4-2020

Spatial Patterns in Socio-Economic Factors and Rapid Repeat Pregnancies in Kalamazoo County, MI

Dennis Donkor
Western Michigan University, denn233@gmail.com

SPATIAL PATTERNS IN SOCIO-ECONOMIC FACTORS AND RAPID REPEAT PREGNANCIES IN KALAMAZOO COUNTY, MI

by

Dennis Donkor

A thesis submitted to the Graduate College in
partial fulfilment of the requirements
for the degree of Master of Science
Geography
Western Michigan University
April 2020

Thesis Committee:

Kathleen Baker, Ph. D, Chair
Catherine Kothari, Ph.D.
Benjamin Ofori-Amoah, Ph.D.

SPATIAL PATTERNS IN SOCIO-ECONOMIC FACTORS AND RAPID REPEAT PREGNANCIES IN KALAMAZOO COUNTY, MI

Dennis Donkor, M.S.

Western Michigan University, 2020

Rapid repeat pregnancy (RRP) refers to a pregnancy that occurs less than 24 months after a live birth. In the United States, several studies have focused on factors that influence women to rapidly repeat pregnancies at the national and state level. As a result, this study explores spatial patterns in RRP in Kalamazoo County at the block group local level using birth records of moms in the county from 2008 to 2014. The study further investigates individual and neighborhood factors influencing RRP. Results from the hotspot (Getis Ord G*) revealed that block groups in eastside Kalamazoo township are significant hotspots for rapid repeat moms in the county. At the individual level, women who had their index birth as teenagers as well as moms that had spouse named on birth certificate and women of color were at higher odds of rapidly repeating pregnancies. At the block group neighborhood level, RRP moms lived in two main neighborhoods. However, moms living in neighborhoods with characteristics of higher population, more black women, women aged 20-24 and more renters are more associated with rapid repeat pregnancy in Kalamazoo County, MI.

ACKNOWLEDGEMENTS

I would like to start by first thanking my wife Dr. Owusua Yamoah and my son Micah Ivyn Donkor both of whom have been my inspiration to want to pursue this degree and carry out this work in the first place. I also thank the all mighty God who has continuously granted with great strength to go on this adventure.

I would also want to express a heartfelt gratitude to my committee chair, Dr. Kathleen Baker. From the day I took her class in Introduction to GIS 5010, she has remained a mentor, a professor and my advisor. She has consistently urged me on to hit the highest heights possible while whipping in me an enthusiasm to study problems in health using GIS. I am grateful for always keeping your doors opened to me for discussions and counsel. I further want to thank Dr Kothari, particularly for believing in me and making this data available to me for analysis, also to Dr Benjamin Ofori-Amoah for your helpful advice and encouragement throughout pursuant of this degree and in carrying out this research. Finally, I thank the entire Western Michigan University Geography Department, for admitting me on the program and providing a loving and inspiring atmosphere to carry out this research. I am grateful.

Dennis Donkor

TABLE OF CONTENTS

Table of Contents - Continued

Table of Contents - Continued

APPENDICES

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Background

Maternal health is prioritized globally as complications associated with pregnancy can lead to maternal and infant death (World Health Organization, 2016). Pregnancies can have dire physical and emotional consequences on couples and families and in a larger sense, communities and nations suffer socially and economically if pregnancy and its associated problems are not managed. Sina (2017) in studying pregnancy and the global disease burden mentions that women go through unique physiological changes during pregnancy that can be detrimental to their health and hence the need for more research to address issues confronting pregnant women. Women and their fetus can suffer from conditions such as gestational diabetes mellitus, preeclampsia, intrauterine growth retardation, poor birth outcomes (preterm delivery, low birth weight), cardiovascular disease and mental disorders (Fraser et al., 2012). In the United States, it is estimated that about 500 women die annually from pregnancy related causes with about 50,000 suffering from severe maternal morbidity (severe pregnancy complications) (Center for Disease Control, 2017).

Central to the issue of pregnancy and maternal health, is the use of contraceptives, which are recommended for reducing unwanted or unplanned pregnancies and slowing the rate of sequential pregnancies (De Bocanegra et al., 2013; Loree et al., 2018). Studies show that having a longer interval between pregnancies has the potential to reduce the number of pregnancies and associated pregnancy related health risks to a woman (Damle et al., 2015; Isquick et al., 2017). The long-term role of contraception as a key intervention strategy highlights the acknowledgement of timing and spacing as important considerations to tackle adverse issues of pregnancy globally.

In addressing the twin problems of pregnancy relating to timing and spacing, the World Health Organization (WHO) (2006) recommended that women increase the interval between a live birth and another pregnancy to at least 24 months. Women should wait at least six months to become pregnant after a miscarriage or induced abortion. This increased interval will help to minimize adverse pregnancy related health problems and poor birth outcomes and ensure a healthy time and space between births (Regan et al., 2019). This recommendation sparked research on the critical nature of the concept of healthy spacing between births or pregnancies to understand the many different reasons that women give birth contrary to the 24 months interval; an issue studies have termed "Rapid Repeat Pregnancy" (RRP) (Aslam et al., 2017; Conroy et al., 2016; Pfitzner et al., 2003).

In the United States, many studies have been done to understand rapid repeat pregnancy and its effects on birth outcomes as well as factors that influence women to rapidly repeat pregnancies (Bennett et al., 2006; de Bocanegra et al., 2014; Reese and Halpern, 2016). For instance, Gemmill and Lindberg (2013) studying 12, 279 women from the National Survey of Family Growth data between 2006 to 2010 shows that 35% representing one third of that sample of women in the United States conceived within 18 months of a prior birth, a result that is consistent with data from the Healthy People 2020 by the U.S. Department of Health and Human Services. In the United States, research on rapid repeat pregnancy has predominantly focused on: risk factors of RRP, trends in rapid repeat births, at-risk groups, adolescent rapid repeat pregnancy, effects on birth outcomes, and effective intervention measures to address the issue (Tocce et al., 2012; Langston, et al., 2014; Maravilla, et al., 2017; Norton et al., 2017). These studies, through their design and approach, provide a multifaceted view to the issue of RRP in the United States.

The range of literature makes it clear that socio-economic background of individuals is a factor that influences mothers to rapidly repeat pregnancies (Bell et al., 2013; Levanthal and Brooks-Gun, 2000). Delara et al., (2018) in a retrospective cohort study of interpregnancy intervals of women living in California mention that African American, Latinos and other minority groups have an increased risk of giving birth at short intervals as compared to white Americans. Zhang et al., (2019) also in a retrospective cohort study to investigate risk and associated risk factors of short birth-to-pregnancy intervals among African-born black women (immigrants) in the state of Washington reveals that African born black women are at a higher risk of giving birth at short intervals than African Americans and Whites. A substantial amount of rapid repeat pregnancy studies in the United States also focus on young people 15-24 as a high-risk population that needs attention and propose numerous interventions targeted at adolescents as they are sexually active, engage in more risky sexual lifestyles and not likely to use contraceptives (Collier, 2009). However, few studies have focused their lens on understanding the socio-economic dynamics and trends by considering geospatial characteristics and using Geographical Information Science (GIS) methods to analyze patterns and spatial variation of rapid repeat pregnancy for women from diverse backgrounds in order to foster unique and effective location specific interventions.

Problem Statement

Rapid repeat pregnancy is a major health problem in the United States of America. Despite numerous studies on pregnancy and birth intervals, about a third of all births in the

United States are not under 24 months (Cha et al., 2015). The high incidence of rapid repeat pregnancies in a developed country like the United States is not surprising though, as rapid repeat pregnancy has been downplayed for not having a significant effect on the health of women in the developed world. For instance, a systematic review of rapid repeat pregnancy studies in the United States, Canada, Australia and other European countries concludes that rapid repeat pregnancies does not have a significant association with adverse maternal outcomes in developed countries as compared to developing countries whose women have low nutritional status and access to contraceptives. It is thus not surprising, that there are no federal guidelines for birth spacing in the United States (Hutcheon et al., 2018). This suggests that rapid repeat pregnancy is a problem but only a problem for the developing world. Contrastingly, it must be noted that adverse maternal outcomes are not only poor birth outcomes but also include maternal morbidity, a problem that is prevalent in the United States. Conde-Agudelo (2012) in a systematic study buttressed this point by establishing that there was significant association between inter-pregnancy intervals of less than 18 months and an increased risk of premature membrane rupture, utero-placental bleeding disorders and uterine rupture in women attempting vaginal birth after an earlier caesarean delivery in the United States.

Statistically, Thoma and Kirmeyer, (2016) in a study in the United States revealed that 33.9% of mothers gave birth to a subsequent baby after a live birth in less than 18 months. Also, Copen et al., (2015) in analyzing births to residents of the 36 states and the District of Columbia that implemented the 2003 revision of the birth certificate as of January 1, 2011 notes that about 32% of births to women with a second or higher order singleton birth in Michigan happened in less than 18 months to a previous birth. In Kalamazoo, 2018 records of the Michigan Birth Certificate Registry show that multiparous mothers gave birth at intervals less than 12 months and 12 to 35 months at rates of 170.3 and 486.2 per 1000 live birth respectively (Michigan Department of Health & Human Services, 2018).

Though these statistics paint a dire picture of the problem, it must be noted that most of these studies of Rapid repeat pregnancy in the United States are done at either national or state level (Appareddy et al., 2017; White et al., 2015). These scales of analysis are very important, for policy formulation at a higher level. However, it is important to also conduct studies within localized levels such as neighborhoods. Local neighborhood analysis can improve insight into the design of location specific interventions to address the issue of rapid repeat pregnancy. With the national Healthy People 2020 program calling for 10% improvement of the 33.1% of pregnancies conceived within 18 months of a previous birth, studies at local to regional scales have the potential to provide a new dimension to understanding RRP.

Against this backdrop, this study used secondary data of reported pregnancies in Kalamazoo County from 2008 to 2014, with 2010 as the base year, from the Kalamazoo County Vital Statistics database to analyze patterns of rapid repeat pregnancies in the County. Three main hypotheses were investigated. The first is, there is spatial variation in rapid repeat pregnancy in Kalamazoo County, MI. The second is there are individual effects of rapid repeat pregnancy in Kalamazoo County, MI. Lastly, there are neighborhood effects of rapid repeat pregnancy in Kalamazoo County, MI.

Study Objectives

The purpose of this study was generally to understand the problem of rapid repeat pregnancy through a case study of Kalamazoo County, MI from 2008 to 2014. Geographical Information

Systems and other statistical techniques were used to understand the issue. Specifically, the study sought to:

1. Understand spatial variations on the issue of rapid repeat pregnancies in Kalamazoo County, MI.
2. Explore individual socio-economic factors that influence rapid repeat pregnancy (second pregnancy ≤ 24 months after the first) amongst women in Kalamazoo County, MI.
3. Investigate block group level neighborhood factors that influence rapid repeat pregnancy (second pregnancy ≤ 24 months after the first) amongst women in Kalamazoo County, MI.

Significance of the Study

There are several studies on rapid repeat pregnancies in the US; however, most do not situate discussions in a geographical context to gain understanding of the variations, or spatial patterns, of the problem. Thus, this study aimed to fill the method gap on the topic and highlight the significance of geography in understanding public health issues.

The study also added to the body of knowledge about RRP with analysis at the block group. With most studies focused at state and national level in the United States, individual and neighborhood block group information is an additional dimension to existing knowledge on RRP.

Study Area

Kalamazoo County can be found in southwestern Michigan 40 mile east of Lake Michigan shown in (Figure 1).

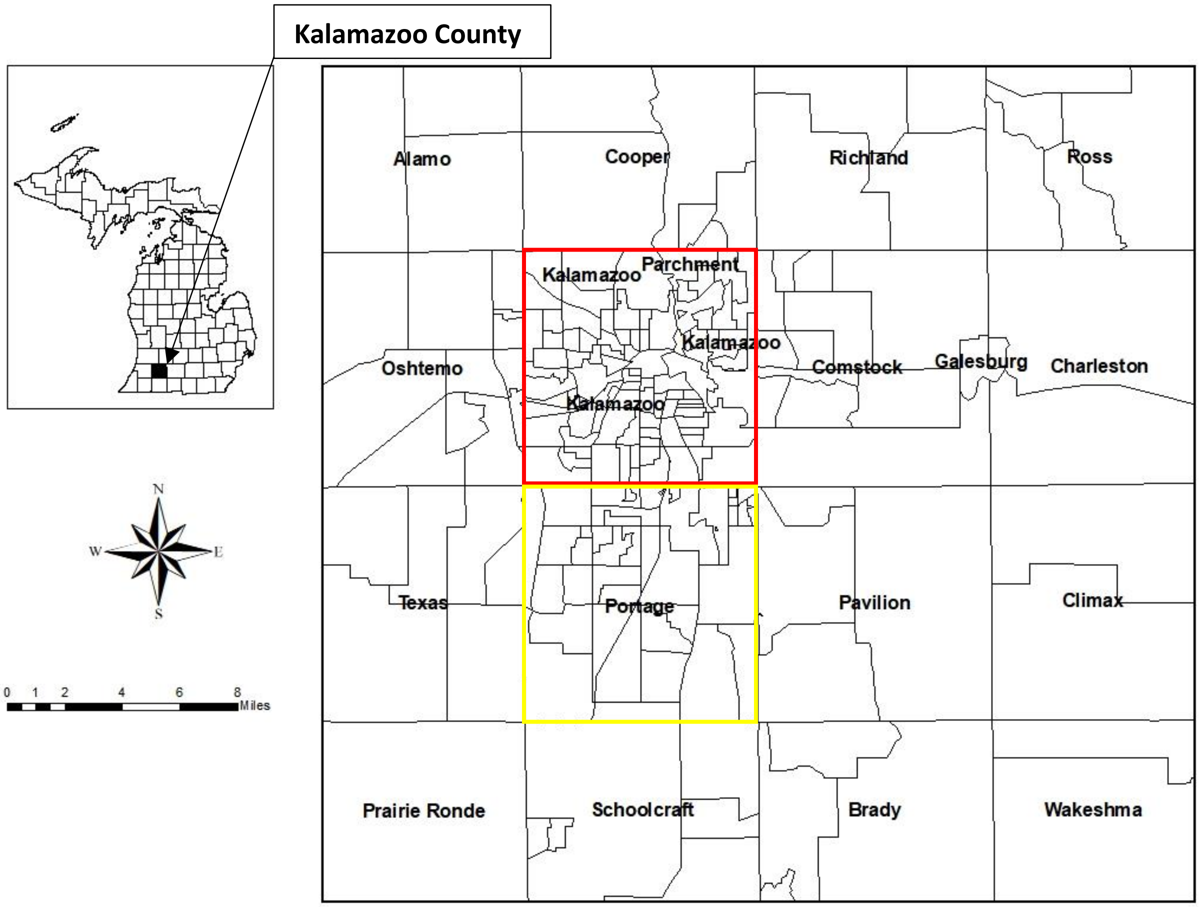

Figure 1. Block Groups in Kalamazoo County, MI
Source: Created by Author

The county has an estimated total population of 262,985 with 51% and 49% of females and males respectively (United States Census Bureau, 2019) making it the 9th most populous county in the State of Michigan. The major cities in the county are Portage and Kalamazoo located in the core of the county, with other areas designated as townships and villages. Figure 1 shows a map of the major cities and Townships and the 89 block groups in Kalamazoo County with a locator map to show the location of the county in Michigan.

CHAPTER II

LITERATURE REVIEW

This section of the thesis provides the conceptual and theoretical bases for the study by reviewing previous literature about rapid repeat pregnancy. This literature review is divided into 3 parts which include:

1. Discussing on intentions and the power (social and economic) related factors that influence RRP.

2. Highlighting the importance of studying health issues in geographical and spatial terms.

3. Exploring debates about health disparities in the United States through studies that have been done in this area to deepen understanding on the socio-economic differences and patterns that affect people's health individually and within areas they find themselves.

Intentions and Power

A key issue of the RRP discourse is about the causal factors. Studies have pointed out different causes spanning social and economic characteristics that vary across individuals and neighborhoods. At the heart of these factors is women's power role in intentionally or accidentally getting pregnant. Boardman et al. (2006), using a polytomous multiple logistic regression model to predict risk factors of repeat pregnancy in the United States, points out that the intention behind a pregnancy is critical in the timing of pregnancy and relative pregnancy interval. Baldwin & Edelman, (2013) argue that pregnancy amongst adolescents, particularly, are unintended. In fact, approximately two thirds of adolescent RRPs reported in the United States are unintended. However, women are now intentionally avoiding early childbirth and purposely initiating pregnancies at an older age and a faster rate. For instance, Haight (2018) in examining association between short inter-pregnancy intervals and adverse outcomes by maternal age among U.S. women argues that with several studies establishing association between adolescent RRP and adverse health outcomes, average age at first birth is increasing and older women are intentionally giving birth to their children at shorter intervals before it is too late.

In terms of economic factors that influences intensions to rapidly repeat, Ranieri and Wiemann (2007) in studying socio-ecological predictors of RRP amongst adolescent girls in Texas found girls living in low income communities and not enrolled in school were associated with giving birth at an early age. These young women also tended to have shorter time to a subsequent baby than their counterparts in high income communities.

Intentions concerning pregnancy can also be argued to evolve overtime as some studies show. Despite the emotional and psychological stress women go through during pregnancy the reward (baby) sometimes serve as a motivation to go through it again. This is affirmed succinctly in the responses of two women in a qualitative study reviewed by Aslam et al., (2017): “My baby needs a brother or sister it is too sad to see him growing up without someone to play with”. A second woman also said: “Now that I have had one, I should just finish it, you know, before going back to school and dropping out all over again.” These statements suggest reformed intentions after giving birth to a first baby. Hence though a first baby may be unintended the subsequent ones could be intended, and the first baby sometimes creates the urge to immediately have another baby.

Socially, a systematic study in the United States in 1998 revealed that first pregnancies are often intended by women. Husbands or male partners often have higher influence in intention and decisions on subsequent pregnancies after an index birth. This highlights the influence of men in pregnancy decision making a factor that is often not considered in RRP studies (Baldwin & Edelman, 2013). Cha et al., (2016) affirms this in a study on pregnancy intentions amongst couples and rapid repeat pregnancy by suggesting that RRP is strongly influenced by paternal rather than maternal pregnancy intentions as subsequent pregnancy and timing are often influenced by partners of the women.

The flip side to the intention conundrum is the issue of individual power to make decisions or implement an intention. Power dynamics are critical in understanding health behaviors of people. Power can be defined as the degree of control over material, human, intellectual and financial resources exercised by different sections of society with some individuals and groups having greater control over the sources of power and others having little or no control (Corbin et al., 2017). In this regard, it can be said that power dynamics, seen and unseen, between men and women influence decisions regarding number of children and intervals for giving birth, etc. As confirmed earlier by Baldwin & Edelman, 2013; Cha et al., 2016, partners of women who repeat pregnancies have a key role in determining subsequent pregnancies and the intervals of these pregnancies.

The social and economic intensions and power discussed above also influence contraception which is a key method and strategy for addressing RRP. As a primary intervention for RRP, contraceptive use is based on power of an individual to obtain and use contraceptives (economic power) as well as the ability to negotiate contraceptive use with a partner (social power). It must be noted that contraceptive use varies based on access, cost and counselling. Mestad et al., (2011) in a method choice study of 5086 women in Washington University, St Louis school of Medicine to investigate contraceptive choice, age and cost on the contraceptive CHOICE project which provided long term contraceptives at no cost found out that, of the 5086 women enrolled on the program, 70% chose Long Acting Reversible Contraceptives (LARC) as a preferred contraceptive. This at face value shows the role of cost in disempowering women to access the contraceptive of choice. LARCs are the preferred choice of protection from rapid repeat pregnancy as it has a higher Couples Year of Protection (CYP). However, this option is often not available to most women particularly the minority and low socio-economic status women. Lack of power to control intentions and decision making is evident in many studies on the subject. In Canada, a study of girls with intellectual and development disabilities as compared with a sample of those with no disabilities revealed that girls with disabilities who lived in low-income neighborhoods and received social assistance were at a higher risk of RRP (Brown et al., 2018). Furthermore, education on the importance of contraceptives before pregnancies is a good option for reducing rapid repeat pregnancies.

However, discussions on the importance of contraceptives often places less emphasis on the side effects which is a major challenge of this option. Aslam et al., (2017), in a systematic review of programs aimed at addressing RRP issues found that many women experienced side-effects with more reliable methods of contraception and often experience a repeated pregnancy when they stop to switch to another form of contraception. Knowledge, cost and side effects could thus be motivating or inhibiting power factors that influence RRP. This sets the premise that women need to be power balanced to be able to make decisions regarding sexual and reproductive health. Consequently, any situation that moves women from this position of being able to make decisions in a safe setting enhances vulnerability and adversely affects their reproductive health particularly for minority and disadvantaged women who are vulnerable to getting pregnant and giving birth at shorter inter-pregnancy intervals.

Geography and Health

A popular statement published in public health studies about the relationship between where people live and health is the words of Rossi (1972) in his book titled “The Human Meaning of Social Change” where he describes the local community we have our life course in as “. . . it supplies to its individual citizens the medical facilities in which he is born, the schools in which he is taught, the housing in which he lives, the social milieu in which he finds his mate and sets up his household, the factories and businesses in which he finds employment and finally the cemetery in which he is buried” p. 89. Public health studies have for some time focused on understanding patterns in terms of the physical environment, pollution, water sources and how these things affect health. Other studies have also paid attention to social, economic, cultural and spatial inequalities with regards to socio-economic inequalities, inequities and poverty.

The relationship between place, space, environment and health has its roots in the history and philosophy of geography. Geographers have for a long time associated different places and the environment with health. From theories such as environmental determinism that related the environment to diseases in different locations, continuing to studies of human ecology that concerned relationship between social and economic structures in what is known as social epidemiology, geography has played a key role in understanding health (Cresswell, 2013). Indeed, where one is born, live, work, the social and built environment are explained to be associated with one’s health (Dummer, 2008). Geographical studies have thus focused on these parameters to provide better understanding to health studies through statistical geographical analysis and mapping. A significant historical study on using mapping to identify and address a public health problem is traced to John Snow a British geographer who was able to establish the relationship between drinking water and cholera cases in 1936.

Some studies have focused on patterns in terms of the physical environment, pollution, water sources and how these things affect health. Other studies have also paid attention to social and spatial inequalities with regards to socio-economic inequalities, inequities and poverty. Aside the built environment, the relationships among individuals in a small location such as a community or neighborhood affects health. For instance, Diez Roux and Mair (2010), highlight the importance of neighborhoods in contemporary public health studies as they have social and physical characteristics which do not only explain individual characteristics but also characteristics of similar group of race, families and neighborhoods. Tobler's first law of geography states that everything is related but closer things are more related in space hence in order to understand patterns of health, distribution of diseases and causes of ill health it is important to understand an individual from the environment they live as well as the characteristics of people who engage and live close together in a geographical unit such as a county, census tract, school district, census block etc. Cozier, (2017) suggests that it is important for researchers to consider the purpose of a study before choosing the scale, adding that studies at the county level will be useful for policy setting or economic structure. Otherwise, block groups provide a more homogenous frame for analysis of the socio-economic environment of residents and very useful for disparity studies covered at the census tract level.

Health studies have now embraced the use of Geographic and Information Systems (GIS) with focus on spatial distribution and neighborhood effects on health etc. GIS is used in collecting, analyzing and mapping health data to find trends and spatial distributions for better intervention measures (Fradelos et al., 2014). Methods such as spatial autocorrelation and geographically weighted regression, point patterns and overlay analysis are a few of the methods used in understanding health data and differences that may exist. The instruments supporting this field include GIS, disease surveillance, big data, and analytical approaches like the Geographical Analysis Machine (GAM), Dynamic Continuous Area Space Time Analysis (DYCAST), cellular automata, agent-based modeling, spatial statistics and self-organizing maps (Musa et al., 2013).

Relevant Methods in Geospatial Analysis of Public Health Outcomes

Geospatial analysis focused on Public health outcomes adapt some key methods to analyze patterns as well as establish relationship between determinants of health and health outcomes. Some key methods adapted in health literature and relevant in this study are also discussed. These include hotspot analysis (Getis Ord G*), binary logistic regression, simple linear regression and principal component analysis.

Hotspot Maps (Getis Ord G*)

The Getis Ord G* statistic developed by Getis and Ord (1992) is a local spatial autocorrelation based on the premise that spatial associations are locally heterogeneous. Generally, a feature with a high value is interesting; however, it may not be statistically significant. Statistically significant features are not identified based only on their individual value with the Getis Ord G*, but by the analysis of z-scores and p-values of features with high values or low values that are surrounded by other features with high or low values, respectively. High-value features with high-value neighbors are hotspots, while low-value features surrounded by low-value features are cold spots (Mitchell, 2005). The method has been used extensively in research to identify the clustering of populations, diseases, health care availability, crime incidence, food retailing, etc. In health studies, for instance, Wang et al., (2012) used the method to identify localized cluster patterns of late-stage breast cancer in the State of Illinois, United States. Geographical patterns of end-stage renal disease and kidney transplant at the county level in 11 states in the Midwestern US have also been analyzed using Getis Ord G* to measure the local spatial clustering tendency of end-stage renal disease rates. (Cao et al., 2016). In recent times, Stopka et al., (2018) in identifying and characterizing hepatitis C virus hotspots in Massachusetts, also used the Getis Ord G* to identify the location of statistically significant clusters of census tracts with higher (or lower) values for HCV cases and infection rates. Though these studies provide evidence of the frequent use of the method in health studies that focused on identifying clusters, a thorough search of studies that have applied this method in studying rapid repeat pregnancies did not reveal any in the United States.

Regressions (Linear and Binary)

Regression analysis is a basic statistical method which is used in finding relationships amongst variables. Often a variable that needs to be explained (dependent) is related to other variables (independent) to find out how best these variables can explain the dependent (Campbell and Campbell, 2008). In a simple mathematical model, it can be represented as the relationship between the dependent variable Y and independent variable X shown below.

$$Y = \beta_0 + \beta_1 X + \varepsilon_i$$

β_1 gives the magnitude and direction of the slope with β_1 as the intercept and ε_i as the error term of the amount of variation not accounted for by the intercept and slope terms. This mathematical formula is a straight line and hence this represents a linear relationship. Linear regressions however are based on five key assumptions which are:

1. There is a linear relationship between dependent and independent variables
2. For the ıth level of the independent Xı the expected value of the error component is equal to zero

3. The variance of the error component ε_i is constant for all levels of X (homoscedastic)
4. The values of the error component for any two ε_i are pairwise uncorrelated
5. Error components are normally distributed.

As a result, these assumptions should underlie statistical analysis that find linear relationships amongst outcomes. In Public Health studies, regression is commonly explored as an important statistical tool to establish the relationship of a response with explanatory variables (Liang and Zeger, 1993). For instance, Kothari et al., (2016) used a multilevel regression to test the relationship between race and socio-economic status at the individual and neighborhood in Kalamazoo County. However, in some cases the assumptions listed above may be violated in a regression model and there will be a need to use alternative regression methods. Particularly in cases where the dependent variable is categorical or binary, the assumptions of a simple linear regression including linearity, normality and continuity are violated and hence logistic regression becomes a preferred alternative (Abedin et al., 2016). Logistic regression, unlike a linear regression, estimates the probability of an event occurring or not occurring by fitting data to a logistic curve. There are generally two logistic models that is binary and multinomial logistic regressions. The binary type is used when the dependent variable is dichotomous, and the independent variables are categorical or continuous (Park, 2013). However, if the dependent variable consists of more than one category, a multinomial logistic regression is most appropriate.

Mathematically, the logistic regression fits a regression curve y =f(x), which consists of binary coded (0, 1, eg. Yes, or No). Because binary data can cause the predicted outcomes to exceed 1, the logistic model uses a function called the Sigmoid or Logistic function ($\frac{1}{1+e^{-x}}$) to squash the output of the curve into an S shape to fit between 0 and 1.

Table 1. Logistic Regression and Rapid Repeat Pregnancy Studies

Author	Title	Year	Method
Ranieri and Wiemann	Social Ecological Predictors of Repeat Adolescent Pregnancy	2007	Logistic Regression
Tocce et al.,	Rapid repeat pregnancy in adolescents: do immediate postpartum contraceptive implants make a difference?	2012	Univariable analysis and multivariable logistic regression
Reese and Halpern	Attachment to Conventional Institutions and Adolescent Rapid Repeat Pregnancy: A Longitudinal National Study Among Adolescents in the United States	2017	multivariate logistic regression
Loree et al.,	Postpartum Contraceptive use and Rapid Repeat Pregnancy Among Women who use Substances	2018	Stepwise logistic regression
Cha et al.,	Discordant pregnancy intentions in couples and rapid repeat pregnancy	2016	Multiple logistic regression

Source: Created by Author

Like linear regressions, logistic regressions also have assumptions which include:

1. Dependent variable needs to be discrete preferably dichotomous
2. Dependent variables should be coded as the probability of an event
3. Model needs to be fitted correctly
4. Requires each observation to be independent (little or no multicollinearity)
5. Independent variables are linearly related to the log odds of an event.
6. Works better with large dataset

In health research, most commonly in rapid repeat pregnancy studies, the logistic regression is most preferred. Table 1 gives a sample list of rapid repeat pregnancy studies that have used logistic regression.

The plethora of studies on the subject using this method probably highlights the multifaceted nature of rapid repeat pregnancy. It's difficult to model a linear relationship but possibly better to analyze the various groups of rapid repeaters and the factors that influence them.

Principal Component Analysis

Principal component analysis (PCA) is a statistical technique that can reduce dimensionality of datasets and extract relevant information from a large dataset while reducing information loss (Shlens, 2014). The PCA method uses orthogonal transformation to represent independent variables that are correlated with principal components or factors that are not correlated linearly. Scholars have found it particularly useful in regression analysis as a tool that provides the right uncorrelated independent variables to be included in a model in order not to violate the assumption of independence in regression (Zhang and Castello). The analysis can be done in SPSS by loading the variables of interest in the PCA tool and specifying the essential parameters expected in the output. The PCA reports the sum of squares within each component as the components variance 'eigenvalue' which is the explanatory strength of the component. Eigen values that are often greater than one is retained and components with the greatest eigen values are the principal components. The benefits of using a PCA is that it can provide patterns in a dataset by loading the logically reduced number of components or variables that explain variance. However, an important limitation to note is the structure of the dataset. Lever et al.,

(2017) observe a limitation with PCAs and that is, if data at different scales are inputted in a PCA, the PCA will only recover data with higher magnitude and thus very important to standardize data and check the structure of the dataset to avoid wrong outputs. In Public health, the principal component analysis is a powerful tool in understanding socioeconomic factors that cluster at the neighborhood level. Friesen et al., (2016) mention that the PCA is a powerful tool in developing area level socioeconomic indices that are often mapped to provide a visual understanding of differences in neighborhoods in terms of health and inform public health resource allocation, service delivery, and program dissemination as it provides a more comprehensive understanding of communities' levels of disadvantage in relation to one another.

Health Disparities

The differences that exist in the health status and healthcare across gender, race or ethnicity, education, income, disability, geographic location and sexual orientation is what is termed as health disparities Riley, (2012). For instance, Thomas et al., (2009) in a study to understand the neighborhood factors affecting rates of sexually transmitted diseases in Chicago using survey data collected from the 1995 Program on Human Development in Chicago and homicide rates found out from the binomial regression model that neighborhoods with high rates of incarceration had higher rates of STIs (chlamydia and gonorrhea) when compared to those with low incarceration. Health disparities can also be viewed from the perspective of health inequality and inequity discussed below

Inequality and Inequity

Some studies mention that differences in health could reflect the inequality that exist in social systems or the sheer systemic, intentional, avoidable and unfair distribution of health resources that is (health equity) (Meyer et al., 2013; Graham 2004). These may be due to policies, structures and the general environment that has an adverse effect on the health of some groups or neighborhoods particularly those considered as minority. Health disparity research have evolved overtime from being basically descriptive and trying to establish association between inequalities and inequities in socio-economic status and health and unfair health distribution in health as well as mechanisms that links these factors rather than focus on the interactions amongst the factors to know what actually causes health disparities in different places (Adler and Stewart, 2010; Omrani-Khoo, et al, 2013). Diez Roux and Mair, (2010) calls for a change in the methods used in understanding health disparities recommending a mixed method approach which provides details beyond the statistical figures to really ascertain perspectives of minorities and the disadvantaged to help provide effective evidence-based interventions to address health inequalities and inequities. Krieger 2014 in a review of articles about health discrimination, concludes that disparities in health are structural and caused by global and local agencies and governments systematically creating disadvantaged minority groups through their activities and actions. Some of these are embedded racism and segregation that an individual must go through the course of life. These structures limit economic participation which is a key tool for racial discrimination perpetuated by institutions and agencies in places like the United States (Bloome, 2014).

In the United States, poverty is experienced 2.6 times more amongst Blacks than whites while racially segregated Black neighborhoods have greater concentrations of poverty compared to white neighborhoods as a result of lack of jobs, and the intergenerational transfer of poverty as a result of the systems and structures that Black have historically lived in (Krieger, 2014; Acevedo-Garcia, 2009). Living in segregated, high poverty communities further increases the effect of individual poverty through exposure to distressed physical environments (pollution, dilapidated housing, zoning), fragmented social networks (social support, norms, crime, political power) and limited health-related resources (health care, nutrition, recreation, transportation) (Cook et al., 2009). These traces are evident in redlined communities even after it was banned (MacQuillan et al., 2019).

Contrastingly, some studies subtly argue against the inequality and inequity assertion particularly racial ones, arguing that there is no linear relationship and that race and ethnic backgrounds do not always predict health or health outcomes. If a study done by the Michigan Department of Health Statistic is something to go by then this argument may really be valid. Based on birth statistics for 2013 from the Michigan Department of Health Services, mothers living in poverty-level census tracts in Michigan have a significant lower incidence of gestational diabetes mellitus (4.6%) compared to mothers living in upper-middle class census tracts (5.3%). This shows that though racial disparities are evident as proven by several studies, the health issue at hand and the causal factors also differ in the level of the disparity. Satel and Klick (2006) in their book on "The Health Disparities Myth: Diagnosing the Treatment Gap", argue that health disparity studies pay too much attention to race and ignore geography which really is the causal factor in the differences between blacks and whites, suggesting that geography independent of racism determines the quality of healthcare, and black people happen to live in locations where healthcare is the worst. This is because people of color are often deprived of health equity as they often live in areas that suffer from unjust distribution of social, economic, political, and environmental conditions that determine health. Another argument against the racial inequality and inequity debate argue racial disparities to be as a result of genetic differences that yield to vulnerability of diseases and poor health outcomes. Most of these studies never even examined the genotypes of research subjects; they inferred a genetic source of racial differences when they failed to find another explanation (Roberts, 2012). A study by a team of obstetric researchers to examine the hypothesis that black race independent of other factors increases the risk for extreme preterm birth and its frequency of recurrence at a similar gestational age in using the Missouri Department of Health's maternally linked database of all births in Missouri between 1989 and 1997 for factors associated with recurrent preterm delivery report from the logistic regression analysis that black women are more likely not only to deliver preterm babies but also to have preterm births in subsequent pregnancies. This results still occurred when medical and socioeconomic factors were controlled prompting their findings were suggestive of a possible genetic component that underlies the often-studied public health problem of racial disparities and health in this case preterm births (Zachary et al., 2007). Although conceding that they may have overlooked ‘‘hidden variables’’ that also contribute, they nevertheless speculated about an unproven genetic mechanism operating in ‘‘the black race’’.

In the United States, racial health disparity is of critical concern even while other disparity indicators have seen improvements. A Risk Factor Survey of 30 communities in the United States by the Racial and Ethnic Approaches to Community Health program a subsidiary of the Centre for Disease Control in 2013 found that populations in minority communities continue to have lower socio-economic status, poor access to health care, greater risks for, and burden of, disease compared with populations in other communities in the same county or state. Lynch and Perera, (2017) also highlight gaps that persist between different races (Black American, White, Latinos, and American Indians), rural and urban areas, college degree holders and people with less than high school education.

Considering that the World Health Organization considers health as a human right and most nations including the United States considers it as a constitutional right, it is worrying that numerous studies find relationships between diseases and general health of some groups considered minority populations. A 2011 report by the Centre for Disease Control and Prevention on health disparities and inequalities amongst Americans outlined disparities amongst ethnic and racial groups, states, gender, different income groups, age groups, rural and urban areas. Adler and Rehkopf, (2008) in examining definitions and studies on health disparities, show that there are consistent disparities for individuals with less resources, blacks and ethnic minorities, however the variation in health disparities should also be considered based on the different classes in the same group. There exist social-class health differences amongst people in the same neighborhood and groups. Earnshaw et al., (2017) explore experiences of discrimination within six low-resource neighborhoods of New Haven, CT, that experience social and health inequities in comparison to residents of neighboring communities using the intersectionality model found that community members who are socio demographically similar may have diverse discrimination experiences. These debates clearly show that inequalities and social justice are the underpinning variables for health disparity. Roberts (2012) notes that countries that control this have better health. For instance, people in Japan, Sweden, and Norway live longer, are less obese, and have fewer teenage births than people in the United States, the United Kingdom, and Australia, because their societies are more equal.

Rural-Urban Health Disparities

Rural-Urban health disparities are also a key part of the United States health system and largely discussed in several studies (Ricketts, 2000; Hart et al., 2005; Douthit, 2015). Access to health care including physical and economic access has being an issue that has been extensively discussed as causing the gap or difference between rural and urban areas. Lu et al., (2010) in a study of health insurance coverage and patterns amongst population aged 16 to 64 years in Kentucky mention that there is a huge rural-urban disparity not only in the number of people who have or do not have health insurance but also differences in the specific types of insurance an individual can afford citing differences in patterns of employment, and population characteristics as pattern indicators. A report by the North Carolina Rural Health Research Program (2017) alludes to the disparity in rural and urban areas accessing health care suggesting that rural folks who live in ranches, farms and frontiers and have relatively lower income status often battle with travelling longer distances to access quality health care. Williams et al., (2015) in a study of breast cancer in 19 counties in Missouri reports that rural residents had to travel 45 minutes one way to access mammography services thus leading to women in 19 counties having higher rates of late stage cancer compared to women in the urban areas. In terms of access, the need for health and the health seeking behavior which obviously is influenced by education (knowledge of health) and (income) the willingness to invest in health differ (Ziller & Lenardson, 2009). Population studies have revealed that lower education and income levels among some rural residents may increase reluctance to seek healthcare services (Ricketts, 2000; Hart et al., 2005). Both the 2012 National Healthcare Disparities Report and the 2012 National Healthcare Quality Report found that almost none of the disparities in access to care are improving. In addition, quality of care varies not only across types of care but also across parts of the country.

Socioeconomic Status and Health in Michigan

Michigan is a Midwest state and predominantly a rural area. According to the Citizens Research Council of Michigan (2018), though Michigan is regarded as a rural area, the characteristics of rural and urban Michigan are not significantly different. Health Data Statistics from the Kaiser Family Foundation shows that, the average per capita income of rural Michigan is relatively low ($37,936) compared to the State average of $46,201. The American Community Survey also reports that the poverty rate in rural Michigan is 13.9%, compared with 14.2% in urban areas of the state. Additionally, 9.8% of the rural population has not completed high school, while 9.8% of the urban population lacks a high school diploma according to 2013-2017 ACS data. Unemployment rate in rural Michigan is at 5.7% while in urban Michigan it is at 4.4% (USDA-ERS, 2017). Rural Michigan is, on average, older than urban Michigan and rural residents are more likely to be military veterans, to be married, to own the homes they live in, and the length of time they live in their homes means that often they grow to be established members of their community. Aside the diverse racial differences in Michigan, urban Michigan also has a large immigrant population (Citizens Research Council, 2015). The differences in the socio-economic status of Michigan residents has been studied extensively in relationship with health. A Committee on Community-Based Solutions to Promote Health Equity in the United States reports that social, economic and environmental factors have caused a number of health issues in Michigan including lead water contamination poisoning of children in Flint and some other parts of Michigan, opioid drug crises amongst low-income rural communities are setting to draw attention to socio-economic status and health (National Academy of Sciences Engineering, and Medicine, 2017). El-Sayed (2015) examining socio-economic position, health behaviors and racial disparities in infant mortality in Michigan analyzed about 2,087,191 mother child dyads between 1989 and 2005. Using multivariable Poisson regression models of infant mortality, adjusting for socio economic position and maternal risk behaviors explained nearly a third of the disparity in infant mortality overall, and over 25% of disparities in several specific causes including homicide, accident, sudden infant death syndrome, and respiratory distress syndrome. Socio-economic position and maternal risk however, had little influence on disparities in other specific causes such as septicemia and congenital anomalies. Again, socio-economic position and maternal risk might not influence health all the time but the differences in health and socio-economic characteristics is a critical issue in health.

Access to Health Care in Michigan

Physical and economic access to health care is also an interesting dimension of the health system in Michigan. Meden et al., (2002) in examining the relationship between travel time and utilization of breast cancer in rural Northern Michigan reviewed 81 medical records of patients treated for breast cancer from 1999 to 2002 and note that association between travel distances to radiation treatment and the utilization of BCT in rural region of Michigan where the nearest radiation oncology center was about 150 miles from patient's homes. Buttressing the notion that rural dwellers travel long distances in order to access health care. A 2015 study by the Citizens Research Council of Michigan in exploring the Michigan rural and urban divide found four rural counties in Michigan including Cass, Keweenaw, Lake and Oscoda consistently fell below recommended ratios of primary care physicians to population. Additionally, the Center for Health Workforce Studies of the Association of American Medical Colleges projects a shortage in Michigan of 4,400 doctors - including both primary care doctors and specialists by 2020.

In terms of economic access to health, insurance is an important part of health for people in Michigan. Data from the American Community Survey 2010 to 2015 suggest that the proportion of Michigan population without health insurance has reduced from 12.4 to 6.1 with urban residents likely to purchase private insurance than rural residents opting for public insurance (Citizen Research Council of Michigan, 2018).

Bourgi et al., (2016) in a study in Health Disparities in Hepatitis C Screening and Linkage to Care at an Integrated Health System in Southeast Michigan using univariate analytical methods notes that, Medicaid beneficiaries were significantly less likely to be treated than Medicare and commercial insurances (10% vs. 35%, $P < 0.05$). This highlights the important role one's insurance status has on access to health care. In the light of the struggle to get an appropriate health insurance coverage to access health, work can also be a limitation in qualifying for Medicaid. For example, proposals for Medicaid work requirements will cause many low-income adults to lose health coverage, including people who are working or are unable to work due to mental illness, opioid or other substance use disorders, or serious chronic physical conditions, but who cannot overcome various bureaucratic hurdles to document that they either meet work requirements or qualify for an exemption from them. These coverage losses will not only reduce access to care and worsen health outcomes but will likely make it more difficult for many people to find or keep a job. Thus, Medicaid work requirements may be self-defeating on their own terms. People who live in counties with higher unemployment rates above 8.5 percent are exempted from the requirement. That is likely to lead in practice, as Kaffer observes, to rural whiter counties, where unemployment is higher, getting a break from these work requirements while urban areas with a higher share of black residents would still be subjected to them. Which means that black Medicaid enrollees would be more likely to lose their health insurance. Tipirneni et al., (2018) studying geographic variation in Medicaid acceptance across Michigan care practices in the era of the affordable care act examines geographic variation in Medicaid acceptance among Michigan primary care practices before and after

Medicaid expansion in the state, using data from a simulated patient study of primary care practices. Using logistic regression analysis with time indicators to assess regional changes in Medicaid acceptance over time found that Geographic regions with lower baseline (<50%) Medicaid acceptance had significant increases in Medicaid acceptance at 4 and 8 months post expansion, while regions with higher baseline (≥50%) Medicaid acceptance did not experience significant changes in Medicaid acceptance. The expansion of Medicaid seems to hence have benefited areas that had low coverage hitherto the affordable act. Contrasting to the increase in coverage, MacQuillan et al., (2019) in a study of Geospatial Analysis of Birth Records to Target Programming for Mothers with Gestational Diabetes Mellitus in Michigan, 2013 reports that there is no difference in Gestational Diabetes Mellitus risks of women on paid Medicaid and non-paid Medicaid. Further mentioning that, the introduction of the Affordable Care Act and Medicaid expansion and adequacy of prenatal care among low-income women in Michigan has increased and disparities between women who are users and non-users of Medicaid has been eliminated. It is evident that there are challenges in accessing health and health status of Michigan residents emanating from the diversity of the population, structural and systemic challenges as well differences in the characteristics of different places which needs studies to provide critical targeted interventions.

Summary

It is clear from literature that there are socio-economic characteristics are associated with RRP. Additionally, other undercurrent factors such as intensions and power also influence RRP. It is important to hence consider variables from these perspectives to understand factors that affect RRP at the individual level. However, RRP as a health problem do not occur only at an individual level, the geographical setting and characteristics around an individual also influence health. With geographical theorist suggesting that closer things are more related than farther things it is important not only to study individuals but also to understand patterns and the differences that exist in different places and how this can affect health in this context RRP.

CHAPTER III

METHODOLOGY

This section details the research design, data, and methods for the study. It specifically describes the data and variables that are analyzed, statistical analytical measures, and methods. It also clearly defines rapid repeat pregnancy in the study context, as well as GIS methods for the spatial analysis. Ethical issues for the study are also discussed.

Data

Data is an essential part of the study and ensuring a good data pipeline enhances the success of any research. This study employed the OSEMN data science approach which involves obtaining, scrubbing, exploring, modeling and interpretation of the data. Table 2 details how this concept was applied to this study and subsequent sections will further elaborate on it.

Table 2. Data Process

Obtain	Scrub/Clean	Explore	Modeling	Interpretation
• Individual data obtained from State of Michigan Statistics for Kalamazoo County • 2013 ACS data downloaded for neighborhood	• Geocoded points were plotted on a Kalamazoo shapefile for accuracy • Data aggregation with SPSS and ARCGIS	• Basic mapping to visualize data distribution • Basic statistics • Choropleth mapping	• Getis Ord • Binary Logistic • Regression • Linear regression • Principal Component Analysis	Enabling connections with findings and literature

Source: Created by Author

This study employed a retrospective approach to analyze secondary databases containing birth records from State of Michigan Vital Statistics for Kalamazoo County and received through Western Michigan University Homer Stryker School of Medicine (IRB#: Wmed-2017-0179). The initial purpose of his data was to conduct a study for the Kalamazoo Healthy Babies-Healthy Start program as part of evaluation efforts aimed at understanding the interaction of socioeconomic factors and race in predicting poor birth outcomes and infant mortality. The initial study involved linkage and analyses of the following datasets:

1. Kalamazoo County birth records dataset (2006-2015), with identifiers
2. Kalamazoo County linked birth-death dataset (2006-2015), with identifiers
3. Kalamazoo County infant death certificates (2016-2017), and matched birth certificates
4. Kalamazoo County Prosecuting Attorney Adult Case Tracking administrative database of charging requests (2005-2017)

However, this study makes use of the Kalamazoo birth records dataset (2008 to 2014), with identifiers. Only singleton births, as opposed to multiple births (i.e. twins, triplets etc.), are considered. The study population include:

(1) Mothers who delivered singleton births while residents of Kalamazoo County MI and delivered at least one baby in 2010; (2) All singleton births born to these mothers from 2008 to 2014. From the data, 2861 women gave birth in 2010, giving birth to a total of 4745 babies during the period 2008 to 2014. The highest number of children born by a mother within the period being 5 and the lowest 1. The birth records dataset contains much useful information, as shown in Table 3 below.

Table 3. Birth Records Data Variables for Kalamazoo County, MI

Data set	Variable(s)
Birth Records Database & Birth Certificate information	Identifiers: maternal first & last name, maternal date of birth, paternal first & last name, paternal date of birth, infant first and last name, infant date of birth, birth certificate number Demographics: *Infant:* gender, race, ethnicity *Maternal:* race, ethnicity, age, marital status, education, Medicaid-paid birth, or not. *Paternal*: race, ethnicity, age, education, named on the birth certificate (yes or no) Birth characteristics: plurality, gestation, birth weight, Apgar scores, infant medical risks, delivery risks, NICU admit, infant seizure/injury/ventilation, abnormal conditions Maternal obstetric hx: previous pregnancies/births, prenatal care hx, maternal residence, maternal health risk factors, obesity, prenatal weight gain, smoking (maternal, quit, household) Maternal plans related to infant care: breastfeeding, WIC Geocode address: longitude, latitude

Source: Created by Author

For the purpose of this study, age, race, education, Medicaid-paid birth or not, and identification of paternity on birth certificate are the essential variables considered for this study to help establish a relationship between socio-economic factors and rapid repeat pregnancy at the individual and neighborhood levels. For useful analysis at the block group level, the variables of interest in the dataset are categorized in a manner that ensures that data can be aggregated to the block group. Values of 0 and 1 are assigned to each variable based on the categorization. This was done by sorting the maximum value of each of the variables by the study period and assigning values to them. Table 4 below shows the meaning of the assigned values and categorization.

Table 4. Coding of Study Variables

Variable	Groups	Coding
Age	Ever less than 20 years during any of the births	(1)
	Never less than 20 during any of the births	(0)
Race	Ever non-white	(1)
	Never non-white	(0)
Education	Ever any college	(1)
	Never any college	(0)
Insurance	Ever Medicaid paid	(1)
	Never Medicaid paid	(0)
Paternity identified on birth certificate	Ever paternity identified (Dad named)	(1)
	Never paternity identified (Never dad named	(0)

Source: Created by Author

After the categorization of the variables, Python Shell 2.7.14 is used in sorting data by mom ID to clearly see each mom and the number of babies they had. Inter-pregnancy intervals between births were then calculated by finding the differences between birthdays of first and second order child and subsequent births per mom ID. This was done in days and then converted to months. The file was then saved as a CSV file and imported in ArcGIS 10.7.

The data received for the study was already geocoded. It is probable the data was not received for mothers with missing address information or cases where post-office boxes were reported in place of residential addresses as reported in other studies. The geographical coordinates of each mom were plotted and overlaid on a Kalamazoo county street base map to validate if the points fell within Kalamazoo County. These plotted points were projected in North American Datum 1983. The points were exported into a shapefile and overlaid on a Block Group shapefile of Kalamazoo County, which is the unit of analysis. A spatial join was done to locate the Block Group of the various coordinates plotted. The coordinates were deleted to avoid identifying the specific location of the moms. This shapefile was again exported as a different shapefile containing the same data at the Block Group. The next step was data aggregation; data is aggregated at the block group level by three mom groups, including rapid repeat moms, slow repeat moms, and single birth moms. For the purpose of this study, the various groupings of moms are defined as follows. Rapid Repeat Moms: Considered as women that gave birth in 2010 and had all previous or successive singleton births with an inter-pregnancy interval less than 24 months; Slow Repeat Moms: Considered as women that gave birth in 2010 and had successive singleton births with an inter-pregnancy interval more than 24 months; Single Birth Moms: women who only gave birth in 2010 and did not repeat a birth after or had no prior birth from 2008 to 2010. Data aggregation was done with ArcGIS 10.7.1. In ArcMap, the attribute table of the joined block group shapefile of Kalamazoo and CSV file containing all mom's dataset was opened. The dataset was summarized using the ‘summarize' function in the attributes table. Based on the GEOID_DATA field, which is the field indicating the unique identity of the block groups, all other fields of variables and relevant data were summarized. The variables were found using either sum, maximum, standard deviation, average, minimum, or a combination of some of these operations when relevant. Appendix 3 shows the kind of operation done on each field and the output derived from it in the summary table for all moms. The same operation was repeated for the 3 groups of moms. The Rapid Repeat moms group derived from the maximum sum of repeat moms was summarized to get a summary table for all RRP moms. A query was done based on the count of mom IDs. Mom IDs were equated to 1 to get moms that gave birth once within the study period. The variables in that Table were also summarized. Lastly, another query was done to derive slow repeat moms using fields of mom IDs and RRP moms. The expression built was count_M_ID>2 & allRRP<1 (count of mom IDs>2 and All RRPs<1).

Block Group Neighborhood Variables

Secondary data was also obtained for the neighborhood level analysis. For the purposes of this analysis, the block group is considered the local neighborhood for each mother. These data were downloaded from the American Fact Finder website. Data downloaded were based on variables that were present in the individual dataset and factors that were identified in literature to affect rapid repeat pregnancy. For consistency, data were based on the American Community Survey 2013 (3-year estimates). Data obtained included educational status of females, race of females, median household income of the entire block group, total population of the block group, marital status of women in the block group, median age of the block group and ages of women in the block group. All data were downloaded in a CSV format and subsequently cleaned by getting rid of unwanted fields, renaming the fields to provide better understanding, as well as combining data into one simple file for analysis. Data was then subsequently joined in ArcMAP by the block group name field to ensure that data was joined to the appropriate block group.

Ethics

Privacy and confidentiality in health research are paramount, as studies often involve subjects whose details researchers have a responsibility to protect (Stevens, 2013). Usually, studies use methods that deidentify locations of subject or altering point locations of individual-level data to avoid reidentification upon release of data or by experts (geographic masking) (Zandbergen, 2014). In the United States, data are considered deidentified based on the HIPPA privacy rule if the data do not "identify an individual and if the covered entity has no reasonable basis to believe it can be used to identify an individual" (Haley et al., 2016). Against this backdrop, privacy and confidentiality were ensured by first removing specific longitudes and latitude points of mothers after plotting them to see specific points before data aggregation. Additionally, data is aggregated to the block group to avoid the identification of individuals by people who are familiar with the study area or experts.

Furthermore, there are no paper formats of the data as all data after aggregation are stored on an encrypted external hard drive, always under lock and key. No part of the data was also shared online, and data was only available to the study team, including the principal investigator and advisor. All processes required by the Human Subject Institutional Review Board (HSIRB) of Western Michigan University and the Homer Stryker School of Medicine were duly followed to adequately protect the human subjects being studied and to ensure that the study benefits the subjects being studied rather than harm them (IRB#: Wmed-2017-0179). The researcher also completed the Research Ethics and Compliance Training of the CITI program.

Conceptual Model

The conceptual basis of the project is shown pictorially in Figure 2. The listed individual characteristics were examined for a relationship with rapid repeat pregnancy. In this study, the Getis Ord G* method found in the toolbox in ArcMap 10.7.1 was used to identify spatial clusters of moms with rapid repeat pregnancy in Kalamazoo and to identify hotspots of the independent variables under investigation. The conceptualization of spatial relationships in the statistic was determined using the ‘continuity edges and corners' to be able to test spatial relationships amongst features sharing borders by edges or corners. In this regard, hotspots of rapid repeat moms were considered at the scale of U.S. Census block groups. Hence block groups that had high number of rapid repeat moms are considered in relation with the number of moms that rapidly repeated a pregnancy in its neighboring edge or corner block groups. Cold spots are block groups with low number of rapid repeat moms and are surrounded by similar block groups with low number of rapid repeat moms.

The block group variables were used to examine the local neighborhood and spatial component of aggregated characteristics. This is because moms are not independent and, as spatial statistic models suggest an objects’ neighbors can influence individual characteristics or behavior. Statistically, individual or neighborhood factors can collectively or independently help to predict rapid repeat pregnancy for moms in Kalamazoo County, MI.

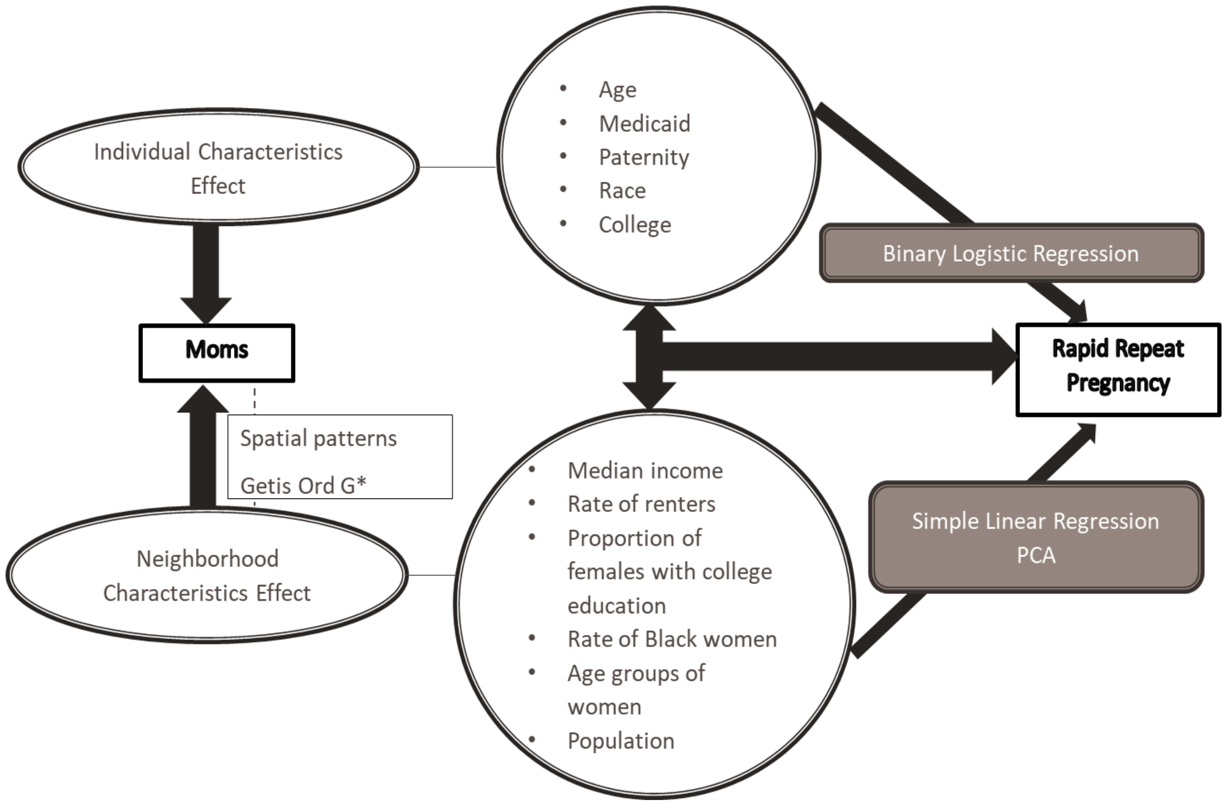

Figure 2. Individual and Neighborhood Factors Influencing Rapid Repeat Pregnancy in Kalamazoo County, MI
Source: Created by Author

For this study the effects of the individual level variables were measured using binary logistic regression. In order to be able to predict at the individual level characteristics of moms that related to the occurrence of rapid repeat pregnancy, the binary logistic regression model was adopted. Unlike a simple regression, which models a linear relationship between a variable of interest (dependent variable) and predictor variables, binary logistic regression estimates the probability of an occurrence and can be particularly useful when the dependent variable is dichotomous. For this analysis the dependent variable was a binary encoding of the occurrence of a mom with a rapid repeat pregnancy (1) and a mom who did not rapidly repeat pregnancy (included moms with multiple births) (0). The model thus estimated the probability of rapidly repeating or not rapidly repeating a pregnancy based on the variables under consideration. The logistic regression model independent variables were decided based on index birth characteristic of each individual mom. All moms in the sample were thus selected as a single case in SPSS. The independent variables included in the model were also all binary in nature and included: whether the mom was of color, ever attended college, ever listed paternity on a birth certificate, gave birth as a teen. Neighborhood level variables including rate of black females in the block groups and the proportion of women 20-24 per block group in Kalamazoo County were also included in the individual model.

At the block group neighborhood level, a linear relationship between the dependent variable (rapid repeat pregnancy/total number of repeating moms) was analyzed using a simple linear regression method. Data downloaded from the American Community Survey 2013 that included data on percent of females who were in college, median age of each block group, rate of black women, proportion of women between 15-19, 20-24 and 35-39 years, rate of renters, log of population and log of income were included in this analysis. To control for population density, block groups with repeaters greater than six were selected as the cases to be included in this analysis. However, a Principal Component Analysis (PCA) a variable reduction tool, was first used to analyze the variables before putting the variables in a linear regression model. The PCA generated orthogonal components to remove any issues of multicollinearity. These components were included in the model to find the best combination of explanatory variables. The simple linear regression was thus the last step of analysis that was used to develop a predictive model based on the components generated from the PCA as the independent variables and rapid repeat pregnancy/total number of repeating moms as the dependent variable. The results of the PCA were also mapped using choropleth mapping in ARCGIS to understand the neighborhood patterns of the generated components.

CHAPTER IV

ANALYSIS AND RESULTS

In this chapter, the results and analysis based on the objectives of the study are presented. This presentation is broken up into three phases which include descriptive statistics of the sample and the variables considered. The second phase is presentation of the spatial patterns of rapid repeat pregnancy and the aggregated variable characteristics of moms using hot spots. Finally, the models that aim to predict rapid repeat pregnancies both at the individual and neighborhood level are presented.

Descriptive Analysis of Sample Moms

As mentioned earlier, between 2008 and 2014 with 2010 as the baseline, a total of 2861 women gave birth (singleton births) in Kalamazoo County, MI and are included in the sample for this study. Amongst these women, some of them repeated births, others rapidly repeated births with a majority giving birth ones. Table 5 provides an extensive description of these women and their characteristics. From Table 5, there were 1398 (49%) moms that did not repeat a pregnancy meaning they only gave birth once in 2010; 646 (22%) moms slow repeated pregnancy with 817 (29%) moms rapidly repeating a pregnancy between 2008 to 2014. This shows that about 51% of moms had more than one pregnancy with a higher proportion of repeaters rapidly repeating a pregnancy.

Table 5 also gives an insight into the characterisitics of the moms. The data showed that RRP moms in comparison with other moms proportionaly, were more likely to be teens at least for one of the births; RRP moms were more likely to be women of color but less likely to have being in college at least for one of the births. Additionally, RRP moms were more likely to be on

Table 5. Demographic Characteristics of Individual Moms in Kalamazoo County Michigan, 2008 to 2014

Individual Characteristics		All Moms (N=2861)	Non-Repeat Moms (N=1398)	Slow Repeat Moms (N=646)	Rapid Repeat Moms (N=817)
		n (%)	n (%)	n (%)	n (%)
Age	Ever <18	333 (12)	119 (9)	77 (12)	137 (17)
	Never <18	2528 (88)	1279 (91)	569 (88)	680 (83)
Color	Ever non-white	586 (20)	257 (18)	121 (19)	208 (25)
	All others	2275 (80)	1141 (82)	525 (81)	609 (75)
College	Ever any college	1716 (60)	759 (54)	453 (70)	504 (62)
	All others	1145 (40)	639 (46)	193 (30)	313 (38)
Insurance	Ever Medicaid	1452 (51)	714 (51)	284 (44)	454 (56)
	Self-pay	1409 (49)	684 (49)	362 (56)	363 (44)
Dad Named	Ever dad named	2510 (88)	1156 (83)	599 (93)	755 (93)
	All others	351 (12)	242 (17)	47 (7)	62 (7)

Source: Created by Author

Mediciad but on the other hand more likely to have dads named on the birth certificate. This distribution of rapid repeat moms follows a similar trend for the total moms sampled. Aside the description of the moms, it is also important to know about the children that were born over the period and Table 6 provides information about the babies.

From Table 6, 48.9 percent of moms in the sample gave birth once in 2010 with 8 (0.3%) moms giving birth to 5 babies between the period of 2008 to 2014. In all a total of 4745 babies (singleton births) were born to the sample moms between 2008 to 2014 with a mean of 1.66 babies.

Table 6. Number of Babies Born to Moms

Number of Babies	Moms	Percent Moms	Total Number of Babies
1	1398	48.9	1398
2	1109	38.8	2218
3	295	10.3	885
4	51	1.8	204
5	8	0.3	40
Total	2861	100	4745

Source: Created by Author

Sample Spatial Distribution

Another key component of this analysis was to describe the spatial distribution of moms in Kalamazoo County by block groups. To do this, data on moms was mapped in 87 block groups in Kalamazoo County, MI area . Using the natural breaks classification (based on natural grouping inherent in the data), the distribution of moms was mapped as a choropleth map to show the number of moms per block group. Layers were also created for block groups with greater than 6 rapid repeat moms and less than 6 rapid repeat moms and also overlaid over the choropleth map to show the spatial distribution of rapid repeating moms. Figure 4 shows this distribution.

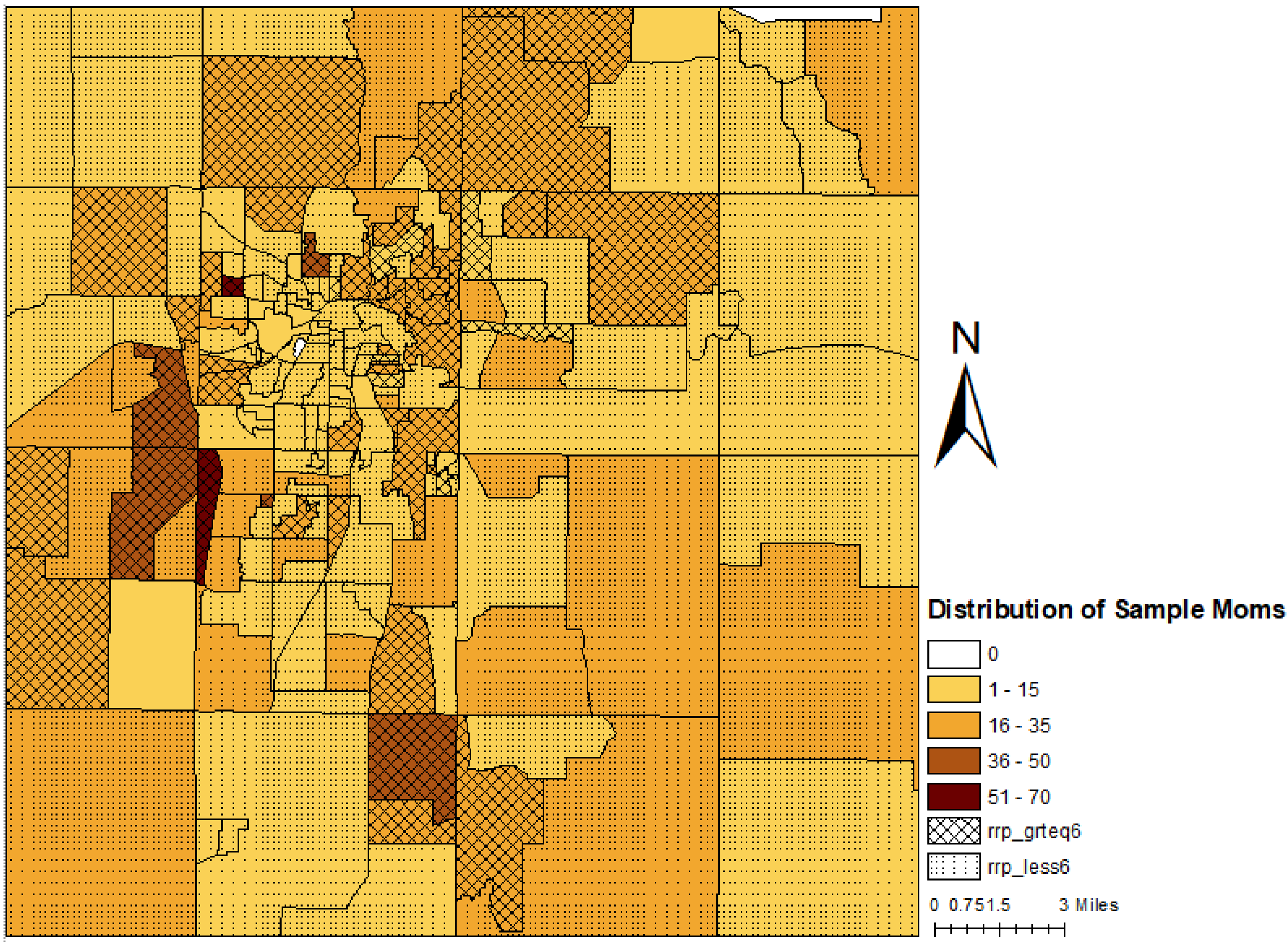

Figure 3. Distribution of Sample Moms by Block Group
Source: Created by Author

From figure 3, the map shows that there were 2 block groups in Kalamazoo County that did not have moms giving birth in Kalamazoo County in 2010; these can be found in the northeastern side of the county and one in the West of Kalamazoo township. Two block groups fall within the highest class (51-70) with a lot of block groups falling within the lowest block group (1-15). Block groups with more than 6 rapid repeat moms are predominantly found in the core of Kalamazoo township, parts of Comstock, Parchment, Cooper, Richlands and Oshtemo. This is synonymous with areas with high population with a sub-urban make up. In the Southern part of the county, block groups in Texas township, Portage and particularly Vicksburg also have moms greater than 6 rapidly repeating pregnancies. Other areas also had block groups of moms less than 6 rapidly repeating pregnancies; such areas predominantly include block groups in the townships of Ross, Charleston, Climax, Wakeshma, Schoolcraft, Prairie Ronde and Alamo. These areas also show a sharp contrast with areas where more rapid repeating moms live as they are areas with relatively low population and typically rural.

Geographic Patterns on Hotspot Maps

The first objective of this study is to understand spatial variation and patterns of rapid repeat moms in Kalamazoo County, MI. The hotspot method is used to find patterns in the living patters of repeating moms as well as patterns in the characteristics of moms that are rapidly repeating pregnancies in Kalamazoo County. The hotspot shows block groups with high values of the variables being mapped in relation to neighboring block groups. The alternate which are cold spots which are block groups with low values and neighboring values are also shown. Statistically significant hotspots and cold spots are shown at a 99% confidence interval.

The first analysis shown in Figure 4 was done to find the clustering pattern of all rapid repeaters as a proportion of all moms in each block group. The map shows the statistically significant highest percent of moms that are rapid repeaters by controlling for the distribution of all moms.

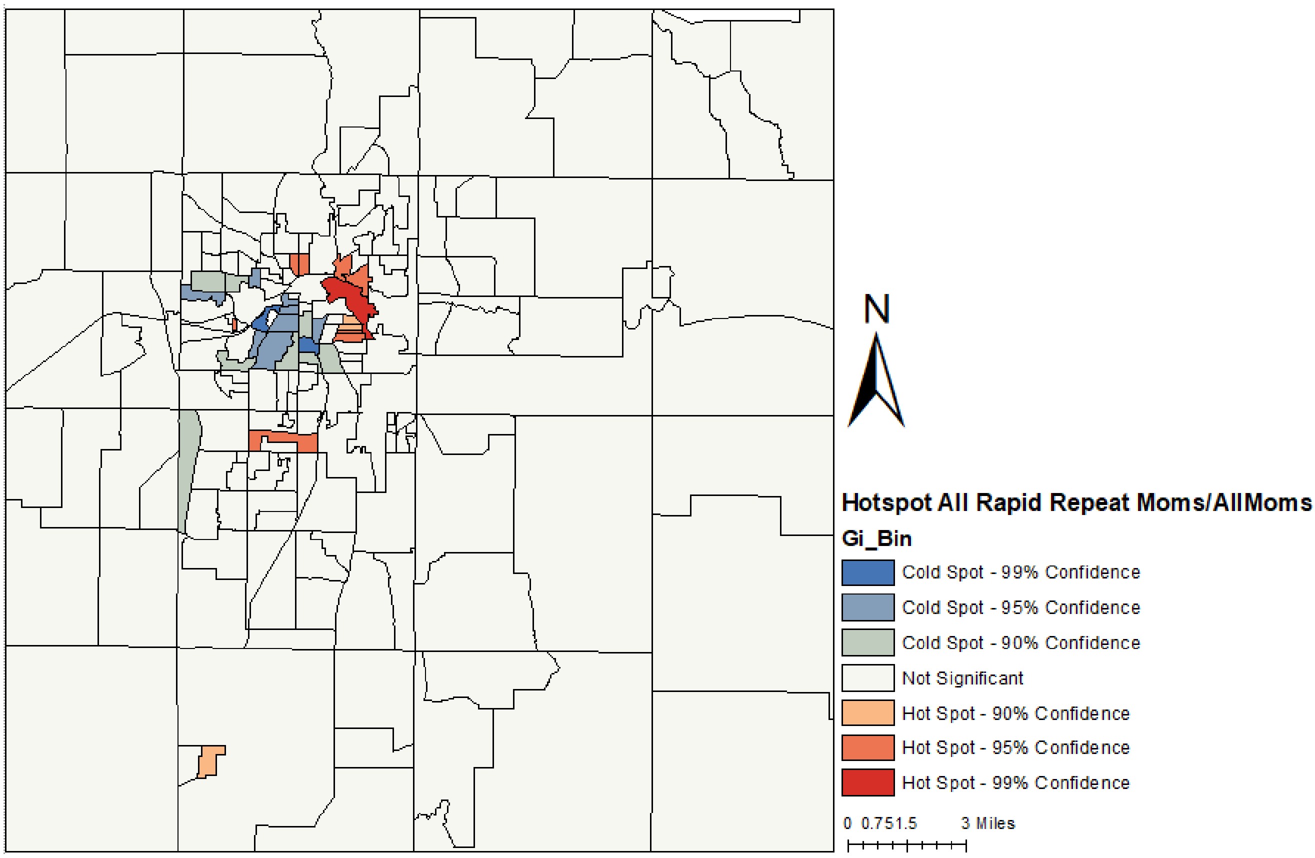

Figure 4. Hotspot of Rapid Repeat Moms in Kalamazoo County, MI
Source: Created by Author

The analysis reveals statistically significant hotspots in block groups clustered in the central eastern part of Kalamazoo Township. There are also some single hotspot outlier block groups each in the North of Portage and Schoolcraft. Contrastingly, there are statistically significant cold spots in block groups on the West of Kalamazoo township. A look at the hot and cold spot shows a sharp contrast in Kalamazoo township which requires further analysis to understand these patterns. Aside the block groups discussed it is also interesting to note that other areas in the county do not show any statistically significant hotspot or cold spots. Hence while rapid repeat moms live in these other block groups, it is statistically not significant as a hot or cold spot.

After knowing the hotspots of rapid repeat moms in the block groups, it is essential to break down this analysis by the various kinds of rapid repeat moms to understand the spatial variations in the characteristics of rapid repeat moms. Figure 5 shows hotspot areas of RRP women of color as a proportion of all moms that rapidly repeated pregnancies. Again, this map shows similar hot spot patterns of block groups in the eastern side of Kalamazoo Township compared with figure 5. Interestingly, the hot spots spread into further block groups in the north as compared to hotspots identified for all rapid repeaters as a proportion of all moms in each block group. There are few outlying hotspots with one in the border of Kalamazoo Township and Oshtemo Township.

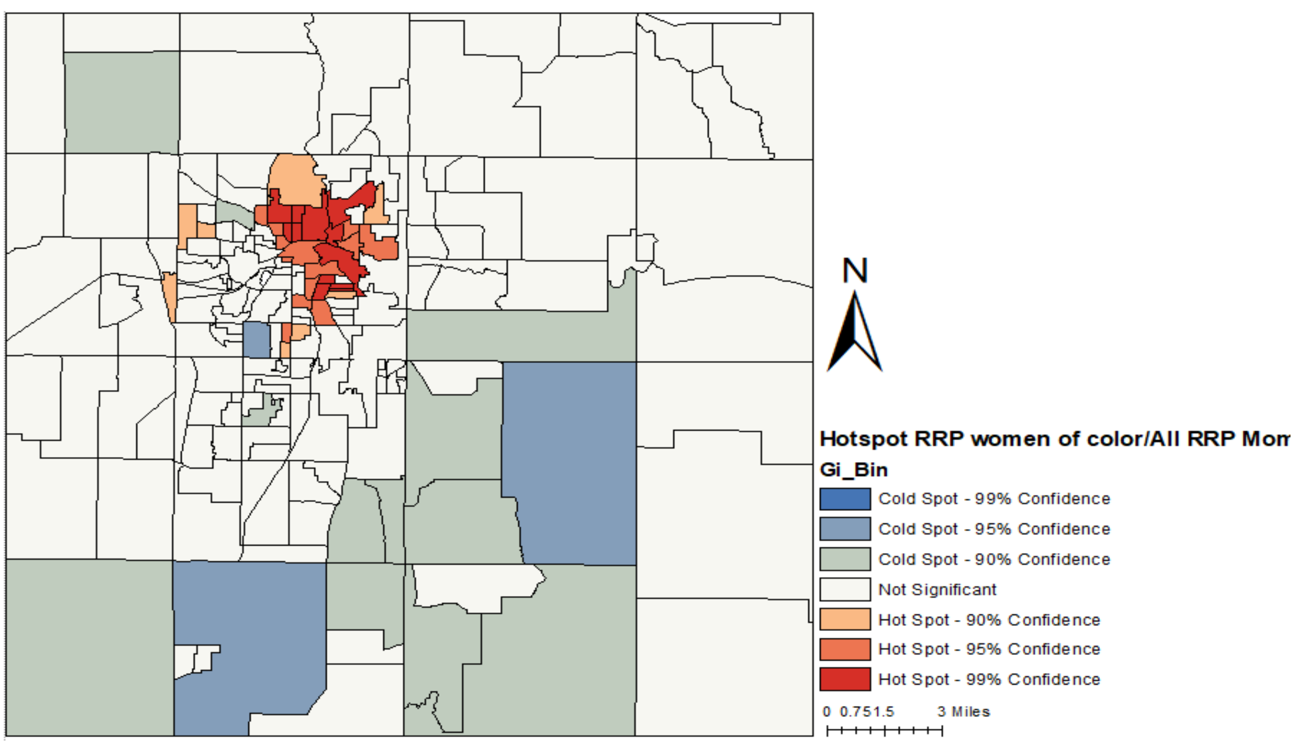

Figure 5. Hotspot of Rapid Repeat Moms of Color in Kalamazoo County, MI
Source: Created by Author

The second variable analyzed with the Getis Ord shown in Figure 6 is the sum of moms that rapidly repeated pregnancies and were ever on Medicaid as a proportion of sum of all rapid repeating moms in each block group.

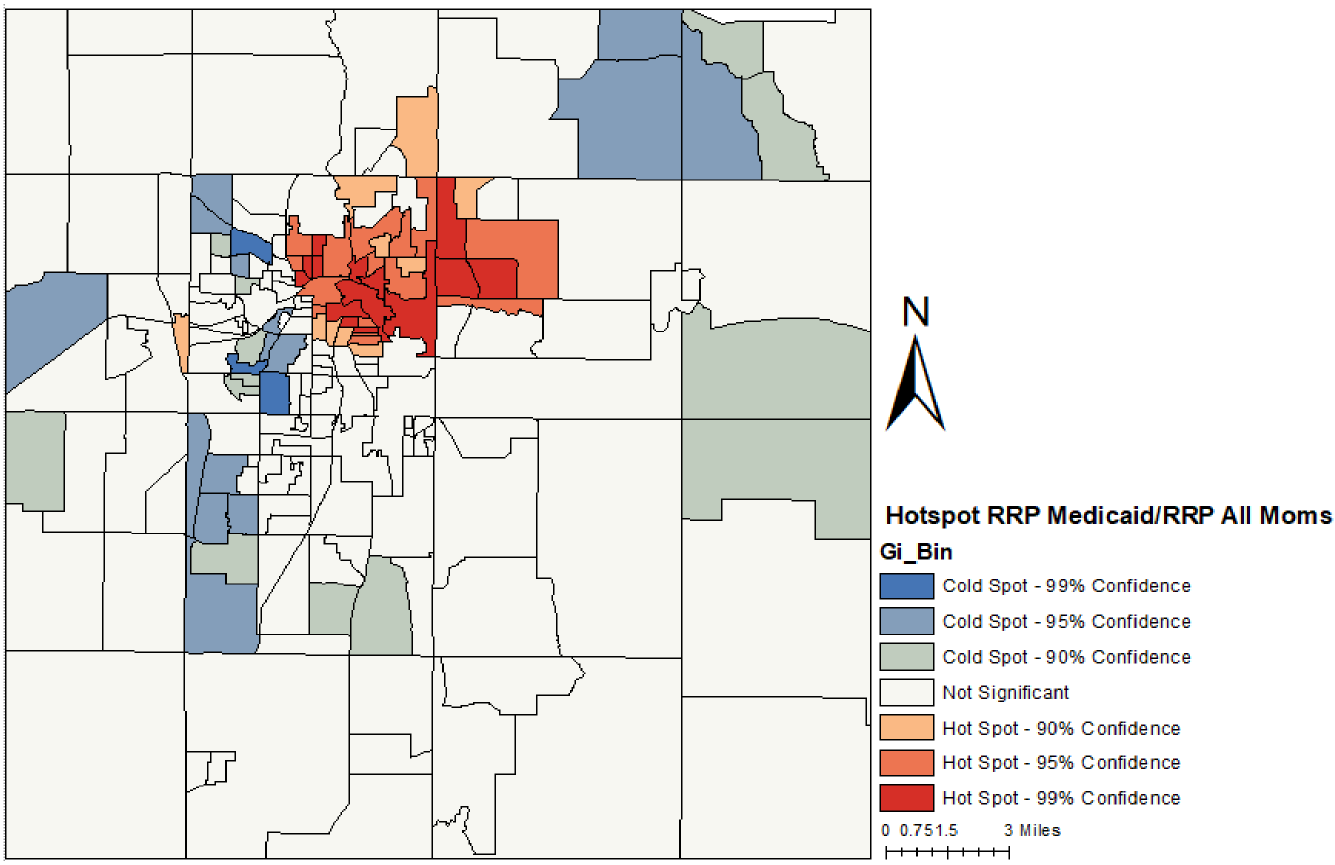

Figure 6. Hotspot of RRP Moms on Medicaid, Kalamazoo County, MI
Source: Created by Author

The core of the significant hotspot block group clustering can still be seen in the east side of Kalamazoo Township with growing number of block groups in this area showing a significant hotspot. Interestingly, the north-western side of Comstock that borders Kalamazoo Township on the east also shows significant block group clustering of women that rapidly repeated pregnancy on Medicaid. Additionally, the same single block group that showed a significant hotspot of RRP women of color in Oshtemo Township also showed significant hotspot of RRP moms on Medicaid.

The third variable under analysis in the hotspot analysis is the sum of RRP moms that ever-had college education within the study period as a proportion of all moms that rapidly repeated a pregnancy shown in figure 7.

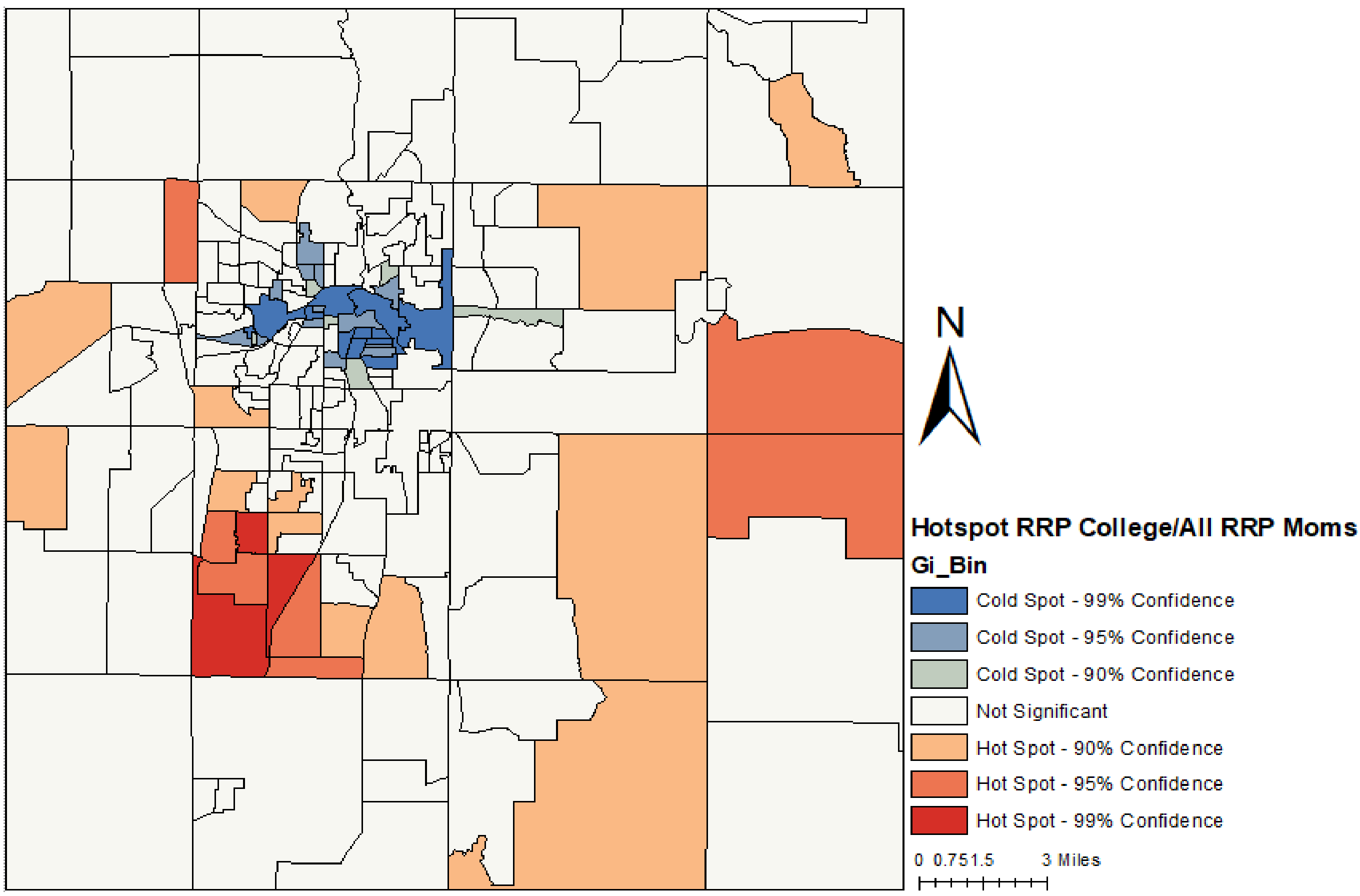

Figure 7. Hotspot of RRP Moms Having Any College in Kalamazoo County
Source: Created by Author

The pattern of hotspot changed from the usual as seen in the hotspot analysis of RRP moms, RRP moms of color and Medicaid. There are only two significant block groups of RRP college moms in the north west and the southwest of Kalamazoo township. Central Kalamazoo

Township particularly the east shows cold spots. Typically implying that rapid repeat moms are not women that have ever been to college. Portage township has the greatest number of block groups showing significant hotspot clusters. This is expected as the contrast to the RRP conundrum, as there are established women who are wealthy and have high level education who would want to give birth at rapid rate and either return to work or focus on a career. These hotspots show patterns in moms exhibiting different characteristics who rapidly repeat pregnancies. Other hotspot areas are block groups in east of Pavillion township, east of Brady township, northern climax, southern Charleston, northeast of Texas township, northeast Comstock, northeast and south east Oshtemo and single block group in Ross township.

Sum of RRP moms that had dad named on birth certificate as a proportion of sum of all RRP moms in each block group is next in the hotspot analysis. As can be seen in Figure 8 no block group showed significant hotspot clustering of block groups. Few block groups in the west of Kalamazoo township showed cold spots with the remaining block groups not being significantly clustered. Interestingly most of the block groups that showed cold spot of rapid repeat moms as a proportion of all moms also showed significant cold spots of dad named on birth certificate primarily because there are less moms in these block groups as shown in the distribution moms and hence less rapid repeaters and also few moms having their partner names on the birth certificate.

Lastly, the sum of RRP moms that ever rapidly repeated a pregnancy as a teen was analyzed as a proportion of sum of all moms that rapidly repeated a pregnancy in each block group. As seen in figure 9, the block groups in the core of the east side Kalamazoo again shows significant hotspots as earlier shown in the Medicaid and Color maps. Additional block groups in the west of Comstock, southeast of Parchment and southwest of Richland townships also showed block groups that had a significant hotspot of teen RRP moms.

These hotspot maps essentially present two spatial patterns based on the characteristics of the mom that is rapidly repeating a pregnancy. There are hotspots of rapid repeat moms in block groups in the eastern core of Kalamazoo Township. These moms are moms that are women of color, teens and use Medicaid. There is also a hotspot of rapid repeat moms who have ever been to college, are not likely to be teens and not on Medicaid who live in block groups in Portage and other periphery areas far away from the center of the County. These maps thus clearly show the distinction in living patterns of rapid repeat moms.

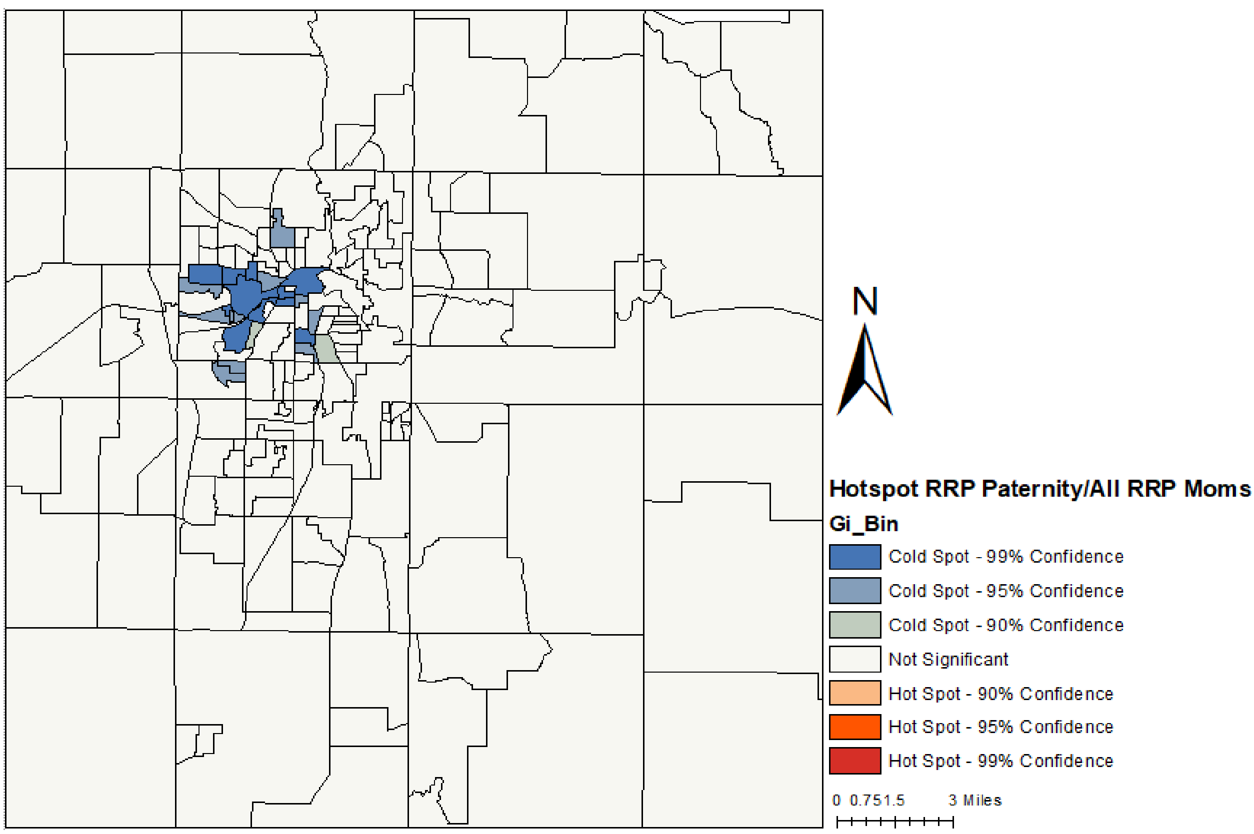

Figure 8. Hotspots of RRP Cases with Dad Name on Birth Certificate in Kalamazoo County
Source: Created by Author

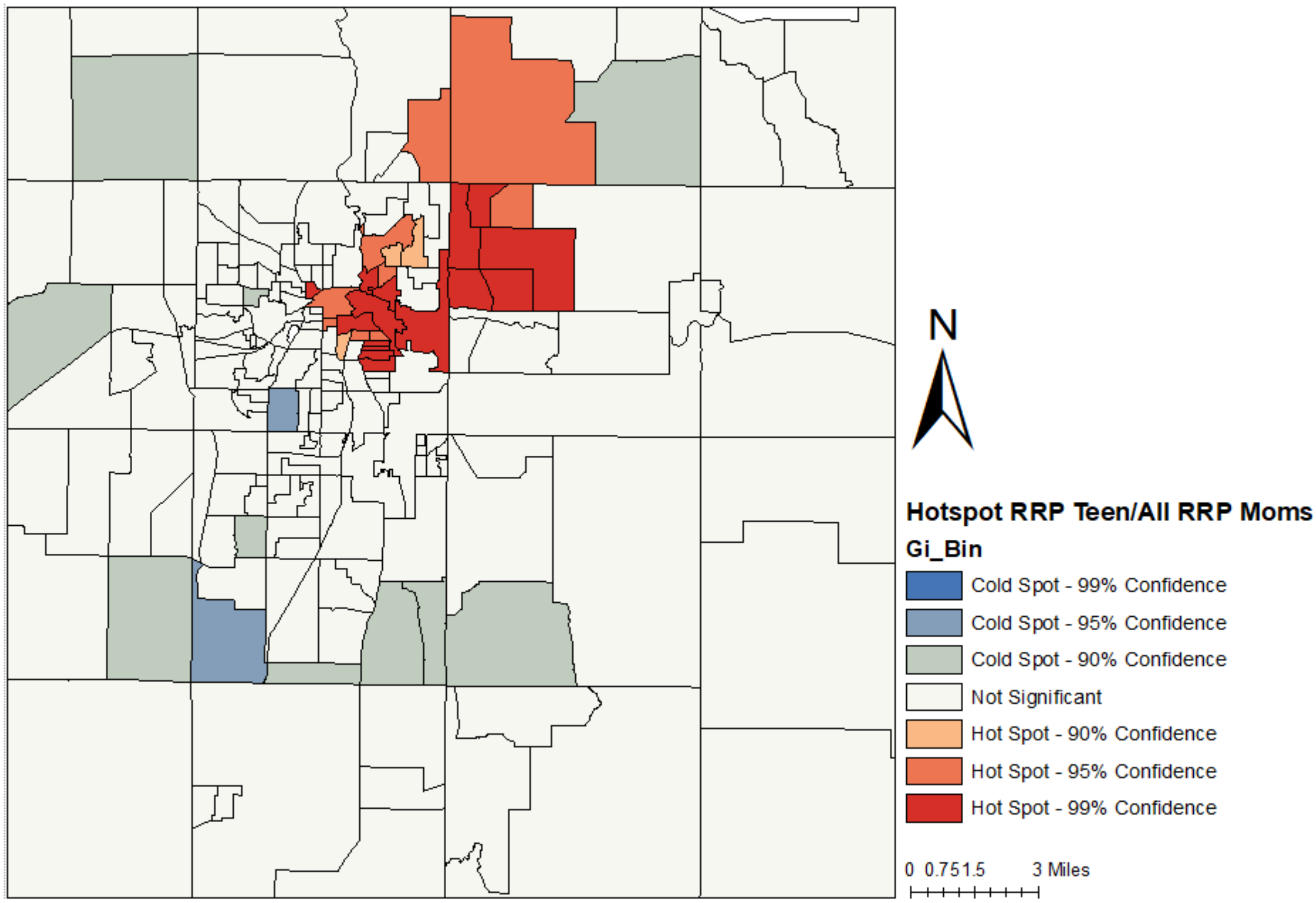

Figure 9. Hotspots of Teen RRP Moms in Kalamazoo County
Source: Created by Author

Individual Statistical Analysis

As indicated earlier, the binary logistic regression was used to understand the second objective of the study which is to explore the individual level factors that influence women to rapidly repeat pregnancies in Kalamazoo County, MI. The binary nature of all 5 independent variables (color, Medicaid, dad named on birth certificate, college education) and the dependent variable (rapid repeat pregnancy or no rapid repeat pregnancy) makes the use of the binary logistic regression appropriate. Table 7 shows the results of the binary logistic regression predicting rapid repeat.

Table 7. Binary Logistic Regression Results Predicting Rapid Repeat Versus Not Rapid Repeating Pregnancy

Variables	B	S.E.	Wald	Df	Sig.	Exp (B)	95% C.I for EXP(B)	
							Lower	Upper
Color	0.465	0.113	16	1	0.000	1.592***	1.276	1.987
Medicaid	0.286	0.100	8.173	1	0.004	1.331**	1.094	1.620
College	0.292	0.098	8.822	1	0.003	1.340**	1.105	1.625
Paternity	1.055	0.159	43.735	1	0.000	2.871***	2.100	3.925
Teen	0.674	0.135	24.720	1	0.000	1.961***	1.504	2.558
Constant	-2.373	0.184	167.068	1	0.000	0.093***		

* $p<0.05$; ** $p<0.01$; *** $p<0.001$

As shown in Table 7, all individual level variables were statistically significant in predicting rapid repeat pregnancy, meaning that controlling for all the variables each individual predictor variable is associated with chances of a woman rapidly repeating a pregnancy. The most significant values are color, paternity and teen. From table 8, women of color were 1.6 times more likely to rapidly repeat a pregnancy than women who were not of color ($p<0.001$, CI, 1.276, 1.987). Moms who gave birth to their first baby as teenagers were 1.9 more likely to rapidly repeat a pregnancy compared to non-teen moms ($p<0.001$, CI, 1.504, 2.558). Women who had the name of their baby’s father on the first child’s birth certificate were 2.9 times more likely to rapidly repeat a pregnancy ($p<0.001$, CI, 2.1, 3.925). Women who are on Medicaid and with College education are 1.3 and 1.3 times more likely to rapidly repeat a pregnancy than their compatriots who are not respectively ($p<.005$). To test how well the model explains variation in the dependent variable, Cox & Snell R square and Nagelkerke R square are generated in SPSS for binary logistics regression. Though these are pseudo R squares intended to perform the same function as the R squares in a linear regression, it is often expected to be low and many scholars warn that it is interpreted with caution. That notwithstanding, for this model, the explained variation ranges from 0.035 to 0.050 when Cox & Snell R square or Nagelkerke R square are referenced respectively as shown in Table 8.

Table 8. Model Summary for Logistic Regression

Log likelihood	Cox & Snell R Square	Nagelkerke R Square
3321.708[a]	.035	.050

The model thus explained 5% (Nagelkerke R square) of the variance in rapid repeat pregnancy and correctly classified 72.0% of cases.

To improve the model fit and to find out if there are other variables (continuous) that predict rapid repeat. Neighborhood level factors were added to the variables in the model. The variables included were median age, percent cohort ages of moms, median household income, percentage of women who have a college degree, percentage of married people and the individual ages of all moms. Rate of black women, population density. All these variables are downloaded from American Community Survey 3-year estimates, 2013. Initially, all variables were pooled together with the individual variables but were gradually dropped one at a time based on its significance level. Table 9 shows the results of the second model.

Table 9. Model 2, Pooled Binary Logistic Regression Results Predicting Rapid Repeat Versus Not Rapid Repeating Pregnancy with Individual and Neighborhood Level Variables

Variables	B	S.E.	Wald	df	Sig.	Exp (B)	95% C.I for EXP(B)	
							Lower	Upper
Color	0.402	0.122	10.796	1	0.001***	1.495	1.176	1.900
Medicaid	0.281	0.102	7.553	1	0.006**	1.324	1.084	1.618
College	0.313	0.099	9.931	1	0.002**	1.368	1.126	1.622
Paternity	1.071	0.161	44.433	1	0.000***	2.918	2.130	3.997
Teen	0.683	0.136	25.065	1	0.000***	1.979	1.515	2.586
Rate of Black women in Block Groups	0.005	0.002	5.270	1	0.022**	1.005	1.001	1.009
Proportion of females 20-24 years in block groups	-0.14	0.004	9.744	1	0.002**	0.986	0.978	0.995
Constant	-2.325	0.190	90.705	1	0.000***	.098		

* p<0.05; ** p<0.01; *** p<0.001

In the second model, again all individual level variables are significant. Two neighborhood variables maintained in the model included rate of Black females in mom block groups and proportion of females aged 20-24 years are significant at (p<0.05) respectively. It is interesting to note that this new model does not change the odd ratios (Exp B) by much for the individual level characteristics. Odds ratio of continuous data are interpreted slightly different from categorical variables. For continuous data, odds ratios that are greater than 1 indicate that the event is more likely to occur as the predictor increases. Odds ratios that are less than 1 indicate that the event is less likely to occur as the predictor increases. Hence from table 10 it can be ascertained that when rate of black women in the block groups of moms increases, moms associated with those block groups individually have a higher odd (1.005) of rapidly repeating a pregnancy at ($p<0.05$, CI, 1.001, 1.009). However, when moms living in block groups with higher proportions of 20-24 aged females are considered, there is less chance (0.986) that moms associated with these block groups will individually rapidly repeat a pregnancy ($p<0.05$, CI, 0.978, 0.995). These two variables highlight the importance of color and age of mom's neighborhoods on individual moms rapidly repeating a pregnancy. This second model of slightly improved the model R squared as this model explained variation ranges from 0.040 to 0.057 when Cox & Snell R square or Nagelkerke R square are referenced respectively as shown in Table 10. The model thus explained 5.7% (Nagelkerke R square) of the variance in rapid repeat pregnancy and correctly classified 71.4% of cases.

Table 10. Model 2 Summary for Logistic Regression

Log likelihood	Cox & Snell R Square	Nagelkerke R Square
3305.548[a]	.040	.057

Neighborhood Statistical Analysis

First, the dependent variable was tested to see if it met the assumption of normal distribution in a regression model by doing some descriptive statistics on the dependent variable. The results showed that the mean of the dependent variable (proportion of rapid repeat moms from repeat moms) was 55.59 The median of this data was also 55.7. Since the skewness value (-0.003) was less than three times the skewness standard error (0.237) of the dependent variable data, the data was found to be normally distributed as shown in the histogram in Figure 10.

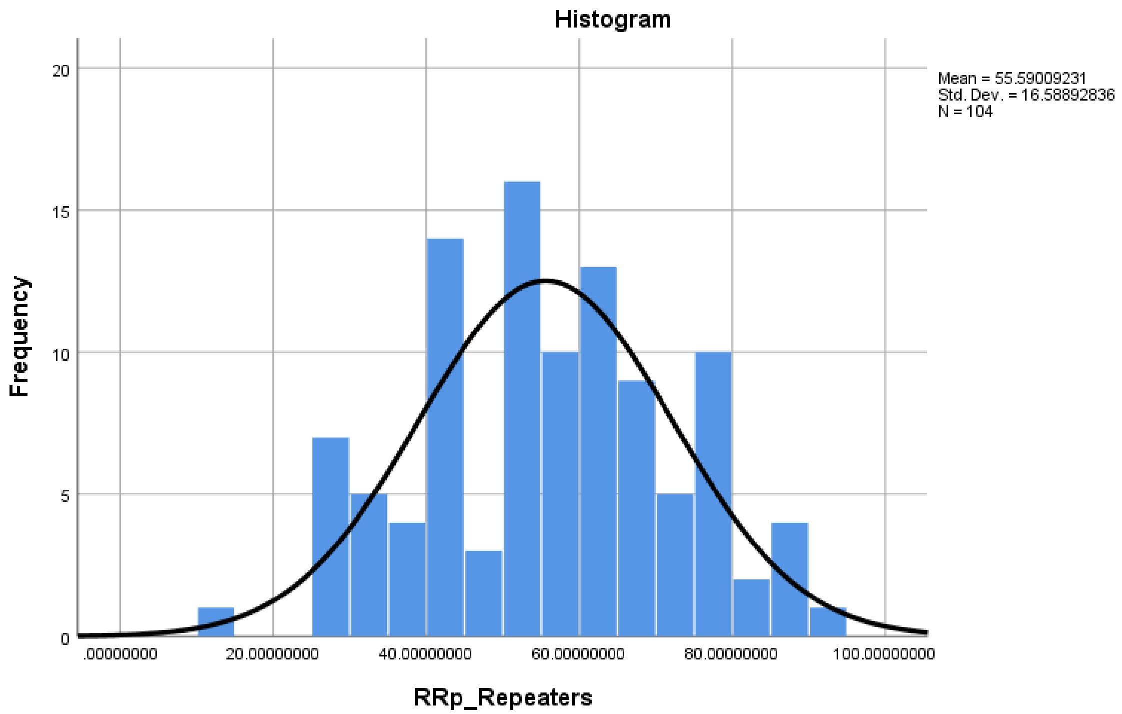

Figure 10. Normal Distribution of Dependent Variable
Source: Created by Author

The variables are first fitted in a factor analysis model to identify the principal component factors that can explain the variation in the dependent variable which is the proportion to rapid repeaters from repeating moms. Principal component analysis is a statistical tool useful in identifying important factors that are useful in explaining the variation in a dataset. In a large dataset, the PCA can identify the most important factors with no multicollinearity that are helpful in explaining an outcome variable. The results of the PCA using extraction of components with eigen values greater than one. After varimax rotation, four principal components (factors) extracted and shown in Table 12. These show the relative contribution of each variable to the components identified. A scree plot of the Eigen values against their principal components are also showed in Figure 11. With components with Eigen values greater than 1 representing good components, the scree plot clearly shows why 4 factors were selected as they had an eigen values greater than 1.

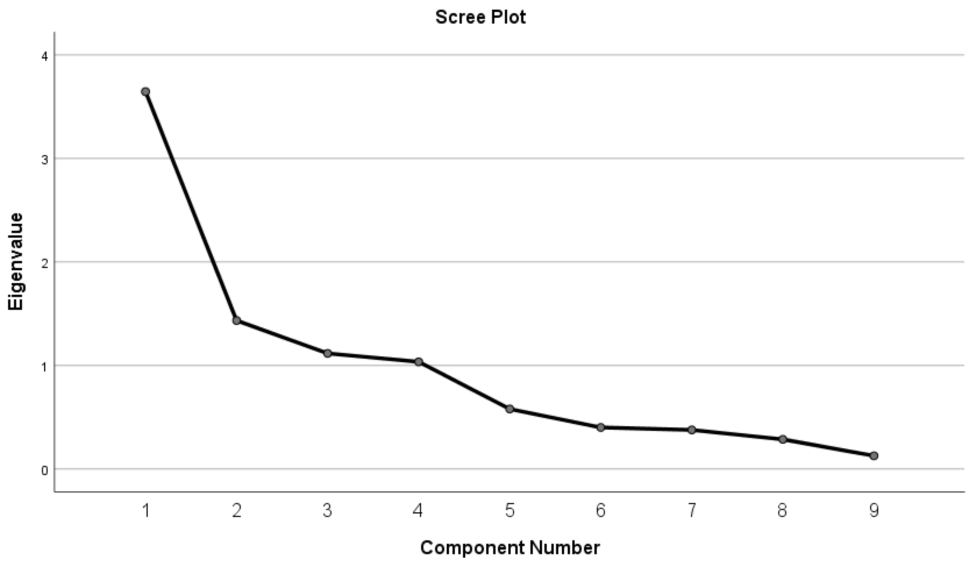

Figure 11. Scree Plot of Components
Source: Created by Author

In terms of the reliability of the PCA a Barlett's test was done and showed the suitability of the data to principal component analysis was highly significant (chi square = 391.932, P = 0.00). Overall, the four factors extracted contributed 80.33 percent of the total variability of the studied variables, with the first factor explaining 32.76%, second factor 19.37%, third factor 15.85% and the fourth factor 12.37% of the total variance. Components 1,2,3 and 4 had eigen values of 3.64, 1.43, 1.11 and 1.04 respectively.

Table 11. Variable Loadings of PCA

Variables 80.3% variance explained	Factor Loadings	Initial Eigenvalue Total	 % of Variance
Component 1: Young black women low socioeconomic status neighborhoods		**3.645**	**40.948**
Rate of black women	0.595		
Proportion of females 20-24 years	0.467		
Rate of renters	0.798		
Log of total population	0.846		
Component 2: Educated high income women neighborhoods		**1.434**	**15.933**
Proportion of females with higher education	0.941		
Log of median income	0.695		
Component 3: Mature women neighborhoods		**1.116**	**12.404**
Proportion of females 35-39 years	0.837		
Component 4: Teen women neighborhoods			
Proportion of females 15-19	0.950	**1.035**	**11.497**

Table 11 shows the loading of the variables on their components. Component 1 is characterized by positive loadings (correlations between the component and the variables) on proportion of females 20-24 (0.50), rate of black (0.59), rate of renters (0.79) and log of total population (0.84). These show one group of women that live in block groups with a lot of young (20-24), black women, living in rented housing and having big populations that can be predicted in terms of rapid repeat pregnancy. A look at the variables that loaded highly on the second component also shows factors that describe a different group of women that live in block groups that have a lot of college (0.94) women and possibly have a higher income (0.69) that can be estimated in terms of rapid repeat pregnancy. The third and fourth components all indicated one variable that highly positively loaded on them including Age 34-39 (0.837) and Age 15-19 (0.914) respectively. The PCA has thus been able to compile logically smaller groups of variables that can be used in predicting rapid repeat pregnancy. From the PCA, age seems to be very important for all the components, implying different age groups are likely to have different motivation to rapidly repeat a pregnancy. However, component 1 and 2 highlights the women neighborhoods exhibiting different characteristics and intentions for repeating a pregnancy. Component one highlights young black low socio-economic women that rapidly repeat pregnancies while component 2 corroborates an erudite population likely with high incomes and jobs that rapidly repeat to be able to focus on careers or other aspect of their life. Component 1 supports Ranieri and Wiemann (2007) that revealed that often young girls who are out of school or with low educational and socio-economic status are likely to rapidly repeat pregnancy. Component 2 also corroborates Haight (2018) finding that not only adolescents with low socioeconomic status are likely to rapidly repeat but there are older, socioeconomically sound women with careers that may want to rapidly repeat in their older years and get back to jobs. From these groups, the PCA makes it clear that there is not one set of distinct variables that can predict RRP, but an interaction of variables will showcase groups of moms in the sample that can rapidly repeat pregnancy. The two key components from the PCA are mapped as choropleths with a natural break classification to find out how these characteristics play out based on the factors generated for each block group shown in Figure 12 and 13.

From figure 12, there are patterns of the higher values of component 1 in Kalamazoo township, both in the East and West. The West block groups match up with block groups from the Getis Ord indicating that rapid repeat pregnancy these block groups have women with component 1 characteristics that are correctly predicted to rapidly repeat pregnancy. Generally, relatively young, low socioeconomic block groups that repeat pregnancy.

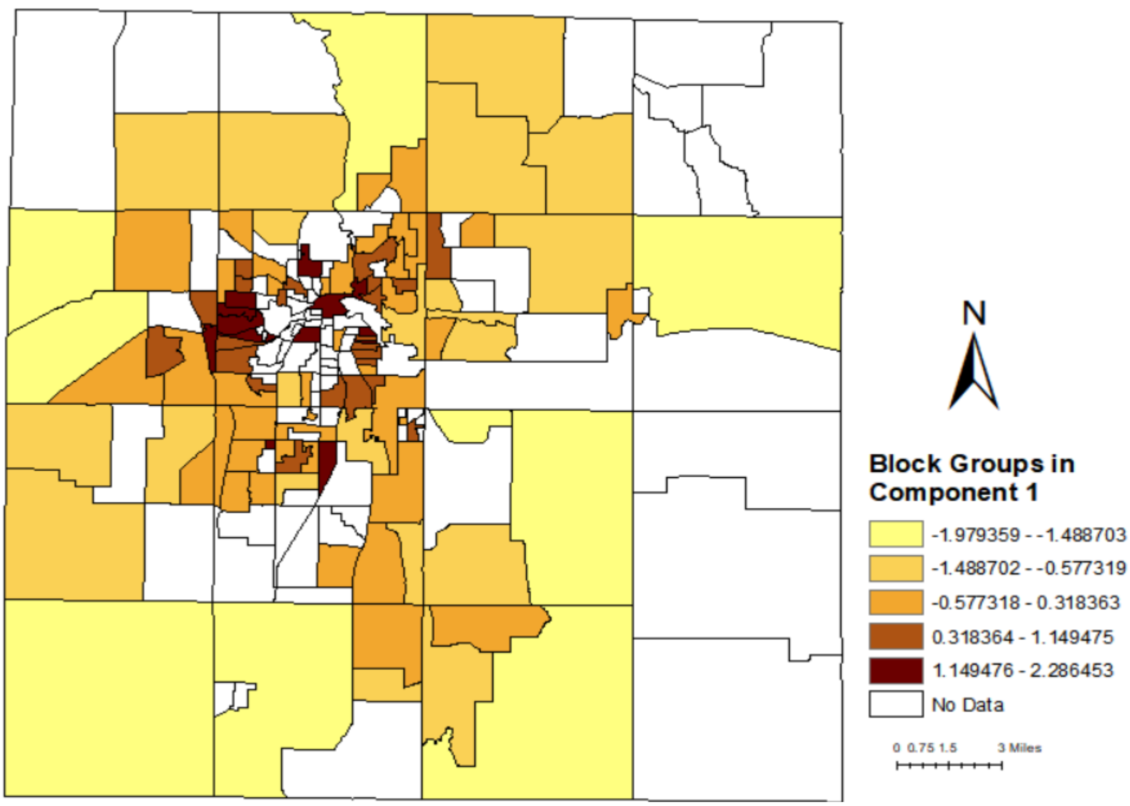

Figure 12. Block Group of Component 1 Characteristics
Source: Created by Author

Component 2 takes the pattern away from Kalamazoo township to block groups in areas that have a relatively higher socioeconomic status such as Portage as shown in figure 13.

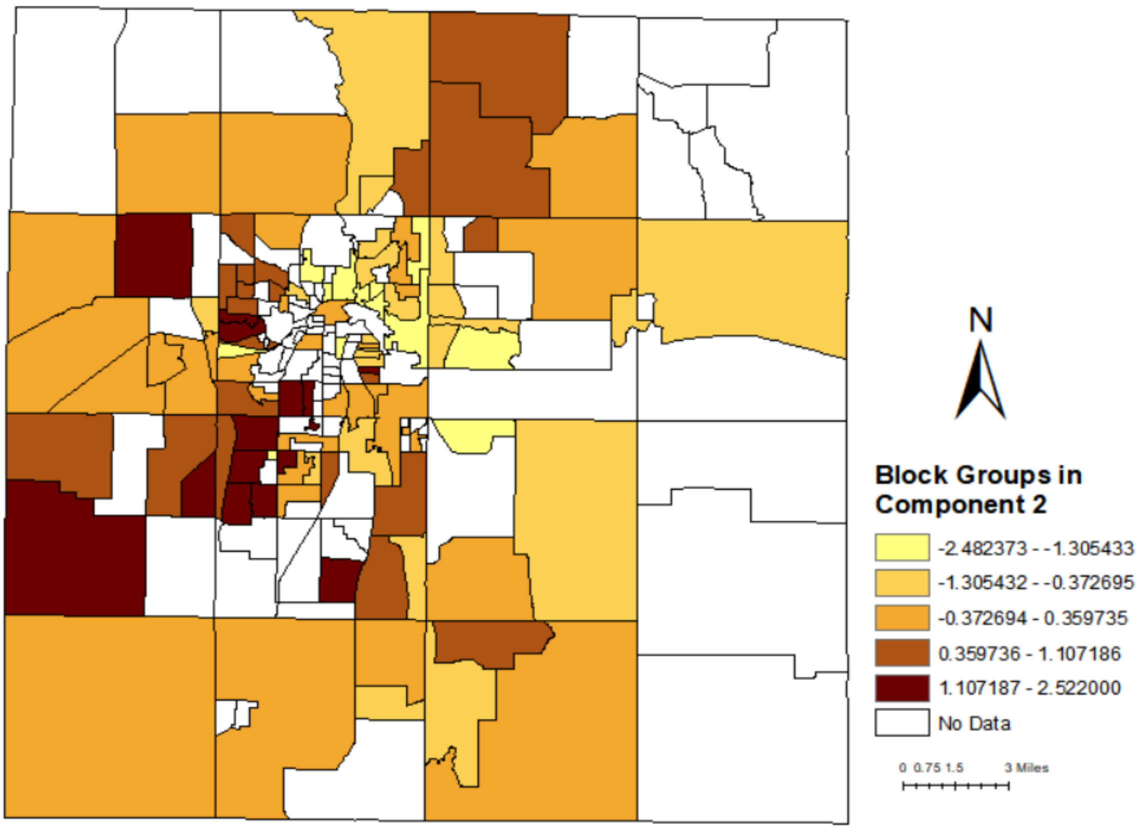

Figure 13. Block Group of Component 2 Characteristics
Source: Created by Author

This results again clearly corroborates the results of the Getis Ord where a characteristic of hotspot of rapid repeat moms with college were found outside of Kalamazoo townships but in relatively higher socioeconomic status in Portage and other suburban areas.

Simple Linear Regression

The components developed from the PCA were used as independent variables in a linear regression to explain the variability in the dependent variable (rapid repeat moms/repeat moms). Again, cases of block groups with greater than 6 repeaters are selected for the prediction to control for population density. Summary of the regression model is presented in Table 12. From the table, the result of the simple linear regression model explains 16% of the variability in the dependent variable (proportion of moms that rapidly repeat pregnancy from number of repeating moms). Variance in the model was checked using the F statistic that assumes a null hypothesis that, the model has no predictive power or the coefficients in the independent variables are equal to zero. However, the results from this test shows that the F statistic is significant thus rejecting the Null hypothesis. The model is thus statistically significant in predicting the dependent variable at α=0.000 shown in Table 13.

Table 12. Summary of Simple Linear Regression

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.442[a]	.196	.163	15.17568863409

Table 13. Anova Test of Model Predictive Power

Model		Sum of square	Df	Mean Square	F	Sig
1	Regression	5544.981	4	1386.245	6.019	.000[b]
	Residual	22799.851	99	230.302		
	Total	28344.832	103			

The model revealed that only component 1(Rate of black women in block group, Proportion of females 20-24 years, Rate of Renters by Block group and Log Population) and component 2 (Proportion of females with college education, Log median income) are statistically significant at predicting rapid repeat pregnancy at (β=.188, $p<0.005$) and (β= -.380, $p<0.001$) respectively. This shows that component two is more statistically significant in predicting rapid repeat pregnancy than component 1 as shown in Table 14.

Table 14. Estimated Coefficients at 5% Level of Significance of Independent Variables in the Model

	Unstandardized Coefficients		Standardized Coefficients	T	Sig.
	B	Std. Error	Beta		
Constant	55.590	1.488		37.356	.000
Component 1	3.122	1.495	.188	2.088	.039*
Component 2	-6.296	1.495	-.380	-4.211	.000***
Component 3	2.091	1.495	.126	1.399	.165
Component 4	.268	1.495	.016	.179	.858

* $p<0.05$; ** $p<0.01$; *** $p<0.001$

Neighborhoods with moms that have college education and higher income are more related to rapid repeat pregnancy than moms in block groups with characteristics of component 1. The value of the standardized coefficient beta showed that component 1 had a higher predictive power than component 2.

CHAPTER V

DISCUSSION, LIMITATIONS AND FUTURE RESEARCH

Discussion

This chapter focuses on discussion of the results and linking results to previous studies to either confirm or reject earlier assertions. The chapter also discusses the limitations of this research and highlights areas for future investigation. This study provides a different dimension to rapid repeat pregnancy studies in the United States than most available literature. While most research on the issue of rapid repeat pregnancy has been conducted at the state and national level (Apparedy et al., 2017; White et al., 2015), this study considers individuals and block groups in a single moderately populated county in southwestern Michigan to examine patterns at that scale.

In terms of results, 28% of sampled moms who gave birth in 2010 had a rapid repeat pregnancy between 2008 to 2014. This represents about a third of sampled moms; this proportion is consistent with a study by Lindberg (2013) using the National Survey of Family Growth data and a study by Copen (2015) based on 36 states in the United States. Additionally, the data from this study reveals that RRP moms in comparison with other moms are more likely to be teens, women of color and on Medicaid (Table 5). This finding corroborates assertions by scholars such as Zhang 2019; Delara 2018 and Tocce et al., 2012; all of whom found that the rate of rapid repeat pregnancy is particularly high among minorities (women of color/black) and young women between the ages of 15 and 24. Indeed 41% of the total teen moms in the sample rapidly repeated pregnancies which is higher than the 20% national figure (20%) in the United States as reported by Boardman (2006). These findings are not new in health inequalities studies where health risks are purported to be associated with young women with low socioeconomic status.

The data also revealed that RRP moms in comparison with other moms are more likely to have the dad of their child named on birth certificates, a phenomenon that has received less attention in the RRP literature review but gives credence to an assertion by Baldwin (2013) that, for women that live with their partners or are still involved with their partners, the birth intervals are likely to be influenced by their partners.

With regards to the first objective of this study which was to understand spatial variations in rapid repeat pregnancies in Kalamazoo County, hotspot maps were used to understand patterns in terms of the total number of RRP moms in each block group as a proportion of all moms. It is immediately evident from figure 4 that the east side Kalamazoo represents the core hotspot of RRP moms. The urban core of Kalamazoo township particularly the eastside is also an identified hotspot for maternal health problems in other studies. For instance, Kothari et al (2016) also found significant hotspots of low birth weight, poverty and minority population in the same neighborhoods. Additionally, these block groups in the east side of Kalamazoo are hotspots for rapid repeat moms on Medicaid (Figure. 6), women of color (Figure. 5) and, teen rapid repeat moms. These patterns are also observed further north to the North Western parts of Comstock and parts of Richland township (Figure 9). Contrastingly, Figure 7 shows that rapid repeat moms with college education are not clustered in the urban core of Kalamazoo township but in block groups in Portage, North east Comstock and in block groups in rural Charleston, Pavilion, Oshtemo and Brady townships. From the hotspot maps, there are clear patterns of young women of color with low socio-economic status moms clustered in east Kalamazoo township with an elite group of moms in Portage and the peripheries who have had at least one rapid repeat pregnancy within the study period.

Hart (2005) mentioned that rural population with less education and lower socioeconomic status are more reluctant to seek medical care including family planning. Results from this study however do not show significant difference between urban and rural areas. Rural and urban women with different characteristics are rapidly repeating pregnancies as a result of the factor of intention. The two groups of rapid repeat population exhibit different characteristics making it difficult to interpret the patterns based solely on socio economic characteristics. Patterns identified from the hotspot maps provokes thought on the issues of risk and choice. Though RRP studies identify the risk factors of RRP, the patterns from these maps do not indicate RRP as a health risk but a health issue that result from different choices.

Boardman (2006) and Haight (2018) found that RRP can be intended or unintended; intensions are particularly based on choice. For women who have college education and high income, RRP is more of a decision and not a risk. The outcome of RRP can be a risk for them but RRP is not a health risk for women intentionally deciding to rapidly repeat pregnancies. Additionally, for young women with low socioeconomic status, a qualitative study by Aslam et al. (2017) found that while the index pregnancy maybe unintended, subsequent rapid repeat pregnancies are usually intentional base on family goal for family size and birth intervals. In this light, the results from this study conform with studies that argue that intensions and decisions are important factors in rapid repeating pregnancies.

The second objective of the study was to explore individual socio-economic factors that influence rapid repeat pregnancy in Kalamazoo county. All individual variables including age, color, Medicaid, paternity and college were statistically significant in predicting rapid repeat pregnancies (Table 10). The most significant variables are color, paternity and teen. In fact, moms with other parent's named on listed on child's birth certificate and teen moms were 2.9

and 1.9 times more likely to rapid repeat respectively. This again is consistent with findings from Baldwin (2013) who found teens of lower low income to be at a higher risk of rapidly repeating a pregnancy. Also, Collier (2009) reveals that moms who give birth to their first baby in as teenagers are a high-risk group for rapid repeat pregnancies as they are sexually active, engage in more risky lifestyles and may not use contraceptives. Having a college degree was also significantly associated with rapid repeat which makes sense when contrasted with the finding of Bennett et al., (2006) who found that unintended rapid repeat pregnancy is associated with low economic status. However, referring to the hotspots, RRP moms with college education are not young moms and of low economic status and as a result are probably intending their frequency and intervals in birth and not rapidly repeating unintentionally. In my opinion because these group of women are part of the sample; college is associated with such women. Interestingly, moms whose partners were listed on child's birth certificate were more likely to rapid repeat pregnancy. This affirms Cha et al., (2016) claim that RRP is strongly influenced by paternal pregnancy intentions. If dads are named on birth certificates, then their influence on the timing of the pregnancy cannot be ignored. Aside, these individual level factors that influence rapid repeat pregnancies, the proportion of black women population and the proportion of females 20-24 years in each block group were also found to be statistically significant in relation the level of rapid repeat pregnancies. This is no surprise particularly with young people as there is a greater likelihood of peer influence when it comes to pregnancy.

The results from this study also gives credence to discussions of risk, intension and choice in the rapid repeat literature. This is because all variables included in the logistic regression model were significant. This again emphasizes the diversity in the rapid repeat population and the role of choice and intention. Intended or unintended, different characteristics of women are associated with RRP. Women from varying backgrounds are making choices on birth patterns and intervals.

The last objective of this study was to explore block group level factors that influence rapid repeating pregnancy in Kalamazoo county. Here again a model was developed based on block group level data from the American Community Surveys (ACS). Variables included educational status, age, proportion of black women in each block group, total population and income level. The principal component analysis provided a logical break down of the key variables that are related to the level of rapid repeat at the block group level. Based on factor loadings from the variables, 4 key components were derived, and the first component had the most variable loading. This component included variables on proportion of black women in the block group, proportion of females between ages 20 and 24, proportion of renters and the log of block group population (Table 12). The second component had the second highest loadings included the proportion of females with college education by block group and the log of median household income (Table 12). These two components have a striking resemblance to the spatial patterns of RRP discussed from the Getis-Ord statistic earlier. Again, block groups dominated by black women, young population and low income or high renter's rate formed a component that correlates with rapid repeat. Another component that correlates with RRP at the block group level just like in the spatial patterns' maps were areas with high educated female population with potentially high-income level. Another interesting finding from the PCA is the two choropleth maps that again buttress the Getis-Ord maps. Factors loading for each block group were mapped using choropleth map with a natural break classification. It clearly shows two distinct patterns: high factor loadings for component 1 are found in the eastside of Kalamazoo township (Figure 13) while high factor loadings of component two are found in parts of Portage and other periphery block groups but outside of eastside Kalamazoo township. Though scholars have identified adolescence and low socioeconomic status as factors that influence RRP, factors like whether moms have ever had abortion, their contraceptive use or substance abuse are also considered critical factors in determining the level of rapid repeat pregnancies. This study however predicts RRP without considering any of the risk condition considered by other studies. It is thus not surprising that the explanatory variables in the model when fitted in a linear regression only explained 16% of the variability in the dependent variable suggesting that these variables are associated with rapid repeat pregnancy but have a low predictive power in terms of predicting a linear relationship.

Limitations

This study is limited in some forms and this is discussed in this section. First, the regression models used in this study recorded very low R squares but highly significant values which is a major problem that has been discussed in studies by Lo et al., (2014). Often in research, the focus is on both significance and predictive power of the model in explaining the variation in an outcome variable of interest. However, when all variables are significant it explains the dependent variable in terms of association and not necessarily correlation. For a health problem like rapid repeat pregnancy, it requires numerous studies over an area to understand the dynamics and obtain variables that possess predictive power in explaining the outcome. RRP is a multi-faceted problem that requires deeper exploration in order to be able understand variables that together have a linear relationship with rapid repeat pregnancy particularly in an area like Kalamazoo County.

Another limitation of this study is the scale of analysis. While most studies in the literature were conducted at national level, this study utilizes data at the block group and individual level. Applying the methods utilized in this study at a larger scale may produce different results. In Geography, the issue of the Modifiable Area Unit Problem is always mentioned in health research when individual data must be aggregated to a larger spatial scale to ensure privacy. Aggregation of data can lead to loss of information from the data that was present at the individual level. Additionally, there are underlying uncertainties and inaccuracies in scale when data is modelled in ArcGIS.

Further on the issue of data, this research is based on secondary data from the Homer Stryker School of Medicine. There were gaps in the data with respect to addresses; while some addresses were missing others had typos. Furthermore, all data were coded as binary which limits the array of analysis that can be done. More continuous data are required at the individual level to support the development of future multilevel regression models that looks at interactions of variables at both individual and neighborhood levels. This will help understand the problem as a whole and help understand factors that influence the different group of RRP moms. A review of literature also suggests that the data did not include variables that would have had more predictive power in the model such as income levels of moms and contraceptive usage.

Finally, it is clear from the analysis that intensions are a major component of rapid repeat pregnancy analysis. The available quantitative is unable to capture intensions in the model used. Including qualitative data in this research would have enabled us understand intension more succinctly. For instance, a study by Aslam et al., (2017) provides another perspective to RRP studies by looking at the problem from the perspective of mothers through qualitative data. The emic perspective of moms will give better understand of the issue of intension and choice.

The World Health Organization's (WHO) definition of rapid repeat pregnancy include mons with twin births. In this study such moms are excluded from the sample. This could have influenced the finding from this study. We however believe that the proportion of moms who had more than one birth at the time was very low.

Future Studies

The study provided a narrow focus on rapid repeat moms in Kalamazoo County; however future research would benefit from understanding the different interpregnancy intervals associated with rapid repeat, that is moms that repeat pregnancies in less than 18 months and between 18 and 24 months. A more detailed analysis at the individual level could also increase understanding the results from the Principal Component Analysis. We assume that the factors that influence high socioeconomic status moms to rapidly repeat differ from those factors influencing young low socioeconomic status women.

Additionally, future studies will benefit from testing the models derived from this study across different scales of analysis to measure the changes or consistency in results. Testing the model at state and national levels as well as other localized levels such as census tract and county level can determine the robustness of the models developed. Also, another dimension to this issue will be a focus on rural and urban difference in RRP using the same data. Lastly, future studies can focus on linking RRP to known health risks such as morbidity and mortality. Such studies will be able to assess the issue of RRP as a risk. It is difficult to intervene in RRP as a health problem if the issue is not tied to health risks. RRP is an action that has health implications. Understanding the risk of moms' actions can inform interventions and the acceptance of those interventions.

Another interesting study will be to check the contraceptive use status for these moms. Contraceptives have been known to increase the interval between births and reduce fertility rates. It would be interesting to know the contraceptive usage of moms and to find out if there is an association between RRP and contraceptives and how failure rates and lack of use is influencing patterns of RRP in Kalamazoo county.

REFERENCES

Abedin, T., Chowdhury, Z., Afzal, A., R., Yeasmin, F., & Turin C. (2016) Application of Binary Logistic Regression in Clinical Research. *JNFHB.* 2, 8-11.

Adler, N., & Rehkopf, D. (2008). U.S. disparities in health: Descriptions, causes, and mechanisms. *Annual Review of Public Health*, 29, 235-52. s DOI: 10.1146/annurev.publhealth.29.020907.090852

Appareddy, S., Pryor, J., & Bailey, B. (2017) Inter-pregnancy interval and adverse outcomes: Evidence for an additional risk in health disparate populations, *Journal of Maternal-Fetal & Neonatal Medicine*, 30:21, 2640-2644, DOI: 10.1080/14767058.2016.1260115

Aslam, R. W., Hendry, M., Booth, A., Carter, B., Charles, J. M., Craine, N., … Whitaker, R. (2017). Intervention Now to Eliminate Repeat Unintended Pregnancy in Teenagers (INTERUPT): a systematic review of intervention effectiveness and cost-effectiveness, and qualitative and realist synthesis of implementation factors and user engagement. *BMC Medicine*, *15*(1), 1-13. https://doi.org/10.1186/s12916-017-0904-7

Baldwin, M. K., & Edelman, A. B. (2013). The Effect of Long-Acting Reversible Contraception on Rapid Repeat Pregnancy in Adolescents: A Review. *Journal of Adolescent Health*, 52(4), 47–53. https://doi.org/10.1016/j.jadohealth.2012.10.278

Bell, S., Wilson, K., Bassinette, L., & Shah, T. (2013). Access to Primary Health Care: Does Neighborhood of Residence Matter*? Annals of the Association of American Geographers*, 103(1), 85-105. DOI:10.1016/j.ajog.2005.10.193.

Bennett, Culhane, Mccollum, & Elo, T. (2006). Unintended rapid repeat pregnancy and low education status: Any role for depression and contraceptive use? *American Journal of Obstetrics and Gynecology*, 194(3), 749-754.

Bloome, D. (2014). Racial inequality trends and the intergenerational persistence of income and family structure. *American Sociological Review*, 79(6), 1196. DOI: 10.1177/0003122414554947

Boardman, L. A., Allsworth, J., Phipps, M. G., & Lapane, K. L. (2006). Risk Factors for Unintended Versus Intended Rapid Repeat Pregnancies among Adolescents. *Journal of Adolescent Health*, 39(4), 597-605. https://doi.org/10.1016/j.jadohealth.2006.03.017

Bourgi, K., Brar, I., & Baker-Genaw, K. (2016). Health disparities in hepatitis C screening and linkage to care at an integrated health system in southeast Michigan. *PloS one*, *11*(8).

Brown, H. K., Ray, J. G., Liu, N., Lunsky, Y., & Vigod, S. N. (2018). Rapid repeat pregnancy among women with intellectual and developmental disabilities: a population-based cohort study. *Canadian Medical Association Journal*, 190(32), 949–956. https://doi.org/10.1503/cmaj.170932

Campbell, D., & Campbell, S. (2008, October). Introduction to regression and data analysis. *In StatLab Workshop Series*. 1-14

Cao, Y., Stewart, K., & Kalil, R. (2016). Geographic patterns of end-stage renal disease and kidney transplants in the Midwestern United States. *Applied geography*, 71, 133-143.

Center for Disease Control (2017) At A Glance 2016 Maternal Health Advancing the Health of Mothers in the 21st Century. https://www.cdc.gov/chronicdisease/resources/publications/aag/pdf/2016/aag-maternal-health.pdf

Cha, S., Chapman, D. A., Wan, W., Burton, C. W., & Masho, S. W. (2016). Discordant pregnancy intentions in couples and rapid repeat pregnancy. *American Journal of Obstetrics and Gynecology*, 214 (4), 494-506. https://doi.org/10.1016/j.ajog.2015.10.149

Citizens Research Council of Michigan (2018) Exploring Michigan's Rural/Urban Divide. http://www.crcmich.org/

Collier, J. (2009). The rising proportion of repeat teenage pregnancies in young women presenting for termination of pregnancy from 1991 to 2007. *Contraception*, 79(5), 393-396. doi:10.1016/j.contraception.2008.11.014

Conde-Agudelo, A., Rosas-Bermudez, A., Castaño, F., & Norton, M. (2012). Effects of Birth Spacing on Maternal, Perinatal, Infant, and Child Health: A Systematic Review of Causal Mechanisms. *Studies in Family Planning*, 43(2), 93-114. https://doi.org/10.1111/j.1728-4465.2012.00308.x

Conroy, K. N., Engelhart, T. G., Martins, Y., Huntington, N. L., Snyder, A. F., Coletti, K. D., & Cox, J. E. (2016). The Enigma of Rapid Repeat Pregnancy: A Qualitative Study of Teen Mothers. *Journal of Pediatric and Adolescent Gynecology*, 29(3), 312–317. https://doi.org/10.1016/j.jpag.2015.12.003

Cook B.L., McGuire T.G., & Zuvekas S.H. Measuring trends in racial/ethnic health care disparities. Medical Care Research and Review. 2009; *66*(1):23.

Copen, C. E., Thoma, M. E., & Kirmeyer, S. (2015). Interpregnancy intervals in the United States: data from the birth certificate and the national survey of family growth. *National vital statistics reports: from the Centers for Disease Control and Prevention, National Center for Health Statistics, National Vital Statistics System*, 64(3), 1-10.

Corbin, C., Janz, K., & Baptista, F. (2017). Good Health: The Power of Power. *Journal of Physical Education, Recreation & Dance*, 88(9), 28-35. https://doi.org/10.1080/07303084.2017.1367742

Cozier, Y. (2017). Invited Commentary: The Enduring Role of "Place" in Health—A Historic Perspective. *American Journal of Epidemiology*, 185(11), 1203-1205. https://doi.org/10.1093/aje/kwx085

Cresswell, T. (2013) Geographic Thought. ISBN 978-1-4051-6939-4

Crittenden, C. P., Boris, N. W., Rice, J. C., Taylor, C. A., & Olds, D. L. (2009). The Role of Mental Health Factors, Behavioral Factors, and Past Experiences in the Prediction of Rapid Repeat Pregnancy in Adolescence. *J Adolesc Health*. 44(1): 25–32. doi:10.1016/j.jadohealth.2008.06.003

De Bocanegra, H. T., Chang, R., Menz, M., Howell, M., & Darney, P. (2013). Postpartum contraception in publicly funded programs and interpregnancy intervals. *Obstetrics & Gynecology*, 122(2 PART 1), 296-303. doi: 10.1097/AOG.0b013e3182991db6

Damle, L. F., Gohari, A. C., McEvoy, A. K., Desale, S. Y., & Gomez-Lobo, V. (2015). Early Initiation of Postpartum Contraception: Does It Decrease Rapid Repeat Pregnancy in Adolescents? *Journal of Pediatric and Adolescent Gynecology*, 28(1), 57–62. https://doi.org/10.1016/j.jpag.2014.04.005

Delara, R., Madden, E., & Bryant A. S., (2018) Short interpregnancy intervals and associated risk of preterm birth in Asians and Pacific Islanders, *The Journal of Maternal-Fetal & Neonatal Medicine*, 31:14, 1894-1899, DOI: 10.1080/14767058.2017.1331431

Diez Roux, A., and Mair, C. (2010). Neighborhoods and health. *Annals of the New York Academy of Sciences*, 11861(1), 125-145.

Douthit, N., Kiv, S., Dwolatzky, T., & Biswas, S. (2015). Exposing some important barriers to health care access in the rural USA. *Public Health*, 129 (6), 611-620. https://doi.org/10.1016/j.puhe.2015.04.001

Dummer, T. J. B. (2008). Health geography: supporting public health policy and planning. *Canadian Medical Association Journal,* 178 (9), 1177-1180. doi:10.1503/cmaj.071783

Earnshaw, V., Rosenthal, L., Gilstad-Hayden, K., Carroll-Scott, A., Kershaw, T., Santilli, A., & Ickovics, J. (2018). Intersectional experiences of discrimination in a low-resource urban community: An exploratory latent class analysis. *Journal of Community & Applied Social Psychology*, 28(2), 80-93. doi-org.libproxy.library.wmich.edu/10.1002/casp.2342

El-Sayed, A. M., Finkton Jr, D. W., Paczkowski, M., Keyes, K. M., & Galea, S. (2015). Socioeconomic position, health behaviors, and racial disparities in cause-specific infant mortality in Michigan, USA. *Preventive medicine*, *76*, 8-13.

Fraser, A., Nelson, S. M., Macdonald-Wallis, C., Cherry, L., Butler, E., Sattar, N., & Lawlor, D. A. (2012). Associations of Pregnancy Complications with Calculated Cardiovascular Disease Risk and Cardiovascular Risk Factors in Middle Age: The Avon Longitudinal Study of Parents and Children. *Circulation*, 125(11), 1367–1380. https://doi.org/10.1161/CIRCULATIONAHA.111.044784

Fradelos, E., Papathanasiou, I., Mitsi, D., Tsaras, K., Kleisiaris, C., & Kourkouta, L. (2014). Health Based Geographic Information Systems (GIS) and their Applications. *Acta Informatica Medica*. 22(6), 402-405. DOI: 10.5455/aim.2014.22.402-405

Friesen, C. E., Seliske, P., & Papadopoulos, A. (2016). Using principal component analysis to identify priority Neighbourhoods for health services delivery by ranking socioeconomic status. *Online journal of public health informatics*, 8(2). DOI: 10.5210/ojphi.v8i2.6733

Gemmill, A., & Lindberg, D.L., (2013) Short Interpregnancy Intervals in the United States. *Obstet Gynecol*. 122(1): 64–71. DOI:10.1097/AOG.0b013e3182955e58.

Getis, A., & Ord, K. (1992) The Analysis of Spatial Association by Use of Distance Statistics. *Geographical Analysis* 24:189-206.

Graham H. (2004) Social determinants and their unequal distribution: Clarifying policy understandings. *Milbank Q*. 82(1):101–124. DOI: 10.1111/j.0887-378X.2004. 00303.x

Haley, D. F., Matthews, S. A., Cooper, H. L., Haardörfer, R., Adimora, A. A., Wingood, G. M., & Kramer, M. R. (2016). Confidentiality considerations for use of social-spatial data on the social determinants of health: Sexual and reproductive health case study. *Social Science & Medicine*, *166*, 49-56.

Hart, L. G., Larson, E. H., & Lishner D. M. (2005). Rural definitions for health policy and research. *American Journal of Public Health*, 95(11):49–55. doi: 10.2105/AJPH.2004.042432

Haight, S. C., Hogue, S. J., Raskind-Hood, C. L., & Ahrens, K. A., (2018) Short interpregnancy intervals and adverse pregnancy outcomes by maternal age in the United States. *Annals of Epidemiology*. *31*, 38-44. https://doi.org/10.1016/j.annepidem.2018.12.002.

Hutcheon J. A., Nelson H. D., Stidd, R, Moskosky S, & Ahrens K.A., (2018) Short interpregnancy intervals and adverse maternal outcomes in high resource settings: an updated systematic review. *Paediatr Perinat Epidemiol*. 33: O48–59. DOI:10.1111/ppe.12518

Isquick, S., Chang, R., Thiel de Bocanegra, H., Chabot, M., & Brindis, C. D. (2017). Postpartum Contraception and Interpregnancy Intervals Among Adolescent Mothers Accessing Public Services in California. *Maternal and Child Health Journal*, 21(4), 752–759. https://doi.org/10.1007/s10995-016-2164-0

Krieger, N (2014) Discrimination & Health Inequities. *International Journal of Health Services*, 44 (4), 643–710. DOI: http://dx.doi.org/10.2190/HS.44.4.b

Langston, A. M., Joslin-Roher, S. L., & Westhoff, C. L. (2014). Immediate post abortion access to IUDs, implants, and DMPA reduces repeat pregnancy within 1 year in a New York City practice. *Contraception*, 89(2), 103–108. https://doi.org/10.1016/j.contraception.2013.10.014

Lenardson, J. D., Ziller, E. C., Coburn, A. F., & Anderson, N. J. (2009). Profile of rural health insurance coverage: A chartbook. Portland, ME: University of Southern Maine, Muskie School of Public Service, Maine Rural Health Research Center. Retrieved from muskie.usm.maine.edu/Publications/rural/Rural-Health-Insurance-Chartbook-2009.pdf

Lever, J., Krzywinski, M., & Altman, N. (2017). Points of significance: Principal component analysis. *Nature of Methods*. 14(7)

Leventhal, T., and Brooks-Gunn, J. (2000). The neighborhoods they live in: The effects of neighborhood residence on child and adolescent outcomes. *Psychological Bulletin*, 126(2), 309-337. http://dx.doi.org/10.1037/0033-2909.126.2.309

Liang, K. Y., & Zeger, S. L. (1993). Regression analysis for correlated data. *Annual review of public health*, 14(1), 43-68.

Loree, A. M., Gariepy, A., Ruger, J. P., & Yonkers, K. A. (2018). Postpartum Contraceptive use and Rapid Repeat Pregnancy Among Women who use Substances. *Substance Use & Misuse*, 53(1), 162–169. https://doi.org/10.1080/10826084.2017.1327976

Lu, N., Samuels, M., Kletke, P., & Whitler, E. (2010). Rural-Urban Differences in Health Insurance Coverage and Patterns Among Working-Age Adults in Kentucky. *Journal of Rural Health*, 26(2), 129-138. DOI: 10.1111/j.1748-0361.2010.00274.x

Lynch, J., & Perera, I. (2017). Framing Health Equity: US Health Disparities in Comparative Perspective. *Journal of Health Politics, Policy, and Law*, 42(5), 803-839.

Macquillan, E., Curtis, A., Baker, K., & Paul, R. (2018). Geospatial Analysis of Birth Records to Target Programming for Mothers with Gestational Diabetes Mellitus in Michigan, 2013. *Public Health Reports*, 134(1), 27-35. DOI: 10.1177/0033354918815183

Maravilla, J. C., Betts, K. S., Couto e Cruz, C., & Alati, R. (2017). Factors influencing repeated teenage pregnancy: a review and meta-analysis. *American Journal of Obstetrics and Gynecology*, 217(5), 527-545.e31. https://doi.org/10.1016/j.ajog.2017.04.021

Mestad, R., Secura, G., Allsworth, J. E., Madden, T., Zhao, Q., & Peipert, J. F. (2011). Acceptance of long-acting reversible contraceptive methods by adolescent participants in the Contraceptive CHOICE Project. *Contraception*, 84(5), 493–498. https://doi.org/10.1016/j.contraception.2011.03.001

Meden, T., John-Larkin, C. S., Hermes, D., & Sommerschield, S. (2002). Relationship between travel distance and utilization of breast cancer treatment in rural northern Michigan. *Jama*, *287*(1), 111-111.

MacQuillan, E., Curtis, A., Baker, K., & Paul, R. (2019). Geospatial Analysis of Birth Records to Target Programming for Mothers with Gestational Diabetes Mellitus in Michigan, 2013. *Public Health Reports*, 134(1), 27-35.

Musa, G., Chiang, P., Sylk, T., Bavley, R., Keating, W., Lakew, B., Hui-Chen, T., and Hoven, C. (2013). Use of GIS Mapping as a Public Health Tool–From Cholera to Cancer. *Health Services Insights*, 6, 111-1116. DOI:10.4137/HSI.S10471

Michigan Department of Health & Human Services (2018). 1990-2018 Geocoded Michigan Birth Certificate Registries. Retrieved: https://www.mdch.state.mi.us/pha/osr/chi/fullscreen.asp?MyTarget=https%3A//www.mdch.state.mi.us/pha/osr/chi/births14/Trends/BirthTrends.asp%3FDatasetCopy%3D1%26Dataset%3D3%26ActiveChar_1%3D208%26ActiveCol_1%3DKS%26TrendType%3D1%26Age_1%3DSTANDARD2%26DxId%3D0%26ActivePNC%3DKessner%26ActiveAge%3DSTANDARD2%26ActiveCG%3DLowRisk%26ActiveChar%3DIPI%26Column%3DCG%26AreaCode%3D39%26AreaType%3DC%26CI%3DN%26Stat%3DR%26Average%3DN

Mitchel, A. (2005). The ESRI Guide to GIS analysis, Volume 2: Spatial measurements and statistics. ESRI Guide to GIS analysis.

National Academies of Sciences, Engineering, and Medicine. (2017). *Communities in action: Pathways to health equity*. National Academies Press.

Norton, M., Chandra-Mouli, V., & Lane, C. (2017). Interventions for Preventing Unintended, Rapid Repeat Pregnancy Among Adolescents: A Review of the Evidence and Lessons from High-Quality Evaluations. *Global Health: Science and Practice*, 5(4), 547–570. https://doi.org/10.9745/GHSP-D-17-00131

Omrani-Khoo, H., Lotfi, F., Safari, H., Zargar Balaye Jame, S., Moghri, J., & Shafii, M. (2013). Equity in Distribution of Health Care Resources; Assessment of Need and Access, Using Three Practical Indicators. *Iranian Journal of Public Health*, 42(11), 1299-308

Park, H. (2013). An introduction to logistic regression: from basic concepts to interpretation with attention to nursing domain. *Journal of Korean Academy of Nursing*, *43*(2), 154-164.

Pfitzner, M. A., Hoff, C., & McElligott, K. (2003). Predictors of Repeat Pregnancy in a Program for Pregnant Teens. *Journal of Pediatric and Adolescent Gynecology*, 16(2), 77–81. https://doi.org/10.1016/S1083-3188(03)00011-1

Raneri, L. G., & Wiemann, C. M. (2007). Social Ecological Predictors of Repeat Adolescent Pregnancy. *Perspectives on Sexual and Reproductive Health*, 39(1), 39–47. https://doi.org/10.1363/3903907

Reese, B. M., & Halpern, C. T. (2017). Attachment to conventional institutions and adolescent rapid repeat pregnancy: A longitudinal national study among adolescents in the United States. *Maternal and child health journal*, 21(1), 58-67. DOI 10.1007/s10995-016-2093-y

Regan, A. K., Ball, S. J., Warren, J. L., Malacova, E., Marston, C., Nassar, N., ... & Pereira, G. (2019). Reply to "Sibling Comparison Design in Birth-Spacing Studies". *American journal of epidemiology*, 188(1), 22-23. https://doi.org/10.1093/aje/kwy187

Rossi P. (1972) Community social indicators. In: Campbell A, Converse P, eds. The Human Meaning of Social Change. New York, NY: Russell Sage Foundation; 87–126.

Ricketts, T.C. (2000). The changing nature of rural health care. Annual Review of Public Health. 21:639–57.

Riley, W. (2012) Health Disparities: Gaps in Access, Quality, and Affordability of Medical Care . VOL. 123

Satel, S., & Klick, J. (2006). The Health Disparities Myth: Diagnosing the Treatment Gap. *Washington, DC: American Enterprise Institute*.

Shlens, J. (2014). A tutorial on principal component analysis. arXiv preprint arXiv:1404.1100.

Sina, B. J. (2017). Pregnancy and the global disease burden. *Reproductive Health,* 14(3), 60-64 https://doi.org/10.1186/s12978-017-0420-4

Stopka, T. J., Goulart, M. A., Meyers, D. J., Hutcheson, M., Barton, K., Onofrey, S., ... & Chui, K. K. (2017). Identifying and characterizing hepatitis C virus hotspots in Massachusetts: a spatial epidemiological approach. *BMC infectious diseases*, 17(1), 294.

Thoma, M. E., Copen, C., & Kirmeyer, S. (2016). Short Interpregnancy Intervals in 2014: Differences by Maternal Demographic Characteristics. *NCHS Data Brief*, (240), 1-8.

Thomas, C. J., Torrone, A. E., & Browning, R. C. (2009). Neighborhood Factors Affecting Rates of Sexually Transmitted Diseases in Chicago. *Journal of Urban Health*, 1(87), 102-112. DOI: 10.1007/s11524-009-9410-3

Tipirneni, R., Rhodes, K. V., Hayward, R. A., Lichtenstein, R. L., Choi, H., Arntson, E. K., ... & Davis, M. M. (2018). Geographic variation in Medicaid acceptance across Michigan primary care practices in the era of the Affordable Care Act. *Medical Care Research and Review*, *75*(5), 633-650.

Tocce, K. M., Sheeder, J. L., & Teal, S. B. (2012). Rapid repeat pregnancy in adolescents: do immediate postpartum contraceptive implants make a difference? *American Journal of Obstetrics and Gynecology*, 206 (6), 481-487. https://doi.org/10.1016/j.ajog.2012.04.015

Wang, F., Guo, D., & McLafferty, S. (2012). Constructing geographic areas for cancer data analysis: a case study on late-stage breast cancer risk in Illinois. *Applied Geography*, *35*(1-2), 1-11.

White, K., Teal, S. B., & Potter, J. E. (2015). Contraception after delivery and short interpregnancy intervals among women in the United States. *Obstetrics and gynecology*, 125(6), 1471. DOI: 10.1097/AOG.0000000000000841

Williams, F., Jeanetta, S., Obrien, D., Fresen, J. (2015). Rural-urban difference in female breast cancer diagnosis in Missouri. *Rural and Remote Health*, 15 (3).

World Health Organization (WHO). 2006. "Report of a WHO technical consultation on birth spacing. Geneva, Switzerland 13–15 June2005." Geneva.

World Health Organization (2016) Disease Burden and Mortality Estimates. www.who.int/healthinfo/global_burden_disease/estimates/en/index1.html

Zandbergen, P. A. (2014). Ensuring confidentiality of geocoded health data: Assessing geographic masking strategies for individual-level data. *Advances in medicine*, *2014*.

Zhang, Z., & Castelló, A. (2017). Principal components analysis in clinical studies. *Annals of translational medicine*, 5(17).

Zhang, Y., Quist, A., & Enquobahrie., D., (2019) Short birth-to-pregnancy intervals among African-born black women in Washington State, *The Journal of Maternal-Fetal & Neonatal Medicine*, 32:6, 947-953, DOI: 10.1080/14767058.2017.1395850

APPENDIX A

HSIRB LETTER

WESTERN MICHIGAN UNIVERSITY
Homer Stryker M.D.
SCHOOL OF MEDICINE

APPROVAL OF RESEARCH - Correction Letter

March 10, 2020

Catherine Kothari, Ph.D.
Western Michigan University Homer Stryker M.D. School of Medicine
Department of Epidemiology and Biostats
1000 Oakland Drive
Kalamazoo, MI 49008

TYPE OF REVIEW: Modifications, Non-Committee Review

IRB#: Wmed-2017-0179 *(please reference this number in all correspondence with the IRB)*

PROTOCOL TITLE: Using Public Health Data to Examine Infant Mortality in Kalamazoo, Michigan (FIMR study)

Dear Dr. Kothari:

This letter serves as a correction to the previous approval letter dated 02/24/2020 which had the incorrect spelling of the name name for Dennis Donkor. The request for modification of approved human research and associated materials were reviewed and approved on 01/30/2020.

As a reminder, IRB approval for this research expires on 09/10/2020.

The IRB reviewed the following documents related to the approval of the modification:

- Modification of Approved research form submitted 01/23/2020 requesting the addition of *Dennis Donkor as key research personnel.
- Agreement for Unpaid Research Associate signed 12/20/2019.

* Dennis Donkor, a Western Michigan University student, is conducting this research in his capacity as a research intern at WMed.

If you have any questions, please contact the WMed IRB office at 269-337-4345 or email irb@med.wmich.edu.

Sincerely,

Kelly M. Quesnelle

Kelly M. Quesnelle, PhD
IRB Chair
Western Michigan University Homer Stryker M.D. School of Medicine
1000 Oakland Drive
Kalamazoo, MI 49008-8012

Completion Date	23-Oct-2018
Expiration Date	22-Oct-2021
Record ID	29085351

This is to certify that:

DENNIS DONKOR

Has completed the following CITI Program course:

Human Research	(Curriculum Group)
Group 1 Social & Behavioral Sciences Researchers.	(Course Learner Group)
2 - Refresher Course	(Stage)

Under requirements set by:

Western Michigan University

CITI
Collaborative Institutional Training Initiative

Verify at www.citiprogram.org/verify/?wb25ba7a0-f0d5-431b-a546-252cee7b287e-29085351

APPENDIX B

FIELD DEFINITIONS

Field	Meaning
GEOID	Block Group
Count_Geoid_Data	Number of Moms Total per BG
Sum_Max_repeat	Number of Moms who repeat, per BG
Average_Max_Number	The average number of children per mom per BG
Sum_Max_number	Total number of children per BG
StdDev_Max_Number	The standard deviation of the average number of children per mom per BG
Sum_Max_Color	Number of Moms of Color
Sum_Max_medica	Number of Total moms with Medicaid paid births
Sum_Max_mom_an	Number of moms ever attending any college
Sum_Max_patern	Number of moms whoever listed paternity on birth certificate
Sum_Max_teen_y	Number of teen moms
Average_Ave_IPI_Da	Average 'average' IPI of all moms in BG by DAY
StdDev_Ave_IPI_Da	Standard deviation of ‘average' IPI of all moms in BG by DAY
Average_Ave_IPI_Mo	Average 'average' IPI of all moms in BG by MONTH
StdDev_Ave_IPI_Mo	Standard deviation of ‘average' IPI of all moms in BG by MONTH
Sum_Max_RRP	Number of rapid repeat moms
Minimum_ALAND	Land area of BG (area without water features)