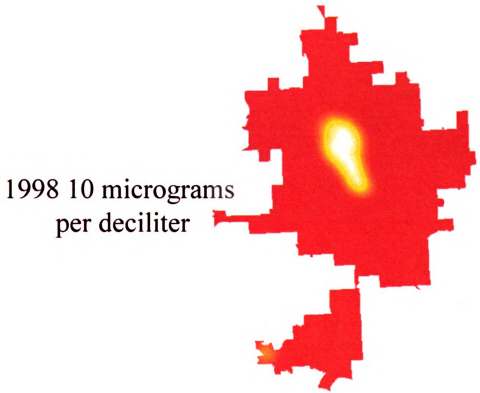


Figure 26: The 1998 GAM map of Flint for the 10 µg/dL threshold

3.1.5 Genesee

The Genesee study area includes the counties of Shiawassee, Lapeer, and all of Genesee County that is not in the Flint Urban Aid Boundary (Figure 27). It is a mostly rural study area that does not have any large cities. The main towns are Lapeer, Owosso, and Perry. The Flint study region divides the Genesee HSA in half, and the number of

blood tests in the Genesee study region is about one-third of the number of tests in the Flint study region. The total blood tests is below 500 for each of the years 1998-2003, followed by a sharp increase in 2004 to around 900 and more than 1,300 in 2005.

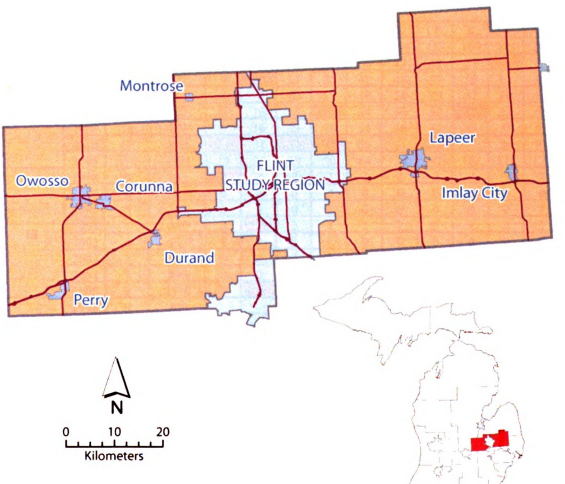


Figure 27: Map of the Genesee study region

The Cuzick-Edwards statistic tests revealed no consistent significant clustering of lead poisoning cases at any level (Table 9). At each threshold level, the number of case-case nearest neighbors does not fall far from what would be expected by chance. This is a stark contrast to the more urban areas of the state, but in line with other regions that lack a major city. The years 2004 in the 5 µg/dL threshold and 2001 in the 10 µg/dL

threshold are the only individual years that indicate clustering is present. In a nearest neighbor test such as Cuzick-Edwards, distance is not a factor. However, there is seemingly little clustering at any level.

				K									
Threshold	Year	Cases	Controls	1	2	3	4	5	6	7	8	9	10
5	1998	87	322	23	48	68	95	108	134	153	169	185	202
	1999	56	291	14	24	35	49	60	73	79	87	103	115
	2000	46	327	9	20	27	35	42	48	53	59	62	69
	2001	44	414	4	9	10	13	19	24	29	37	40	42
	2002	45	402	6	10	17	27	34	36	39	49	51	55
	2003	53	484	9	16	23	28	31	36	42	43	50	53
	2004	67	830	11	21	31	44	53	61	72	77	86	94
	2005	119	1204	15	29	39	50	64	76	91	107	115	127
10	1998	6	403	2	2	2	2	2	2	2	2	2	2
	1999	3	344	0	0	0	0	0	0	0	0	0	0
	2000	5	368	0	0	0	0	0	0	0	0	0	1
	2001	3	455	0	2	2	2	2	2	2	2	2	2
	2002	4	443	0	0	0	0	0	0	0	0	0	0
	2003	2	535	0	0	0	0	0	0	0	0	0	0
	2004	7	890	0	0	1	1	2	3	3	3	4	4
	2005	5	1318	0	1	1	1	1	2	2	2	2	2
25	1998	0	409										
	1999	2	345	0	0	0	0	0	0	0	0	0	0
	2000	1	372										
	2001	0	458										
	2002	0	447										
	2003	0	537										
	2004	0	897										
	2005	1	1322										

Table 9: Cuzick-Edwards results for Genesee

Difference of K results confirms the lack of clustering of elevated BLL. At every threshold level, the difference of K values at every distance is within the simulation envelopes. There is not a year where the K values of any of the three threshold levels rise above the upper bound of the simulation envelopes. Figure 28 shows the difference of K for 2002 at the 5 µg/dL threshold, and the K values stay around zero and fall well within the simulation envelopes. The second graph shows the difference of K values never

exceeded 60% of the upper bound of the simulation envelope, a sign that the pattern of cases does is not significantly different from the results of the random simulations.

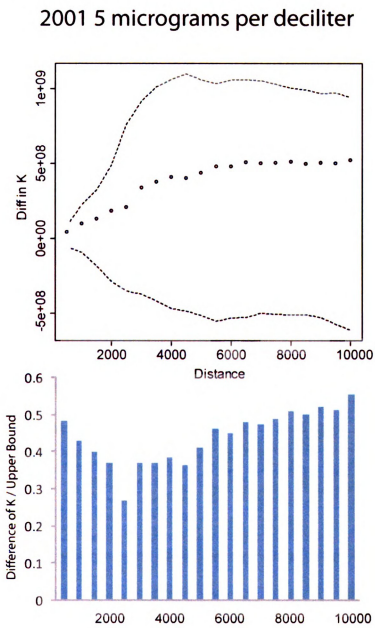


Figure 28: The 2002 Genesee difference of K graph for the 5 µg/dL threshold

Despite the lack of any small or large clusters in the study area, the GAM maps for the Genesee can be useful to show a general pattern of cases. At the 5 µg/dL threshold level, this pattern seems to be that many cases are located in Shiawassee

County around the city of Owosso. But the problem with rural areas is that without a large number of cases, individual cases show up as hotspots. Shiawassee County seems to have the most cases in the region, like in figure 29, but the hotspots change year to year without any consistency. At the 10 and 25 μg/dL thresholds, the dearth of cases makes it difficult to find any discernable pattern.

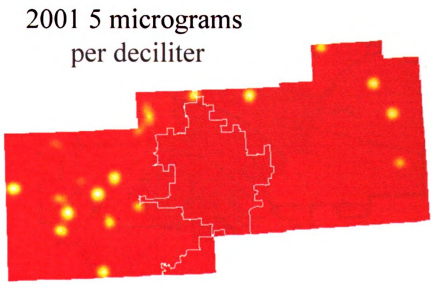


Figure 29: The 2001 GAM map of Genesee for the 5 μg/dL threshold

3.1.6 Lansing

The Lansing study area consists of the Lansing Federal Urban Aid Boundary. The study region is situated around the city of Lansing (Figure 30). Surrounding cities within this area are East Lansing, Grand Ledge, Okemos, and Mason. The area is a developed urban area. The number of yearly blood tests in the Lansing study area range from 1,300 to 1,800 in the years 1998-2004, followed by a increase to over 2,100 in 2005.

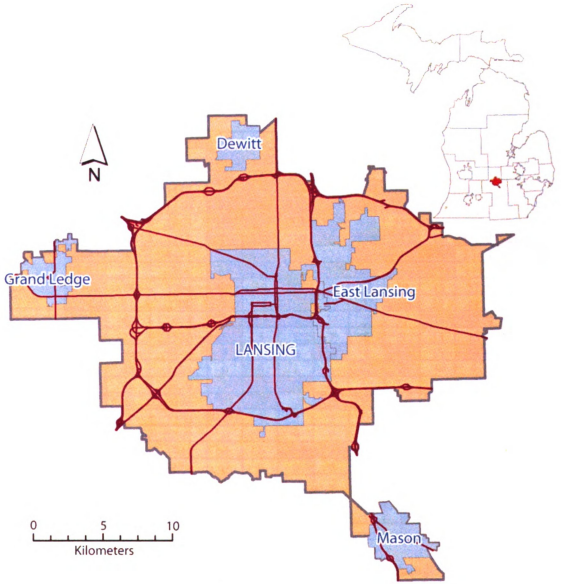


Figure 30: Map of the Lansing study region

The 5 μg/dL threshold Cuzick-Edwards statistics reveal clustering within the Lansing area (Table 10). As the k value is increased, the number of case neighbors continues to grow nearly every year. This would indicate that the clusters of elevated BLL are fairly large within the Lansing area. With the 10 μg/dL threshold, the results changed slightly. At lower k values, the significance was high, but little growth in the test statistic occurred at k values higher than 3 or 4. Still, nearly every year had

significant clustering at the 10 µg/dL threshold according to the Bonferroni p-value. Since this continues through all years within the database, it likely indicates a sustained risk exposure. The 25 µg/dL threshold indicated no clustering except in the year 2000.

				K									
Threshold	Year	Cases	Controls	1	2	3	4	5	6	7	8	9	10
5	1998	396	942	173	313	455	600	723	864	1010	1150	1265	1409
	1999	417	1434	161	298	427	532	632	747	861	966	1081	1206
	2000	355	1110	139	243	349	452	555	682	817	938	1049	1161
	2001	313	1310	75	159	238	313	394	482	568	637	723	799
	2002	332	1453	108	214	291	379	446	512	604	704	795	892
	2003	274	1169	90	168	230	304	375	441	503	572	654	743
	2004	284	1411	75	143	212	271	344	414	476	534	603	669
	2005	278	1910	69	148	204	264	314	366	429	471	524	579
10	1998	74	1264	15	19	27	29	35	40	43	51	57	63
	1999	61	1790	9	14	23	25	27	33	37	41	42	47
	2000	46	1419	5	12	22	26	29	31	39	47	50	52
	2001	35	1588	5	5	7	7	9	9	11	13	17	19
	2002	39	1746	3	4	4	8	8	13	13	17	19	21
	2003	35	1408	0	0	1	1	4	8	9	9	11	13
	2004	46	1649	7	12	16	17	17	19	20	22	25	27
	2005	42	2146	10	16	18	23	30	31	31	33	33	33
25	1998	8	1330	2	2	2	2	3	3	4	4	4	4
	1999	2	1849	0	0	0	0	0	0	0	0	0	0
	2000	4	1461	0	0	1	2	2	3	5	6	6	6
	2001	5	1618	0	0	0	0	0	0	0	0	0	0
	2002	4	1781	0	0	0	0	0	0	0	0	0	0
	2003	0	1443										
	2004	3	1692	0	0	0	0	0	0	0	0	0	0
	2005	4	2184	0	0	0	0	0	0	0	0	0	0

Table 10: Cuzick-Edwards results for Lansing

The difference of K values in the Lansing study area are surprisingly inconsistent. At the 5 µg/dL threshold, the K value each year rises quickly at short distance and falls beyond six kilometers. The results are surprisingly inconsistent, with a couple years exhibiting significant clustering while other years do not. The trend seems to be that the amount of clustering dissipates over time, suggesting that the cluster might weaken. Another interesting fact is that 10 µg/dL threshold graphs show clustering across all years. The graphs all show an early rise in the K values at short distances, then fall below

the upper bound of the simulation envelopes like in figure 31. The peak around four kilometers in the difference of K graph coincides with the K values being 3 times as large as the upper bound of the simulation envelope, making four kilometers the likely diameter of the cluster. At the 25 μg/dL threshold, the k values never fall outside the simulation envelopes.

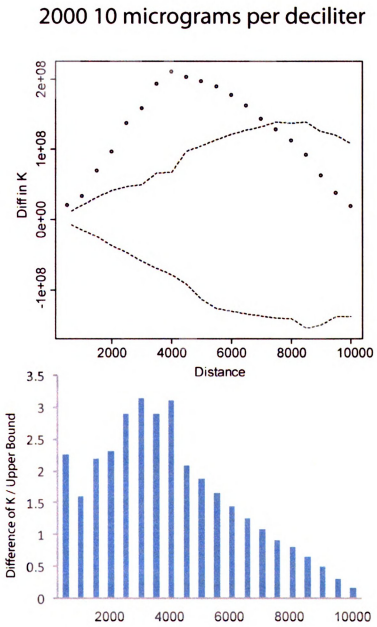


Figure 31: The 2000 Lansing difference of K graph for the 10 μg/dL threshold

The GAM maps show a clear cluster of BLL cases within the Lansing study region. The main cluster in nearly all of the maps is the area around downtown Lansing. The neighborhoods between downtown and the eastern edge of the city of Lansing are a hotspot for elevated BLL every year. This pattern manifests itself in both the 5 and 10 µg/dL threshold levels and can be seen in figure 32.

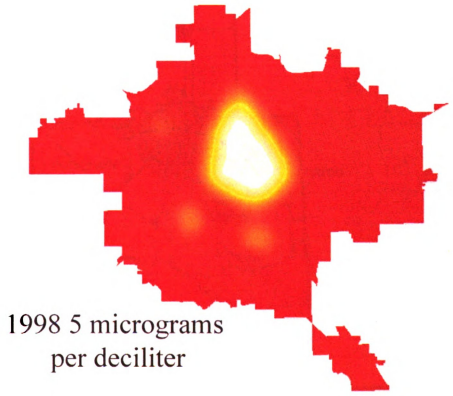


Figure 32: The 1998 GAM map of Lansing for the 5 µg/dL threshold

3.1.7 Mid-South

The Mid-South study area covers all of the Mid-South HSA not within the boundaries of the Lansing study region (Figure 33). This is a mostly rural study area, and includes the counties of Clinton, Eaton, Ingham, Jackson, Hillsdale, and Lenawee. There

are several cities within the Mid-South area such as Jackson, Adrian, Hillsdale, and Charlotte. The number of blood tests in the region shows a decrease from over 1,600 in 1998 to under 700 in 2000. This initial decrease is offset in 2004, where the yearly number of tests more than doubled from less than 1,300 the previous year to over 2,800. The larger number of tests in 2004 and 2005 has an effect on the results of each test.

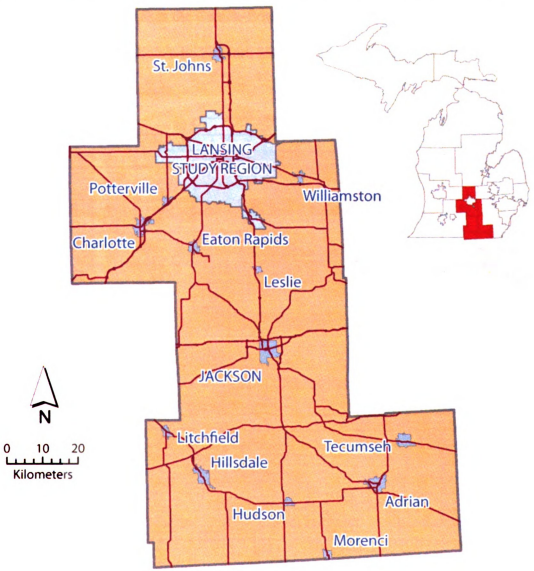


Figure 33: Map of the Mid-South study region

The Cuzick-Edwards tests reveal clustering of 5 µg/dL threshold cases across nearly all years (Table 11). An interesting pattern is the huge increase in the number of tests in 2004 and 2005. This greatly increases the Cuzick-Edwards statistic at all k values for those two years. At the 10 µg/dL threshold, most years have a significant Bonferroni p-value due to initial clustering at the k = 1 or k = 2 levels. The low number of cases at the 25 µg/dL threshold makes the Cuzick-Edwards test ineffective. The years 2000, 2002 and 2005 have two neighbors who both are 25 µg/dL threshold cases, but these could be siblings in the same household.

				K									
Threshold	Year	Cases	Controls	1	2	3	4	5	6	7	8	9	10
5	1998	474	1163	199	384	569	725	890	1060	1251	1456	1617	1775
	1999	205	892	76	119	176	228	284	331	376	428	476	531
	2000	136	560	49	76	117	151	188	230	271	314	351	390
	2001	152	625	57	105	147	193	241	286	325	359	394	433
	2002	194	831	48	92	134	177	214	264	303	341	379	423
	2003	226	1045	73	130	187	229	281	333	384	436	489	521
	2004	791	2041	373	697	1042	1346	1648	1960	2255	2565	2887	3207
	2005	708	2291	274	506	758	1010	1254	1498	1718	1955	2189	2422
10	1998	84	1553	16	31	44	51	60	66	69	83	91	101
	1999	38	1059	4	5	6	9	12	13	14	17	20	24
	2000	27	669	7	15	23	28	34	37	40	42	43	45
	2001	25	752	0	2	3	4	5	6	7	8	9	10
	2002	22	1003	5	7	7	7	8	10	11	11	11	12
	2003	30	1241	3	9	12	14	19	21	23	25	26	27
	2004	61	2771	3	12	17	21	25	29	34	37	41	47
	2005	81	2918	11	13	16	17	20	23	25	28	32	36
25	1998	6	1631	0	0	0	0	0	0	0	0	0	0
	1999	0	1097										
	2000	4	692	2	2	2	2	2	2	2	2	2	2
	2001	0	777										
	2002	4	1021	2	2	2	2	2	2	2	2	2	2
	2003	1	1270										
	2004	6	2826	0	0	0	0	0	0	0	0	0	0
	2005	5	2994	2	2	2	2	2	2	2	2	2	2

Table 11: Cuzick-Edwards results for Mid-South

Results from the Cuzick-Edwards test were confirmed by the difference of K graphs. The K value remains well above the simulation envelopes every year for the 5

μg/dL threshold such as figure 34, indicating strong clustering. The K values remain between 3 and 4 times as large as the upper bounds of the simulation envelope as the result of strong initial clustering and no edge effects. After about three kilometers, the K values stay at around the same value, an indication that they are no longer increasing cases. This is unusual for a mostly rural region, indicating a strong cluster likely exists somewhere in the study area. The 10 μg/dL threshold graphs have K values which remain above the upper bounds of the simulation envelope as well. For the 25 μg/dL threshold, there seems to be little clustering due to lack of cases.

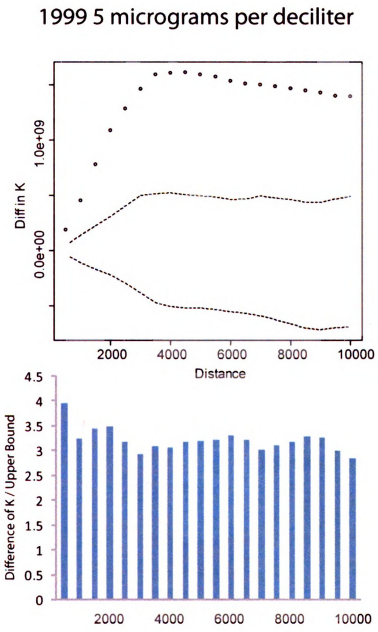


Figure 34: The 1999 Mid-South difference of K graph for the 5 µg/dL threshold

The GAM maps reveal interesting patterns. At the 5 µg/dL threshold, two major factors stand out. First is the reoccurring cluster in the city of Jackson. This result is similar to other urban areas across the state. It is likely that the city of Jackson is the source of the consistent cluster seen in the difference of K graphs. The second is the high number of cases in Lenawee County in 2004 and 2005. This was seen in the Cuzick-

Edwards table, and it seems that many of the cases were found in this county, particularly in the city of Adrian. Each of these two clusters can be seen in figure 35, as well as many constellations of individual cases. This pattern dissipates at the 10 µg/dL threshold level, and the city of Jackson becomes more apparent. No pattern can be found at the 25 µg/dL threshold level.

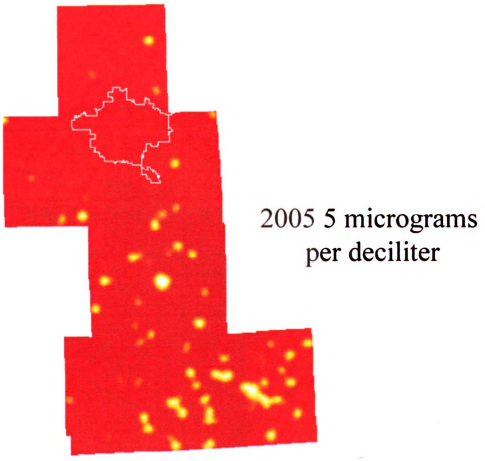


Figure 35: The 2005 GAM map of the Mid-South for the 5 µg/dL threshold

3.1.8 Battle Creek

The Battle Creek study region includes all area within the Battle Creek Urban Aid Boundary (Figure 36). This is a fairly small study area that includes the cities of Battle Creek and Springfield, as well as some areas to the north and east of the cities. It is the smallest of the 19 study regions in this thesis in terms of area size. The number of blood tests in a year does not exceed 1,000 except for the year 2005.

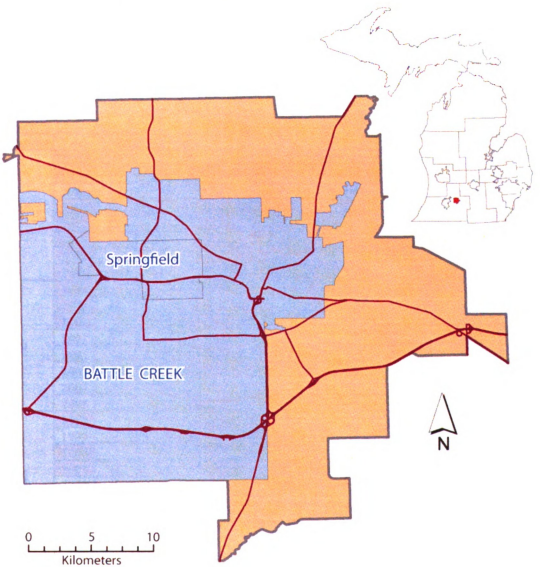


Figure 36: Map of the Battle Creek study area

Battle Creek shows a pattern of Cuzick-Edwards results which is similar to other mid-sized cities (Table 12). At the 5 µg/dL threshold, the results show consistent clustering across all years in the database. The values increase fairly slowly at the higher k values, indicating that any clusters within the study area are smaller than in other cities. The 10 µg/dL threshold results show that in earlier years, there is strong clustering fed by several k = 1 neighbors, but this pattern seems to fade over time. The 25 µg/dL results show a couple years where two k = 1 neighbors both were 25 µg/dL threshold cases. This is interesting considering the low number of total cases at the 25 µg/dL threshold level.

				K									
Threshold	Year	Cases	Controls	1	2	3	4	5	6	7	8	9	10
5	1998	184	226	92	187	289	385	479	585	678	765	858	941
	1999	165	370	81	157	230	298	364	440	506	577	631	682
	2000	160	351	63	120	182	250	315	365	423	464	515	572
	2001	154	402	58	113	159	208	262	326	377	426	480	526
	2002	162	471	57	114	180	248	287	334	394	446	507	568
	2003	110	510	39	77	101	133	160	190	220	251	283	304
	2004	162	815	42	87	123	161	188	228	264	289	326	360
	2005	201	957	57	105	163	217	269	311	360	411	450	492
10	1998	45	365	12	20	27	40	44	53	64	71	75	83
	1999	40	495	10	18	21	30	37	42	43	46	49	53
	2000	33	478	8	10	12	16	19	22	22	22	24	24
	2001	37	519	6	8	9	9	14	20	23	29	32	37
	2002	36	597	6	10	13	20	21	21	23	28	32	34
	2003	30	590	6	8	11	13	16	17	21	24	24	25
	2004	23	954	0	2	2	4	5	6	6	6	6	6
	2005	21	1137	3	3	4	4	5	6	6	6	6	6
25	1998	5	405	2	2	2	2	2	2	2	2	2	2
	1999	5	530	0	0	0	0	0	1	1	1	1	2
	2000	3	508	2	2	2	2	2	2	2	2	2	2
	2001	4	552	2	2	2	2	2	2	2	2	2	2
	2002	0	633										
	2003	2	618	0	0	0	0	0	0	0	0	0	0
	2004	2	975	0	0	0	0	0	0	0	0	0	0
	2005	1	1157										

Table 12: Cuzick-Edwards results for Battle Creek

Results from the Cuzick-Edwards test are confirmed by the difference of K graphs. The 5 µg/dL threshold K values show up immediate sharp jump above the simulation envelopes like in figure 37. The K values rise to around 3.5 times the upper bound of the simulation envelope by two kilometers and continue to add cases until around four kilometers. In each graph around four kilometers, the K values begin a rapid decline. The consistency of this drop indicates the edge of the cluster, but could also be related to edge effects of the small study area. A similar pattern is repeated at the 10 µg/dL threshold level in earlier years, but only in the early years of the database. There is no real change in the 25 µg/dL threshold results.

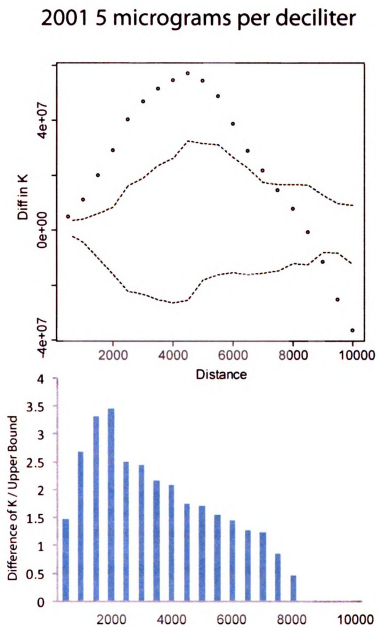


Figure 37: The 2001 Battle Creek difference of K graph for the 5 μg/dL threshold

The GAM results show that the 5 μg/dL threshold cases are concentrated in downtown Battle Creek. A closer analysis shows that the strongest hotspots across all years appear to be on the eastern side of downtown. The 10 μg/dL threshold results show a similar pattern to the 5 μg/dL threshold. Though the hotspot is not the same every year,

the downtown area seen in figure 38 is central to the hotspot. At the 25 µg/dL threshold, the low number of cases makes GAM analysis less reliable.

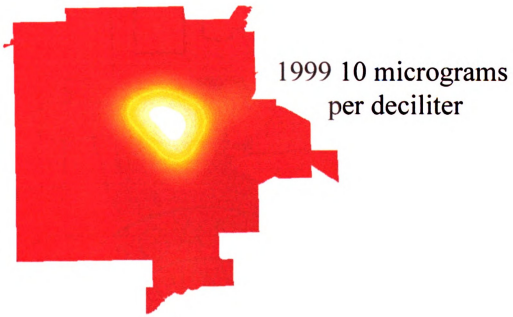


Figure 38: The 1999 GAM map of Battle Creek for the 10 µg/dL threshold

3.1.9 Kalamazoo

The Kalamazoo study area covers the Federal Urban Aid Boundary around the aforementioned metro area (Figure 39). This is a mostly developed district that surrounds the city of Kalamazoo, as well as the cities of Portage and Galesburg. The study area also includes some rural area around the cities. Similar to several other study areas, there is a large increase in blood lead tests in 2004 and 2005 compared to previous years. There were over 1,200 blood tests in 2004 and 2005, while none of the other years exceeded 850.

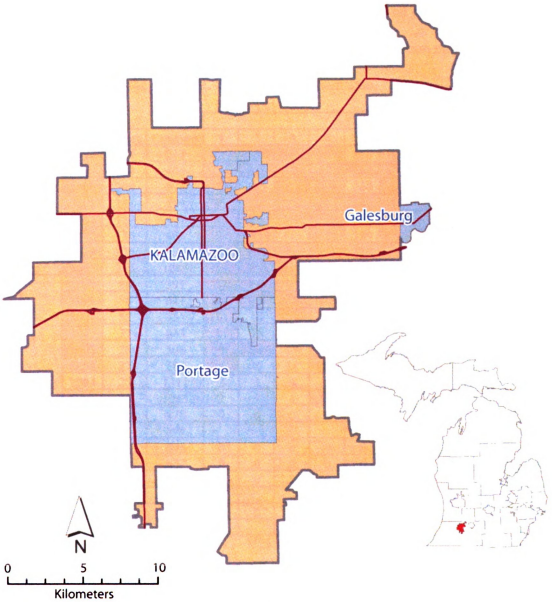


Figure 39: Map of the Kalamazoo study area

The pattern seen in the Cuzick-Edwards results is similar to other mid-sized cities (Table 13). The 5 µg/dL threshold has significant clustering of cases across all years according to the Bonferroni p-values. It appears that the clusters of cases are fairly large as well, as the total case-case count continues to steadily rise as the number of nearest neighbors is increased. At the 10 µg/dL threshold, strong initial clustering exists, but it

does not continue to grow at a significant rate as k increases. The clustering at the 10 µg/dL threshold seems to fade over time, possibly due to remediation efforts. There is no apparent clustering at the 25 µg/dL threshold for Kalamazoo.

				K									
Threshold	Year	Cases	Controls	1	2	3	4	5	6	7	8	9	10
5	1998	271	214	180	361	529	709	866	1023	1192	1368	1551	1736
	1999	234	465	122	235	339	452	562	678	790	895	1007	1127
	2000	209	598	73	135	212	295	360	440	516	590	646	728
	2001	181	499	75	140	209	283	349	417	491	564	627	701
	2002	127	407	45	85	131	175	214	254	295	333	377	410
	2003	136	686	54	85	115	159	189	223	261	286	319	355
	2004	226	1110	70	131	198	258	316	374	438	498	566	644
	2005	215	1083	75	136	188	264	317	360	407	455	524	589
10	1998	88	397	32	66	90	113	143	173	197	227	254	287
	1999	58	641	21	35	46	62	73	76	86	94	105	117
	2000	52	755	11	14	24	34	43	51	58	65	70	77
	2001	37	643	11	15	17	20	29	32	36	39	40	47
	2002	28	506	9	14	17	22	24	29	33	34	36	36
	2003	27	785	6	8	11	17	19	23	25	28	30	33
	2004	25	1311	2	3	4	5	8	10	10	10	11	11
	2005	33	1265	2	2	4	9	12	14	14	16	18	18
25	1998	7	478	0	0	0	0	0	0	0	0	0	0
	1999	1	698										
	2000	2	805	0	0	0	0	0	0	0	0	0	0
	2001	3	677	0	0	0	0	0	0	0	0	0	0
	2002	5	529	0	0	0	0	0	0	0	0	0	0
	2003	1	821										
	2004	2	1334	0	0	0	0	0	0	0	0	0	0
	2005	1	1297										

Table 13: Cuzick-Edwards results for Kalamazoo

Similar to Cuzick-Edwards, the difference of K results in Kalamazoo show patterns of clustering similar to other mid-sized cities within Michigan. At the 5 µg/dL threshold level, K values immediately jump up at short distances. There is no doubt that significant clustering of 5 µg/dL threshold cases exists within Kalamazoo. At the 10 µg/dL threshold, results show strong clustering at short distances as well. The K values rise well above the upper bound of the simulation envelopes, and then fall back at around six kilometers such as in figure 40. The peak of the K values occurs around four

kilometers where the difference of K is 2.5 times as high as the upper bound of the simulation envelope. The rapid decline of K values afterwards indicates four kilometers is the likely diameter of the cluster. This pattern persists across all years without fading, possibly indicating the consistent underlying threat. The 25 µg/dL threshold K values were not significant.

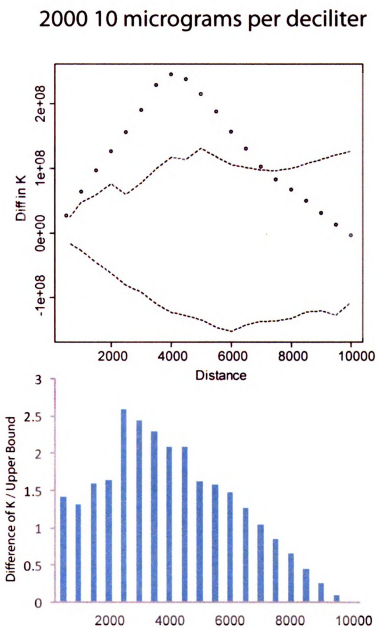


Figure 40: The 2000 Kalamazoo difference of K graph for the 10 µg/dL threshold

The GAM results for Kalamazoo show a consistent pattern of hotspots. At each of the threshold levels, the corresponding hotspot is located around the central business district of the city of Kalamazoo. This hotspot stretches from there down to the southeast through the nearby neighborhoods, shown in figure 41. The neighborhoods directly to the north of downtown Kalamazoo are affected as well. These areas are the most likely source of the clustering seen in earlier tests.

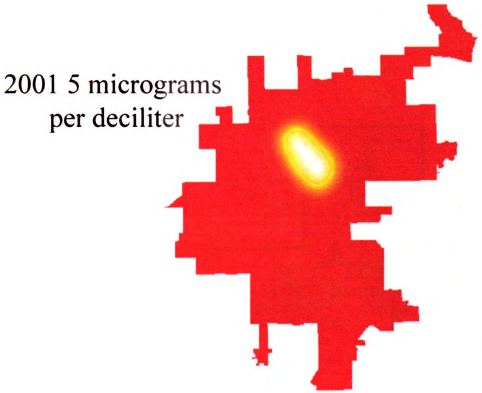


Figure 41: The 2001 GAM map of Kalamazoo for the 5 μg/dL threshold

3.1.10 Southwest

The region of Southwest Michigan covers the similarly named HSA with the exception of the Kalamazoo and Battle Creek study areas (Figure 42). With these cities

removed, the study region is more rural in composition. It covers the counties of Berrien, Van Buren, Cass, St. Joseph, Branch, Calhoun, Barry, and all of Kalamazoo County that does not fall within the Kalamazoo study area. While the Southwest Michigan region is more rural with some of the cities removed, there are still several smaller cities and towns. These include Benton Harbor, Niles, Sturgis, and Coldwater. The number of yearly blood lead tests is typically between 2,000 and 2,500, but there is an increase to over 4,000 in 2004 and 2005.

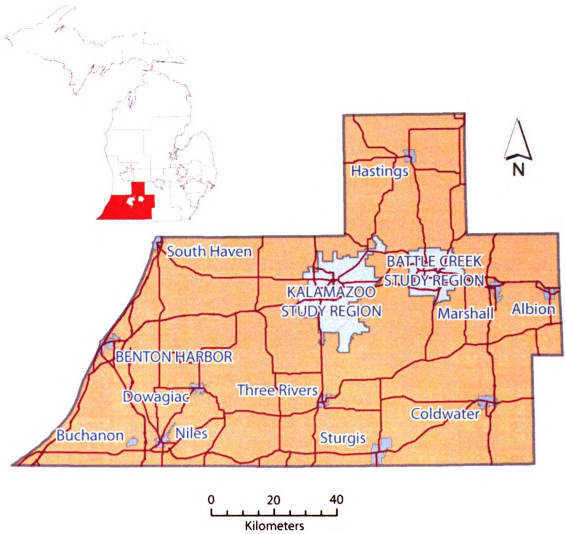


Figure 42: Map of the Southwest study area

Despite the more rural nature of the study region, the Southwest area Cuzick-Edwards results display strong clustering across all years at the 5 and 10 µg/dL thresholds and several instances at the 25 µg/dL threshold (Table 14). With the 5 and 10 µg/dL thresholds, the Bonferroni p-values indicate clustering across all k sizes. This is the highest amount of clustering found for a HSA-based study area, indicating that there is a real hotspot in the region. The Monte Carlo simulations reveal that the steady growth of case-case neighbors continues to steadily increase as k gets larger. The 25 µg/dL threshold has significant clustering in several years as well, but it is more inconsistent.

				K									
Threshold	**Year**	**Cases**	**Controls**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
5	**1998**	900	1122	532	1008	1514	2021	2545	3036	3528	4015	4495	5023
	1999	841	1426	458	895	1299	1730	2128	2547	2982	3371	3800	4198
	2000	646	1340	316	617	884	1175	1434	1701	1979	2263	2544	2806
	2001	630	1763	276	503	721	962	1213	1443	1671	1905	2159	2413
	2002	639	2023	250	449	652	869	1070	1279	1472	1650	1876	2089
	2003	477	2092	152	293	453	586	716	847	992	1121	1258	1400
	2004	702	3720	184	351	514	685	859	1043	1178	1328	1492	1629
	2005	771	3918	209	379	567	731	907	1099	1260	1429	1595	1762
10	**1998**	244	1778	83	148	219	277	345	387	444	493	547	616
	1999	221	2046	80	134	193	246	294	355	407	459	503	543
	2000	146	1840	30	56	70	91	110	129	149	168	198	236
	2001	118	2275	27	51	72	97	120	148	169	187	209	235
	2002	136	2526	32	52	77	106	125	146	169	191	217	244
	2003	71	2498	7	15	21	25	31	35	40	46	50	56
	2004	86	4336	11	16	18	26	32	41	45	47	50	53
	2005	124	4565	16	30	41	49	55	66	76	82	84	89
25	**1998**	18	2004	2	5	7	7	8	8	8	8	8	9
	1999	16	2251	0	0	0	1	1	4	4	4	5	7
	2000	7	1979	0	0	0	0	0	1	1	1	2	2
	2001	8	2385	0	0	0	0	0	0	0	2	2	2
	2002	9	2653	1	2	2	2	2	2	2	3	3	4
	2003	6	2563	0	0	0	0	0	0	0	0	0	0
	2004	8	4414	0	0	0	0	0	0	0	1	1	1
	2005	5	4684	0	0	0	0	0	0	0	0	0	0

Table 14: Cuzick-Edwards results for Southwest Michigan

The difference of K graphs for Southwest Michigan confirm the earlier results that there is strong clustering of elevated BLL at every threshold level. For the 5 µg/dL

threshold level, the K values rise far above the upper bound of the simulation envelope. This is also true for the 10 μg/dL threshold. At the 25 μg/dL threshold, the K values stay above the upper bounds of the simulation envelopes for most years in the database like in figure 43. The K values increase very quickly to over three times the value of the upper bound of the simulation envelope, and then levels off at two kilometers. This is rare for a region this large and likely indicates areas of unusually high BLL rates. Both Cuzick-Edwards and difference of K seem to point to a very strong cluster in the region.

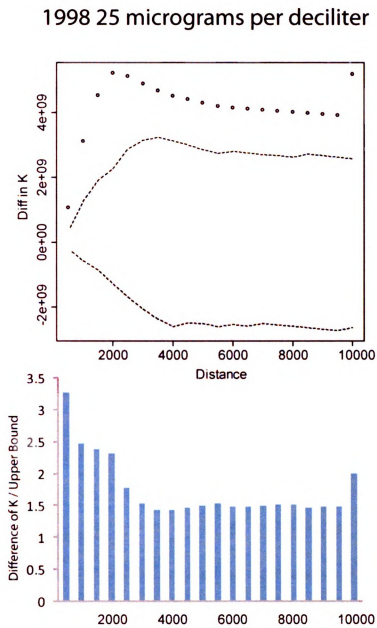


Figure 43: The 1998 Southwest Michigan difference of K graph for the 25 µg/dL threshold

The GAM results reveal that the Benton Harbor area is the likely source of the high clustering. The city is present on every threshold level map through all years of the database. At the 5 µg/dL threshold level, this city is present, but there is also a constellation of smaller hotspots. It is difficult to determine whether or not these

represent significant clusters. At the 10 µg/dL threshold, the primacy of the Benton Harbor area becomes more apparent. The 25 µg/dL threshold GAM maps show only Benton Harbor, which can be seen in figure 44.

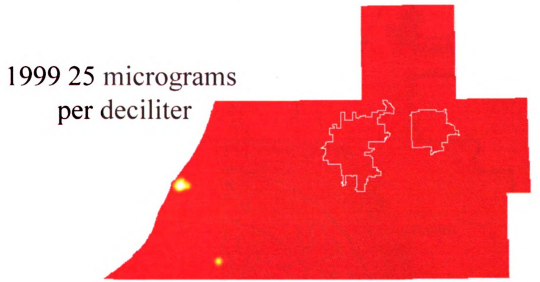


Figure 44: The 1999 GAM map of Southwest Michigan for the 25 µg/dL threshold. Other study regions outlined in white

3.1.11 Grand Rapids

The Grand Rapids study region covers the city's Federal Urban Aid Boundary (Figure 45). This is the second most populous area of the state after Detroit. Several cities are included within the Grand Rapids study area. They are Grand Rapids, Wyoming, Kentwood, and Walker. The number of yearly blood lead tests range from 3,500 to 6,000.

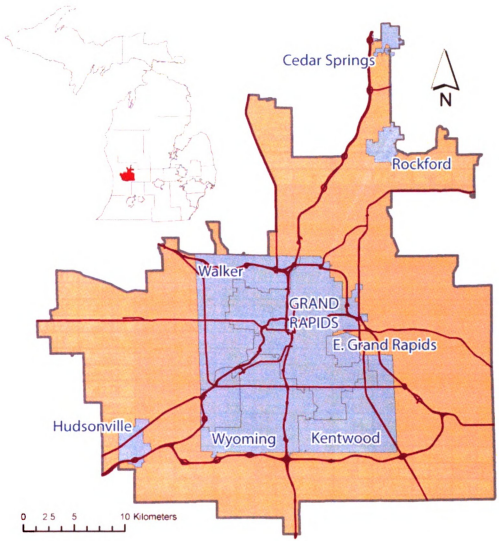


Figure 45: Map of the Grand Rapids study region

The Cuzick-Edwards results reveal the Grand Rapids region has large clusters at all threshold levels (Table 15). Given the large population and results in other Michigan urban areas, this is not a surprise. At the 5 μg/dL threshold level, there is strong clustering across all years in the database. The number of case-case neighbors continues to grow at a prodigious rate as k values climbs in value, leading to the conclusion that the cluster or clusters are large. The 10 μg/dL threshold shows very large spatial clustering as well. This is different from many other cities within Michigan and is evidence of the

extent of the problem in Grand Rapids. Strong initial clustering with the 25 µg/dL threshold can also be seen in the study area. Much of it is linked to a small number of cases at the k = 1 level, but the Bonferroni p-value indicates it is significant in several years.

				K									
	Year	Cases	Controls	1	2	3	4	5	6	7	8	9	10
5 Threshold	1998	2882	1686	2087	4163	6285	8383	10438	12529	14564	16613	18641	20698
	1999	2187	1349	1641	3237	4804	6379	7961	9459	10992	12530	14072	15644
	2000	2116	1380	1527	3024	4488	5951	7450	8961	10437	11918	13392	14876
	2001	2425	2791	1512	2946	4381	5731	7144	8615	10005	11419	12837	14253
	2002	2241	2480	1371	2680	4019	5312	6626	7925	9287	10628	11918	13243
	2003	1551	2652	847	1602	2386	3174	3973	4735	5472	6251	7047	7857
	2004	2057	3897	1022	1941	2836	3790	4763	5701	6591	7499	8402	9285
	2005	1127	4837	480	898	1245	1609	1987	2353	2730	3092	3451	3818
10 Threshold	1998	1177	3391	525	1003	1420	1820	2240	2676	3102	3508	3966	4421
	1999	911	2625	417	754	1117	1462	1788	2125	2466	2808	3156	3529
	2000	819	2677	307	619	913	1176	1450	1731	2010	2302	2558	2806
	2001	769	4447	257	441	642	811	1010	1208	1391	1542	1719	1887
	2002	483	4238	144	225	305	400	495	577	670	764	851	935
	2003	315	3888	80	144	196	250	316	361	400	448	493	548
	2004	298	5656	59	89	113	136	174	194	222	251	278	312
	2005	262	5702	62	92	129	158	198	231	258	296	328	370
25 Threshold	1998	97	4471	11	20	23	29	35	39	45	46	49	56
	1999	69	3467	3	7	9	13	13	15	15	15	20	21
	2000	67	3429	2	2	5	5	7	8	9	11	14	16
	2001	42	5174	3	4	4	4	5	5	5	5	6	6
	2002	32	4689	5	6	6	6	6	8	9	9	9	10
	2003	28	4175	0	0	0	0	0	0	0	0	0	0
	2004	21	5933	4	4	4	4	4	4	4	4	4	4
	2005	32	5932	6	6	6	6	6	6	6	6	6	6

Table 15: Cuzick-Edwards results for Grand Rapids

The difference of K graphs confirms the strong clustering of elevated BLL cases at all threshold levels within the study area of Grand Rapids. At both the 5 and 10 µg/dL thresholds, the K values rise far above the upper bounds of the simulation envelope. The elevated BLL cases at both thresholds appear to be in large clusters. There is a consistent drop off after about seven kilometers at the 5µg/dL threshold level and six kilometers at the 10 µg/dL threshold level (Figure46). These are fairly sizable cluster diameters.

Despite the drop after six kilometers, the K values remain twice as high as the upper bound even at ten kilometers. The 25 µg/dL threshold also shows clustering. The drop off in K values is lower, around four kilometers. Overall, the region shows strong, large clusters at each threshold level.

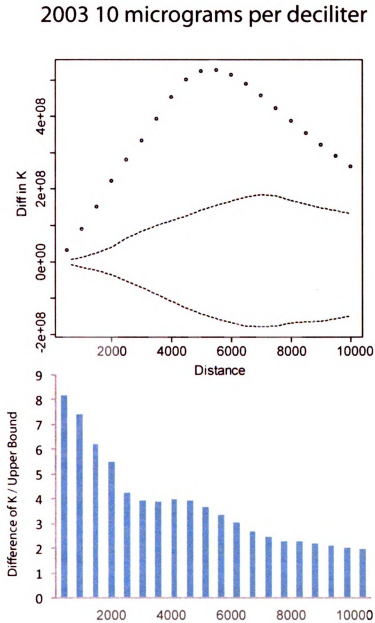


Figure 46: The 2003 Grand Rapids difference of K graph for the 10 µg/dL threshold

GAM analysis reveals a strong concentration of elevated BLL cases in central Grand Rapids. Figure 47, representative of the pattern across all thresholds, shows the hotspot of BLL in downtown Grand Rapids. The prime area of clustering of elevated BLL seems to be on the eastern side of the city. Similar to other urban study areas, the central downtown area overwhelms other cities within the region.

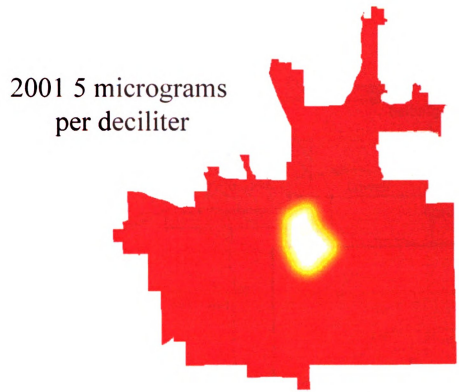


Figure 47: The 2001 GAM map of Grand Rapids for the 5 μg/dL threshold

3.1.12 Lower Coast

The study region titled "Lower Coast" represents the lower half of the West HSA excluding the Grand Rapids urban aid boundary (Figure 48). This includes the counties of Ionia, Kent, Allegan, Ottawa, and Muskegon. The study region is a majority rural area, but several cities are located within the area. A couple of examples are Muskegon,

Holland, Ionia, Grand Haven, and Zeeland. The number of blood lead tests in a year within the study area falls between 1,800 and 2,200 for the years 1998-2003, followed by an increase to nearly 4,000 in 2004 and over 5,000 in 2005.

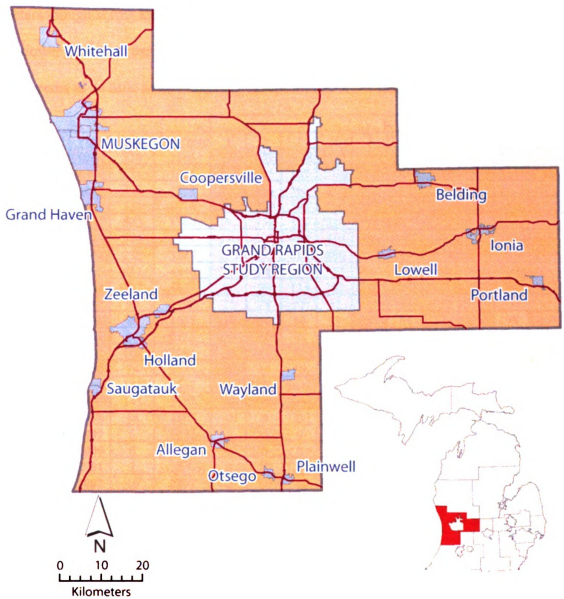


Figure 48: Map of the Lower Coast study region

The Lower Coast study area exhibits clustering tendencies of elevated BLL cases at the 5 and 10 µg/dL thresholds levels according to Cuzick-Edwards (Table 16). Across

all years in the database, the 5 μg/dL threshold has both significant overall clustering according to the Bonferroni p-value and clustering at many levels of k. The 10 μg/dL threshold contains clustering across k values for every year as well. The size of these clusters though seems to be small. The Monte Carlo tests reveal strong initial clustering, but slower growth to the total case neighbors as k grows. At the 25 μg/dL threshold, there seems to be little to no clustering except for two k = 1 neighbors in 2004.

				K									
Threshold	Year	Cases	Controls	1	2	3	4	5	6	7	8	9	10
5	1998	716	1105	398	777	1158	1531	1884	2225	2588	2939	3309	3654
	1999	571	1326	266	505	762	1016	1255	1460	1680	1895	2145	2384
	2000	492	1238	191	384	563	742	946	1125	1321	1504	1702	1876
	2001	525	1716	201	387	563	725	896	1069	1233	1383	1546	1720
	2002	502	1700	192	358	527	720	896	1073	1227	1366	1543	1712
	2003	514	2111	180	302	463	610	753	907	1072	1221	1344	1486
	2004	732	3217	230	435	667	880	1088	1295	1511	1741	1950	2126
	2005	832	4440	285	478	684	913	1103	1319	1531	1740	1990	2217
10	1998	187	1634	56	92	138	182	225	270	313	357	396	441
	1999	99	1798	26	39	51	62	74	91	103	118	131	147
	2000	85	1645	10	26	34	41	57	61	76	90	98	107
	2001	86	2155	13	26	32	37	48	55	57	58	60	65
	2002	108	2094	30	45	60	74	91	107	118	133	145	158
	2003	74	2551	12	16	19	23	29	31	40	45	52	55
	2004	102	3847	20	32	41	46	51	57	70	79	92	98
	2005	145	5127	24	36	49	61	73	84	98	113	129	143
25	1998	14	1807	0	0	0	0	0	0	0	0	0	0
	1999	6	1891	0	0	0	0	0	0	0	0	0	0
	2000	8	1722	0	0	0	0	0	0	0	0	0	0
	2001	7	2234	0	0	0	0	0	0	0	0	0	0
	2002	10	2192	0	0	0	0	0	0	0	0	0	0
	2003	2	2623	0	0	0	0	0	0	0	0	0	0
	2004	7	3942	2	2	2	2	2	2	2	2	2	2
	2005	7	5265	0	0	0	0	0	0	0	0	0	0

Table 16: Cuzick-Edwards results for the Lower Coast

Difference of K results reveal clustering in the cases at both the 5 and 10 μg/dL thresholds. At both of these levels, there is a quick rise in K values until about four kilometers, where the values level out and begin a slow decline. Still, the K values remain above the upper bounds of the simulation envelope in every year. This pattern

can be seen in figure 49. The K values are 4 times as high as the upper bound of simulation envelope, indicating the concentration of cases within the region in a cluster. Similar to the Cuzick-Edwards results, the difference of K graphs indicate at least one very strong cluster of cases at both the 5 and 10 μg/dL threshold.

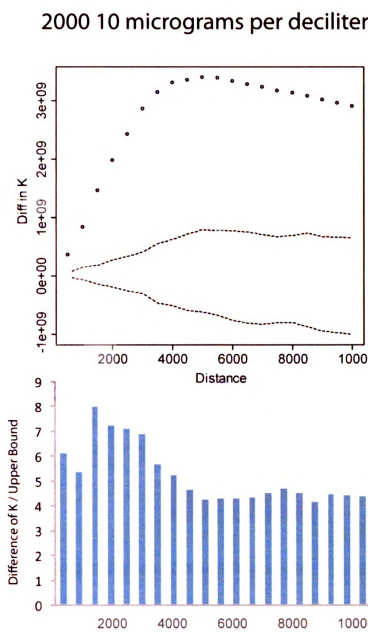


Figure 49: The 2000 Lower Coast difference of K graph for the 10 μg/dL threshold

The GAM maps point to the source of the clustering in several locations. The most obvious source is the coastal city of Muskegon. This area shows up in every yearly map at every threshold level. In figure 50, The Muskegon area is the obvious source of the cluster seen in the Cuzick-Edwards and difference of K tests. Another hotspot that factors into the clustering seen earlier is the city of Holland. It is not as consistently a hotspot, but the city could be the source of clustering in addition to Muskegon. At the 5 µg/dL threshold level, there are a large number of hotspots that do not appear regularly. These are likely single cases. In all likelihood, Muskegon is the source of the strong clustering seen in earlier tests.

Figure 50: The 2002 GAM map of Lower Coast for the 10 µg/dL threshold

3.1.13 Mid Coast

The region labeled "Mid Coast" represents the upper half of the West HSA (Figure 51). The mostly rural region includes the counties of Mason, Oceana, Lake, Newaygo, Osceola, Mecosta, and Montcalm. There are not too many built up areas within the region. A couple of the cities are Big Rapids, Ludington, Reed City, and Newaygo. Blood lead test numbers range from 800 to 1,000 in most the years, but quickly rise towards 1,500 and 2,000 in 2004 and 2005.

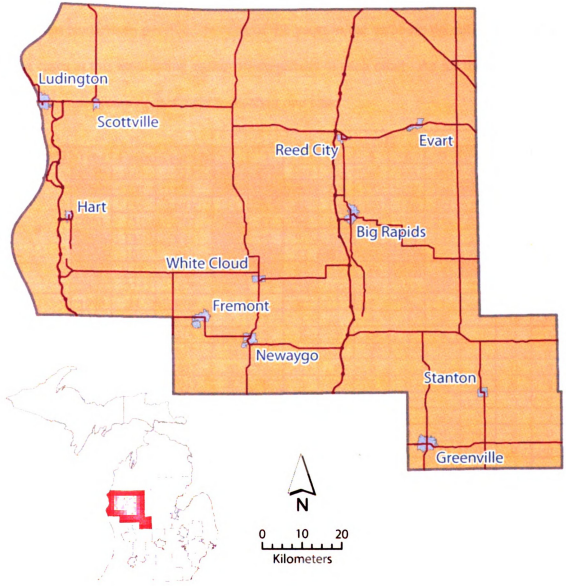


Figure 51: Map of the Mid Coast study region

The Cuzick-Edwards results for the Mid Coast region tend to show clustering only at the 5 µg/dL threshold level (Table 17). In all years in the database, it seems that initial clustering is present and provides a significant Bonferroni p-value for the overall test. The Monte Carlo results for the 5 µg/dL threshold reveal that these clusters are small and involve mostly low k values. With the 10 µg/dL threshold level, some years provide two neighbors next to each other, but none of the years in the database show a significant Bonferroni p-value. Several of the years in the database do not even show any of the cases at this level being within 10 neighbors of each other. As for the 25 µg/dL threshold, most years do not have more than one case.

				K									
Threshold	Year	Cases	Controls	1	2	3	4	5	6	7	8	9	10
5	1998	220	777	71	130	180	233	287	333	380	428	477	515
	1999	200	694	60	126	184	234	273	320	364	422	457	508
	2000	135	692	24	60	81	108	126	148	171	191	211	234
	2001	151	700	36	64	96	130	160	193	224	258	284	323
	2002	144	894	40	65	91	109	135	158	181	206	225	249
	2003	101	940	21	37	47	73	93	111	122	136	148	159
	2004	148	1438	28	50	64	79	99	125	143	163	186	198
	2005	225	1870	47	73	110	143	166	193	228	257	275	300
10	1998	21	976	0	0	0	2	3	3	3	3	3	3
	1999	21	873	0	2	3	3	4	4	6	7	10	10
	2000	12	815	2	2	2	2	3	3	3	3	3	3
	2001	16	835	1	2	2	2	2	2	2	2	2	2
	2002	13	1025	0	0	0	0	0	0	0	0	0	0
	2003	4	1037	0	0	0	0	0	0	0	0	0	0
	2004	11	1575	0	0	0	0	0	0	0	0	0	0
	2005	20	2075	0	2	2	2	2	2	2	2	2	2
25	1998	2	995	0	0	0	0	0	0	0	0	0	0
	1999	2	892	0	0	0	0	0	0	0	0	0	0
	2000	0	827										
	2001	0	851										
	2002	1	1037										
	2003	0	1041										
	2004	0	1586										
	2005	1	2094										

Table 17: Cuzick-Edwards results for the Mid-Coast

The difference of K results for the Mid-Coast region do not reveal strong clustering. Nearly every year, even at the 5 µg/dL threshold, has K values that fall within the simulation envelopes (Figure52). The difference of K values never rise above 60% of the upper bound of the simulation envelope. At the 10 µg/dL threshold level, the number of cases is so low that the K values do not show much vertical movement.

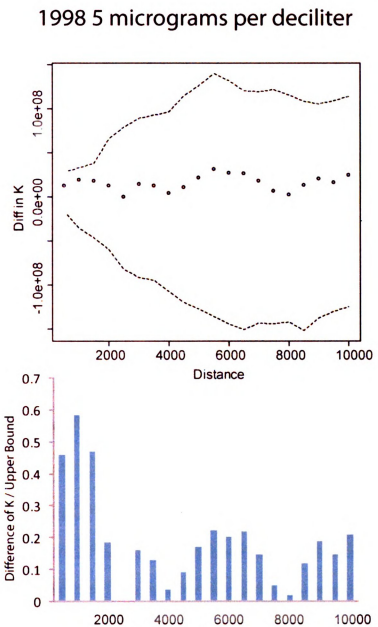


Figure 52: The 1998 Mid-Coast difference of K graph for the 5 µg/dL threshold

With the lack of clustering in the region, the GAM maps mostly reveal the locations of single cases. As with other rural areas, it is difficult to discern any pattern in the results. The spots appear as constellations that seem to differ in patterns every year like in figure 53. While the Cuzick-Edwards indicated clustering at the 5 µg/dL threshold, it is possible that the neighbors are spread out far enough that they appear only as single cases in GAM and not a large hotspot. It is therefore nearly impossible to find an underlying pattern in the GAM maps for the Mid-Coast.

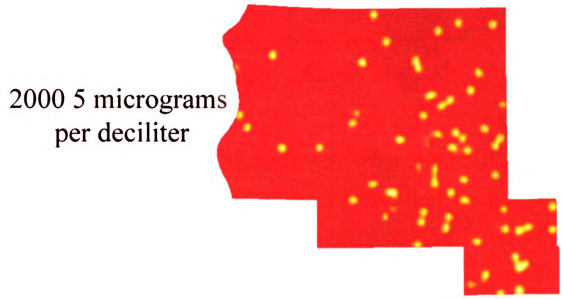


Figure 53: The 2000 GAM map of Mid-Coast for the 5 µg/dL threshold

3.1.14 Saginaw/Bay City

The Saginaw/Bay City study region represents the Federal Urban Aid Boundary around the two cities (Figure 54). It runs from the city of Saginaw and its surrounding environs down a thin connecting strip of land to Bay City and the Saginaw Bay coastline. The region is urban and developed. There is a steady increase in the number of blood

lead tests in the Saginaw/Bay City study region in the years of the database, from under 650 in 1998 to over 2,500 in 2005.

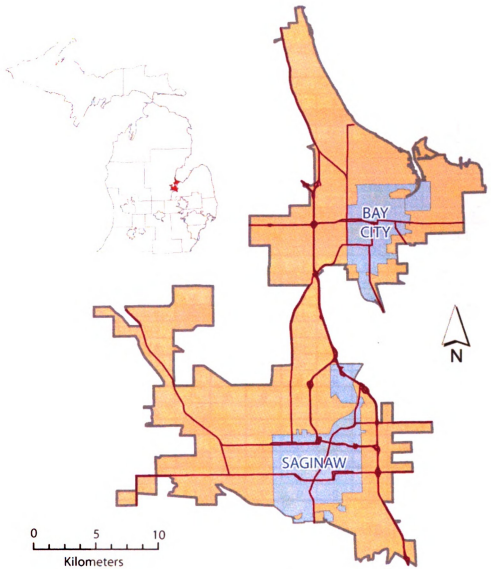


Figure 54: Map of the Saginaw/Bay City study region

The Cuzick-Edwards results for the Saginaw/Bay City region tend to follow a typical pattern for mid-to-large sized cities within Michigan (Table 18). The 5 μg/dL threshold level shows large clusters, a strong Bonferroni p-value, and continued growth of the total case neighbors as k rises. The 10 μg/dL threshold also shows a pattern seen

in other urban study areas. There is strong initial clustering that gives the region a strong Bonferroni p-value, but the growth slows at larger k values and indicates the small size of the clusters. There are not enough cases at the 25 µg/dL threshold level to distinguish real clusters, though some years have two neighbors at the k = 1 level.

				K									
Threshold	Year	Cases	Controls	1	2	3	4	5	6	7	8	9	10
5	1998	290	343	155	305	460	629	798	964	1124	1280	1444	1610
	1999	278	574	128	258	367	470	567	664	766	874	974	1094
	2000	353	784	184	343	479	639	772	912	1060	1212	1365	1508
	2001	386	820	167	339	501	656	810	964	1109	1282	1444	1595
	2002	451	1100	192	374	532	695	877	1064	1236	1393	1575	1724
	2003	359	1036	124	251	352	456	558	667	785	896	1008	1140
	2004	526	1764	203	374	554	719	881	1049	1219	1388	1571	1750
	2005	537	2004	163	300	446	595	731	865	1031	1200	1348	1480
10	1998	73	560	20	29	44	63	77	93	102	118	124	138
	1999	60	792	14	26	31	37	44	51	60	70	78	86
	2000	73	1064	20	34	44	55	60	65	71	81	88	98
	2001	65	1141	4	9	14	19	25	29	36	41	44	55
	2002	82	1469	16	24	37	41	50	60	70	85	94	104
	2003	56	1339	6	8	11	12	15	19	22	25	30	38
	2004	74	2216	17	21	26	30	35	42	47	53	64	73
	2005	76	2465	12	18	22	26	30	34	39	42	47	47
25	1998	7	626	0	0	0	0	0	0	0	0	0	0
	1999	4	848	0	0	0	0	0	0	0	0	0	0
	2000	1	1136										
	2001	9	1197	2	2	2	2	2	2	2	2	2	2
	2002	5	1546	0	0	0	0	0	0	0	0	0	0
	2003	6	1389	2	2	2	2	2	2	2	2	2	2
	2004	5	2285	0	0	0	0	0	0	0	0	0	0
	2005	7	2534	0	0	0	0	0	0	0	0	0	0

Table 18: Cuzick-Edwards results for Saginaw/Bay City

The difference of K results in the Saginaw/Bay City region show signs of clustering. At the 5 µg/dL threshold, the K values rise above the simulation envelopes immediately, and then fall back down below after about five kilometers. The yearly consistency in this pattern leads to the possibility that the same underlying area is showing up each year. The 10 µg/dL threshold results show the same early rise in K values, though the drop below the upper bound occurs quickly such as figure 55. The

difference of K values stay around 2 times as high as the upper bound of the simulation envelope, though K values precipitously drop after four kilometers. Given the consistency of the pattern, this region seems to exhibit clustering at the lower thresholds. There is no vertical movement in the K values at the 25 μg/dL threshold.

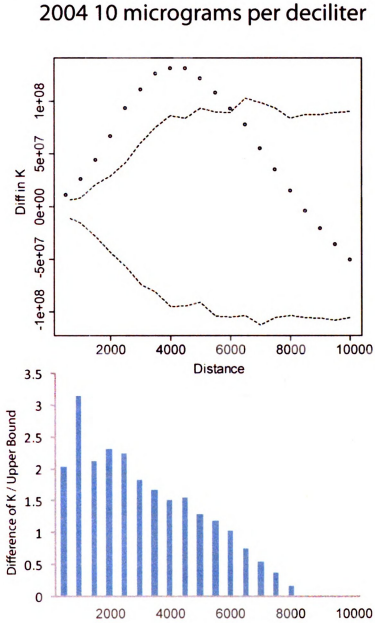


Figure 55: The 2004 Saginaw/Bay City difference of K graph for the 10 μg/dL threshold

GAM results for the region reveal that the clusters of elevated BLL cases occur almost exclusively within the city limits of Saginaw and Bay City. While this is not surprising given similar results around the state, it is still significant. The city of Saginaw exhibits the strongest hotspots such as figure 56. In Saginaw, most of the hotspots appear to occur either near the Saginaw River or on the eastern side of the city. For Bay City, the main yearly hotspots seem to occur on the eastern side of the river.

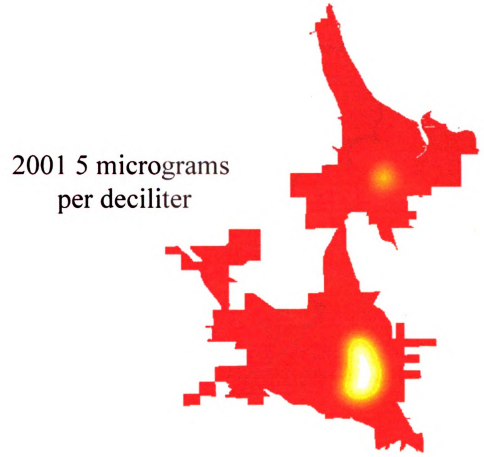


Figure 56: The 2001 GAM map of Saginaw/Bay City for the 5 µg/dL threshold

3.1.15 West Bay

The "West Bay" region represents the western half of the Bay HSA, not including the Saginaw/Bay City study area (Figure 57). The mostly rural region includes the

counties of Iosco, Ogemaw, Roscommon, Clare, Gladwin, Arenac, Isabella, Midland, Gratiot, and the portions of Saginaw and Bay counties that lie to the west of the Shiawassee/Saginaw Rivers. Midland is the main city within the region, but there are other built-up areas such as Mount Pleasant, Alma, and Gladwin. The number of yearly blood lead tests ranges from a low of 571 tests in 1998 to 1,898 tests in 2005.

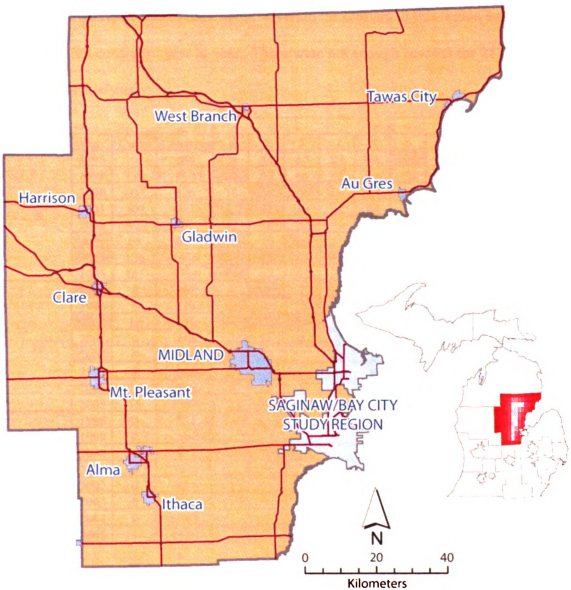


Figure 57: Map of the West Bay study region

The Cuzick-Edwards results for the West Bay region are inconsistent (Table 19). The years of 2003 and 2004 show significant results at the 5 μg/dL threshold level according to the Bonferroni p-values. The clustering seen in these years are a result of case-case neighbors at lower k values. Three different years (1998, 2000, and 2002) have 10 μg/dL threshold Bonferroni p-values which are significant, but this is often entirely due to only two cases next to each other. Overall, the clusters in this region are not very big and are not consistent year to year. There were not enough cases at the 25 μg/dL threshold for analysis.

				K									
Threshold	Year	Cases	Controls	1	2	3	4	5	6	7	8	9	10
5	1998	107	464	24	44	64	81	95	128	147	164	180	201
	1999	134	710	34	52	66	92	120	138	160	180	198	220
	2000	107	826	18	37	56	66	75	90	98	110	119	136
	2001	100	912	17	33	43	47	61	73	86	96	109	117
	2002	92	873	12	17	27	30	33	42	51	60	76	82
	2003	65	897	12	19	28	33	41	47	50	51	56	61
	2004	126	1476	25	49	63	80	95	106	116	126	134	141
	2005	165	1733	26	43	61	79	93	109	134	149	171	182
10	1998	6	565	0	2	2	2	2	2	2	3	3	4
	1999	5	839	0	0	0	0	0	0	0	0	1	1
	2000	12	921	6	10	14	14	14	14	15	15	15	15
	2001	4	1008	0	0	0	0	0	0	0	0	0	0
	2002	6	959	2	2	2	2	2	2	2	2	2	2
	2003	2	960	0	0	0	0	0	0	0	0	0	0
	2004	11	1591	0	0	0	0	1	1	2	2	2	2
	2005	15	1883	0	0	0	0	0	1	1	1	1	1
25	1998	0	571										
	1999	0	844										
	2000	1	932										
	2001	0	1012										
	2002	0	965										
	2003	0	962										
	2004	0	1602										
	2005	1	1897										

Table 19: Cuzick-Edwards results for West Bay

The difference of K graphs reveal no clustering at any distance for any threshold level. This is somewhat surprising given the fact that a city the size of Midland, with a

population around 50,000, is located within the study region (US Census Bureau 2001). At both the 5 and 10 μg/dL threshold levels, the K values fail to clear the upper bounds of the simulation envelopes. In figure 58, this is demonstrated by the lack of vertical movement of the K values. The difference of K values do not even rise above zero until nearly eight kilometers, indicating large distances between the individual cases in the study area. This result leads to the conclusion that the spatial organization of cases to controls is not significantly different than what is produced by the random labeling hypothesis.

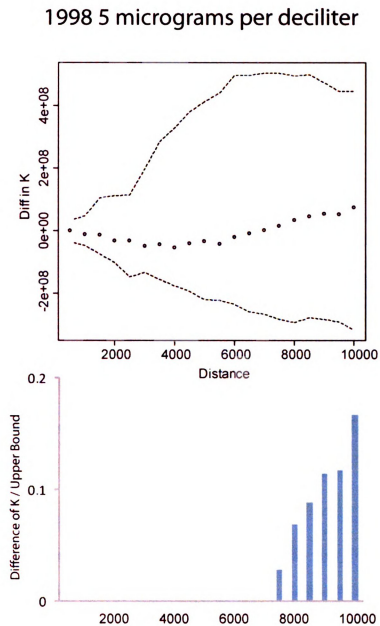


Figure 58: The 1998 West Bay difference of K graph for the 5 µg/dL threshold

GAM results for the West Bay region confirm the earlier analysis showing lack of any clustering. The maps reveal that cases do exist within the region, but no real discernable pattern can be found. Midland does not show up prominently on many of the maps. This is surprising given results seen in other portions of the state where large cities As with other rural areas of the state, the GAM suffers from the low case/control rate

exposing nearly every case as a hotspot. Figure 59 shows individual cases, not necessarily hotspots.

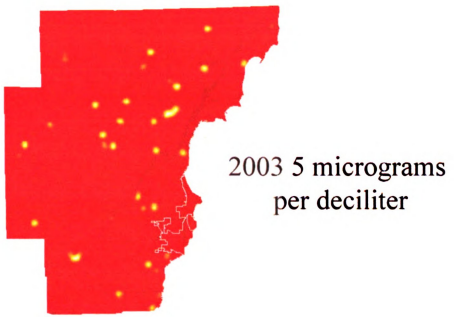


Figure 59: The 2003 GAM map of West Bay for the 5 µg/dL threshold

3.1.16 East Bay

The "East Bay" region represents the eastern half of the Bay HSA with the exception of the Saginaw/Bay City study area (Figure 60). Most of this rural region covers the area of Michigan known as "the thumb" of the state. This includes the counties of Sanilac, Huron, Tuscola, and the parts of Saginaw and Bay counties east of the Shiawassee/Saginaw Rivers. The region has very few towns and developed areas. A few towns within the study area are Bad Axe, Sandusky, Croswell, and Frankenmuth. The number of blood lead tests in the East Bay region ranges from a low of 279 in 1999 to 1,161 in 2005.

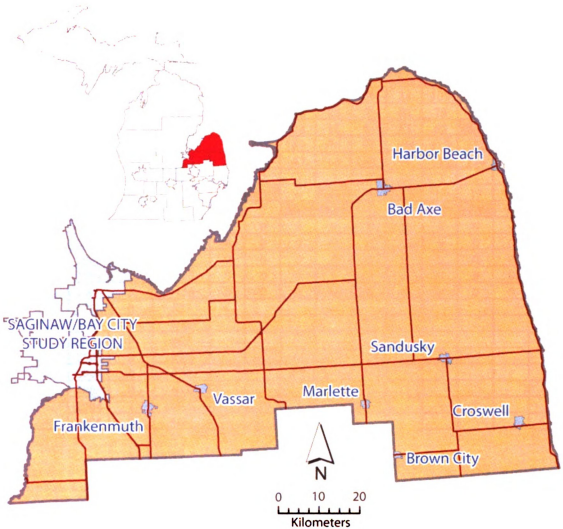


Figure 60: Map of the East Bay study region

Cuzick-Edwards results for the East Bay region reveal on-and-off level of clustering across all years (Table 20). At both the 5 and 10 µg/dL thresholds, the years of 1998-2000have significant levels of clustering according to the Bonferroni p-value while later years, with the exception of 2004, do not. The difference is usually in whether or not there is a large amount of case-case neighbors at the k = 1 level. Overall, the pattern of clustering seems fairly weak. The 25 µg/dL threshold does not have any cases most years to analyze.

				K									
Threshold	Year	Cases	Controls	1	2	3	4	5	6	7	8	9	10
5	1998	129	394	47	84	106	154	185	214	246	285	313	345
	1999	41	238	14	25	34	49	53	62	66	72	79	86
	2000	105	500	22	44	70	99	117	151	181	210	231	256
	2001	89	436	16	38	52	63	79	88	109	132	141	157
	2002	73	397	16	25	36	48	55	64	76	93	104	109
	2003	71	541	13	21	28	33	41	53	65	82	90	101
	2004	129	950	30	47	59	76	92	104	119	138	158	173
	2005	137	1024	24	32	47	66	80	99	119	136	147	171
10	1998	15	508	2	2	2	2	2	2	2	2	2	2
	1999	6	273	3	5	6	8	8	12	12	13	13	13
	2000	12	593	2	2	4	7	7	7	9	11	11	11
	2001	4	521	0	0	0	0	0	0	0	0	0	0
	2002	7	463	0	0	0	0	0	0	0	0	0	0
	2003	2	610	0	0	0	0	0	0	0	0	0	0
	2004	7	1072	0	0	0	0	0	0	0	0	0	0
	2005	5	1156	0	0	0	0	0	0	0	0	0	0
25	1998	0	523										
	1999	1	278										
	2000	0	605										
	2001	0	525										
	2002	0	470										
	2003	0	612										
	2004	1	1078										
	2005	0	1161										

Table 20: Cuzick-Edwards results for East Bay

The difference of K results for the East Bay region exhibit little if any signs of clustering. At the 5 µg/dL threshold, the K values briefly creep above the upper bound of the simulation envelope in the years 1998-2000, but most exhibit no clustering like in figure 61. In this figure, the K values barely rise to 50% of the upper bound of the simulation envelope. Since the simulation envelopes can change slightly with each run, it cannot be confirmed that clustering is visible in any of the graphs. The 10 µg/dL threshold graphs show very little linear movement in the K values. This is the result of a low number of cases at the threshold level in addition to lack of clustering.

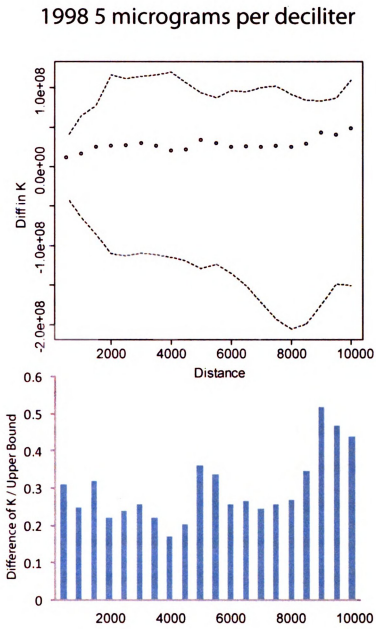


Figure 61: The 1998 East Bay difference of K graph for the 5 µg/dL threshold

Similar to other more rural areas, the GAM maps are hard to read for the East Bay region. The study area's low rates of cases mean that any area with cases at all can show up as a hotspot. On the western side of the study area, there are many single cases in the Vassar area and surrounding environs (see figure 62). Unfortunately, it is difficult to pick up a consistent pattern in the cases year to year.

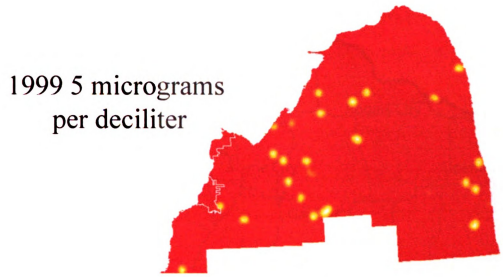


Figure 62: The 1999 GAM map of East Bay for the 5 µg/dL threshold

3.1.17 North Central

The study region of North Central covers the HSA that holds the same name (Figure 63). The mostly rural and natural area covers the northern parts of the Lower Peninsula. The counties included in the North Central study region are Emmet, Cheboygan, Presque Isle, Alpena, Montmorency, Otsego, Charlevoix, Antrim, Leelanau, Benzie, Grand Traverse, Kalkaska, Crawford, Oscoda, Alcona, Missaukee, Wexford, and Manistee. This region has several cities, including Traverse City, Alpena, Cadillac, Cheboygan, and Rogers City. The region has a large increase in the number of blood lead tests over the years covered by the database, from 414 tests in 1998 to 2,408 tests in 2005.

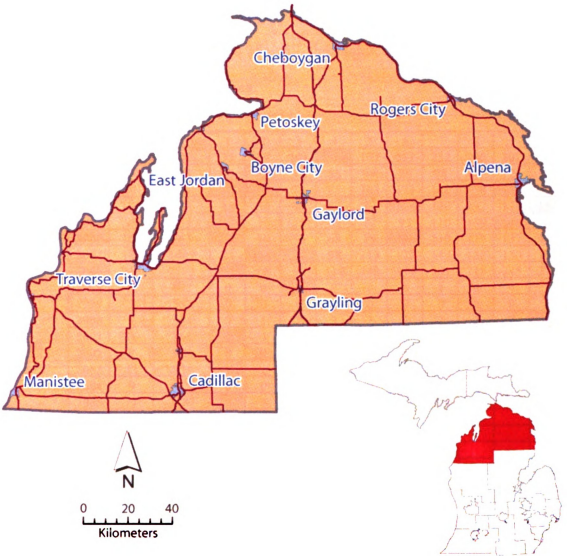


Figure 63: Map of the North Central study region

Cuzick-Edwards results for the North Central region seem to reveal inconsistent results (Table 21). At the 5 μg/dL threshold level, there are as many years where the Bonferroni p-values are not significant as there are significant years. It seems that the number of case neighbors at most k values do not differ from what would be expected by chance given the case/control ratios within the region. There are a couple years where initial clustering at the low k values pushes the Bonferroni p-values into significance.

But the temporal pattern is inconsistent and does not suggest a strengthening or weakening pattern. At both the 10 and 25 µg/dL thresholds, the number of cases is too small to detect any conclusive clustering.

				K									
Threshold	Year	Cases	Controls	1	2	3	4	5	6	7	8	9	10
5	1998	113	301	33	69	101	141	177	211	232	266	302	337
	1999	84	484	7	26	41	56	74	86	98	113	125	137
	2000	91	461	28	55	74	85	103	123	142	158	180	201
	2001	102	682	18	34	48	67	82	104	120	132	138	161
	2002	99	734	19	36	51	69	82	95	110	130	145	163
	2003	102	763	23	40	56	68	79	101	110	119	137	148
	2004	180	1449	21	52	84	98	117	127	143	158	180	197
	2005	235	2173	45	68	96	119	148	179	201	218	245	270
10	1998	12	402	0	0	3	4	4	4	4	4	4	5
	1999	5	563	0	0	0	0	0	0	0	0	0	0
	2000	8	544	0	1	2	2	2	2	2	2	2	4
	2001	16	768	3	4	4	4	5	5	7	7	7	8
	2002	8	825	0	0	1	1	1	1	1	1	2	2
	2003	11	854	2	2	2	2	2	2	2	2	2	2
	2004	9	1620	0	0	0	0	0	0	0	1	1	1
	2005	21	2387	2	2	3	4	4	4	6	6	6	6
25	1998	1	413										
	1999	1	567										
	2000	0	552										
	2001	1	783										
	2002	1	832										
	2003	0	865										
	2004	1	1628										
	2005	2	2406	0	0	0	0	0	0	0	0	0	0

Table 21: Cuzick-Edwards results for North Central

The North Central study region shows no clustering in the difference of K graphs. Figure 64 is a good example. The K values do not jump at all, a good indication of just how scarce cases of elevated BLL are, even at the 5 µg/dL threshold. In figure 64, the K values do not even exceed 50% of the upper bound of the simulation envelope anywhere within the ten kilometers tested. While cases certainly exist within this region, their spatial configuration does not seem particularly clustered.

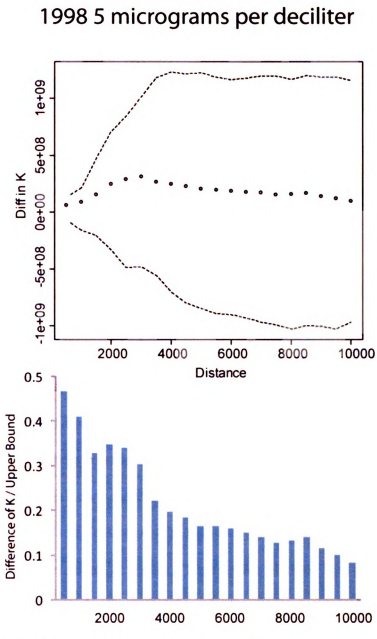


Figure 64: The 1998 North Central difference of K graph for the 5 µg/dL threshold

Similar to other rural regions in the state, the GAM maps for the North Central region do not reveal any specific hotspots year to year. Instead, a collection of individual cases spot the landscape like in figure 65. It is tough to even find a pattern within the individual cases, compounding any attempt to find hotspots. Since GAM is based on grid points, it will not locate individual cases.

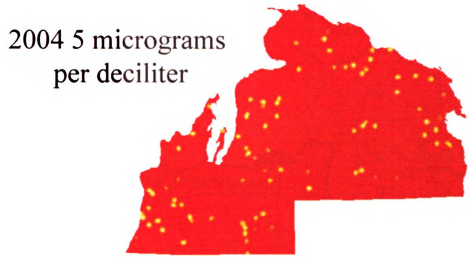


Figure 65: The 2004 GAM map of North Central for the 5 μg/dL threshold

3.1.18 Eastern Upper Peninsula

The study area of Eastern Upper Peninsula includes the three easternmost counties (Figure 66). These counties are Chippewa, Mackinac, and Luce. It is a mostly rural region, but with a fair concentration of people on the route from Sault St. Marie to the Mackinac Bridge. Sault St. Marie is the major city within the region, but there are a few other towns as well such as St. Ignace. The number of blood lead tests in the study area is under 400 every year in the database.

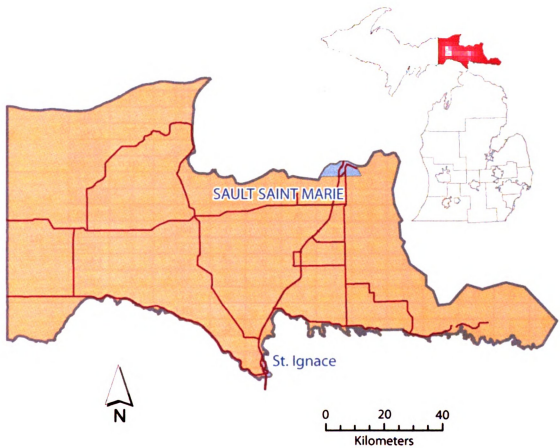


Figure 66: Map of the Eastern Upper Peninsula study region

The Eastern Upper Peninsula region results for the Cuzick-Edwards tests reveal little clustering (Table 22). The 5 µg/dL threshold level does not have significant clustering except for the final two years of 2004 and 2005. The 10 µg/dL threshold numbers reveal no significant clustering only in 1999 and there are not enough cases at the 25 µg/dL threshold. What these numbers could reveal is a lack of testing in this study region. Both 2004 and 2005 were years with a substantial statewide increase in BLL testing. It is possible that these clusters at the 5 µg/dL threshold were not discovered until more tests were done.

				K									
Threshold	Year	Cases	Controls	1	2	3	4	5	6	7	8	9	10
5	1998	41	209	14	26	34	44	53	62	68	74	82	90
	1999	43	212	12	21	30	35	46	53	61	69	75	79
	2000	56	242	13	27	37	47	58	67	76	89	99	113
	2001	47	324	11	19	22	27	31	39	45	50	58	62
	2002	52	274	15	26	34	41	54	64	70	81	91	100
	2003	37	230	7	14	17	22	24	28	35	38	45	50
	2004	56	268	5	24	39	59	73	91	100	114	132	146
	2005	38	345	14	21	23	25	29	32	36	39	42	49
10	1998	5	245	0	0	0	0	0	0	0	1	1	1
	1999	4	251	2	2	2	2	2	2	2	2	2	2
	2000	2	296	0	0	0	0	0	0	0	0	0	0
	2001	7	364	0	0	0	0	0	0	0	0	0	0
	2002	3	323	0	0	0	0	0	0	0	0	0	0
	2003	3	264	0	0	0	0	0	0	0	1	1	1
	2004	6	318	0	1	1	2	2	2	2	2	3	3
	2005	5	378	0	0	0	0	0	0	0	0	0	0
25	1998	0	250										
	1999	2	253	0	0	0	0	0	0	0	0	0	0
	2000	0	298										
	2001	0	371										
	2002	0	326										
	2003	0	267										
	2004	0	324										
	2005	0	383										

Table 22: Cuzick-Edwards results for Eastern Upper Peninsula

In the Eastern Upper Peninsula, the difference of K values show little to no vertical movement at any threshold level, as displayed in figure 67. The years which did show vertical movement did so were nearly entirely within the simulation envelope. The K values do not even exceed 40% of the upper bound of the simulation envelope. Also, the movement did not occur initially, but after one or two kilometers. This cast doubts on any tight urban clusters within the region. This is a somewhat surprising result given that a city as large as Sault St. Marie is located in the study area.

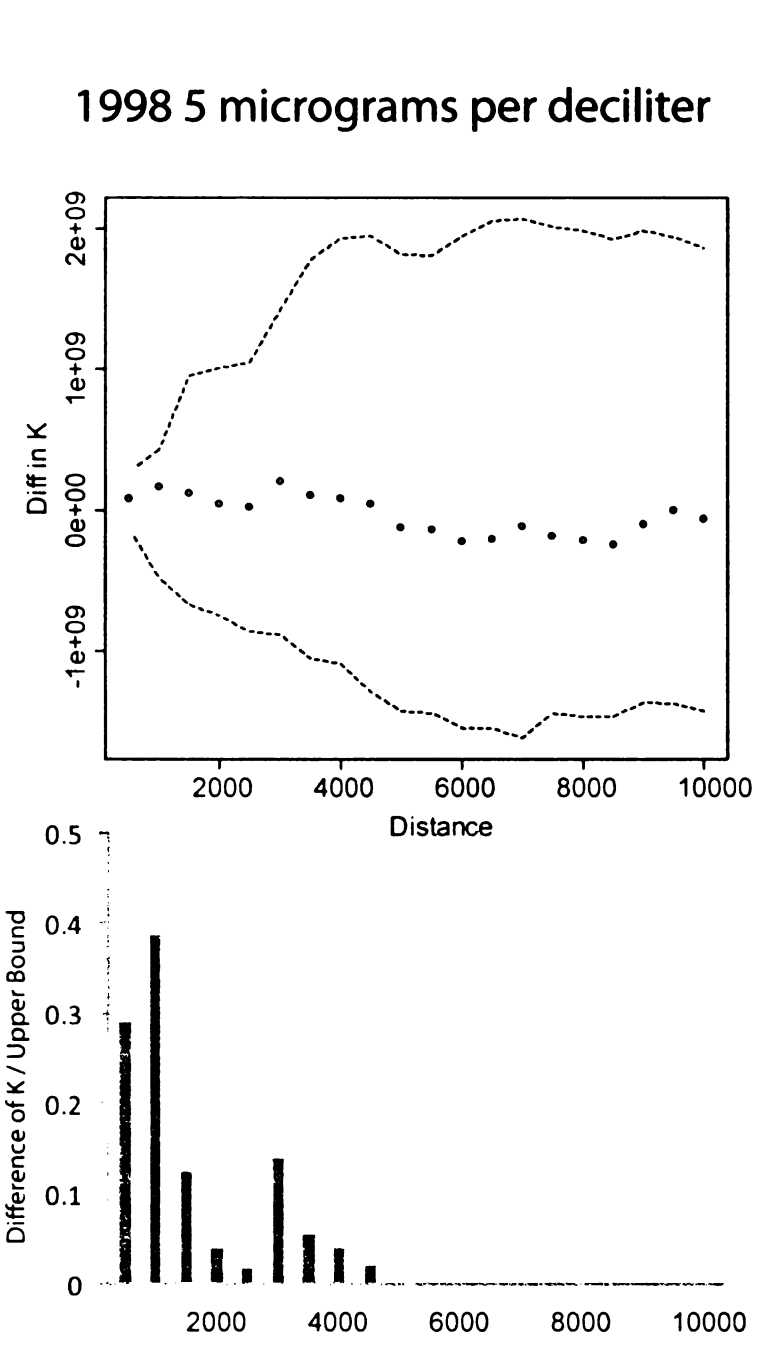


Figure 67: The 1998 Eastern Upper Peninsula difference of K graph for the 5 µg/dL threshold

GAM results for this region, similar to other more rural study areas, are more useful for looking for patterns of cases rather than identifying the location of clusters. One surprising pattern that reemerged across many years was a group of cases in the rural roads directly south of Sault St. Marie. Figure 68 is a good example of this, where there

are several single cases near each other in this rural area. Surprisingly, the pattern is stronger in this area than in Sault St. Marie. This is different from elsewhere in the state, where urban areas consistently exhibited more hotspots than nearby rural areas. Cases at both the 5 and 10 µg/dL thresholds also seem to show up in the western part of the study region as well.

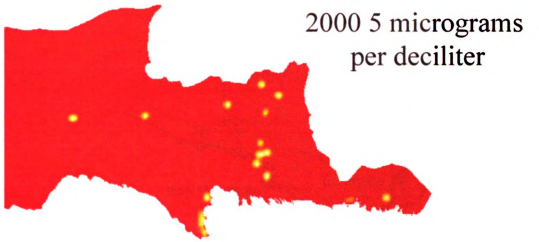


Figure 68: The 2000 GAM map of Eastern Upper Peninsula for the 5 µg/dL threshold

3.1.19 Western Upper Peninsula

The final region covers all of the Upper Peninsula of Michigan except the three easternmost counties (Figure 69). The region of the Western Upper Peninsula covers the counties of Schoolcraft, Alger, Delta, Menominee, Marquette, Dickinson, Iron, Baraga, Gogebic, Ontonagon, Houghton, and Keweenaw. It is mostly rural or natural area, but there are several cities and towns of importance. These include Marquette, Houghton, Escanaba, Ishpeming, Iron Mountain, and Ironwood. The number of yearly blood lead tests in the study region grows from under 500 in 1998 to over 1,300 in 2005.

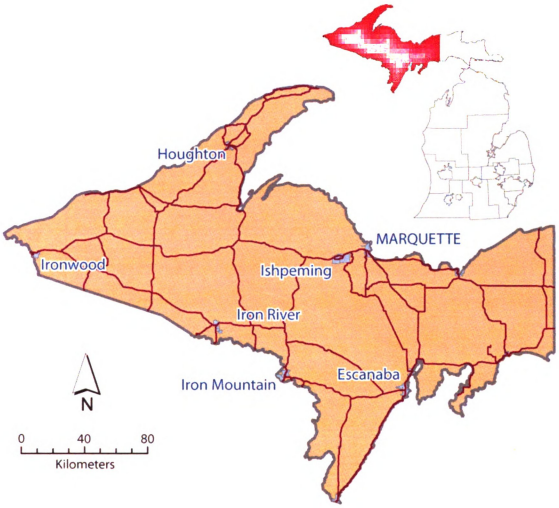


Figure 69: Map of the Western Upper Peninsula study region

The Cuzick-Edwards test results for the Western Upper Peninsula show a similar pattern to the eastern half of the peninsula (Table 23). The results are inconsistent until the large increase in the number of blood tests exhibits clustering in 2004 and 2005. Unlike the eastern part, the Western Upper Peninsula study region does have clustering in 1998. Given that both Upper Peninsula study areas show increased clustering in the last two years of the database, it is possible that this part of the state is conducting more rigorous lead screening.

				K									
Threshold	Year	Cases	Controls	1	2	3	4	5	6	7	8	9	10
5	1998	164	320	84	155	219	291	358	416	488	558	623	691
	1999	236	899	65	118	168	218	273	327	373	431	488	542
	2000	227	743	70	134	207	260	320	375	428	482	548	601
	2001	152	796	30	61	86	122	139	166	196	218	249	278
	2002	140	793	22	45	72	98	130	155	181	204	229	246
	2003	144	665	32	62	101	132	148	181	212	247	276	312
	2004	156	899	34	81	116	155	195	223	255	286	311	331
	2005	133	1217	24	44	56	71	98	114	135	151	171	187
10	1998	25	459	2	9	12	12	15	17	19	22	24	26
	1999	25	1110	2	2	2	2	2	2	2	2	2	2
	2000	27	943	4	5	5	5	5	5	6	6	7	7
	2001	9	939	0	0	0	0	0	0	0	1	1	1
	2002	10	923	0	0	0	0	0	0	0	0	0	0
	2003	13	796	0	0	0	0	0	0	0	0	0	0
	2004	18	1037	0	1	1	3	3	3	3	3	4	4
	2005	10	1340	0	0	0	0	0	0	0	0	0	0
25	1998	1	483										
	1999	1	1134										
	2000	0	970										
	2001	0	948										
	2002	0	933										
	2003	0	809										
	2004	0	1055										
	2005	0	1350										

Table 23: Cuzick-Edwards results for Western Upper Peninsula

The difference of K results for the Western Upper Peninsula study area follows the Cuzick-Edwards findings. There are a few years in the 5 µg/dL threshold results where the K values hug the upper bound of the simulation envelopes such as figure 70. The K values nearly touch reach the upper bounds of the simulation envelope. Since the random simulations would be different each time the difference of K is run, even if the K values had slightly exceeded the upper bound the results would still not prove clustering. At the 10 µg/dL threshold, there is no year where the difference of K values differs greatly from zero. Everything points to little if any confirmed clustering of elevated BLL cases within the region according to difference of K.

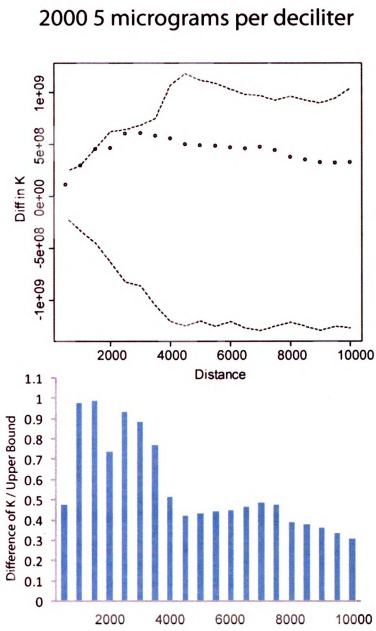


Figure 70: The 2000 Western Upper Peninsula difference of K graph for the 5 μg/dL threshold

Despite the lack of provable clustering, the GAM results do reveal areas of the state that consistently look troublesome. An area in which cases seem to continually crop up is the Ishpeming area. In nearly all of the years examined, cases show up in this area. The Houghton area is also visible on most of the maps as well. Finally, Escanaba and the

surrounding environments look like they could be the home of some cases of elevated BLL (Figure 71). The city of Marquette, the most populated city in the study region, is surprisingly not much of a factor. This goes against the pattern of results for most of the rest of the state for large cities.

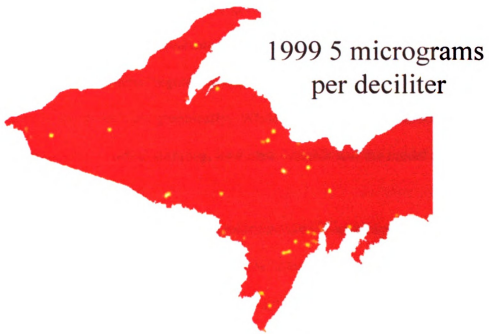


Figure 71: The 1999 GAM map of Western Upper Peninsula for the 5 µg/dL threshold

3.2 Geographically Weighted Regression Results

Regression analysis was employed in this thesis in order to understand and explain the spatial patterns of childhood BLL in Michigan. Linear regression was run on three different areal units: US census tract, zip code, minor civil division. US census block groups were also considered for this analysis, but the small size of the individual

units made the analysis useless for two main reasons. The size often left many units with few if any test results located within, and the huge number of block groups statewide made computing the GWR models impossible for the R software. For the three geographic units utilized, this analysis used linear regression for the creation of a statewide model, hereafter referred to as a global model, of childhood BLL. The linear regression models were used to evaluate the performance of independent variables at a statewide level, but additional regression methods were needed to analyze the performance of the models geographically. While linear regression allows for geographic analysis of error with residual mapping, how each variable and the model as a whole varies over space is unknown.

The second part of the regression analysis used Geographically Weighted Regression (GWR) to examine the effectiveness of the model and its variables across space. GWR models work by conducting the regression analysis on each geographical unit (i.e. each census tract) rather than statewide like the global linear regression. All other observations are weighted in GWR based on their distance to the focal geographical unit. This thesis used a common GWR weighting scheme based on a Gaussian curve, where nearby observations a given more weight than observations further away. To define the shape of the curve, a bandwidth is selected by finding the minimum residual sum of squares for all data points.

The dependent variable in all of the regression models was the mean BLL based on all blood test results within the geographical unit. In the linear regression analysis, the mean BLL of test results for each individual year of the database were also tested as dependent variables in order to evaluate the models over time. All mean BLL values

calculated for this thesis were not weighted by population or the number of test results. In the case of all three different geographic units, the mean BLL numbers were normally distributed and did not require any data transformation.

The ten independent variables shown in table 24 used were chosen based on earlier studies (see tables 2 and 3) as well as availability from the US Census Bureau. Three out of the ten variables had skewed distributions of values in all areal units, and were logarithmically changed to achieve a normal distribution. For each of the three variables, any zero values were changed to 0.00001 to permit logarithmic transformation. To decide which variables to use in each model, linear regression was used to eliminate variables which were not significant ($\alpha = 0.05$) for mean BLL based on all years of blood tests. The remaining significant variables were then used for the yearly and GWR regression models.

Percentage Pre-1940 Housing
Percentage of African-Americans (logged)
Percentage of Latinos (logged)
Percentage of Recent Immigrants (logged)
Percentage under 6 years of age
Percentage of Housing Rented
Percentage of Housing Headed by Females
Percentage of Housing Vacant
Percentage without a high school diploma
Percentage below 185% of the Poverty Line

Table 24: Independent variables tested by regression analysis

Presented in the results section for regression are several different maps and models. The first map is a map of the standard deviation of yearly mean blood lead

levels. The mean BLL for each year of the database (1998-2005) was calculated based on the μg/dL blood lead test results within each individual unit. The standard deviation of the eight yearly mean BLL results was calculated for each geographic unit. This map gives a sense of the yearly volatility in the mean BLL. The second part of the regression results section shows a map of the mean BLL in each unit for all eight years combined, as well as the results of the linear regression model with the eight year mean BLL as the dependent variable. The third section shows the results of linear regression models where the independent variables were used to predict the mean BLL in an individual year. The variables that are significant predictors ($\alpha = 0.05$) are marked in blue in the table, while variables that are not significant are marked in red. The bar graph shows the R^2 values for each yearly model with a line for comparison to the all years model.

The final section contains the GWR results, which are put into a table. The tables show a summary of the coefficients produced for each individual geographic unit divided in quartiles. Also available are the regression diagnostics including the size of the fixed bandwidth in meters, the number of individual geographic units, the effective number of parameters and degrees of freedom, sigma squared (standard error of the estimate), and the Akaike Information Criterion (AIC) which is a measure of the goodness of fit (Fotheringham, Brunsdon, and Charlton 2002). Also listed is the Leung statistic, which was explained in equation 10, a measure of how well the GWR model reduces the residual sum of squares compared to the linear OLS regression. Finally, maps are provided which show how the coefficients of key independent variables change across Michigan.

3.2.1 Minor Civil Division

The first areal unit regression analysis was Minor Civil Divisions (MCD), a term covering all local political boundaries such as city limits and townships. The map of the mean BLL for all years in figure 73 shows a different pattern from the other areal units. The cities such as Detroit and Grand Rapids have the highest mean values, but they have far less influence as single entities. Select rural areas dominate the map, including the southwest portion of the state, the "thumb" of Michigan, and portions of the northern half of the Lower Peninsula. The standard deviations map in figure 72 follows the mean BLL map fairly consistently.

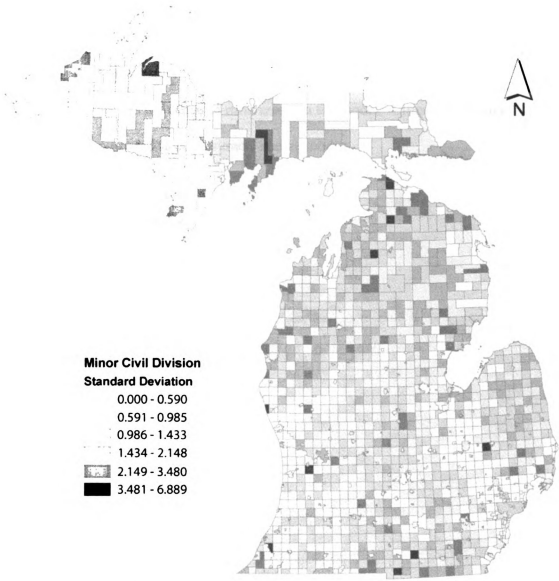
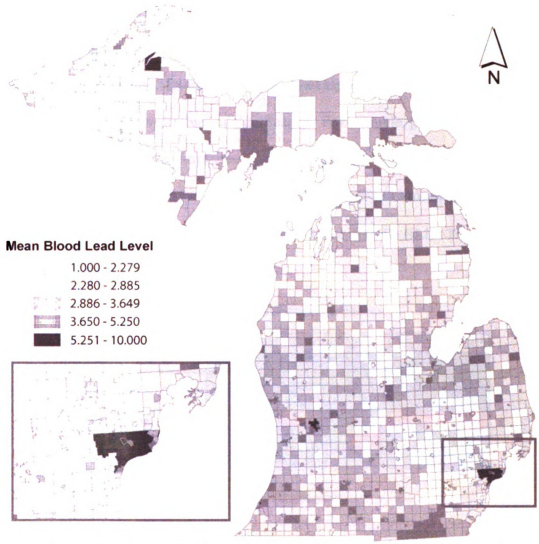


Figure 72: Map of the minor civil division standard deviations of yearly mean BLL

The global regression model in figure 73 shows MCD level analysis to be poor for studying elevated BLL based on the independent variables commonly associated with the ailment. The R^2 for the overall global model is 0.17, very poor when compared with census tracts and zip codes. In the MCD global model, the main independent variable is

again percentage pre-1940 housing. Perhaps the most interesting facet of the MCD all years model is that percentage African-American has a lower t-value than percentage without a high school diploma. This is certainly due to the fact that the cities, such as Detroit, are entire units rather than broken up into sections. The large number of townships increases the influence of rural areas on the model. Cities have far less influence when compared to census tracts and zip codes.



Coefficients	Estimate	Std. Error	t-value	Pr(>|t|)
(Intercept)	2.07544	0.12616	16.451	2E-16
lnBlack	0.03776	0.00592	6.38	2.34E-10
lnLatino	0.04114	0.01053	3.907	9.76E-05
Pct_Pre1940	1.45256	0.14864	9.772	2E-16
Pct_FemaleHeaded	0.78562	0.21061	3.73	0.000198
Pct_No High School	3.01664	0.41777	7.221	8.14E-13
Pct_Under 6	3.74431	1.13606	3.296	0.001004
R2 = 0.1774				
Adjusted R2 = 0.1742				
F-statistic = 54.67 on 6 and 1521 Degrees of Freedom				2E-16

Figure 73: Map of mean BLL by minor civil division and all years global regression results

The yearly global regression model results (Table 25) reinforce the notion that MCD level analysis is not suitable for mean BLL. The significance of each variable oscillates from year to year. Even the variables most associated with mean BLL in the all years model fall below lower values of significance. For example, pre-1940 housing is a better predictor of mean BLL than female-headed households by far in the all years model, but not in the 1998 or 1999 model. Much like the all years model, the individual MCD yearly models do not explain much of the variance in mean BLL. The R^2 range is typically between 0.10 and 0.16.

Yearly Significance Table

Coefficient	1998	1999	2000	2001	2002	2003	2004	2005
lnBlack								
lnLatino								
Pct_Pre1940								
Pct_FemaleHeaded								
Pct_No High School								
Pct_Under 6								

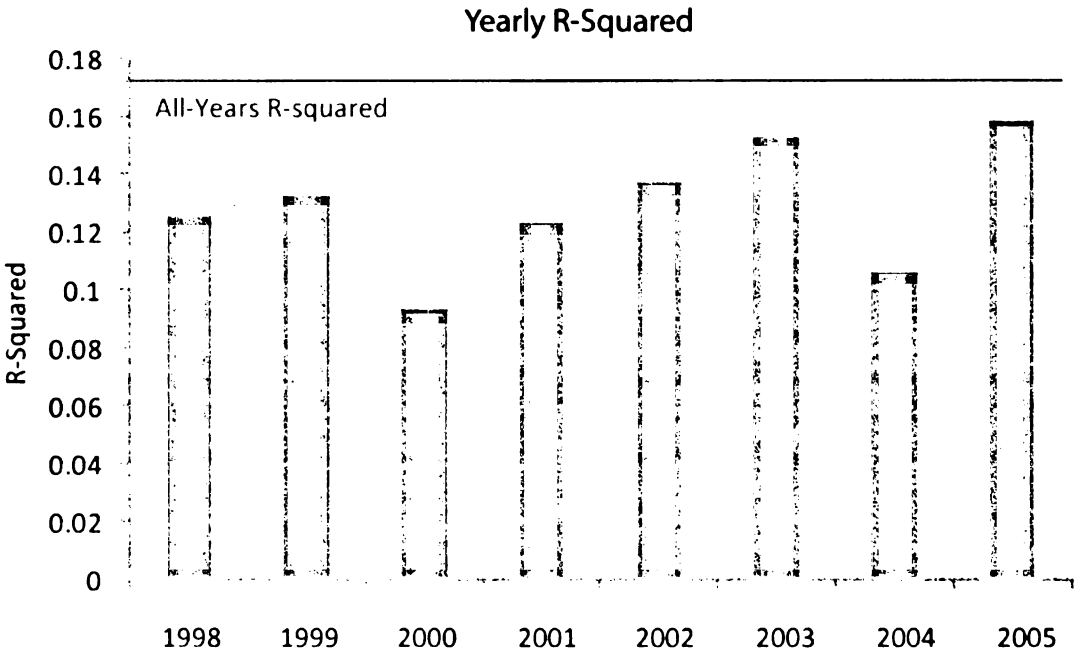


Table 25: Yearly global regression results for minor civil divisions. Light blue represents a significant variable ($\alpha = 0.05$)

A combination of low predictive value and a fairly large bandwidth of around 66 kilometers cause the GWR model for minor civil divisions to be not much of an improvement over the global model. The Leung test in table 26 reveals that the GWR model did significantly reduce the sum of squares of the residuals from 798.17 in the original global model to 611.33. The variable percent under 6 years of age has a very large difference between the median GWR model coefficient value and the coefficient from the global linear model. The likely cause is that some outlier areas of the state may show a very strong link between this variable and mean BLL, but it is less predictive for the state as a whole.

Summary of Regression Coefficients						
	Minimum	1st Quartile	Median	3rd Quartile	Maximum	Global
Intercept	0.03069	2.045	2.651	2.963	3.643	2.0754
lnBlack	-0.01174	0.01537	0.02971	0.04694	0.1001	0.0378
lnLatino	-0.09798	0.02673	0.1022	0.1442	0.1981	0.0411
Pct Pre1940	-0.3829	0.9842	1.235	1.875	3.178	1.4526
Pct FemaleHeaded	-0.7762	0.37	0.9645	1.353	2.894	0.7856
Pct No High School	-6.46	1.292	2.637	3.558	4.633	3.0166
Pct Under 6	-8.573	-4.298	0.8652	3.637	29.72	3.7443

Fixed Bandwidth (meters)	66745.8
Number of Data Points	1528
Effective number of parameters	86.05015
Effective degrees of freedom	1441.95
Sigma Squared	0.4000854
AIC	2998.605
Residual sum of squares	611.3304

Leung Statistic	
OLS Residuals Sum of Squares	798.1792
GWR Residuals Sum of Squares	611.3304
F - Statistic	0.8079
p - value	1.92E-05

Table 26: GWR regression results for minor civil division all years mean BLL

For the minor civil division level model, the GWR maps are of little value. In general, the large bandwidth size resulted in stripe-like patterns across the state. The

pattern across the state for the R^2 is very smooth and not reflecting the pockets of high and low mean BLL that exist (Figure 74). The highest R^2 values appear to be in the southwest corner of Michigan. A likely reason is that the southwestern portion of the state seems to have higher mean BLL values in many of the rural townships. Since cities are single units at the minor civil division level, the rural areas have more influence on the model result.

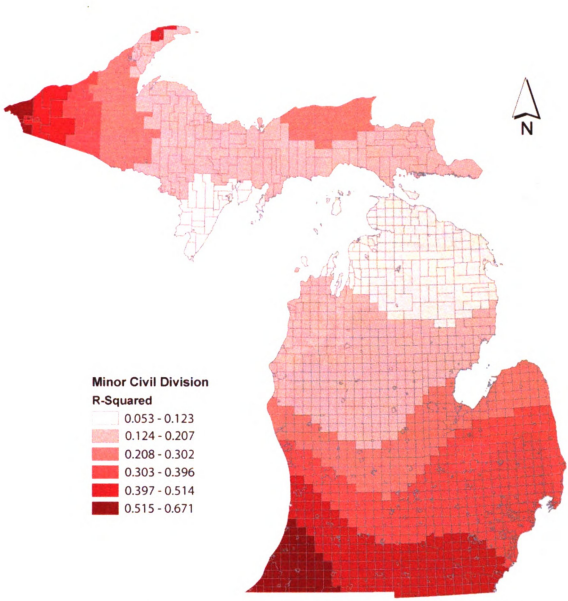


Figure 74: Map of the R-Squared for the minor civil division GWR model

The map of coefficients for the variable percent pre-1940 housing shows the influence of Detroit. The high coefficient values reveal that older housing is having a large amount of influence on the model. The map in figure 75 does not reveal however the variability that likely exists throughout the state. The larger bandwidth size, caused

by the low predictive ability of the variables at the minor civil division level, is causing many likely pockets of the state such as Grand Rapids to be missed.

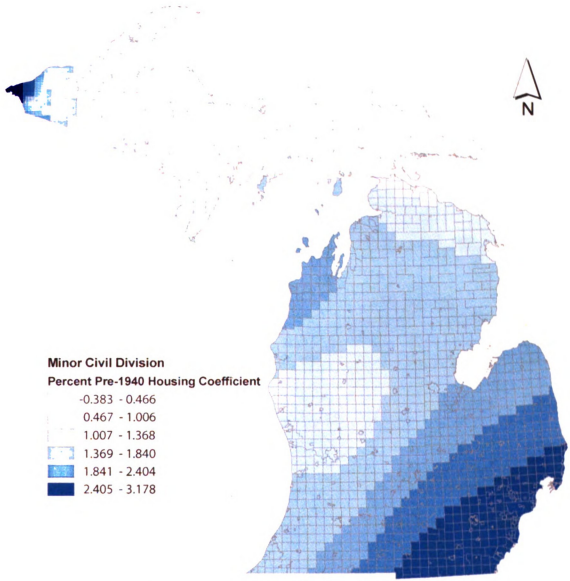


Figure 75: Map of the coefficients from the minor civil division GWR model for pre-1940 housing

3.2.2 Zip Code

The second areal unit regression analysis involved US postal zip codes for Michigan. Similar to Census tracts, the highest mean BLL numbers were found in the urban zip codes. Other prominent areas include the southwest corner of the state as well as parts of the southern border of the Lower Peninsula. The standard deviations map (Figure 76) shows that the rural areas of the state are more volatile year-to-year in mean BLL than the urban areas of the state.

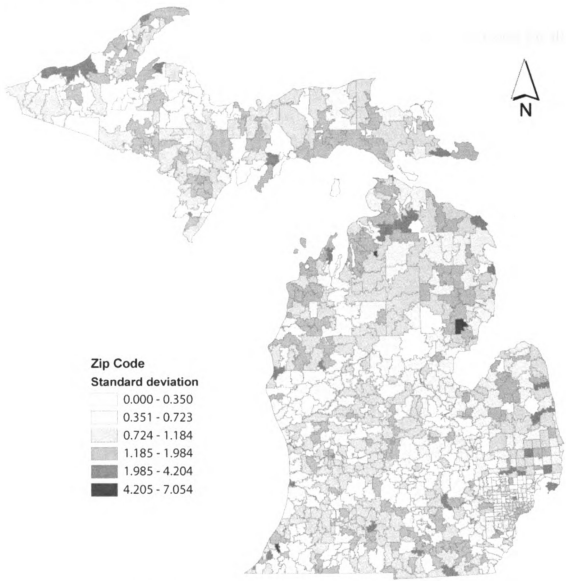
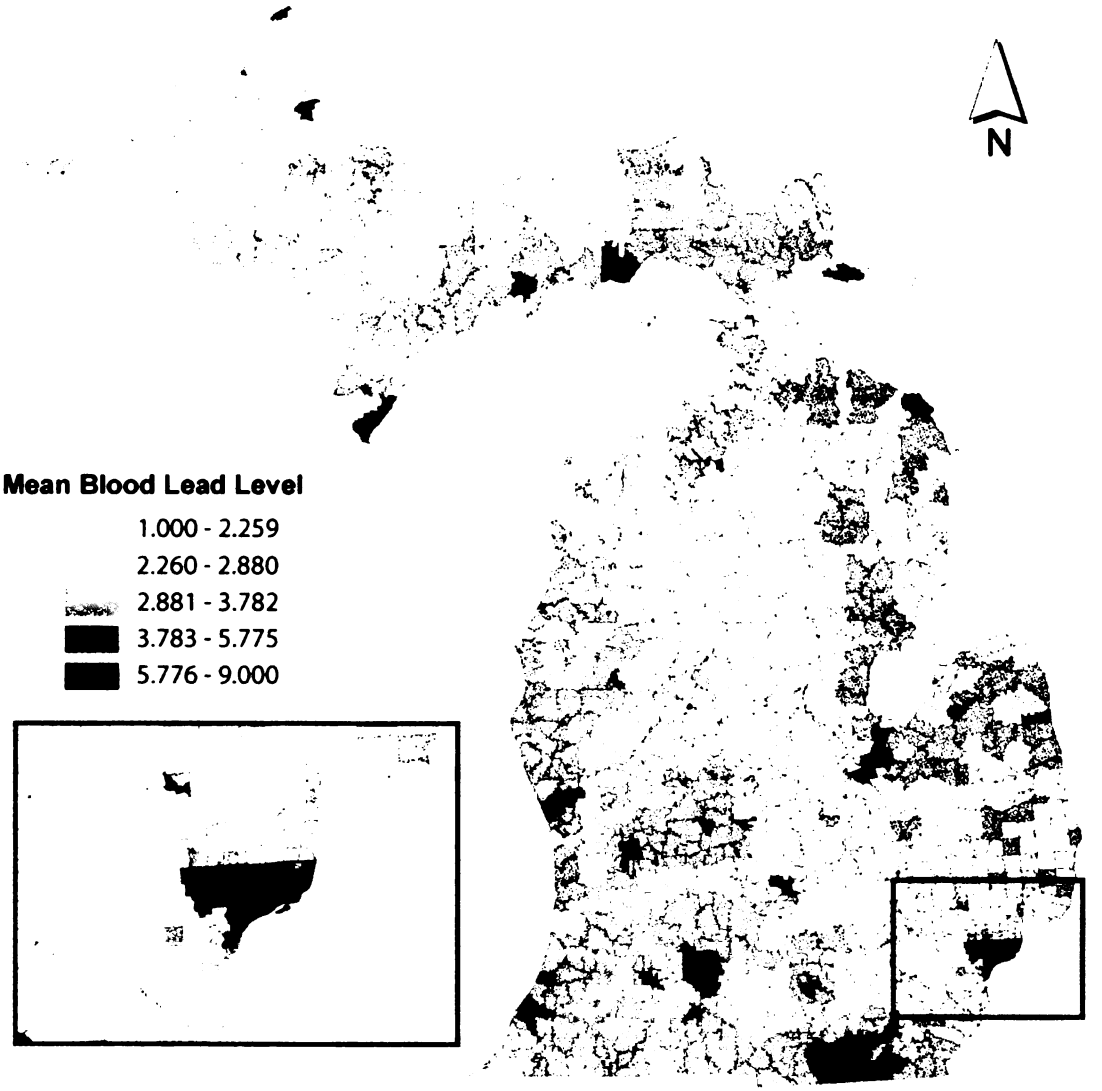


Figure 76: Map of zip code standard deviations of the yearly mean BLL

Nine variables from the original choices were used in the global model for zip codes. Though more variables proved to be significant ($\alpha = 0.05$) than in census tracts, the t-values are not as high. The most significant variable proves to be percentage pre-1940 housing. This is not surprising given similar results seen in other areal units. What is interesting in the t-values is that both Percentage African-American and Percentage Latino are well above the other remaining variables (Figure 77). This could suggest the strength of ethnicity as a strong predictor at the zip code level. Overall, the model for all years had an R^2 value of 0.41.



Coefficients	Estimate	Std. Error	t-value	Pr(>|t|)
(Intercept)	1.99385	0.173767	11.474	2E-16
lnBlack	0.081036	0.010365	7.818	1.18E-14
lnLatino	0.101676	0.013875	7.328	4.30E-13
lnRecent Immigrants	0.030678	0.008648	3.547	0.000404
Pct_Rental	0.973719	0.272797	3.569	0.000372
Pct_Vacant	0.70847	0.177352	3.995	6.88E-05
Pct_Pre1940	2.04307	0.208258	9.81	2E-16
Pct_FemaleHeaded	1.609344	0.282665	5.693	1.57E-08
Pct_No High School	2.815125	0.551952	5.1	3.94E-07
Pct_Under 6	7.864077	1.567739	5.016	6.07E-07
R2 = 0.4164				
Adjusted R2 = 0.4119				
F-statistic = 94.01 on 9 and 1186 Degrees of Freedom				2E-16

Figure 77: Map of mean BLL by zip code and all years global regression results

The yearly models for zip codes proved that independent variables in the all years model may not represent significance on a yearly basis (Table 26). The clearest example is percentage houses rented and percentage houses vacant, which both are significant in the all years model, but are rarely significant in an individual year. Often these variables have opposite positive and negative coefficients, indicating likely colinearity in the individual year's model. Several other variables such as percentage recent immigrants and percentage without a high school diploma show varying levels of significance. The yearly models reinforce the strength of three variables: percentage pre-1940 housing, percentage African-American, and percentage Latino. Similar to the other areal units, the zip code yearly R^2 falls below the all years model, with a range around 0.30-0.38.

Yearly Significance Table

Coefficient	1998	1999	2000	2001	2002	2003	2004	2005
lnBlack								
lnLatino								
lnRecent Immigrants								
Pct_Rental								
Pct_Vacant								
Pct_Pre1940								
Pct_FemaleHeaded								
Pct_No High School								
Pct_Under 6								

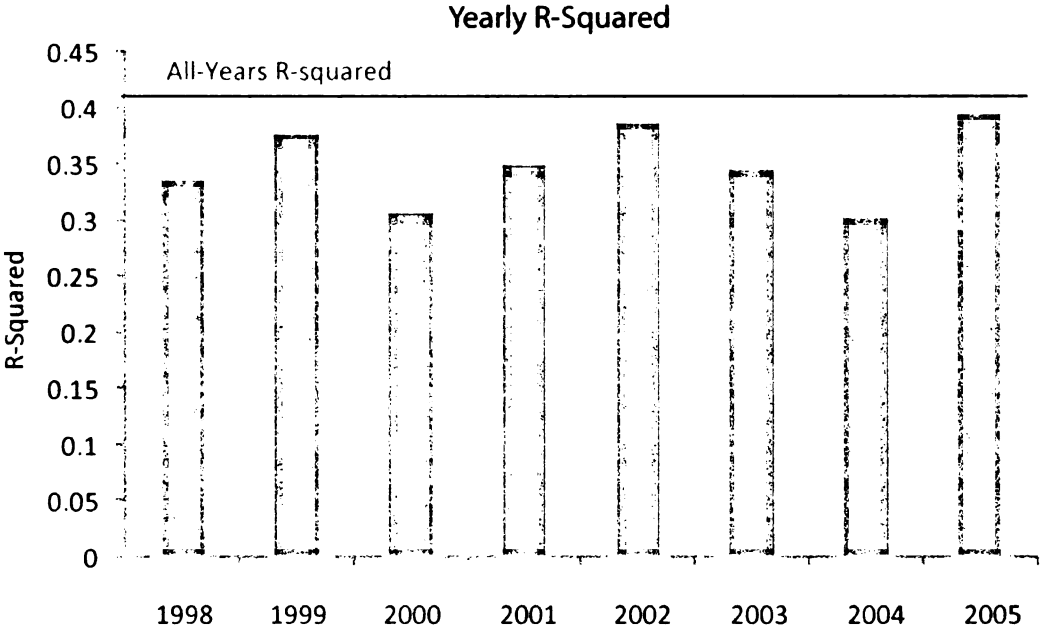


Table 27: Yearly global regression results for zip codes. Light blue represents a significant variable ($\alpha = 0.05$)

The GWR model for zip codes turned out to be a case of a better model does not necessarily improve the analysis capabilities. The Leung test for the GWR model versus the global model showed that using the GWR model significantly reduced the sum of squares of the residuals (Table 28). This would indicate that the model is better at predicting the mean BLL than the global model. What is interesting is that the reduction of the residuals for zip codes was the lowest of any of the three geographic units. In the

summary of coefficients, the large difference between the median GWR coefficient for the variable percentage under 6 years of age and the global linear coefficient.

Summary of Regression Coefficients						
	Minimum	1st Quartile	Median	3rd Quartile	Maximum	Global
Intercept	0.2863	2.448	2.674	2.917	3.516	1.9939
lnBlack	0.04087	0.072	0.08873	0.0924	0.09548	0.081
lnLatino	-0.01687	0.1358	0.1508	0.1713	0.1943	0.1017
Pct_Recent Immigrant	0.009961	0.0255	0.0291	0.031	0.061	0.0307
Pct_Rental	-1.289	0.1662	0.4538	1.111	3.382	0.9737
Pct_Vacant	-0.6116	0.517	0.8137	1.04	2.065	0.7085
Pct_Pre1940	-0.04869	1.537	2.437	2.922	3.325	2.0431
Pct_FemaleHeaded	-0.5557	0.7076	1.919	2.757	3.551	1.6093
Pct_No High School	0.03435	0.844	1.56	2.941	4.759	2.8151
Pct_Under 6	-1.957	-0.5089	1.89	7.457	21.13	7.8641

Fixed Bandwidth (meters)	117372.7
Number of Data Points	1196
Effective number of parameters	51.7817
Effective degrees of freedom	1144.218
Sigma Squared	0.6535902
AIC	2924.042
Residual sum of squares	781.6938

Leung Statistic	
OLS Residuals Sum of Squares	945.0133
GWR Residuals Sum of Squares	781.6938
F - Statistic	0.8574
p - value	0.004238

Table 28: GWR regression results for zip code all years mean BLL

Similar to the minor civil division, the zip code GWR model suffers from a weaker weighting scheme. The bandwidth for the all years model for zip codes was around 117 kilometers, which is twice as high as minor civil divisions and nearly 5 times as high as census tracts. While the cross-validation algorithm chose this bandwidth because reduced the sum of squares to the greatest degree, it provides little sound mapping examples. In the R^2 map in figure 78, the values trend downward as distance from Detroit increases. Similar patterns can be seen in the individual variable maps.

What this indicates is that there is a spatial component to mean BLL at the zip code level and that including a spatial component does improve the predictive power. Unfortunately, the linear nature of this spatial component indicates that the model is not picking up the pockets of spatial variation seen in the census tracts GWR model. In all likelihood, an independent variable based in latitude would likely work as well.

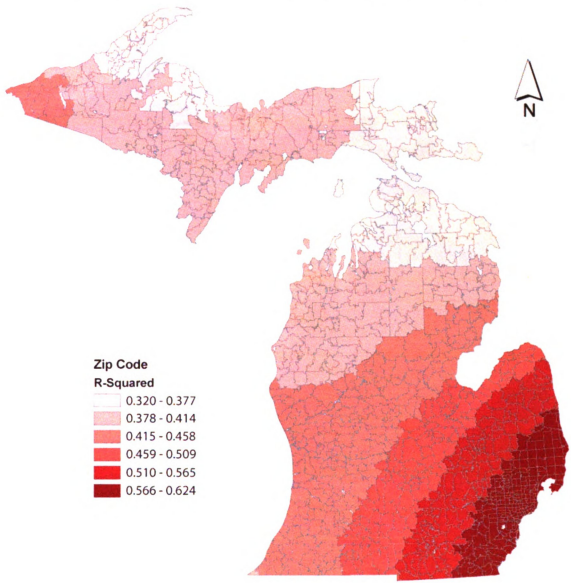


Figure 78: Map of the R-squared for the zip code GWR model

In both zip code GWR models as well as the earlier minor civil division model, the variable percent under 6 years of age produces the widest variability in coefficient values. Figure 79 shows the map for coefficients for the percentage under 6 years of age. The highest coefficients are in the far western areas of the Upper Peninsula. What could be behind the high coefficients is that many other predictive variables such as percentage African-American are not a big factor.

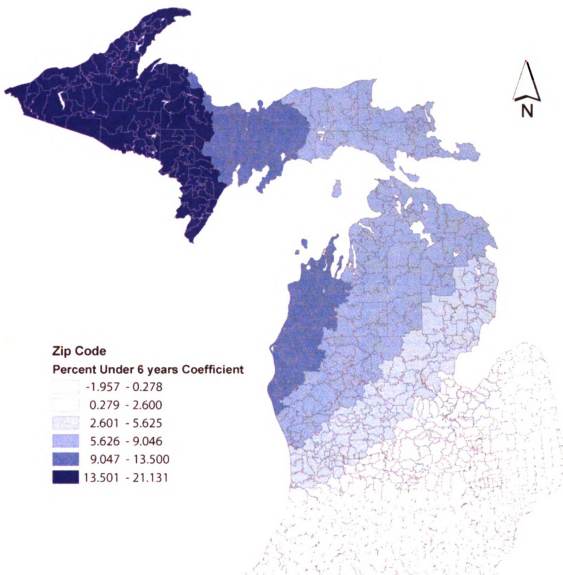


Figure 79: Map of the coefficients from the zip code GWR model for percentage under 6 years of age

3.2.3 Tract

Census tracts were the third areal unit examined by regression analysis. The preference of the US census bureau for relatively homogenous populations when drawing up the boundaries of tracts is a great advantage for regression. There is often a sharp divide between the means in neighboring tracts. Each yearly map of BLL means yields similar results. To test the yearly variability in the mean BLL, the standard deviation was computed for each tract. The resulting map shows the strongest deviations scattered among more rural or suburban tracts (Figure 80). A closer examination showed high standard deviations were usually due to a couple factors: the presence of a high BLL outlier case, a low test population, and generally low BLL test results in the tract.

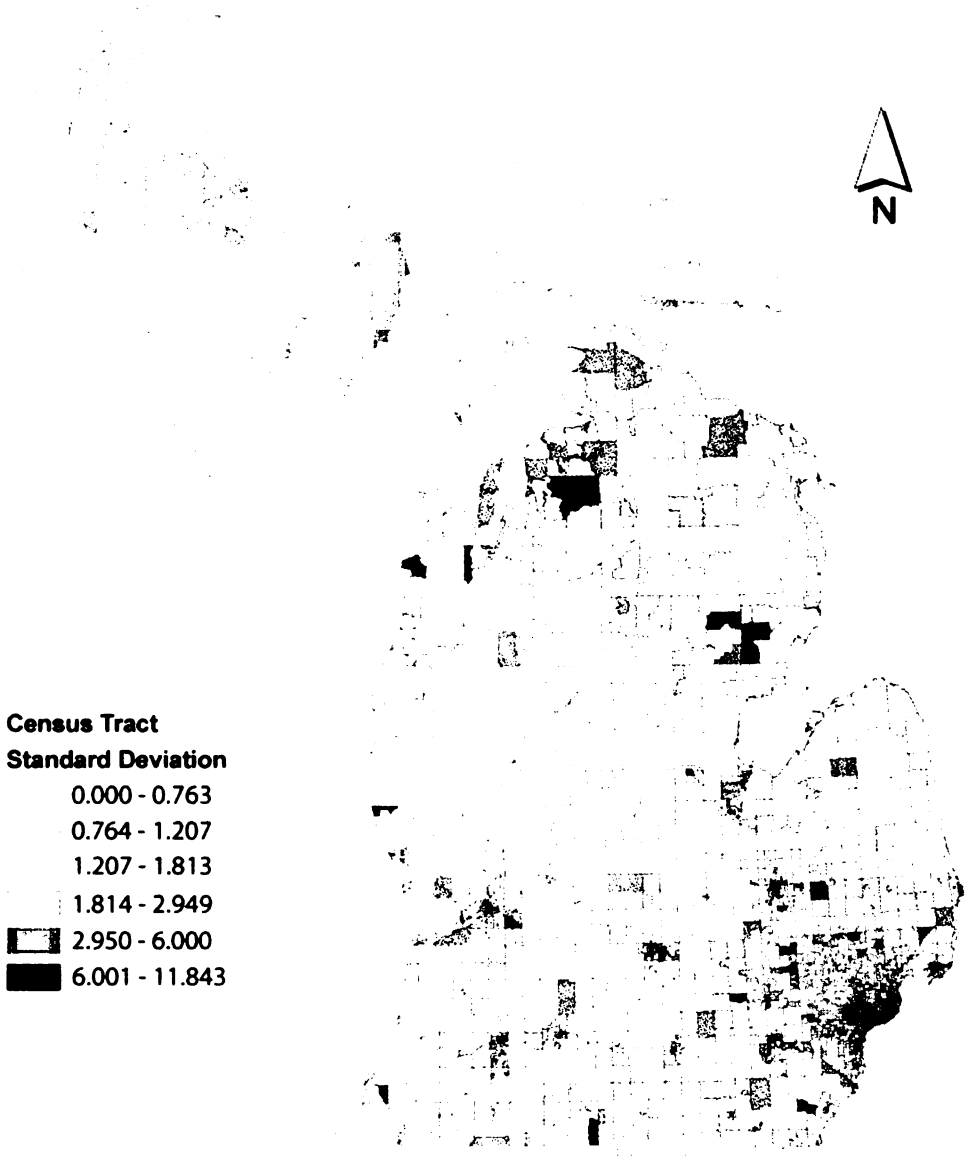
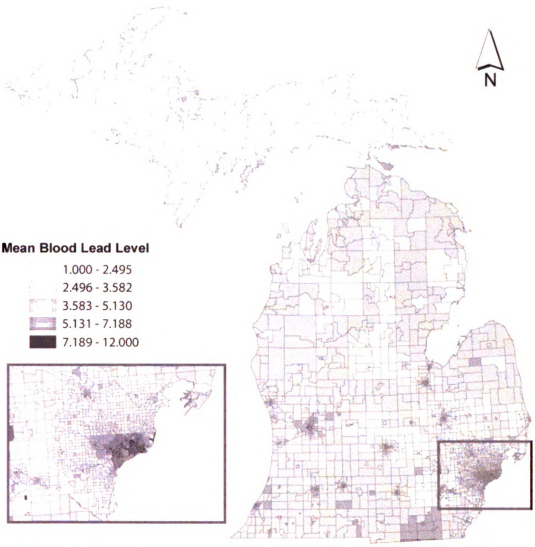


Figure 80: Map of census tract standard deviations of yearly mean BLL

The results of the regression analysis on Census tracts yielded the best and most conclusive results (Figure 81). In the global regression, the eight independent variables yielded an R^2 value of 0.67 for elevated BLL data covering all years. All of the independent variables yielded p-values that were highly significant. Not surprisingly, the percentage of pre-1940 homes within the tract is the most significant variable, with a t-value at 35.6. The percentage of African-American residents and percentage of

households headed by a woman only were also highly significant. Note that at the Census tract level, the percentage of Latino residents had a negative effect on the mean BLL in a tract. This is different from what was found in the MCD or zip code regressions.



Coefficients	Estimate	Std. Error	t-value	Pr(>|t|)
(Intercept)	1.18331	0.13137	9.007	2E-16
lnBlack	0.19279	0.01013	19.022	2E-16
lnLatino	-0.17622	0.01828	-9.639	2E-16
Pct_Rental	-0.54969	0.1212	-4.535	0.000006
Pct_Vacant	0.87723	0.16228	5.406	7.03E-08
Pct_Pre1940	3.88197	0.109	35.616	2E-16
Pct_FemaleHeaded	1.94133	0.16194	11.988	2E-16
Pct_No High School	3.45028	0.34639	9.961	2E-16
Pct_Under 6	6.01755	0.81978	7.34	2.8E-13
R2 = 0.6724				
Adjusted R2 = 0.6714				
F-statistic = 693.2 on 8 and 2702 Degrees of Freedom				2E-16

Figure 81: Map of mean BLL by census tract and all years global regression results

In addition to testing the independent variables against the mean BLL results for all years, the predictors were tested against the mean BLL in the tracts for each year (Table 27). A glimpse at the R^2 across the eight years shows a range of about 0.44 to 0.53. This is below the R^2 for the all years model and likely reveals some volatility in the yearly mean BLL numbers. The global regression analysis by year confirms that both pre-1940 housing and percentage African-American are the strongest predictors. In every year, their p-value is highly significant. The percentage of houses within a tract that are vacant shows itself to be a worst predictor when looking at individual years.

Yearly Significance Table

Coefficient	1998	1999	2000	2001	2002	2003	2004	2005
lnBlack								
lnLatino								
Pct_Rental								
Pct_Vacant								
Pct_Pre1940								
Pct_FemaleHeaded								
Pct_No High School								
Pct_Under 6								

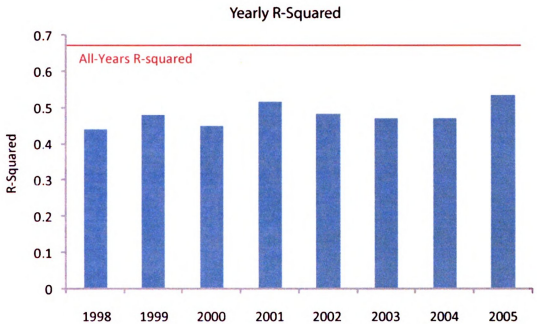


Table 29: Yearly global regression results for census tracts. Light blue represents a significant variable ($\alpha = 0.05$)

The GWR model, where individual regression analyses were run on each tract based on a weighting scheme, performed better at reducing the sum of squares of the residuals than the global model according to the Leung test statistic (Table 30). This statistic showed vast improvement in the predictive capability of the GWR model. This might be linked to the lower bandwidth value, around 25 kilometers. The median coefficient values for all the individual GWR models are similar to the coefficient values

from the global linear model. The largest exception seems to the percentage of vacant houses within the study region. In addition to being the least consistent variable in the yearly global linear models, the effect on mean BLL the percentage of vacant houses is responsible for seems to vary widely across the state.

Summary of Regression Coefficients						
	Minimum	1st Quartile	Median	3rd Quartile	Maximum	Global
Intercept	-3.497	1.106	1.396	2.513	9.952	1.1833
lnBlack	-0.1357	0.09274	0.2029	0.2434	0.4129	0.1928
lnLatino	-0.3762	-0.219	-0.1533	0.03366	0.6571	-0.1762
Pct_Rental	-25.34	-1.136	-0.7168	-0.2082	5.152	-0.5497
Pct_Vacant	-9.62	1.049	3.329	4.834	6.543	0.8772
Pct_Pre1940	-1.074	2.647	4.084	4.549	5.684	3.882
Pct_FemaleHeaded	-6.749	0.858	1.797	2.015	11.52	1.9413
Pct_No High School	-13.36	2.365	2.733	3.054	9.65	3.4503
Pct_Under 6	-28.28	1.533	4.739	5.973	40.57	6.0175

Fixed Bandwidth (meters)	25539.43
Number of Data Points	2711
Effective number of parameters	339.184
Effective degrees of freedom	2371.816
Sigma Squared	0.4891147
AIC	6020.737
Residual sum of squares	1325.99

Leung Statistic	
OLS Residuals Sum of Squares	2115.556
GWR Residuals Sum of Squares	1325.99
F - Statistic	0.714
p - value	2.2E-16

Table 30: GWR regression results for census tract all years mean BLL

The real value of GWR and where the census tract model really shines is the maps of coefficients. A map of the R^2, shown in figure 82, reveals that the model works very well in urban areas, but also in some of the rural areas as well. Grand Rapids stands out as an area where the model is highly effective among the urban areas of Michigan, with Detroit and Flint visible to a lesser degree. The model is also effective on much of the Upper Peninsula, particularly in the far western end as well as the Sault St. Marie area.

Finally, the center of the Lower Peninsula shows rural areas where the model works effectively as well.

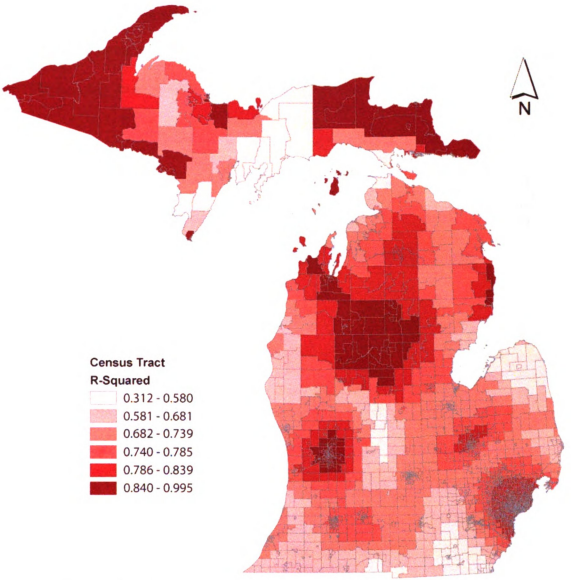


Figure 82: Map of the R-Squared from the census tract GWR model

The maps of the coefficients for each of the variables give an important clue as to what parts of the state each variable is contributing most. For the percentage African-American variable, the Grand Rapids and Detroit areas show the highest positive coefficients (Figure 84). According to this model, in the two largest cities in Michigan,

the areas that have the higher percentages of African-Americans have the higher mean BLL. This pattern is largely repeated in the map of coefficients for percentage houses built before 1940 (Figure 83). Detroit and Grand Rapids continue to stand out well beyond the rest of the state. The two main variables, percentage African-Americans and pre-1940 housing, exert the greatest influence in Michigan's urban areas

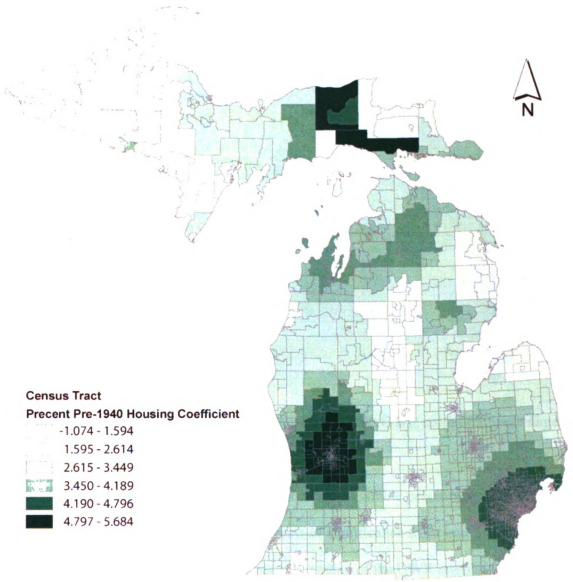


Figure 83: Map of the coefficients from the census tract GWR model for pre-1940 housing

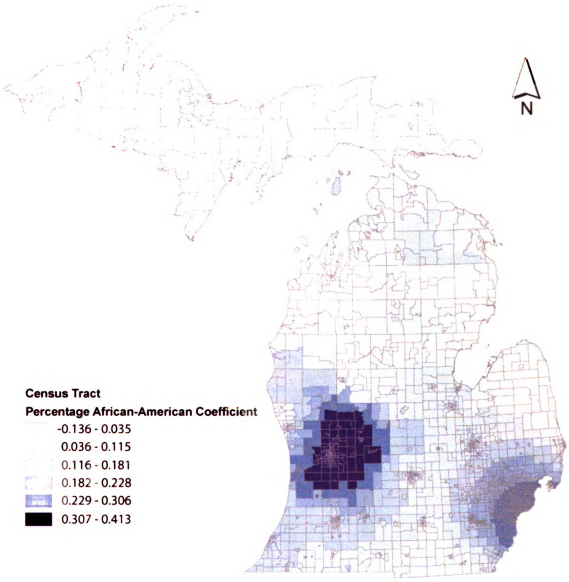


Figure 84: Map of the coefficients from the census tract GWR model for percentage African-American

The final map is the map of coefficients for the variable percentage vacant houses. This was the most inconsistent variable in terms of significance from year to year and the variable that had a large difference between the median of the GWR coefficients and the global coefficient. The map in figure 85 reveals the likely cause of this disparity. Percentage vacant houses seem to have a large effect in the southern areas of Detroit and

extending down to the Ohio border. But in the Grand Rapids area, the variable has no effect. This disparity could be the underlying cause behind the inconsistent performance of vacant houses as a predictor of mean BLL.

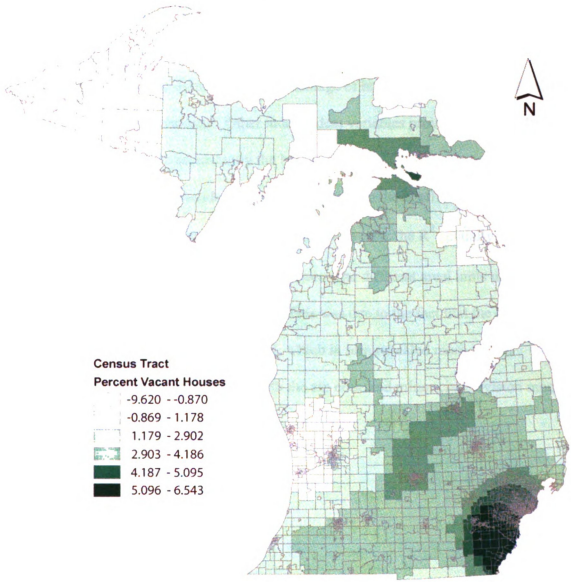


Figure 85: Map of the coefficients from the census tract GWR model for percentage Vacant Houses

The overall results of the regression analysis prove the importance of the unit of analysis as well as the independent variables used. In all three areal units, three of the

variables (percentage African-American, percentage Latino, and percentage recent immigrants) were logged in order to give the data values a normal distribution. Each of the three different areal units tested produced very different outcomes of what census variables were significant and how much of the variance in mean BLL could be explained. One constant throughout the different units of analysis was the two main variables that proved most significant, the percentage of houses built before 1940 and the percentage of African-Americans. Other independent variables proved to be significant as well, but these two were consistently the best predictors.

The GWR analysis provided an opportunity to map the coefficients of each variable in every regression run as well as the chance to view the R^2 spatially. The mapped results showed the great difference between the different areal units used. Census tract analysis proved best for GWR. This was due to the fact that the independent variables were better predictors at this level, which in turn revealed more spatial variation. The low predictive ability of both the zip code and minor civil division models made GWR analysis basically worthless.

4 Conclusions

4.1 Overview

The legacy of commercial lead usage continues to affect Michigan children to this day. The large amount of lead used in early 20th century products made the element accessible to children. Industry pressure and dismissal of medical evidence allowed lead usage in paint and gasoline to continue in the United States much longer than other developed nations. For many years, the warning signs of lead poisoning in children were dismissed and many suffered grievous injury and even death. As lead was phased out of paint and gasoline in the 1970s, the number of serious clinical cases of lead poisoning has dropped.

New research has shown that sub-clinical levels of lead in a child's body cause irreparable harm. Though chelation therapy can be used to slowly cleanse the body, the only sound solution to the problem of lead in the human environment is primary prevention. This tactic has been emphasized within the United States since passage of Title X in 1992. The state government of Michigan responded in 1998 with the Lead Abatement Act, which provided funds for reducing elevated BLL in Michigan through the creation of database of all blood test results of children and eradicating lead from dangerous home environments. Supplemental legislation in 2004 has worked to streamline the testing process and setting a firm goal of eliminating elevated BLL within Michigan by 2010.

This thesis utilized the Michigan Department of Community Health (MDCH) database of child blood lead test results from 1998 to 2005 in order to study the spatial

patterns of distribution. The research was limited to children on Medicaid, two-thirds of the original database, to deal with sampling issues. This database was created by MDCH from all the testing labs in Michigan by law. Information available included the child's address, age, test result (in μg/dL), test type, and the data the blood test occurred.

For all children tested more than once, the highest test result was used. The research examined at both the point patterns based on the children's addresses as well as areal analysis the characteristics of the neighborhoods based on US Census data. Several different clustering techniques were used in order to examine the number of neighbors, size of the cluster in terms of distance, and the likely locations of clusters. Each test was done on the data from every individual year of lead testing in order to look at possible changes over time. Because of computing limitations, the state was divided into nineteen different study areas. In the census-based analysis, variables that had been found to be significant in previous studies of spatial variation in lead poisoning were tested in Michigan. Regression analysis in this thesis was run on three different areal units, all of which were used in previous spatial-based childhood BLL studies. Geographically Weighted Regression was employed to visually understand how well the model works in various portions of the state and how the independent variables changed over space.

A number of conclusions can be drawn from the results of the clustering and regression methods about childhood BLL in Michigan. Listed below is a summary of the major points that emerged:

1. Elevated BLL in children insured by Medicaid is clustered in Michigan.
2. Clusters of elevated BLL are most considerable in the urban areas of the state.

3. The size of clusters is greatest when 5 μg/dL is used as the partition between cases and controls. When 10 μg/dL is used as the divide, the size of the clusters is smaller. Clusters of elevated BLL cases at the 25 μg/dL partition are only common in the more populated study regions such as South Detroit.
4. In Federal Urban Aid Boundary-based study areas, the central city and surrounding neighborhoods display elevated BLL hotspots.
5. Rural study regions that lack a central city do not typically display clustering of elevated BLL regardless of what partition of μg/dL is used.
6. In HSA-based study areas, presence of clustering is dependent on a moderate to large city within the region. The only consistent hotspots in the study region are centered on these cities.
7. The choice of areal unit in regression analysis is critical to the predictive capability of the regression model. With the independent variables used in this thesis, US Census tracts explain the variance in mean BLL to the greatest degree. The same variables at zip code level explain the mean BLL variance to a lesser degree, and have a low predictive ability when aggregated to minor civil divisions.
8. The percentage of an area's housing that was built before 1940 was the best predictor of mean BLL. The next best predictor of mean BLL was percentage of an area of African-American ethnicity.
9. The Geographically Weighted Regression (GWR) model for census tracts confirmed that the Detroit and Grand Rapids had the highest positive coefficients in the state for both the percentage pre-1940 housing and

percentage African-American variables, indicating that these two cities exert the greatest influence over the statewide model.

4.2 Discussion of Results

4.2.1 Clustering

A thorough search of the academic literature found no studies where clustering methods were used to identify areas of lead poisoning. Typically, such techniques are more suited for study of infectious diseases to identify hotspots and clusters where a disease epidemic is occurring. For a chronic disease such as lead poisoning, the hazard is mostly stationary because the lead threat is fixed in the local environment. The clustering methods presented in this thesis as well as others available in the literature have value for evaluating lead poisoning cases.

Three different methods for analyzing point patterns were utilized for this thesis. Each method uncovered a different aspect of the point patterns. Cuzick-Edwards tests were used to reveal the size and significance of clusters of elevated BLL cases based on neighbor analysis. The difference of K graphs was used to understand the size and significance of clusters based on distance. Finally, Geographic Analysis Machine (GAM) maps were created to highlight hotspots where clustering was likely occurring. The results from all three tests reveal distinct patterns of elevated BLL throughout the state of Michigan.

All evidence in the clustering methods points to the severity of lead exposure in urban areas. The Cuzick-Edwards statistic and the difference of K graphs both provided a sort of informal ranking of the study regions as to the severity of elevated BLL. At the

top of this ranking are the metropolitan areas of Detroit (represented by two study areas) and Grand Rapids. Each showed extraordinary amounts of clustering of cases at all three thresholds, evidenced by the highly significant test statistic values in the Cuzick-Edwards statistics as well as the difference of K values which rose quickly above the upper bounds of the simulation envelopes. The GAM maps showed that the hotspots of elevated BLL occurred primarily in the urban core of each city.

A second level of the informal ranking was middle to small-sized cities. These were study areas such as Lansing, Flint, Kalamazoo, Battle Creek, and Saginaw/Bay City. The three clustering techniques revealed as high amount of clustering among the lower thresholds of 5 and 10 μg/dL, but diminished at the 25 μg/dL threshold due to the lack of cases. Often the 5 μg/dL threshold had clustering levels nearly as high as the major cities, but the 10 μg/dL threshold showed a noticeable drop off in the size of the clusters. This is evident in both the Cuzick-Edwards and the difference of K graphs, leading to the conclusion that there are small pockets of lead poisoning cases in urban study regions. The GAM maps demonstrated that the hotspots were in the central sections of the mid-sized cities, similar to Detroit and Grand Rapids but on a smaller scale.

The third level in the ranking was HSA-based areas that had cities or several large towns within them. These included the Southwest, Southeast, Mid-South, and Lower Coast regions. Similar to the smaller cities, these regions displayed clustering at the 5 μg/dL threshold level. At the 10 μg/dL threshold, clustering results are typically much weaker and vary in significance year to year. The GAM maps for these regions were also more difficult to interpret due to the large number of single case hotspots. Having a

lower case/control ratio than the urban study areas causes these hotspots. The resulting maps show a constellation of hotspots that shift from year to year. But in each of the four study regions in this level, one constant is a hotspot centered on an urban area. This primary city is certainly the source of clustering seen throughout the region.

The fourth and final tier of the informal ranking from the clustering analysis was the more rural areas. These were the Upper Peninsula study areas, North Central, West Bay, East Bay, and the Mid Coast. They were characterized by some clustering at the 5 μg/dL threshold, occasionally picked up by the Cuzick-Edwards test. But overall, the regions displayed little if any clustering. GAM maps were less useful in these regions because a hotspot could be just one case. In such instances, investigators would not need to consult clustering maps and would likely not rely on clustering methods.

While these results seem fairly conclusive, there are lingering questions with regards to the point-based clustering analysis. The most important uncertainty is the validity of the sample. This thesis used statewide testing data, numbering in the hundreds of thousands, for analysis. The study was limited to Medicaid-only children, a majority of the MSU database, so that the sample constituted a better representation of the underlying population at risk. Since Medicaid requires recipients to undergo a blood test for lead, this population is more represented in the test results than the Michigan population as a whole. Still, limiting the study to Medicaid-insured children carries biases as well. The population and spatial distribution of children in Michigan may be different than Medicaid-insured children. This difference could complicate clustering and hotspot analysis and lead to false conclusions.

A question or issue that also inevitably arises is the idea that the clustering methods are only showing clusters in cities due to the high number of test results. This idea does lend itself to some credence given the impressive stratification of clustering within the state almost entirely based on population. However, there are some factors to consider. First, the task of looking at lead poisoning across an entire state means that much of the local variation can be missed. The individual clusters picked up in the Cuzick-Edwards and difference of K measures may not perfectly translate to GAM analysis. In GAM, what looks like a hotspot containing an entire city may be a coarser picture of the local spatial variation. But the fact that GAM worked much better in urban areas at pinpointing locations of elevated BLL makes it a useful tool.

The relationship between size of the city and cluster magnitude demonstrates that the highest BLL cases are still in major cities with a few exceptions visible. The 25 μg/dL threshold probably best illustrates the significance of elevated BLL in the major cities. Cases of BLL 25 μg/dL and above are the most indicative of a major problem, and the fact that they are almost exclusively found in the major urban areas negates the assumption that all the clustering was only due to a larger number of samples. The second point is that a few major cities of Michigan did not fit the ranking rule that developed. The most obvious case was Midland, which is in a study region where it is the only major town, but still did not show up as a cluster or hotspot on the GAM maps.

Each individual clustering method that was used has both an upside and downside to implementation. The main upside to the Cuzick-Edwards statistic is that in not considering distance, the results can pick up clusters in both cramped urban areas and spread-out rural study regions. While this is useful, it did not seem to factor into the

results from this thesis. The mostly rural study areas of the state did not seem to display clustering at any level without the presence of a moderate-sized town or city. Meanwhile, even with the larger number of control test results, nearly every urban aid boundary-based study region showed significant clusters at the 5 and 10 μg/dL threshold levels. The downside to the Cuzick-Edwards is related to the upside. The distance between the nearest 20 neighbors is much closer in urban areas than in rural areas. Twenty neighbors in an urban area likely constitute a neighborhood, while twenty neighbors in a more rural area are likely much more dispersed. Since clustering analysis seeks to link cases within a cluster, this can complicate matters in rural areas. For this thesis, the downside of Cuzick-Edwards seems to be mostly mitigated due to the differences in clustering results between urban and rural study regions. The urban areas of the state showed much stronger clustering than the rural areas, leading to the conclusion that certain areas of Michigan cities exhibit high lead exposure risk.

The main drawback to the difference of K method is the problem of edge effects. The study area boundaries can have an effect on the results. There are examples in this thesis. The smallest study area, Battle Creek, has a quick drop in K values right after four or five kilometers. This is not due to the sudden loss of cases as much as the concentric circles extending beyond the boundaries of the region. Another drawback to the difference of K method is difficulty of interpretation. The K values can be inside or outside the simulation envelope depending on the simulation results, a situation that can lead to confusion about significance. In this thesis, clustering was assumed to only be occurring when the difference of K values far exceeded the upper bound of the simulation

envelope. Most study areas with clustering of elevated BLL have difference of K values well above the envelope, leaving the ambiguity problem most mute.

The greatest weakness of the GAM analysis turned out to be the case/control ratio for each study area baseline rate. The ratio of cases to controls in many rural regions of the state was much smaller than in the more urban regions of the state. This meant that the hotspots in rural study areas often only had one case in them. This is significant for remediation, but it does not count as a cluster. This leads to a varying pattern of hotspots year to year. Identifying places with higher threats from lead exposure becomes more difficult. More urban areas that had a larger ratio of cases to controls were more successful at identifying consistent hotspots, but individual cases outside of the main clusters could be missed. This becomes a problem when the area the individual case's area is under-sampled, but contains environmental lead hazards.

4.2.2 Geographically Weighted Regression

The clustering portion of this thesis answers many of the questions as to where the hotspots of elevated BLL were located, but regression analysis can provide insight into why these clusters occur and who is most affected. The results of the regression analysis confirmed that the spatial patterns in Michigan were similar to what was seen in earlier studies of other locations. The main predictor of children's BLL was older housing. This is to be expected. Pre-1940 housing showed up as the main predictor on all three different areal units as well as during almost every individual year. Another variable that was significant was percentage of African-Americans. The positive coefficients associated with the percentage African-American variable around the high mean BLL

cities of Detroit and Grand Rapids suggest that children of this ethnicity are likely the primary victim of lead exposure.

Beyond older housing and percentage of African-Americans, the three different areal unit global regression models diverged in predictive value. The census tract model was by far the best. This is due to the US census bureau attempts to divide areas into tracts with relatively homogeneous populations. Therefore, the ability of independent variables to explain mean BLL in census tracts is superior due to stark differences in socio-economic conditions in different units. This was a great contrast from the minor civil divisions model. In that model, all spatial and socio-economic variation within the urban areas was lost. Zip codes worked slightly better, but not as well as tracts. The conclusion is that the modifiable areal unit problem is significant in the study of BLL. None of the earlier statewide regression studies (Bailey 1994; Sargent 1995; Talbot 1998; Haley 2004) used census tracts, so they all could have missed much of the spatial variation.

The GWR results were only useful at the census tract level. Both the zip code level analysis as well as the minor civil division level analysis yielded coarse results because the independent variables explained less in zip codes and far less in MCD of the variance when compared to census tracts. As a result, the GWR models for these two areal units used larger bandwidth values for the weighting schemes. The reason was that the geographic variation in mean BLL is not explained well in zip codes and minor civil divisions by the independent variables used. Therefore, larger bandwidths giving greater weight to distant observations are needed to explain the spatial pattern. The resulting maps of the coefficients for zip codes and minor civil divisions had a linear striped

pattern. In this case, adding x and y coordinates as independent variables would have worked just as well.

Census tract results for GWR yielded the most insights. The model, according to the R^2 values, explained variance the best in the urban areas, particularly the two main cities of Detroit and Grand Rapids. It is not surprising that the two most significant variables from the global model, percentage pre-1940 housing and percentage African-American, both had coefficient maps that mimicked the R^2 values fairly well. This would lead to the conclusion that these two variables are linked to urban BLL levels. Since urban mean BLL is more stable year to year than suburban or rural areas, older housing and percentage African-American are the best predictors because they are higher in the cities. Coefficient maps for other variables revealed that they were a greater factor in more rural areas. It is more difficult to discern meaning because the rural areas of the state have more unpredictable mean BLL numbers.

A drawback to running regression analysis across eight years is that the US census data is fixed in the year 2000. Any changes that occurred across the eight years, such as migration of people or the building of new homes, is not available for modeling. Unfortunately, many of the census yearly estimates are completed at large geographic levels such as counties or states. Gathering data at the census tract, zip code, and minor civil division level requires waiting for the decennial census.

4.2.3 Research Questions

At the outset of Chapter 1, this thesis presented three research questions relating to the spatial distribution of elevated BLL in Michigan. Each of these three questions

will be discussed in terms of the stated hypothesis and results from the clustering and regression tests.

(1) Are there spatial clusters of elevated BLL in Michigan? At what spatial scales do these patterns manifest?

The hypothesis of this thesis was that spatial clusters of elevated BLL existed in Michigan's older, urban areas. By all measures, this has been confirmed. The Cuzick-Edwards tests and the Difference of K graphs both confirmed a clustering hierarchy in Michigan. Each found the greatest amount of clustering occurred in urban areas, such as Detroit and Grand Rapids. Smaller urban areas, such as Flint, Lansing, and Kalamazoo, all showed strong signs of clustering as well. In the larger study areas based on HSA boundaries, the occurrence of spatial clusters usually depended on the presence of a city or town within the region. GAM analysis confirmed that hotspots occurred most often in urban areas.

The global regression analysis confirmed the significance of older housing on mean BLL. Each regression models for all three areal units revealed the percentage of housing units within an area that date to before 1940 was the best predictor of BLL. The geographically weighted regression model for census tracts confirmed that the coefficients of the pre-1940 housing variable were greatest in the urban core of Michigan, particularly Grand Rapids. These findings, combined with the clustering results, show that clusters of BLL in Michigan are greatest in the older, urban areas.

The spatial scale of the clustering explored in this thesis was slightly different from Griffith et al (1998). In that paper, changes in the spatial scale of elevated BLL

were evaluated through using hierarchical census units. This thesis used three different areal units that are not hierarchical, but were created by three different supervising bodies. The clustering analysis based on point data in this thesis did provide interesting results for the spatial scale of lead poisoning in terms of both distance and severity.

(2) Are socio-demographic and economic variables in the US Census able to predict and explain the geographic variation in elevated blood lead levels in Michigan children?

Socio-economic and demographic data proved to be effective at predicting BLL in Michigan. The hypothesis put forth in this thesis was that lack of education, recent immigration to the US, lower income, and older housing were predictors of the geographic variation of elevated BLL. The results confirmed two out of the four variables. Virtually every regression model run showed that older housing was the best predictor of BLL. The percentage of residents without a high school diploma was also a good predictor in most regression analyses. The other two variables listed in the hypothesis as likely predictors were disappointing. The US census variable percentage under 185% of the poverty line was not a significant predictor of BLL in Michigan in any of the three areal units. Recent immigration was only significant at the zip code level, and not significant for several individual years of that areal unit. Demographic variables that proved to be effective predictors were Percentage African-American and Percentage Latino.

Overall, the results from this study seemed to fit into a pattern found by other researchers who studied BLL through regression analysis. Four of the geographic studies

listed in section 1.2.3 of this thesis were conducted at a statewide level. Bailey (1994) found in Massachusetts that the percentage pre-1940 housing was the best predictor of the number of children above 25 μg/dL, the dependent variable in the study. Similar results were found in Sargent (1995), who found that both percentage pre-1950 housing as well as percentage African-American was significant predictors. These two variables were also the most significant in two regression studies of New York State: Talbot (1998) and Haley (2004). The similarity of the patterns seen in this thesis in Michigan compared to previous studies in Massachusetts and New York reveal the same factors at work. Older urban housing within the cities seems to be the primary source of lead exposure, with African-Americans suffering the most.

(3) Can a model based on US Census socio-demographic and economic variables accurately predict the spatial distribution of elevated BLL in Michigan over time?

The answer to this question is a bit more complicated than the previous two. The hypothesis of this thesis was that a model based on socio-demographic and economic variables would work over time because the same underlying factors were predictive for lead exposure. In the regression portion of this thesis, this assertion turned out to be true for some variables, but not others. For each of the three areal units, several independent variables that were significant when the mean BLL from all years in the database was used turned out to not be significant in several of the individual years. On the other hand, the strongest predictors such as pre-1940 housing turned out to predict mean BLL on a yearly basis as well.

The GWR model for the census tract level also sheds light on this question. The three variables that best predicted mean BLL were percentage pre-1940 housing, percentage African-American, and percentage female-headed households. GWR maps of the coefficients for these variables revealed that they had the highest positive effect in the urban areas of Michigan where mean BLL is higher. The implications are that the variables that predict best in the cities are going to work best on a yearly basis. Variables that characterize suburban or rural areas, where mean BLL is more volatile on a yearly basis according to the standard deviation maps, are less likely to significantly predict mean BLL over a shorter time span. The implication of this is that the temporal length of the research is very important to the outcome. A study that only covers a couple of years within the database may show independent variables as significant or insignificant predictors of mean BLL differently from a study that covers all years of the database. An example is at the census tract level, the variable percentage of housing units vacant is a significant predictor of mean BLL for all eight years of the database. But when tested as a predictor of the mean BLL for each individual year, percentage of vacant houses is only significant in two years, 2000 and 2001.

4.3 Future Research

Spatial epidemiology is a useful tool in understanding and combating the threats posed by health hazards such as lead. With the firm goal of eliminating elevated BLL in Michigan children, future work must take both a research and policy route. These two routes are not mutually exclusive, instead relying heavily upon each other in order to

accomplish meaningful results. Future research involving lead poisoning should involve two different tracks. First, studies from a spatially epidemiological perspective such as this thesis could delve deeper into the issue at a finer spatial scale. A second line of future research could examine the problem through on-site medical investigation of children who have been exposed to lead. This line of inquiry could take on a geographic perspective by determining if different lead-based hazards (paint, water pipes, and atmospheric lead deposition) are responsible for exposure in different areas of Michigan. As for public policy, greater coordination with academia and public health could improve statewide remediation efforts. Spatial epidemiologic approaches to the elevated BLL highlighting hotspots and areas of concern could be a more efficient remediation measure in the long run than targeting houses case by case.

This thesis sought to follow both previous geographic analyses of elevated BLL and commonly used techniques for testing for clusters. In seeking to cover the entire state of Michigan, the analysis in this thesis remained rather coarse. Study areas in this thesis covered either health districts comprising multiple counties or large urban areas. This might not be ideal for micro-targeting problem areas on a limited budget. Future research could focus instead on taking methods such the Geographic Analysis Machine in smaller study areas such as sections of a city to find pockets of consistently high blood lead test results. The statewide analysis in this thesis used a one-kilometer grid, but a study in a smaller study region could use a much smaller grid such as 100 meters since computer processing time would not be an issue. This might reveal neighborhood variation and strongly localized clusters that a statewide or citywide study might miss. In a more localized cluster analysis, it might be possible to obtain a better control dataset as

well. A focus on smaller geographic units for regression analysis might yield better predictive models as well. The regression analysis in this thesis was limited to enumerative units for which census data were available. More locally focused analysis could use a unit of analysis such as tax parcels that would illuminate variation within the neighborhood. Housing information such as the year an individual home was built would greatly aid primary prevention efforts. Such data would likely be difficult to obtain, but the information would be invaluable in building a strong regression model at a parcel level. If these results were combined with survey data collected in the field, a more accurate picture of the local risks could be obtained.

The second line of future research could take a medical investigation approach to ground-level studies elevated BLL in children. While the majority of cases of elevated BLL occurred within urban areas of Michigan, the GAM maps proved that elevated BLL was present as well in more rural areas. An interesting research question would be whether the mechanism of exposure was any different between different parts of state. While many cases in both urban and rural can might still be related to exposure to old paint, it would be compelling if other mechanisms such as old drinking water pipes, nearby smelters, or other paths to exposure were present. Areas where these extra factors were present could then be examined for possible increased incidence of elevated BLL. This could go a long way in explaining areas with anomalously high incidence compared to what might be expected based on housing age. Case investigation could yield the greatest results in rural areas of the state, where individual cases are more likely to go against what the area models predicted. While cluster analysis and spatial regression are powerful tools, the exact cause of exposure can only be inferred from these methods.

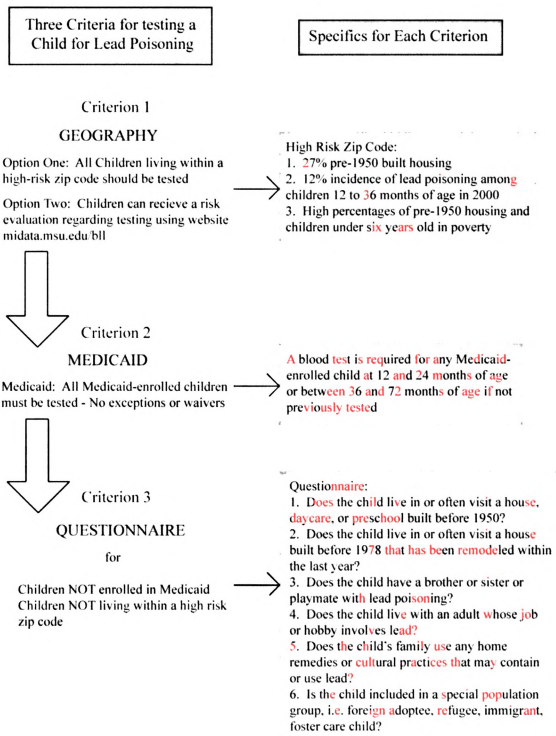
The map in figure 3 showed the zip codes deemed high risk based on the CDC recommendations. The majority of zip codes within Michigan were deemed high risk. This project has while many of the zip codes that have the largest clusters of elevated BLL identified in this thesis are deemed high risk, several areas of the state considered not high risk still show cases. A good example is in the North Central study region in this thesis. The GAM map in figure 65 shows a constellation of cases in areas that are not considered high risk. Other non-high risk areas in other parts of the state show examples of these isolated cases. A comparison of the figure 3 high risk zip code map with the mean BLL zip code map in figure 77 reveals non-high risk areas such as the suburbs around Grand Rapids have as high if not higher mean BLL values than the high risk zip codes. Since this thesis focused on children covered by Medicaid, in theory these kids in non-high risk zip codes would be tested anyway. Still, it is a reminder that even outside of the high risk zip areas, the threat of lead poisoning is present. Kids who are not covered by Medicaid could very easily slip through the testing plan in Appendix 1. To reach the final goal of complete elimination of lead poisoning in Michigan, the best solution might be the most difficult: full screening of children under two years of age and prompt remediation.

In 2004, the Task Force to Eliminate Childhood Lead Poisoning published seven public policy priority recommendations for the government action. These included building effective coalitions to secure funding for community prevention programs, case management for children with elevated BLL, establish a trust to secure stability for lead prevention funding, create a housing registry for pre-1978 homes, develop a public awareness program, coordinate activity statewide, and expand lead remediation in

residential environments (Task Force to Eliminate Childhood Lead Poisoning 2004). The main recommendation that could be added to the list is a closer relationship between the state and the academic community regarding research. A coordinated effort between the state and academia could harness spatial epidemiology studies in order to analyze test results in real time. Such analysis would provide insight into how incoming results fit the overall patterns of BLL within Michigan. Real time spatial epidemiology could find areas that have been overlooked. Perhaps more importantly, such coordination between the state and academia could evaluate the progress of remediation efforts. Only so much can be gleaned for looked at maps and test results without the context of what is being done on the ground. With such a partnership of real-time test results and statistical mapping, remediation of lead-based hazards could take a leap forward and lead poisoning in Michigan children could finally become a relic of an earlier era.

Appendix 1

Michigan Statewide Lead Testing/Lead Screening Plan



Appendix 2

Difference of K code in R

```
# Difference of K function#

library(maptools)
library(spatstat)
library(splancs)

lan<- read.shape("Lansing") #Load study area shapefile
med<- read.shape("Med98L") #Load 1998 Lansing test results shapefile

x<- vector(length=length(med$Shape))#Create empty vector for x coordinates
y<- vector(length=length(med$Shape)) #Create empty vector for y coordinates
for (i in 1:length(med$Shape)) {
x[i] <- med$Shape[[i]]$verts[,1]#Fill x and y vectors with the Michigan
y[i] <- med$Shape[[i]]$verts[,2]#Georef coordinates
  }

wp<- cbind(x, y, med$att.data) #Create data frame with locations and attributes
wp<- subset(wp, select = c(x, y, CC10))#Select out the case/control threshold of 10

cx <- lan$Shape[[1]]$verts[,1]#Create data frame of study area x coordinates
cy <- lan$Shape[[1]]$verts[,2]#Create data frame of study area y coordinates

lan.bdy<- cbind(cx, cy) #Create study area boundary

cases<- wp[wp$CC10==1,] #Select out all cases at the 10 µg/dLthreshold
controls<- wp[wp$CC10==0,]#Select out all controls at the 10 µg/dLthreshold

p.cases <- as.points(cases)#Convert cases to points
p.controls <- as.points(controls)#Convert controls to points

#define distances
dist<- seq(500, 10000, 500)#Define distances of concentric circles

k.case <- khat(p.cases, lan.bdy, s=dist)#Calculate Ripley's K for cases
k.control <- khat(p.controls, lan.bdy, s=dist)#Calculate Ripley's K for controls

K.diff <- k.case - k.control#Calculate the difference of K

# Random Labeling Simulation#
env.lab<- Kenv.label(p.cases, p.controls, bboxx(bbox(lan.bdy)), nsim=19, s=dist)
```

```
#Plot the Results#
plot(dist, K.diff, xlab="Distance", ylab="Diff in K", ylim=range(K.diff-dist,
 + env.lab$lower-dist, env.lab$upper-dist))
lines(dist, env.lab$upper, lty=2)
lines(dist, env.lab$lower, lty=2)
```

Appendix 3

Geographic Analysis Machine code in R

```
#Geographic Analysis Machine#

library(splancs)
library(spatstat)
library(maptools)

lan<- read.shape("Lansing")#Load study area shapefile
med98<- read.shape("Med98L")#Load 1998 Lansing test results shapefile

lx<- lan$Shape[[1]]$verts[,1]#Create data frame of study area x coordinates
ly<- lan$Shape[[1]]$verts[,2] #Create data frame of study area y coordinates

lan.bdy<- cbind(lx, ly)#Create study area boundary

x<- vector(length=length(med98$Shape))#Create empty vector for x coordinates
y<- vector(length=length(med98$Shape))#Create empty vector for y coordinates
for (i in 1:length(med98$Shape)) {
x[i] <- med98$Shape[[i]]$verts[,1]#Fill x and y vectors with the Michigan
y[i] <- med98$Shape[[i]]$verts[,2]#Georef coordinates
  }
medp<- cbind(x, y, med98$att.data)#Create data frame with locations, attributes
medp<- subset(medp, select = c(x, y, CC10))#Select out the case/control threshold
                                            #of 10

distance<- function (x1, y1, x2, y2) {#Create function to calculate distance
euc<- sqrt((x2 -x1)^2 + (y2-y1)^2)
return(euc)
}

backgd.rate <- 0.014147#ENTER BACKGROUND RATE HERE

lan.grid<- gridpts(lan.bdy, xs=1000, ys=1000)  #Create 1 kilometer grid

#Create empty distance matrix
dist.mat<- matrix(nrow=length(lan.grid[,1]), ncol=length(medp$x))

#Create empty matrix for calculation results
close<- matrix(data=0, nrow=length(lan.grid[,1]), ncol=4)

#Calculate Distance between grid points and test results
```

```
for (i in 1:length(mich.grid[,1]))
dist.mat[i,]<-distance(mich.grid[i,1], mich.grid[i,2], medp$x, medp$y)



#Loop to fill calculation matrix with number of points within 1.8 kilometers of the
#grid points, the number of these points that are controls, number that are elevated
#BLL cases, and the expected number of cases
for (i in 1:length(mich.grid[,1])) {
close[i,1] <- sum(dist.mat[i,] < 1800)        # all pts within 1.8km
close[i,2] <- sum(dist.mat[i,medp$CC10==0]<1800) # just control
close[i,3] <- sum(dist.mat[i,medp$CC10==1]<1800) # just lead
close[i,4] <- close[i,1]*backgd.rate  # Expected # cases
}

# Highlight grid points where there is less than a 5% chance of the number of
#elevated BLL cases occurring according to a Poisson distribution with the
#background rate as the mean
v1800.98<- ((ppois(close[,3], (close[,4])) > 0.95) & (close[,3] > 0))

#Run kernel smoother over the resulting grid
k1800.98<- kernel2d(mich.grid[v1800.98,], mich.bdy, h0=1800, nx=500, ny=500)

#Plot final map
polymap(mich.bdy, border="grey")
image(k1800.98, add=TRUE, col=heat.colors(20))
```

Literature Cited

Agency for Toxic Substances & Disease Registry. 2007. *Lead Toxicity - What Are the Physiologic Effects of Lead Exposure* 2007 [cited October 18 2007]. Available from http://www.atsdr.cdc.gov/csem/lead/pbphysiologic_effects2.html.

Akhtar, R. 1982. *The Geography of Health: An Essay and a Bibliography*. New Delhi: Marwah Publications.

American Academy of Pediatrics. 2003. Michigan Medicaid Facts.

Angier, N. 2007. The Pernicious Allure of Lead. *New York Times*, August 21, 2007.

Bailey, A., J. Sargent, and M. Blake. 1998. A Tale of Two Counties: Childhood Lead Poisoning, Industrialization, and Abatement in New England. *Economic Geography* 74:96-111.

Bailey, A., J. Sargent, D. Goodman, J. Freeman, and M. J. Brown. 1994. Poisoned Landscapes: The Epidemiology of Environmental Lead Exposure in Massachusetts Children 1990-1991. *Social Science and Medicine* 19 (6):757-766.

Barboza, D. 2007. Why Lead in Toy Paint? It's Cheaper. *New York Times*, September 11, 2007.

Beam, C. 2007. *Why Do They Put Lead Paint in Toys: It's Bright, Cheap, and Lasts Forever* 2007 [cited October 13 2007]. Available from http://slate.com/id/2172289.

Bellinger, D., and A. Bellinger. 2006. Childhood lead poisoning: the torturous path from science to policy. *The Journal of Clinical Investigation* 116 (4):853-857.

Bellinger, D., and J. Schwartz. 1997. Effects of Lead in Children and Adults. In *Topics in Environmental Epidemiology*, eds. K. Steenland and D. Savitz. New York: Oxford University Press.

Brill, R., and J. Wampler. 1967. Isotope Studies of Ancient Lead. *American Journal of Archaeology* 71 (1):63-77.

Canfield, R., C. Henderson, D. Cory-Slechta, C. Cox, T. Jusko, and B. P. Lanphear. 2003. Intellectual Impairment in Children with Blood Lead Concentrations Below 10 Micrograms per Deciliter *The New England Journal of Medicine* 348 (16):1517-1526.

Centers for Disease Control and Prevention. 2005a. ToxFAQs for Lead, ed. ATSDR.

———. 2005b. Building Blocks for Primary Prevention: Protecting Children from Lead-Based Paint Hazards, ed. H. a. H. Services, 264 p.

Chen, A., K. Dietrich, J. Ware, J. Radcliffe, and W. Rogan. 2005. IQ and Blood Lead from 2 to 7 Years of Age: Are the Effects in Older Children the Residual of High Blood Lead Concentrations in 2-Year Olds. *Environmental Health Perspectives* 113 (5):597-601.

Chisolm, J. 2001. Evolution of the Management and Prevention of Childhood Lead Poisoning: Dependence of Advances in Public Health on Technological Advances in the Determination of Lead and Related Biochemical Indicators of Its Toxicity. *Environmental REsearch Section A* 86 (2):111-121.

Clarkson, T. 1995. Health Effects of Metals: A Role for Evolution? *Environmental Health Perspectives* 103 (Supplement 1):9-12.

Cromley, E. K., and S. L. McLafferty. 2002. *GIS and Public Health*. New York: The Guilford Press.

Daniel, K., M. Sedlis, L. Polk, S. Dowuona-Hammond, B. McCants, and T. Matte. 1990. Childhood Lead Poisoning, New York City, 1988. *The Morbitity and Mortality Weely Report* 39:1-7.

Department of Housing and Urban Development. 1993. Understanding Title X: A Practical Guide to the Residential Lead-Based Paint Hazard Reduction Act of 1992.

———. 2004. History of Lead-Based Paint Legislation.

Dietrich, K., J. Ware, M. Salganik, J. Radcliffe, W. Rogan, G. Rhoads, M. Fay, C. Davoli, M. Denckla, R. Bornschein, D. Schwartz, D. Dockery, S. Adubato, and R. Jones. 2004. Effect of Chelation Therapy on the Neuropsychological and Behavioral Development of Lead-Exposed Children After School Entry. *Pediatrics* 114 (1):19-26.

Dignam, T., A. Evens, E. Eduardo, S. Ramirez, K. Caldwell, N. Kilpatrick, G. Noonan, D. Flanders, P. Meyer, and M. McGeehin. 2004. High-Intensity Targeted Screening for Elevated Blood Lead Levels among Children in 2 Inner-City Chicago Communities. *American Journal of Public Health* 94 (11):1945-1951.

Dockerty, J., K. Sharples, and B. Borman. 1999. An Assessment of Spatial Clustering of Leukaemias and Lymphomas among Young People in New Zealand. *Journal of Epidemiology and Community Health* 53:154-158.

Dolk, H., A. Busby, B. Armstrong, and P. Walls. 1998. Geographical Variation in Anophthalmia and Microphtalmia in England, 1988-94. *British Medical Journal* 317 (7163):905-910.

Environmental Protection Agency. 1996. EPA Takes Final Step in Phaseout of Leaded Gasoline.

———. 2001. Lead Based Paint Prevention in Certain Residential Structures.

Ettinger, A. S. 2007. *Chelation Therapy for Childhood Lead Poisoning: Does Excretion Equal Efficacy?* Harvard School of Public Medicine 1999 [cited October 21 2007]. Available from http://www.hsph.harvard.edu/Organizations/ddil/chelation.htm.

Fee, E. 1990. Public Health in Practice: An Early Confrontation with the 'Silent Epidemic' of Childhood Lead Paint Poisoning. *Journal of the History of Medicine and Allied Sciences* 45 (4):570-606.

Finkelstein, Y., M. Markowitz, and J. Rosen. 1998. Low-Level Lead-Induced Neurotoxicity in Children: An Update on Central Nervous System Effects. *Brain Research Reviews* 27 (2):168-176.

Finn, M. 2007. Health Care Demand in Michigan: An Examination of the Michigan Certificate of Need Acute Care Bed Need Methodology, Geography, Michigan State University, East Lansing.

Flegal, A., and D. Smith. 1992. Lead Levels in Preindustrial Humans. *New England Journal of Medicine* 326 (19):1293-1294.

Foley, J., P. Foley, and J. Madigan. 2001. Spatial Distribution of Seropositivity to the Causative Agent of Granulocytic Ehrlichiosis in Dogs in California. *American Journal of Veterinary Research* 62 (10):1599-1605.

Fotheringham, A., C. Brunsdon, and M. Charlton. 2002. *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. Hoboken, NJ: John Wiley & Sons.

Frost, S. W. 2004. Lead Poisoning in Young Children: Determining Risk Factors and Exposure Sources - An Environmental Justice Approach, Sociology, Michigan State University, East Lansing.

Garza, A., H. Chavez, R. Vega, and E. Soto. 2005. Cellular and Molecular Mechanism of Lead Neurotoxicity. *Salud Mental* 28 (2):48-58.

Gaston, J. 1972. Geography of Lead Poisoning: Development of a Model, Geography, Michigan State University, East Lansing.

Gibson, J. 1904. A Plea for Painted Railings and Painted Walls of Rooms as the Source of Lead Poisoning Amongst Queensland Children. *Australasian Medical Gazette (reprinted in Public Health Reports May-June 2005).*

Gilbert, S., and B. Weiss. 2005. Preventing Neurodevelopment Disorders: The CDC Should Lower the Blood Lead Action Level From 10 to 2 micrograms per deciliter. Paper read at 22nd International Neurotoxicology Conference, September 11-14, at Research Triangle Park, NC.

Goyer, R. 1993. Lead Toxicity: Current Concerns. *Environmental Health Perspectives* 100:177-187.

Griffith, D. A., P. G. Doyle, D. C. Wheeler, and D. L. Johnson. 1998. A Tale of Two Swaths: Urban Childhood Blood-Lead Levels across Syracuse, New York. *Annals of the Association of American Geographers* 88 (4):640-655.

Guthe, W., R. Tucker, E. Murphy, R. England, E. Stevenson, and J. Luckhardt. 1992. Reassessment of Lead Exposure in New Jersey Using GIS Technology. *Environmental Research* 59 (2):318-325.

Haley, V., and T. Talbot. 2004. Geographic Analysis of Blood Lead Levels in New York State Children Born 1994-1997. *Environmental Health Perspectives* 112 (15):1577-1582.

Hernberg, S. 2000. Lead Poisoning in a Historical Perspective. *American Journal of Industrial Medicine* 38 (3):244-254.

Honari, M. 1999. Health Ecology: An Introduction. In *Health Ecology: Health, Culture and Human-Environment Interaction*, eds. M. Honari and T. Boleyn. London: Routledge.

Huang, Y., and Y. Leung. 2002. Analysing Regional Industrialisation in Jiangsu Province using Geographically Weighted Regression. *Journal of Geographical Systems* 4 (2):233-249.

Hunter, D. 1969. *The Diseases of Occupations*. 4th ed. Boston: Little, Brown.

Hunter, J. 1976. Aerosol and Roadside Lead as Environmental Hazard. *Economic Geography* 52 (2):147-160.

Jacobs, D., R. Clickner, J. Zhou, S. Viet, D. Marker, J. Rogers, D. Zeldin, P. Broene, and W. Friedman. 2002. The Prevalence of Lead-Based Paint Hazards in U.S. Housing. *Environmental Health Perspectives* 110 (10):A599-A606.

Jacobziner, H., and H. Raybin. 1962. Epidemiology of Lead Poisoning. *Archives of Pediatrics* 79 (2):72-76.

Jones, K., and G. Moon. 1987. *Health, Disease, and Society.* London: Routledge & Kegan Paul.

Kaplowitz, S., H. Perlstadt, and L. Post. 2007. Predicting Blood Lead Level from Medicaid Eligibility, Race, and Neighborhood Census Data: An Analysis of Michigan Data. East Lansing, MI: Michigan State University.

Kemper, A. R., C. Bordley, and S. Downs. 1998. Cost-Effectiveness Analysis of Lead Poisoning Screening Strategies Following the 1997 Guidelines of the Centers for Disease Control and Prevention. *Archives of Pediatrics & Adolescent Medicine* 152 (12):1202-1208.

Kemper, A. R., and S. Clark. 2005c. Physician Barriers to Lead Testing of Medicaid-Enrolled Children. *Ambulatory Pediatrics* 5 (5):290-293.

Kemper, A. R., L. M. Cohn, K. E. Fant, and K. J. Dombkowski. 2005a. Blood Lead Testing Among Medicaid-Enrolled Children in Michigan. *Archives of Pediatrics & Adolescent Medicine* 159 (7):646-650.

Kemper, A. R., L. M. Cohn, K. E. Fant, K. J. Dombkowski, and S. Hudson. 2005b. Follow-up Testing Among Children With Elevated Screening Blood Lead Levels. *Journal of the American Medical Association* 293 (18):2232-2237.

Kemper, A. R., R. Uren, and S. Hudson. 2007. Childhood Lead Poisoning Prevention Activities within Michigan Local Public Health Departments. *Public Health Reports* 122 (1):88-92.

Kitman, J. L. 2000. The Secret History of Lead: Special Report. *The Nation.*

Kovarik, W. 2005. Ethyl-Leaded Gasoline: How a Classic Occupational Disease Became an International Public Health Disaster. *International Journal of Occupational and Environmental Health* 11 (4):384-397.

Lam, T. 2007. Money on the Way to Fight Lead Poisoning in Homes. *Detroit Free-Press*, October 2, 2007.

Lanphear, B. P. 2005a. Childhood Lead Poisoning: Too Little, Too Late. *Journal of the American Medical Association* 293 (18):2274-2276.

Lanphear, B. P., R. Byrd, P. Auinger, and S. Schaffer. 1998b. Community Characteristics Associated with Elevated Blood Lead Levels in Children. *Pediatrics* 101 (2):264-271.

Lanphear, B. P., R. Hornung, J. Khoury, K. Yolton, P. Baghurst, D. Bellinger, R. L. Canfield, K. N. Dietrich, R. Bornschein, T. Greene, S. J. Rothenberg, H. L. Needleman, L. Schnaas, G. Wasserman, J. Graziano, and R. Roberts. 2005b. Low-Level Environmental Lead Exposure and Children's Intellectual Function: An International Pooled Analysis. *Environmental Health Perspectives* 113 (7):894-899.

Lanphear, B. P., T. Matte, J. Rogers, R. Clickner, B. Dietz, R. Bornschein, P. Succop, K. Mahaffey, S. Dixon, W. Galke, M. Rabinowitz, M. Farfel, C. Rohde, J. Schwartz, P. Ashley, and D. Jacobs. 1998d. The Contribution of Lead-Contaminated House Dust and Residential Soil to Children's Blood Lead Levels: A Pooled Analysis of 12 Epidemiologic Studies. *Environmental Research* 79 (1):51-68.

Leung, Y., C.-L. Mei, and W.-X. Zhang. 2000. Statistical Tests for Spatial Nonstationarity Based on the Geographically Weighted Regression Model. *Environment and Planning A* 32 (1):9-32.

Lidsky, T., and J. Schneider. 2003. Lead Neurotoxicity in Children: Basic Mechanisms and Clinical Correlates. *Brain* 126:5-19.

Litaker, D., C. M. Kippes, T. E. Gallagher, and M. E. O'Connor. 2000. Targeting Lead Screening: The Ohio Lead Risk Score. *Pediatrics* 106 (5):Art. No. e69.

Mahaffey, K., J. Annest, J. Roberts, and R. Murphy. 1982. National Estimates of Blood Lead Levels: United States 1976-1980. *New England Journal of Medicine* 307 (10):573-579.

Markowitz, G., and D. Rosner. 2000. "Cater to the Children": The Role of The Lead Industry in a Public Health Tragedy, 1900-1955. *American Journal of Public Health* 90 (1):36-46.

———. 2002. *Deceit and Denial: The Deadly Politics of Industrial Pollution*. Berkeley: University of California Press.

Mayer, J. 1982. Medical Geography: Some Unsolved Problems. *The Professional Geographer* 34 (3):261-269.

———. 1986. Ecological Associative Analysis. In *Medical Geography: Progress and Prospect*, ed. M. Pacione. London: Croom Helm.

McKnight, K. 2006. Spatial Trends of West Nile Virus in Detroit, Michigan 2002, Geography, Michigan State University, East Lansing.

Meade, M. 1977. Medical Geography as Human Ecology: The Dimension of Population Movement. *Geographical Review* 67 (4):379-393.

Meade, M., and R. Earickson. 2000. *Medical Geography*. New York: The Guilford Press.

Michigan Department of Community Health. 1998. Annual Report on Blood Lead Levels in Michigan.

———. 2001. Annual Report on Blood Lead Levels in Michigan.

———. 2005a. Annual Report on Blood Lead Levels on Adults and Children in Michigan.

———. 2006. Annual Report on Blood Lead Levels on Adults and Children in Michigan.

———. 2007. Statewide Lead Testing/Lead Screening Plan.

Michigan Department of Natural Resources. 2001. GIS/GPS Education.

Mielke, H. 1999. Lead in the Inner Cities. *American Scientist* 87 (1):62-73.

Miranda, M. L., D. Dolinoy, and M. A. Overstreet. 2002. Mapping for Prevention: GIS Models for Directing Childhood Lead Poisoning Prevention Programs. *Environmental Health Perspectives* 110 (9):947-953.

Murray, K., D. Rogers, and M. Kaufman. 2004. Heavy Metals in an Urban Watershed in Southeastern Michigan. *Journal of Environmental Quality* 33 (1):163-172.

Nakaya, T., A. Fotheringham, C. Brunsdon, and M. Charlton. 2005. Geographically Weighted Poisson Regression for Disease Association Mapping. *Statistics in Medicine* 24 (17):2695-2717.

Needleman, H. 1998. Clair Patterson and Robert Kehoe: Two Views of Lead Toxicity. *Environmental REsearch Section A* 78 (2):79-85.

———. 2004. Lead Poisoning. *Annual Review of Health* 55 (1):209-222.

Needleman, H., and D. Bellinger. 1991a. The Health Effects of Low Level Exposure to Lead. *Annual Review of Public Health* 12:111-140.

Nriagu, J. 1983. Saturnine Gout among Roman Aristocrats. Did Lead Poisoning Contribute to the Fall of the Empire. *New England Journal of Medicine* 308 (11):660-663.

———. 1990. The Rise and Fall of Leaded Gasoline. *The Science of the Total Environment* 92:13-28.

———. 1998. Clair Patterson and Robert Kehoe's Paradigm of "Show Me the Data" on Environmental Lead Poisoning. *Environmental REsearch Section A* 78 (2):71-78.

O'Brien, D., J. Kaneene, A. Getis, J. Lloyd, G. Swanson, and R. Leader. 2000. Spatial and Temporal Comparison of Selected Cancers in Dogs and Humans, Michigan, USA, 1964-1994. *Preventive Veterinary Medicine* 47 (3):187-204.

O'Sullivan, D., and D. Unwin. 2003. *Geographic Information Analysis*. Hoboken, NJ: Wiley and Sons, Inc.

Openshaw, S., A. Craft, M. Charlton, and J. Birch. 1988. Investigation of Leukaemia Clusters by use of a Geographic Analysis Machine. *The Lancet* 331 (8580):272-273.

Ozden, T., H. Issever, G. Gokcay, and G. Saner. 2004. Longitudinal Analyses of Blood-Lead Levels and Risk Factors for Lead Poisoning in Healthy Children under Two Years of Age. *Indoor Built Environment* 13:303-308.

Parsons, P., A. Reilly, and D. Esernio-Jenssen. 1997. Screening Children Exposed to Lead: An Assessment of the Capillary Blood Lead Fingerstick Test. *Clinical Chemistry* 43 (2):302-311.

Pirkle, J., R. Kaufmann, D. Brody, T. Hickman, E. Gunter, and D. Paschal. 1998. Exposure of the US Population to Lead, 1991-1994. *Environmental Health Perspectives* 106 (11):745-750.

Prince, M., A. Chetwynd, P. Diggle, M. Jarner, J. Metcalf, and O. James. 2001. The Geographical Distribution of Primary Biliary Cirrhosis in a Well-Defined Cohort. *Hepatology* 34 (6):1083-1088.

Rabin, R. 1989. Warnings Unheeded: A History of Child Lead Poisoning. *American Journal of Public Health* 79 (12):1668-1674.

———. 2008. The Lead Industry and Lead Water Pipes: "A Modest Campaign". *American Journal of Public Health* 98 (9):1584-1592.

Richardson, J. 2005. *The Cost of Being Poor: Poverty, Lead Poisoning, and Policy Implementation*. Westport, CT: Praeger Publishers.

Rosen, J., and P. Mushak. 2001. Primary Prevention of Childhood Lead Poisoning: The Only Solution. *New England Journal of Medicine* 344 (19):1470-1471.

Sargent, J., A. Bailey, P. Simon, M. Blake, and M. Dalton. 1997. Census Tract Analysis of Lead Exposure in Rhode Island Children. *Environmental Research* 74 (2):159-168.

Sargent, J., M. J. Brown, J. Freeman, A. Bailey, D. Goodman, and D. Freeman. 1995. Childhood Lead Poisoning in Massachusetts Communities: Its Association with Sociodemographic and Housing Characteristics. *American Journal of Public Health* 85 (4):528-534.

Shearmur, R., P. Apparicio, P. Lizion, and M. Polese. 2007. Space, Time, and Local Employment Growth: An Application of Spatial Regression Analysis. *Growth and Change* 38 (4):696-722.

Silbergeld, E. 1997. Preventing Lead Poisoning in Children. *Annual Review of Public Health* 18:187-210.

Talbot, T., S. Forand, and V. Haley. 1998. Geographic Analysis of Childhood Lead Exposure in New York State. Paper read at Proceedings of the 3rd National Conference on GIS in Public Health, August 17-20, at San Diego.

Task Force to Eliminate Childhood Lead Poisoning. 2004. Final Report of the Task Force to Eliminate Lead Poisoning.

Tong, S. 1990. Roadside Dusts and Soils Contamination in Cincinnati, Ohio, USA. *Environmental Management* 14 (1):107-113.

Tong, S., Y. Schirnding, and T. Prapamontol. 2000. Environmental Lead Exposure: A Public Health Problem of Global Dimensions. *Bulletin of the World Health Organization* 78 (9):1068-1077.

United States Geological Survey. 2007. *Lead: Statistics and Information* 2007 [cited October 11 2007]. Available from http://minerals.usgs.gov/minerals/pubs/commodity/lead/index.html#myb.

US Census Bureau. 2000. Geographic Areas Reference Manual.

———. 2001. DP-4, Profile of Selected Housing Characteristics: 2000 (Geographic Area: Michigan).

Venables, W., and D. Smith. 2008. An Introduction to R.

Vojnovic, I., C. Jackson-Elmoore, J. Holtrop, and S. Bruch. 2006. The Renewed Interest in Urban Form and Public Health: Promoting Increased Physical Activity in Michigan. *Cities* 23 (1):1-17.

Waldron, H. 1973. Lead Poisoning in the Ancient World. *Medical History* 17 (4):391-399.

Waller, L., and C. Gotway. 2004. *Applied Spatial Statistics for Public Health Data.* Hoboken, NJ: Wiley and Sons, Inc.

Weiss, D., W. Shotyk, and O. Kempf. 1999. Archives of Atmospheric Lead Pollution. *Naturwissenschaften* 86 (6):262-275.

Wheeler, D. C. 2007. A comparison of spatial clustering and cluster detection techniques for childhood leukemia incidence in Ohio, 1996-2003. *International Journal of Health Geographics* 6 (13):2-38.

Yohn, S., D. Long, J. Fett, and L. Patino. 2004. Regional Versus Local Influences on Lead and Cadmium Loading to the Great Lakes Region. *Applied Geochemistry* 19 (7):1157-1175.

Zandbergen, P., and J. Green. 2007. Error and Bias in Determining Exposure Potential of Children at School Locations Using Proximity-Based GIS Techniques. *Environmental Health Perspectives* 115 (9):1363-1369.