GEOGRAPHIC DISPARITIES OF OBESITY AS A PUBLIC HEALTH ISSUE IN SUMMIT COUNTY, OHIO

A dissertation submitted
to Kent State University in partial
fulfillment of the requirements for the
degree of Doctor of Philosophy

by

Mohammad Alnasrallah

August 2015

© Copyright
All rights reserved
Except for previously published material
Dissertation written by

Mohammad Alnasrallah

B.A., Kuwait University, 2008
M.A., Kent State University, 2012
Ph.D., Kent State University, 2015

Approved by

Dr. Jay Lee, Professor, Ph.D., Department of Geography, Doctoral Advisor

Dr. Xinyue Ye, Assistant Professor, Ph.D., Department of Geography, Committee Member

Dr. Milton Harvey, Professor, Ph.D., Department of Geography, Committee Member

Dr. Eric S. Jefferis, Associate Professor, Ph.D., Department of Public Health, Committee Member

Dr. Joel W. Hughes, Associate Professor, Ph.D., Department of Psychology, Graduate Faculty Representative

Accepted by

Dr. Mandy Munro-Stasiuk, Professor, Ph.D., Chair, Department of Geography

Dr. James L. Blank, Professor, Ph.D., Dean, College of Arts and Sciences
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF FIGURES</td>
<td></td>
<td>v</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td></td>
<td>viii</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td></td>
<td>ix</td>
</tr>
<tr>
<td>1</td>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>RESEARCH PROBLEMS</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>LITERATURE REVIEW</td>
<td>6</td>
</tr>
<tr>
<td>3.1</td>
<td>The Built Environment</td>
<td>8</td>
</tr>
<tr>
<td>3.1.1</td>
<td>Access To Parks</td>
<td>8</td>
</tr>
<tr>
<td>3.1.2</td>
<td>Walkability, Sidewalk, and Street Connectivity</td>
<td>10</td>
</tr>
<tr>
<td>3.1.3</td>
<td>Land-use Mix</td>
<td>12</td>
</tr>
<tr>
<td>3.1.4</td>
<td>Fitness Centers</td>
<td>13</td>
</tr>
<tr>
<td>3.1.5</td>
<td>Food Outlets</td>
<td>14</td>
</tr>
<tr>
<td>3.1.6</td>
<td>Socio-Economic Factors (Income, Education, Unemployment), Crime, and Obesity</td>
<td>15</td>
</tr>
<tr>
<td>3.2</td>
<td>Geographic Scales</td>
<td>17</td>
</tr>
<tr>
<td>4</td>
<td>RESEARCH PREPARATION AND METHODOLOGY</td>
<td>20</td>
</tr>
<tr>
<td>4.1</td>
<td>Study Area</td>
<td>20</td>
</tr>
<tr>
<td>4.2</td>
<td>Data Preparation</td>
<td>24</td>
</tr>
<tr>
<td>4.3</td>
<td>Analytic Procedure</td>
<td>27</td>
</tr>
<tr>
<td>4.3.1</td>
<td>Hot spot, Cluster, and Outlier Analysis</td>
<td>30</td>
</tr>
<tr>
<td>4.3.2</td>
<td>Network Analysis (Service Area)</td>
<td>30</td>
</tr>
<tr>
<td>4.3.3</td>
<td>Network Analysis (Closest Facility)</td>
<td>33</td>
</tr>
<tr>
<td>4.3.4</td>
<td>Geographically Weighted Regression (GWR)</td>
<td>35</td>
</tr>
<tr>
<td>4.3.5</td>
<td>Extending Thomas’s Nationwide Model</td>
<td>38</td>
</tr>
</tbody>
</table>
ANALYSIS AND RESULTS.......................................................................................47
5.1 Geography of Obesity in Summit County, Ohio ............................................47
  5.1.1 Ethnicity and Obesity in Summit County, Ohio ....................................57
  5.1.2 Socioeconomic Variables, Land Cover, and Obesity
     In Summit County, Ohio..............................................................................61
5.2 Built Environment Features and Obesity in Summit County, Ohio ..........71
  5.2.1 Network Analyst (Service Area)..............................................................71
  5.2.2 Network Analyst (Closest facility) ...........................................................74
5.3 Results of Applying Geographically Weighted Regression ......................82
5.4 Extension of Thomas’s Dynamic Model ......................................................99
5.5 Discussion ....................................................................................................108
  5.5.1 Advantage / Disadvantage of Census Tracts and Block groups ..........110

CONCLUSION .......................................................................................................112
6.1 Limitations ....................................................................................................114
6.2 Future Directions ........................................................................................116

REFERENCES ......................................................................................................117

APPENDIX............................................................................................................125
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Percentage of ethnicity in Summit County, Ohio</td>
<td>21</td>
</tr>
<tr>
<td>2</td>
<td>Adjacent counties of Summit County, Ohio</td>
<td>22</td>
</tr>
<tr>
<td>3</td>
<td>Cities in Summit County, Ohio</td>
<td>23</td>
</tr>
<tr>
<td>4</td>
<td>Flowchart of the analytical procedure of this study</td>
<td>29</td>
</tr>
<tr>
<td>5</td>
<td>Applying service area around food outlets</td>
<td>33</td>
</tr>
<tr>
<td>6</td>
<td>Applying Closest Facility tool to food outlets</td>
<td>34</td>
</tr>
<tr>
<td>7</td>
<td>GWR typical equation</td>
<td>36</td>
</tr>
<tr>
<td>8</td>
<td>Dynamic Model for predicting the prevalence of obesity by Thomas et al. (2014)</td>
<td>40</td>
</tr>
<tr>
<td>9</td>
<td>Equations of the dynamic model by Thomas et al. (2014) to arrive the obesity prevalence plateau</td>
<td>44</td>
</tr>
<tr>
<td>10</td>
<td>Ratio of Obesity formula</td>
<td>47</td>
</tr>
<tr>
<td>11</td>
<td>Average obesity ratio for Summit County, Ohio cities</td>
<td>49</td>
</tr>
<tr>
<td>12</td>
<td>Moran’s I Index of Obesity ratio in Summit County, Ohio</td>
<td>51</td>
</tr>
<tr>
<td>13</td>
<td>Spatial distribution of obesity ratio in Summit County, Ohio at the block group level</td>
<td>52</td>
</tr>
<tr>
<td>14</td>
<td>Spatial distribution of obesity ratio in Summit County, Ohio at the tract level</td>
<td>53</td>
</tr>
<tr>
<td>15</td>
<td>Hot spot analysis of obesity ratio at the block group level in Summit County, Ohio</td>
<td>55</td>
</tr>
<tr>
<td>16</td>
<td>Hot spot analysis of obesity ratio at the tract level in Summit County, Ohio</td>
<td>56</td>
</tr>
<tr>
<td>17</td>
<td>African American calculation formula</td>
<td>57</td>
</tr>
</tbody>
</table>
Spatial distribution of White population in Summit County, Ohio by block group........59
Spatial distribution of African American in Summit County, Ohio........60
Urban versus green spaces in Summit County, Ohio.................................64
Cluster and outlier analysis of obesity ratio at the block group level...........69
Cluster and outlier analysis of obesity ratio at the block group level.........70
Results of Walkability index in Summit County, Ohio Populations...............73
Distance to grocery stores in Summit County, Ohio at the census block group level......77
Distance to fast food in Summit County, Ohio at the census block group level........78
Distance to parks in Summit County, Ohio at the census block group level........79
Distance to fitness centers in Summit County, Ohio at the census block group level......80
Distance to bus stops in Summit County, Ohio at the census block group level........81
Relationship between ratio of obesity and education attainment at tract and
block group level.........................................................................................85
Relationship between ratio of obesity and median family income at tract
and block group level.......................................................................................86
Relationship between ratio of obesity and Unemployment at tract
and block group level.......................................................................................87
Spatial distribution of residuals for GWR model (1) at the tract level........89
Spatial distribution of residuals for GWR model (1) at the block group level........90
Spatial distribution of residuals for GWR model (2) at the tract level........91
Spatial distribution of residuals for GWR model (2) at the block group level........92
GWR results of model (1) at the census tract level........................................95
<table>
<thead>
<tr>
<th>Page</th>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>37</td>
<td>GWR results of model (1) at the census block group level</td>
<td>96</td>
</tr>
<tr>
<td>38</td>
<td>GWR results of model (2) at the census tract level.</td>
<td>97</td>
</tr>
<tr>
<td>39</td>
<td>GWR results of model (2) at the census block group level</td>
<td>98</td>
</tr>
<tr>
<td>40</td>
<td>Screenshot of the dynamic model parameters and software interface</td>
<td>101</td>
</tr>
<tr>
<td>41</td>
<td>Results of applying the dynamic model at the census block group in Summit County, Ohio</td>
<td>102</td>
</tr>
<tr>
<td>42</td>
<td>First and last block groups that plateaued in Summit County, Ohio</td>
<td>105</td>
</tr>
<tr>
<td>43</td>
<td>Time at which the obesity prevalence would plateau with different birth rates</td>
<td>107</td>
</tr>
<tr>
<td>44</td>
<td>Time at which the obesity prevalence would plateau at different social influence rates</td>
<td>107</td>
</tr>
<tr>
<td>45</td>
<td>The same BMI score with a different body shape</td>
<td>115</td>
</tr>
</tbody>
</table>
## LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>BMI formulas according to CDC (Center for Disease Control and Prevention)</td>
<td>26</td>
</tr>
<tr>
<td>2</td>
<td>BMI categories according to CDC (Center for Disease Control and Prevention)</td>
<td>26</td>
</tr>
<tr>
<td>3</td>
<td>Service area distances in urban and sub-urban areas</td>
<td>31</td>
</tr>
<tr>
<td>4</td>
<td>Walkability score in Summit County, Ohio</td>
<td>32</td>
</tr>
<tr>
<td>5</td>
<td>Number of geocoded points in Summit county, Ohio</td>
<td>45</td>
</tr>
<tr>
<td>6</td>
<td>Comparison between obesity ratio between census tract and block group</td>
<td>50</td>
</tr>
<tr>
<td>7</td>
<td>Average obesity ratio and number of obese people for White and African</td>
<td>56</td>
</tr>
<tr>
<td></td>
<td>American at the census block group level</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Socioeconomic variables and land cover percentage for the top ten-obesity</td>
<td>63</td>
</tr>
<tr>
<td></td>
<td>ratio at the census block group and tract</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Socioeconomic variables and land cover of the high and low clusters of obesity</td>
<td>68</td>
</tr>
<tr>
<td></td>
<td>ratio at the block group and tract level</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Walkability index score, average obesity ratio, and number of obese people</td>
<td>72</td>
</tr>
<tr>
<td></td>
<td>by census block group in Summit County, Ohio</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Average distance between top ten obesity ratio and grocery stores, fast food,</td>
<td>76</td>
</tr>
<tr>
<td></td>
<td>parks, fitness centers, and bus stops</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Results of adjusted $R^2$ and each independent variable</td>
<td>83</td>
</tr>
<tr>
<td>13</td>
<td>Independent variables for the two GWR models</td>
<td>88</td>
</tr>
<tr>
<td>14</td>
<td>Comparison between the first and last block groups that their obesity</td>
<td>104</td>
</tr>
<tr>
<td></td>
<td>prevalence become plateau</td>
<td></td>
</tr>
</tbody>
</table>
ACKNOWLEDGEMENTS

I would like to acknowledge and express my sincere gratitude to my adviser Dr. Jay Lee, my doctoral committee members, Dr. Xinyue Ye, Dr. Milton Harvey from the Department of Geography, and Dr. Eric S. Jefferis from the Department of Public Health at Kent State University for their patience and guidance. The faculties of the Department of Geography have all been encouraging and helpful in my stay here and I would like to thank all of them for their assistance and encouragement through my academic years. In addition, I extend my appreciation to my adviser, Dr. Jay Lee, for his comments and ideas, which helped me to complete this study successfully.

I would also like to express great thanks to my country Kuwait and the Department of Geography at Kuwait University for their financial support, which gave me the opportunity to pursue my postgraduate studies at Kent State University. Additionally, I would like to thank many people who have helped and encouraged me to further my studies.

Finally, I would like to thank all of my family members for their support. I especially want to thank my parents (Abdullah and Romiah) for their constant faith and encouragement that inspired me to pursue my graduate studies in the United States. I would like also to thank my siblings, Sheikha, Fahad, and Hessa Alnasrallah for their support.
CHAPTER 1
INTRODUCTION

In 2010, the prevalence of obesity among adults in the United States was more than 35.7% (Ogden et al., 2013). The prevalence of obesity increased in the last few years but at a slower rate than that over the past 10 years (Flegal et al., 2010). Obesity is a public health issue that can cause many health related problems such as, diabetes, hypertension, high-blood pressure, various cancers, and other chronic diseases (World Health Organization, 2003). Geography and built environment of a locality were argued to play a significant role in the prevalence of obesity.

According to Wakefield (2004), the built environment included aspects of the environment such as homes, schools, workplaces, parks, streets, and infrastructure. Built environment could influence the levels of physical activity of residents in a neighborhood. For example, sidewalks, connectivity, the availability of parks, and the access to healthy food, all of those physical parts in the built environment could contribute to poor health outcomes such as obesity and the aforementioned chronic diseases as complications, of residents in the area.

A dynamic model for predicting overweight, obesity, and extreme obesity prevalence was suggested by Thomas et al. (2014). It was useful when applying the model for large geographic area, such as the entire United States or United Kingdom, by using data available at the country level. Their model was the first to categorize population into subpopulations of normal weight, overweight, obese, and extremely obese. They also pioneered the use of the concept of
susceptible, infected, and recovered (SIR) subpopulation on obesity prevalence, which was applied in modeling the movement of population between subpopulations. This model also allowed different assumptions to be made on death rates, birth rates, and social influences by behavioral patterns of food consumption and physical activities. Through their model, Thomas et al. (2014) showed that predicting future obesity prevalence was feasible and that obesity prevalence would be plateauing in the future.

As a geographic study, this research attempted to address two shortcomings in Thomas et al. (2014):

1. Lack of geographic details: since the model by Thomas et al. (2014) was only a nationwide model, a potential improvement would be to apply the model at neighborhood level so to explore the public health disparities among local communities.

2. Lack of clearly defined approaches to incorporate physical, food, built, and socio-economic environments of the area being modeled. In other words, this research developed a mechanism that incorporated the influence by socio-economic environment on how obesity prevalence changed in different localities.

Studying the spatial distribution of obesity in Summit County, OH, was essential to investigate the neighborhoods of high obesity prevalence and link it to socioeconomic variables and built environment features that could be related to obesity. This research focused on the variables that were associated with obesity in the study area. Those variables include the socio-economic variables, such as income, education, and employment. In other words, this study try to find the factors and variables that were associated with high levels of obesity prevalence in the study area. In addition, this research also focused on the built environment features and their relations to obesity. The built environment features includes, food outlets (grocery stores and fast food),
bus stops, and the percentage of open green spaces and urban. Finding the variables that were associated with obesity were done by using GIS, which was based on quantitative method. The unit of analysis in this research was at the census tract and block group levels. However, the main unit of analysis was at the block group level. The tracts were included in this study to make a comparison between larger and smaller geographic scales. The issue of using different scale units leads to focus on one of the issues that is important in geography, which is Modifiable Area Unit (MAUP). This issue occurred when aggregating the data of smaller geographic units in to larger ones. It occurred also when there were differences in the zoning and grouping of the geographic boundaries.

This study also focused on extending the model that had been applied by Thomas et al. (2014) and applied it at the neighborhood level. In the following chapters, I discuss the research objectives of this study and the literature review regarding the obesity prevalence and geography. Next I discuss the literature review and existing studies in relation to build environment and obesity. Then I discuss the collected data, the analytical methods, the results derived from the analysis, limitation, and finally future directions to pursue in the future.
CHAPTER 2
RESEARCH PROBLEMS

There are many factors that are associated with obesity. These factors could be related to socioeconomic factors, such as education, income, and employment. Another aspect could be related to the culture. For example, some ethnic groups tended to eat food that was high in calories, which might increase the chances for being obese in the long term. The availability of unhealthy food outlets in the neighborhood could increase the chances for people to consume unhealthy food, especially when it comes to the geographic location of the food outlets and the prices of the food they sell. In general, buying fast food costs less than buying healthy and fresh food. Therefore, low-income people tend to buy fast food, which costs much less than healthy food. Several studies suggested that eating from fast food restaurants without doing physical activity regularly increased the fat in the body, which would cause obesity in the long term (Anderson et al., 2011). Other studies found that living in neighborhoods where there were more options for non-fresh food outlets, the chances for being obese were often higher than those who were living in neighborhoods where there was enough access to fresh food outlets (Gallagher, 2006). Access to parks also could increases the chances for people to increase their physical activity.

In study area, this research used a total number of geocoded points of people’s weight aged 16 – 21, which were 33,484 people. The percentage of normal weight people (BMI < 24.9)
according to the data geocoded (age 16 – 21) was 64%, while the percentage of overweight (BMI 25 – 29.9) 23.5%, was obese (BMI 30-39.9) 11.5%, and extremely obese (BMI > 40) was 1%. From these percentages, it seemed that there was high percentage of obesity in the study area, which might still increase in the future because the overweight percentage was higher too. Therefore, this study focused on understanding the spatial distribution of obesity prevalence, how it was related to socioeconomic variables, and built environment features. Socioeconomic variables that were included in this study were income, education attainment, and unemployment. In addition, this study focused on the built environment features that could affect the prevalence of obesity at smaller geographic area. The study was performed at different geographic scales (tracts and block groups) for a comparison purpose and to understand how the scale affects the results of the analysis. The aim of study was also to apply the dynamic model that was developed by Thomas et al. (2014) at the national level to local areas in order to understand how the model could be different while applying it at smaller geographic area.

In summary, social influences of the physical, food, built environment and socioeconomic environments of the study area could play a significant role in affecting obesity prevalence. By applying a large-area model to small geographic areas, this study attempted to explore the feasibility of structuring factors that were related to obesity prevalence by the exploring the environmental factors and socioeconomic variables of the study area. In addition, the study explored the feasibility of modeling obesity prevalence in small geographic areas to reveal area disparities of this important public health issue.
CHAPTER 3

LITERATURE REVIEW

The prevalence of obesity had increased dramatically in last few decades worldwide, such as in the United States, United Kingdom, Eastern Europe, the Middle East, the Pacific Islands, Australasia, and China (World Health Organization, 2003). In 2012, more than one-third (34.9%) of adults in the United States were obese (Ogden et al., 2013). Obesity was a major cause of death in the United States (Center for Disease Control and Prevention, 2013). According to CDC, being overweight or obese increased the risks for several chronic diseases such as, heart disease, type II diabetes, cancers, Hypertension, cholesterol, stroke, liver, and gallbladder disease. In addition, from health consequences of obesity on human health, there were also economic consequences. In 2008, the medical care cost of due to obesity was about $147 billion in the United States (Finkelstein et al., 2009). More than 72 million people in the United States were obese in 2010 (CDC, 2010). People usually did not decide to become overweight or obese. Besides genetic and other biological factors, the built environment of neighborhood, access to fresh food outlets, and the connectivity between sidewalks could contribute to the chances of being obese.

In terms of access to healthy food, there were three main environmental factors that might affect consuming healthy food: (1) informational, (2) economic, and (3) geographical access (McEntee & Agyeman, 2010). The informational factors depended on the level of education and knowledge in order to make healthy food choices. The literature suggested that
high level of education was associated with lower rates of obesity (Webbink et al., 2010). The second factor was the economic factor, which was an important factor because not everyone could afford to buy healthy food due to the normally higher prices. Eating from fast food restaurants was a better choice from low-income people because of the convenience and lower prices, which might explain why many low-income people were obese since they often could consume only low unhealthy food (Miech et al., 2006).

The third factor was the geographical access to healthy food. There were three barriers: informational, economical, and geographical conditions, which could contribute to the levels of access to healthy or unhealthy food. These might change people’s diet overtime. However, the built environment and how the community was designed were a big contributor to the increase of obesity rates. Weight gain was a consequence from living in an environment where unhealthy foods were available and opportunity for physical activity was lacking. In these areas, it might be easier and cheaper for residents to buy less healthy food. Some communities and neighborhoods were built in ways that made it difficult or unsafe for residents to walk or do physical activities. Such situation would reduce the amount of daily physical activity and might cause weight gain. The built environment could play a significant role in changing obesity rates in small communities (Swinburn et al., 1999). Living in an unhealthy environment where there was less access to parks, green spaces, and healthy food could increase the prevalence obesity in the long term. There was no simple solution for obesity epidemic, people need to make better choices, however, these choices needed to be changed in the built environment, schools, and communities.
3.1. The Built Environment

The built environment had changed dramatically over the last century because of the creation of motor vehicles, which changed the design of neighborhoods in cities (Falconer & Giles-Corti, 2008). The built environment formed the behavioral and health patterns of individuals living in a neighborhood. The physical environment was an important factor that could reduce the rates of obesity. The physical environment could include access to parks, walkability, sidewalk, street connectivity, the characteristics of street, mixed land-use, residential density, public transportation, crime etc.

3.1.1 Access to Parks

Park facilities offered a great opportunity for people to do physical activity and exercises during the day. Access to parks could be an important factor in increasing physical activity in such a neighborhood. In the United States, it had been suggested that minority population and low-income neighborhoods had less access to parks and open green space (Byrne et al., 2009), which obviously reduced their physical activity and increased their body mass index.

In the United States, poverty level was negatively associated with distances to parks and high density of green spaces in urban and suburban areas while positively correlated in rural areas (Wen et al., 2013). Other studies found that neighborhoods that were more walkable and had enough access to parks were associated with higher levels of physical activity and lower values of body mass index (Giles-Corti & Donovan, 2002,b; Saelens, Sallis, Black, & Chen, 2003). Wen et al. (2013) found that access to parks was negatively associated with obesity risks. In terms of proximity from parks and physical activity, Roemmich et al. (2006), found that
neighborhoods proximity between homes and high access to parks were associated with high levels of physical activity in young children.

Having access to parks was not enough to increase physical activity because it depended on the characteristics of the park such as the size of the park, safety, and park features. Bedimo-Rung et al. (2005) did a study that focused on the significance of parks to physical activity and they found some park environmental characteristics that might affect the frequencies of residents going to parks:

1. park features – this included the presence of sport fields and walkable sidewalks where people could do exercises.

2. condition of park features – this included maintenance of the parks and equipment in the parks.

3. access to parks – this was related to the distance the parks were located from residential areas.

4. the density of parks in a city – this was related to the use potential of the parks by residents.

5. the aesthetics of the park – this determined how attractive the parks are to residents.

6. the safety of the park to the occurrences of crime.

7. the administrative policies – this defined how residents could use the parks as well as how parks were maintained.

All of these environmental characteristics were interrelated with each other. They would affect frequencies of visiting to parks, which would determine the level physical activity of the neighborhood.
3.1.2 Walkability, Sidewalk, and Street Connectivity

Another aspect in the physical environment was walkability. The term, a walkable neighborhood, referred to a neighborhood whose physical layout helped residents to engage pedestrian activities such as walking or cycling to parks, to access open green spaces for leisure activity, to go to places such as work, or to take public transportation instead of driving private vehicles (Pearce, 2012). More walkable neighborhoods were found to be associated with higher levels of physical activity by their residents.

Frank and his colleagues found that lower BMI was associated with more walkable neighborhoods (Frank et al., 2006). Levels of walkability in a neighborhood, in general, was different among neighborhoods of different income levels. A study on Baltimore, Maryland, found that individuals living in areas of high socioeconomic status and with a predominately-white population lived in highly walkable neighborhoods. This finding was associated with lower prevalence of obesity in those neighborhoods, as compared with poor walkable neighborhoods (Casagrande et al., 2011). In that study, the author and his colleagues found four factors that could affect the walkability in such a neighborhood,

1. the connection to other sidewalks,
2. the availability of stop signs,
3. obstructions in the sidewalk, and
4. the design of the crosswalks.

Even though, other studies found that the relationship between walkable neighborhoods was not significantly associated with lower BMI (Saelens et al., 2003b; Doyle et al., 2006). Recent study in Austria did find that individuals living in high walkable areas were less likely to be obese and
had fewer cases of type-2 diabetes at a buffer of 800 meters than those for the individuals living in less walkable areas (Muller et al., 2013).

Sidewalk and street connectivity in a neighborhood was also associated with higher amounts of physical activity, which in turn could reduce the BMI (Body Mass Index) of its residents (Ross, 2001). A study in Houston, TX found that physical activity, resource attributes and neighborhood sidewalk connectivity were related to high BMI and body fat among low-income African Americans living in low-income housing developments (McAlexander et al., 2009). In contrast, Grafova et al. (2008) found no correlation associated with street connectivity and overweight or obese men.

For women, street connectivity was not associated with obesity; however, women living in areas with high street connectivity were less likely to be overweight or obese. The characteristics of streets might also be an important factor that could contribute to physical activity and obesity. The characteristics of street included different aspects of streets, such as (Cleland et al., 2008, Hong & Farley 2008, Nagel et al., 2008, Ogilvie et al., 2008; Owen et al., 2004).

- the type of streets: whether it is highway or suburban road,
- the traffic volume on the street, the speed of traffic on the street,
- availability of sidewalks,
- the quality of those sidewalks whether street have lights,
- street safety

Joshu et al. (2008) found that the chances for obesity were higher among respondents who viewed traffic as a problem in their neighborhoods. However, there was no correlation between
streetlight availability and obesity. Giles-Corti et al. (2003) examined the association between environmental and lifestyle factors and overweight or obesity. They found that overweight people were associated with neighborhoods adjacent to highways or in neighborhoods with no sidewalks or just one sidewalk. All street and sidewalk characteristics played an important role in increasing the physical activity of a neighborhood.

3.1.3 Land-use Mix

Another aspect that could contribute to how physical environment was related to obesity was land-use mix. Land-use mix was the diversity of land uses such as how residential, commercial, industrial, and agricultural land uses mixed in an area. A variety of land use mixes at neighborhood levels were found to be correlated with short travel distances between places (Gebel et al., 2005). This indicated that people could move to different places by walking on the sidewalks, which increased the physical activities. Residents were more likely to walk if they lived in high residential density neighborhoods, neighborhoods with high street connectivity, or neighborhoods with a greater land use mix (Frank et al., 2008). High degree of land use mix could contribute to reducing the prevalence of overweight or obesity because of the walkability in high land-use mix areas. This, however, was not conclusive because several studies found a negative association between land-use mix and body weight, which supports the idea of being obese, is more likely if living in neighborhoods with high land-use mix (Frank et al., 2004, Mobley et al., 2006; Rundle, et al., 2007). On the confirmative side, a study that on 120 neighborhoods in Portland, Oregon, found that each increase in land-use mix was correlated with lower prevalence of obesity (Li et al., 2008).
Land-use mix offered a variety of options for residents to walk between different places such as, commercial and residential, which could increase the access to supermarkets and different options of healthy food (Gebel et al., 2005). Areas with high land-use mix also tended to have high densities of public transit. Neighborhoods with high density of public transportation stations were found to be associated with more walking by residents comparing to those with lower density of public transit (Li et al., 2008).

Rutt and Coleman (2005) found a positive correlation between mixed land-use with body mass index values in neighborhoods that mixed commercial and residential uses in El Paso, Texas. The positive correlation was among Hispanic people and low-income community. The relationship was positive because of the individual socioeconomic status was included in the analyses which suggested that land-use mix and built environment influenced BMI among low to moderate income and minorities.

3.1.4 Fitness Centers

Physical activity and the distribution of recreational centers could have a significant impact or could increase the prevalence of obesity (Gebel et al., 2007). The proximity to fitness centers could influence the prevalence of overweight and obesity in some neighborhoods. Several studies reported a positive association between access to recreational environment and physical activity for both adults and children (Rundle et al., 2007; Owen et al., 2004). Going to recreational centers regularly could increase physical activity; therefore, lower rates of obesity and overweight could be expected in neighborhoods with sufficient access to fitness centers.

Mobley et al. (2006) found that there was a lower average BMI in areas with more fitness centers. In addition, Boehmer et al. (2007) reported that having fewer fitness centers within close
proximity was associated with a higher likelihood of obesity among women, but not men. As opposed to this, there was a study that found no relationship between distance to fitness centers and BMI (Rutt & Coleman, 2005). Several studies showed that going to the fitness centers increased when the distance between home and facilities decreased (Gebel et al., 2005; Giles-Corti et al., 2005). To improve long-term health benefits, people should focus on improving fitness by increasing physical activity rather than simply by relying only on dieting for weight control (Lee et al., 1999). Going to fitness centers maybe a critical behavior, but there were multiple factors that might discourage or encourage this key behavior such as the price of membership, and geographical access (distances).

3.1.5 Food Outlets

The built environment and the distribution of food outlets were important factors that contributed to the increase or decrease of the prevalence of overweight or obesity in urban areas. In the last few years, a new term had emerged: food deserts. Food deserts were urban areas that had little access to healthy and fresh food and at the same time had a high density of fast food and unhealthy food options (Cummins & Macintyre, 2002; Gallagher, 2006). In the last few decades, fast food restaurants became familiar and affordable to all people because the convenience they offered and the low prices of food they offered (Pearce, 2012). The causes that might affect the consumption of fast food were the price of the food, the walking or driving distance to such outlets, and different cultural, behavioral, or environmental factors (McEntee & Agyeman, 2009; Anderson et al., 2011). Fast food meals provided more than one-third of the recommended daily energy intake, which might cause unnecessary weight gain (Bowman & Vinyard, 2004). In addition, some of the fast food restaurants such as Wendy’s and Burger King
had increased their portion sizes. To that end, it was found that fast food portion sizes were larger in the US than in Europe (Young & Nestle, 2007). This could explain why there were higher rates of obesity in the United States than those in European countries.

High density of Fast food restaurants and absence of supermarkets in a neighborhood (where fresh produces tended to be offered) could be an environmental factor for obesity. A study that had been done in three cities in Maryland, North Carolina, and New York found that the absence of supermarkets near household was associated with lower chances for consuming healthy diets (Moore et al., 2008). The availability of fast food and high density of unhealthy food were associated with higher rates of obesity rates in African Americans and other low-income minorities (Gallagher, 2006). Block and his colleagues (2004) found in New Orleans, Louisiana that predominately African American neighborhoods had 2.4 fast food outlets per square mile as compared to 1.5 fast food outlets in predominantly White neighborhoods.

On the other hand, a neighborhood that had supermarkets and large grocery stores that sold fresh food such as vegetables and fruits were associated with lower chances for being obese or overweight (Gallagher, 2006). Supermarkets in the United States offered a wide variety of healthy food (Sallis et al., 1986). Therefore, higher chances for eating healthy food would be around large supermarkets because they afforded healthy food options. Morland et al. (2006) found that the availability of supermarkets that sell healthy food was associated with lower prevalence of obesity and overweight in urban areas.

3.1.6 Socio-Economic Factors (Income, Education, Unemployment), Crime, and Obesity

Socioeconomic status (SES) and obesity rates were linked to each other. The level of income was an important factor in determining the types of food people consumed. The literature
on this issue suggested that poor and low-income people had high rates of obesity because they tended to choose to consume fast food or unhealthy food due to the convenience and the low prices (US Department of Health and Human Services, 2012). In addition, high-income families were associated with increased physical activities (Gordon-Larsen et al., 2000).

Recent study found the density of fast food restaurants within 0.5, 1, and 2 miles from residential areas was positively associated with BMI values among participants with low-income. In addition, close proximity to fast food restaurants was correlated to high BMI values of people of low-income status (Reitzel et al., 2014). High numbers of fast food purchasing was associated with low household income, lower education level, and often being a blue-collar employee (Thornton et al., 2011).

Another study among US adults found that Americans with high SES had a better Eating Healthy Index scores (Wang & Chen, 2011). In the United Kingdom, a recent study found that the prevalence of overweight and obesity among English adults is higher with lower SES (Howel et al., 2013). The education level of people could be an important factor in explaining the prevalence of obesity. In the United States, higher prevalence of obesity was among people with lower educational level, lower-income level, and in more labor-intensive occupations (Singh et al., 2011). A study in India found that people with low educational, occupational and socioeconomic status had a greater prevalence of obesity (Gupta et al., 2012). Another study on Education and obesity at age 40 among American Adults found that among Whites college graduates were less likely to be obese than high school graduates; However, the obesity prevalence and educational attainment among African American or Hispanic populations were not statistically significant (Cohen et al., 2013). Unemployment was one of the social factors that
could be related to obesity. A study among childhood overweight and unemployment found that children of unemployed mothers who did not receive unemployment benefits were more likely to have higher BMI when compared with children with unemployed mothers who received the unemployment benefits (Stewart et al., 2012).

Another factor that may contribute to the decrease or increase of physical activity was the safety of the neighborhood and crime occurrences in the neighborhood. In urban areas, safety and the occurrence of crime could be an important source of stress (Doyle et al., 2006). Low-crime rates in neighborhoods were associated with increases in physical activities (Ferreira et al., 2007). A study found that low rates of crime were associated with lower BMI values among adults aged 18 years and older (Doyle et al., 2006). Women feared crime more than men did (Hollander, 2001). That meant that women in general tended not to walk in a neighborhood where chances for crime occurrence were high. A study that had been done on obesity and low-income women found that crime was positively associated with BMI (Mobley et al., 2006). That meant the high rates of crime were associated the high level of BMI among women. The hypothesis about Active Community Environments (ACES) model demonstrated that active community needed to have two characteristics, walkability, and safety of the community in order to have a greater physical activity that would lower obesity and prevent weight related chronic diseases, which would ultimately make better overall health.

3.2 Geographic Scales

Several studies had used GIS technology to analyze and investigate the spatial distribution of overweight and obesity (Duncan et al., 2012; Lamichhane et al., 2012; Christian et al., 2011; Norman et al., 2013; Reitzel et al., 2014). GIS was a useful tool for investigating
spatial distribution of obesity prevalence in space and how that changed over time. GIS could also be used to measure the association between built environment and BMI in order to see the correlation between built environment and prevalence of obesity (Gebel et al., 2007; Swinburn et al., 1999).

On the issue of geographic scale, there were just few obesity studies that had been done at small geographic areas, such as at the level of census tract or block group (Gordon-Larsen et al., 2006; Greves Grow et al., 2010). It had been suggested that the smallest geographical scale was the best to measure environmental factors in urban cities because of the level of data aggregation (Leslie et al., 2007). A study that had been done using census block groups found that low- socioeconomic and high density of minority population census block groups had fewer facilities, which would decrease the levels of physical activity in that block group and might increase the prevalence of obesity in that census block group (Gordon-Larsen et al., 2006).

Yamada et al. (2012) studied the issues of neighborhood scales and measuring built environment by exploring the relationship between BMI and four types of mixed land use at three different geographic scales for defining the neighborhood. They found that street-network buffer was the best for defining the neighborhood. Another study on the relationship between neighborhood food environment and obesity risk, found that the scale of analysis and how the neighborhood was defined in terms of the geographic area matters and it could change the outcome of the analysis (Fan et al., 2014).

In conclusion, most of the existing studies were carried out using large geographical areas due to the data availability and confidentiality. It was, however, preferred, if smaller geographic areas could be used as the unit of analysis for topics such as obesity prevalence and built
environment. This was because smaller geographic areas reveal more details and reduce the adverse effects of generalization due to data aggregation.
CHAPTER 4
RESEARCH PREPARATION AND METHODOLOGY

This chapter explains the process of the data preparation and analytic methods that were used in this study. A variety of GIS techniques were used in this research to explore the geographic disparities of obesity prevalence and detect the variables that are associated with obesity.

4.1 Study Area

Summit County is one of the counties that is located in northeast Ohio. It is surrounded by six counties: Cuyahoga and Geauga from the north side, Portage on the east side, Stark and Wayne from the south, and Medina on the west side. Akron is the largest city in Summit County and it is the major metropolitan area in the county (See Figures 2 & 3). The total area of Summit County is 412.75 square miles with a population density of 1,312.6 persons per square mile (US Census Bureau, 2012). According to the US Census, in 2012, the total population in Summit County was 540,811. Regarding ethnicity composition, white alone accounted for 80.6%, African American for 14.6%, Asian for 2.4%, Hispanic for 1.8%, and others for 0.6% (See Figure 2). About 36% of the county’s population lived in Akron city. Most of the minorities and low-income people lived in Akron city. This study focused on Summit County, OH, also for the reason that its demographic profile was very similar to the average demographic profile of the United States as a whole.
Figure 1. Percentage of ethnicity in Summit County, Ohio
Figure 2. Adjacent counties of Summit County, Ohio
Figure 3. Cities in Summit County, Ohio
4.2 Data Preparation

There were several data sets that were used in this study. First of all, data from the US Census Bureau were assembled for the demographics, population counts, ethnicity, ages, gender, education level, employment, and level of income of Summit County at tracts and block groups level. In addition, boundary files of census tracts, block groups, and roads were extracted from the TIGER/Line files and formulated to shapefiles for use in ArcGIS. Furthermore, parks shapefile were digitized and downloaded from Summit Metro Parks website (http://www.summitmetro.parks.org) and from Google earth.

Secondly, locations and attribute information of different types of food outlets in the county, which included fast food restaurants, non-fresh food outlets, and large small chain supermarkets, were downloaded from “A to Z databases website” (http://www.atozdatabases.com). This website provided data for 30 million businesses in the United States. It was one of the most comprehensive data sources, which offer the most complete coverage of business establishments in the US.

Finally, body mass index data for Summit County, Ohio were calculated using the self-reported weights and heights recorded on drivers licenses. The driver license data were obtained from Ohio Bureau of Motor Vehicles for five years from 2008 to 2012. The study used individual records for driver licenses holders for adults aged 16 to 21. Targeting at population aged between 16 and 21 was because adults at this age group put their most recent weights and heights when they first applied for their driver’s licenses. Therefore, the heights and weights could be closest to their real weights and heights. In most cases, people would not update their weights and heights when they renewed their drivers licenses every five years. Therefore, using this age group hold result in better data than using all age groups.
This dataset was not 100% coverage for all population adults in Summit County, Ohio. It
did not include those who chose not to acquire driver’s license and those who failed to renew
licenses. Moreover, it was possible that heights and weights might be underreported for driver’s
licenses. For example, a person’s height and weight at age 16 when first acquiring his/her
driver’s license was likely lower than his/her height and weight at age 20 when renewing the
license at a later age. A study on assessing driver’s licenses as a source of data on height and
weight, did a comparison between Washington state driver’s license records (height and weight)
and control women enrolled in a cancer etiology study (height and weight), they found that there
was a close relation between height but not between weights (Ossiander et al., 2004). That meant
people tended to put their accurate heights but not their actual weights. However, BMI data as
derived from heights and weights reported from the Bureau of Motor Vehicles were probably the
best data that was available for each coverage.

The BMI data was calculated from the heights and weights in each driver’s licenses in
Summit County, Ohio. BMI is an index for measuring the body fatness. The formula used to
calculate BMI was shown in Table 1. The calculated BMI describes a person’s body features as
shown in Table 2. This data source included only adults who held driver’s licenses in the county
between 2008 and 2012. This data did not include population age 15 and below or those who did
not hold driver’s license.
Table 1. BMI formulas according to CDC (Center for Disease Control and Prevention)

<table>
<thead>
<tr>
<th>Measurement Units</th>
<th>Formula and Calculation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kilograms and meters (or centimeters)</td>
<td>Formula: weight (kg) / [height (m)]&lt;sup&gt;2&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>Example: Weight = 68 kg, Height = 165 cm (1.65 m)</td>
</tr>
<tr>
<td></td>
<td>Calculation: 68 ÷ (1.65)&lt;sup&gt;2&lt;/sup&gt; = 24.98</td>
</tr>
<tr>
<td>Pounds and inches</td>
<td>Formula: weight (lbs) / [height (in)]&lt;sup&gt;2&lt;/sup&gt; x 703</td>
</tr>
<tr>
<td></td>
<td>Example: Weight = 150 lbs, Height = 5’5”</td>
</tr>
<tr>
<td></td>
<td>(65”)Calculation: [150 ÷ (65)&lt;sup&gt;2&lt;/sup&gt;]X 703 = 24.98</td>
</tr>
</tbody>
</table>

Table 2. BMI categories according to CDC (Center for Disease Control and Prevention)

<table>
<thead>
<tr>
<th>BMI</th>
<th>Weight Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Below 18.5</td>
<td>Underweight</td>
</tr>
<tr>
<td>18.5 – 24.9</td>
<td>Normal</td>
</tr>
<tr>
<td>25.0 – 29.9</td>
<td>Overweight</td>
</tr>
<tr>
<td>30.0 -39.9</td>
<td>Obese</td>
</tr>
<tr>
<td>Above 40</td>
<td>Extremely obese</td>
</tr>
</tbody>
</table>
4.3 Analytic Procedure

The approach to analyzing used in this study was based on GIS technology. The analysis was based on a quantitative approach. The research measured the walkability to several facilities and built environment features such as parks, fitness centers, and food outlets (non-fresh versus fresh food) then compared the measured walkability with ratios of obesity at the census tracts and block groups levels. The analysis was carried out for both the census tracts and block groups in order to detect the differences between the results by using the two geographical scales. Several studies had used GIS as a method for measuring walkability (Hajna et al., 2013; Neckerman et al., 2009; Lwin & Murayama et al., 2011). This study measured walkability by applying service area tool in Network Analyst of ArcGIS, which is to be discussed in the next section.

This study focused on finding variables, such as the geographical access to parks, fitness centers, food outlets (non-fresh versus fresh food), and bus stops that could be associated with obesity. The food outlets were categorized into non-fresh versus fresh food outlets. Non-fresh food outlets included fast food restaurants, food carry out and, pizza restaurants. Fresh food outlets included small grocery stores and large chain supermarkets.

In addition, the socioeconomic variables, such as income, level of education, and unemployment, were included in the analysis to find if there was any existed association with obesity prevalence. The level of education and income could change people’s diet in the long term. The presence of sidewalk and street connectivity could change people’s habits by increasing physical activity (Ross, 2001). Street connectivity was determined by calculating how many street intersections fell inside each census tract or block group. In addition, street density was used as a good indicator of street connectivity.
The study also focused on formulating the measurement of the social influences on obesity in Summit County, Ohio. A recent study by Thomas et al. (2014) created a dynamic model that predicted the percentage or the level at which the prevalence of obesity would plateau. The model also described the transition from the overweight classification to a healthy BMI. This model was formulated and tested for large geographic areas, such as countries or states, while in this study it was applied at the neighborhood level.

Figure 4, shows the analytical procedure for this study and how the analysis was done. The first step was collecting the data from different sources, such as the US Census Bureau and the BMI data from the BMV (Bureau of Motor Vehicles). After that, the data were cleaned for any errors. Especially the errors in BMI data, such as missing addresses, or with other typo errors. In addition, some driver licenses records did not have a BMI score. In this case, those records were excluded. The next step was importing the data into GIS and geocode the physical addresses of BMI data. Furthermore, aggregate the data at the tracts and block groups levels. The next step was the analysis part, which included, network analysis, hotspot analysis, geographically weighted regression, and the extended dynamic model of obesity. The results then were represented in tables and maps. The last procedure was for the recommended solutions and suggestions to reduce the prevalence of obesity in the study area. The following sections explain how each tool was performed in this study.
Figure 4. Flowchart of the analytical procedure of this study

Data collection
- BMI data
- Socioeconomic variables
- Built environment features

Data editing
- Cleaning the data
- Errors

Importing data to GIS
- BMI geocoding
- Aggregate at the tract and block group level (Point to polygon)

GIS analysis
- Network analysis
- Hot spot analysis
- GWR
- Extended dynamic model of obesity

Results
- Tables and maps

Recommended solutions
- Policy makers
4.3.1 Hot spot, Cluster, and Outlier Analysis

Hot spot analysis was one of the widely available tools used in GIS. This tool identified areas of hot spots and cold spots of such attribute values, which is in this case the prevalence of obesity. This tool allowed us to find the areas of high prevalence of obesity (hot spots) and areas of low prevalence of obesity (cold spots). Hot spot analysis was one of the powerful tools for mapping clusters of attribute values. This tool showed hot spots, which represented the clusters of high values that were surrounded by high values. In the case of this research, hot spots were locations where locations of high obesity ratios were surrounded by neighboring that also had high obesity ratios. In addition, outcomes from applying hot spot analysis showed the clusters of low values that were surrounded by low values in the neighboring areas. In this case, such locations were called cold spots. Another tool that was used in this study was the Cluster and Outlier analysis. This tool was used to detect the Clusters and Outliers of obesity ratios in Summit County, at the block groups and tracts level. This tool was used to identify the locations of high concentrations values and the concentrations of low values. After applying the cluster and outlier analysis on obesity ratio, socioeconomic variables was defined in areas of high clusters and low clusters obesity ratio.

4.3.2 Network Analysis (Service Area)

The service area tool in the Network Analyst module of ArcGIS was one of the common tools that had been used in the analysis of spatial configuration of streets or road segments. This tool measured the service area around a facility using the road network in the study area (Figure 5). I chose to use this tool rather than using the buffer analysis because this tool measured distances along the street segments in the road network where the buffer analysis just drew circles around the facilities and calculated distances by using straight Euclidean distances.
between two points. Real street distances were needed for better realistic analysis because people moved from places to places using the road network, not via straight lines between locations.

In this study, walkability in a neighborhood (Census tracts of census block groups) was measured based on the service areas of food outlets, fitness centers, parks, and bus stops. By using the Service Area tool in ArcGIS/Network Analyst, I created a walkability index by measuring the distances around different features in the neighborhood such as, grocery stores, fitness center, parks, and bus stops. Each facility was given a distance to measure the walkability by using the network analysis (Service Analysis) tool. The distances in the urban area, which in this case, was the City of Akron was set in the following order as shown in Table 3.

Table 3. Service area distances in urban and sub-urban areas

<table>
<thead>
<tr>
<th></th>
<th>Distances in Urban Areas (Miles)</th>
<th>Distance in Sub-urban Areas (Miles)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grocery stores</td>
<td>0.5</td>
<td>1.5</td>
</tr>
<tr>
<td>Parks</td>
<td>0.5</td>
<td>1.5</td>
</tr>
<tr>
<td>Bus stops</td>
<td>0.5</td>
<td>1.5</td>
</tr>
<tr>
<td>Service Plaza</td>
<td>0.5</td>
<td>1.5</td>
</tr>
</tbody>
</table>

After applying these distances around those features, each census block group was given a set of scores for measuring its walkability according to the following criteria as shown in Table 4. In Table 4, the scores were determined by a scale ranged from 0 to 1. The scores were set by dividing the number of features that had access in each block group by four. For example, if the census block group had access to three features, then it would be divided by the number of 4 to get score of walkability index of 0.75 (Table 4). In other words, the denominator is the number
of features which was 4 and the numerator would be the number of features that were accessible from such a neighborhood.

Table 4. Walkability score in Summit County, Ohio

<table>
<thead>
<tr>
<th>Walkability features</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grocery stores, parks, bus stops, service plaza</td>
<td>1</td>
</tr>
<tr>
<td>Grocery stores, parks, bus stops</td>
<td>0.75</td>
</tr>
<tr>
<td>Grocery stores, parks:</td>
<td>0.50</td>
</tr>
<tr>
<td>Parks</td>
<td>0.25</td>
</tr>
<tr>
<td>No access</td>
<td>0</td>
</tr>
</tbody>
</table>

After applying the network analysis (service area tool) in the study area, the most and least walkable areas in Summit County were identified according to the distances specified. Furthermore, I compared the results with the distribution of obesity ratio and the walkability scores at the census block groups level. Network Analyst service area is an important tool for measuring the walkability in the study area. In addition, I looked at the demographics and compared it with the most and least walkable areas. Results of these comparisons are presented in the next chapter (Chapter 5).
4.3.3 Network Analysis (Closest Facility)

Closest facility tool was another useful tool used in this research. This tool calculated the shortest distance between two points by using the road network (Figure 6). In the research, the distance calculation was carried out for each census tract and block group. The distance was
calculated between the centroid of each tract and block group to the built environment features, such as parks, food outlets, fast food, and grocery stores, and bus stops. The results of distance calculation were used as a dependent variable in the geographically weighted regression tool, which will be explained in the next section.

Figure 6. Applying closest facility tool to food outlets
4.3.4 Geographically Weighted Regression (GWR)

Geographically weighted regression (GWR) is a spatial regression modeling tool that shows the spatial relationship between the dependent and independent variables that accounts for local variations of the variables (Charlton et al., 2009). The traditional liner regression models assumed stationarity in the relationship between dependent and independent variables. In ordinary least square regression models, the regression coefficients were assumed to be the same for all of the study. As opposed to that, GWR assumed that the relationship between the variables was to vary over space (Brunsdon et al., 1996). GWR was first proposed in the geography literature by Brunsdon et al. (1996), they found that the traditional regression models did not take into account the locations in the analysis of the relationships between dependent and independent variables (Brunsdon et al., 1998).

The use of GWR in this research revealed new details in terms of spatial trends of how independent variables were associated with dependent variables. Conventional regression models such as ordinary least square regression models cannot uncover the spatial trends of the relationships between dependent and independent variables. GWR had been used in several fields, such as crime, disasters, and hazards (Light & Harris, 2012; Schmidtlein et al., 2008).

Chi et al. (2013) did a national obesity study in the United States by using GWR in order to understand the relationship between prevalence of obesity and socio-economic factors. Chi et al. (2013) used obesity rates as the dependent variable and seven explanatory variables

1. ratio of convenience- to- grocery stores,
2. ratio of fast-food-to full- service restaurants,
3. households with no car living more than 1 mile from any grocery store,
4. natural amenity score,
(5) poverty rate,

(6) percentage of white people, and

(7) urbanization.

The result from applying GWR was that higher ratios of convenience to grocery stores, poverty rate, and urban environments were strongly associated of high chances for being obese in the United States. On the other hand, areas with better physical environment were negatively associated with obesity.

Chalkias et al. (2013) used GWR as a tool in order to investigate the relationship between childhood obesity and socioeconomic variables in Athens, Greece. In their study, the dependent variable was childhood obesity, while the explanatory variables were parent’s education level, average annual household income, population density, and green-recreation cover. The results showed that parents’ education level was the highest correlated variable among the four variables. That meant lower education level was associated with the increased chances for high rates of childhood obesity. The typical formula of GWR model was shown in Figure 7 (Charlton et al., 2009). As shown in Figure 7, \( y_i(u) \) represents the dependent variable at a specific location \( u \), and \( m \) represents the independent variables at the same location \( x \), while \( \beta \) shows the relationship around the location \( u \) around the dependent variable (Chalkias et al., 2013).

\[
y_i(u) = \beta_{0i}(u) + \beta_{1i}(u)x_{1i} + \beta_{2i}(u)x_{2i} + \cdots + \beta_{mi}(u)x_{mi}
\]

*Figure 7. GWR typical equation*

Geographically weighted regression (GWR) model is a spatial regression model that can identify the locations of high or low correlation between the dependent variable and independent variables. One of the advantages of using this tool was that GWR allowed the correlations of the
variables to vary over space. GWR created a localized regression model for each feature in the data set. In the case of this research, the features were the census tracts and block groups. The dependent variable in GWR was the ratios of obesity. The independent variables included, for tracts and block groups, were

- education attainment,
- unemployment,
- median household income,
- population density per square mile of people who were above or below poverty level,
- distances to food outlets, parks, and fitness centers, distance to bus stops,
- percentage of urban and green spaces,
- average time to work, and
- number of vehicles.

All of these independent variables were assumed to be highly correlated and associated with obesity ratios, as reviewed in the literature review. All of these independent variables were included in the model in order to know the spatial relationship between the ratio of obesity and the independent variables. GWR was applied at two different geographic scales (Block groups & tracts) to find out if geographic scales affected the outcomes. The results of applying this tool was presented in maps that show the adjusted $R^2$ which helped to find whether the correlation between the variables were strong or weak. The range of $R^2$ was from 0 to 1. The closer that number was to 1, the stronger the correlation between the dependent and independent variables would be. Adjusted $R^2$ helps to distinguish how much of the variation in the dependent variable
could be explained by the variation in the independent variables. GWR was used individually for each independent variable to see what independent variable had high adjusted $R^2$ score or not.

By running the GWR for the study area, each polygon in the study area was given a value of $R^2$. The $R^2$, or the coefficient of determination, which was an index that showed how strong the relationship was between the dependent and the independent variables. The dependent-independent relationships might vary using data at different geographical scales (the so-called scale effect under the modifiable areal unit problem, MAUP) (Wong, 2009). The study used the analysis at both census tracts and census block groups in order to do comparisons between two different geographical scales. The outcome from GWR were maps that showed the distribution of regression coefficients for the independent variables. In these maps, it would be possible to visualize the geographic variation of how each independent variable influence obesity prevalence in different parts of the county, as indicated by the varying values of the regression coefficient for that variable.

**4.3.5 Extending Thomas’s nationwide model**

Several studies found that living in obesogenic environment increased the chances for being overweight or obese, especially in socioeconomically disadvantaged area (Giskes et al., 2011). An obesogenic environment was defined as an unhealthy environment where access to fresh food is not affordable for everyone or areas where there were no access to parks where people can do some physical activities (Swinburn et al., 1999; Gauthier & Krajicek, 2013). In other words, it is an environment, which may contribute to increasing the prevalence of obesity in the long term.

Thomas et al. (2014) created a dynamic model for predicting overweight, obesity, and extreme obesity prevalence trends by using countries as the unit of analysis. The aim of their
study was to predict a model based on different population parameters with changes in BMI in order to know when the prevalence of obesity would plateau (level off). Thomas et al. (2014) found that prevalence of obesity was associated with birthrate and the chances of being born in obesogenic environment.

As described in Figure 8, this dynamic model divided the population into two groups: infected (overweight, obesity, and extreme obesity) and non-infected (normal weight). The normal weight population had a body mass index below 25, overweight population have their BMI range from 25 to 30, obese population have their BMI range from 30 to 40, and the extreme obese population have their BMI greater than 40. Figure 8, shows how people move between different BMI classes. The Figure shows how people’s BMI could be changed in such environment. The first part of the cycle as shown in Figure 8, is susceptible people (BMI<25). This category is normal weight people, which had not been exposed to the social and non-social influences of obesity. Once the susceptible people were influenced by social and non-social influences, they moved to the exposed category (BMI<25). The exposed people are normal weight people; however, they are exposed to the social and non-social influences of obesity. Exposed people would have two options: (1) to move to the overweight category (25 ≤BMI< 30), (2) to move to the recovered category and stay at normal weight (BMI<25). In the overweight category, people might move to the recovered category or might move to the obese category (30 ≤BMI< 40). Once people moved to the obese category, they would also have two options: (1) move to the overweight category, (2) move to the extreme obese category (BMI< 40). The extreme obese category could just move to the obese category, then overweight category, then the recovered or exposed category. In each BMI category, there would be deaths with different rates. For example, the death rates in the advance stages of obesity (extreme obese) are higher
than lower stages of obesity and normal weight category. Thomas et al. (2014) assumed that people moved between different BMI classes at different rates. For example, normal weight people who had never been obese, their chances for being obese were less than overweight people, which was the first step for being obese. In other words, the chances for being obese for the overweight people were higher than normal weight people that had never been overweight.

![Dynamic Model for predicting the prevalence of obesity by Thomas et al. (2014)](image)

**Figure 8.** Dynamic Model for predicting the prevalence of obesity by Thomas et al. (2014)

In short, the model suggested that being born in an obesogenic environment increased the chances to be infected with obesity because of the social influences. This model helped to understand when the prevalence of obesity would plateau and in which conditions. Thomas et al.
(2014) concluded that the prevalence of obesity would plateau by about 28% for obesity, 32% overweight, and 9% for extreme, possibly in 2030 for the United States. This model was done on large geographic area (a whole country such as the entire United States) but it had not been used on smaller scales such as the county level.

The nationwide model created by Thomas et al. (2014) predicted when the prevalence of obesity would plateau by using the SIR model (i.e., Susceptible, infected, and recovered). SIR model divided the population into these infected groups, which are

1. overweight population \((25 \leq \text{BMI} < 30)\),
2. obese population \((30 \leq \text{BMI} < 40)\), and
3. extremely obese population \((\text{BMI} \geq 40)\),

while the non-infected population was the normal weight group with less than the BMI score of 25 and had never been overweight. The SIR model also described the recovered population, which was overweight population that lost weight and recovered back to normal weight. In other words, susceptible population were normal weight people with lower than the BMI score of 25, infected population are three groups were three groups, overweight, obese, and extremely obese population. Recovered subpopulation as mentioned before was those who lost weights and transitioned back from overweight to normal weight.

Thomas et al. (2014) mentioned that the social influence of being born in obesogenic environment was an important factor for predicting the prevalence of obesity. In their model, they mentioned birth rates as another predictor for obesity prevalence. People who were born in obesogenic environment (unhealthy environment) and were socially influenced by the environment would more likely become overweight or obese in the long term. Therefore, birth
rates and social influences, such as, being born in obesogenic environment, could help to predict the prevalence of obesity.

Thomas et al. (2014) created their model to predict when the prevalence of obesity would plateau by using several parameters

(1) probability of being born in obesogenic environment,
(2) birth rate,
(3) death rate of normal weight,
(4) death rate of extremely obese,
(5) transition rate from normal weight to overweight category,
(6) transition rate from overweight to obese category,
(7) transition rate from obese to extremely obese category,
(8) transition rate of weight gain in the exposed category (normal weight),
(9) social influence of overweight people,
(10) social influence of obese people,
(11) transition rate from extremely obese to obese category,
(12) transition rate from obese to overweight category, and
(13) transition rate from overweight to normal weight,

The rate at which each group moved to another category was based on using rates estimated for the national trends of the US. This study applied the model to the local neighborhoods in Summit County, Ohio. Thomas et al. (2014) concluded that by using those parameters we could predict when the prevalence of obesity would plateau. In a similar way, this research extended their model to predict the trends of obesity prevalence at neighborhood levels so that geographical disparity among the neighborhood can be revealed.
The dynamic model that was proposed by Thomas et al. (2014) was re-programmed so that the extended model could be used for any geographic units to do the simulations. At the census block group level, the extended model was used in order to find when obesity prevalence in Summit County, would plateau based on the available data. According to Thomas et al. (2014), the model of obesity prevalence was based on several assumptions. The first assumption was that the immune population (those who would never gain weight) was not included in the analysis and could be ignored. The second assumption was dividing the population into six components, which were

1. susceptible population, which represents individuals with a BMI < 25,
2. exposed population (BMI < 25),
3. overweight population (25 < BMI < 30),
4. obese population (30 < BMI < 40),
5. extremely obese population (BMI > 40), and
6. recovered population which have returned to normal weight.

The third assumption assumed that social influence of extremely obese people could not influence population weight gain. The fourth assumption assumed that the death rates of normal weight, overweight, obese, exposed, and recovered populations were lower than the death rate of extremely obese population. The fifth assumption assumed that recovered individuals were exposed to weight gain more than individuals who had never been overweight and the transition was based on social influence. The sixth assumption assumed that the law of mass action affected social interactions between populations. The seventh assumption was that a fraction of the population had higher chances for becoming obese because they were born in obeseogenic environment. The eighth assumption was that not all movements between BMI classes were
affected by social influence. The last assumption was that the population is consistent for all the years in the study and they should remain the same. Thomas et al. (2014) developed the assumptions and the formula to arrive at the obesity prevalence that was based on the following equations (Figure 9 and Appendix):

\[
S'(t) = (1 - p) \mu N(t) - DS(t) - \frac{K_1 I_1(t) S(t)}{N(t)} - \frac{K_2 I_2(t) S(t)}{N(t)} - \alpha S(t)
\]

\[
E'(t) = \rho \mu N(t) - DE(t) - \alpha E(t) + \frac{K_1 I_1(t) S(t)}{N(t)} + \frac{K_2 I_2(t) S(t)}{N(t)} + \rho R(t) + \alpha S(t)
\]

\[
I_1'(t) = -D_0 I_1(t) + \alpha E(t) - \alpha I_1(t) - \rho_1 I_1(t) + \beta_2 I_2(t)
\]

\[
I_2'(t) = -D_0 I_2(t) + \alpha_1 I_1(t) - \alpha_2 I_2(t) - \beta_1 I_1(t) + \beta_3 I_3(t)
\]

\[
I_3'(t) = -D_0 I_3(t) + \alpha_2 I_2(t) - \beta_3 I_3(t)
\]

\[
R'(t) = -DR(t) + \rho_1 I_1(t) - \rho R(t)
\]

*Figure 9. Equations of the dynamic model by Thomas et al. (2014) to arrive the obesity prevalence plateau*

Where \(S'(t)\) stands for the susceptible population (BMI<25), \(t\) the year, \((1 - p)\) probability of being born in obesogenic environment, \((\mu)\) birth rate in the susceptible population, \((D)\) death rate, \((K_1)\) social influence on overweight people, \((K_2)\) social influence on obese people, \((N)\) number of population, \((\alpha)\) rate of weight gain, \((\rho R)\), rate of recovered people, \((I_1')\) infected people, which are overweight population (30 ≤BMI<40), \((\alpha_1)\) rate of weight gain from overweight to obese, \((\rho_1)\) rate of weight loss from overweight to normal weight, \((\beta_2)\) rate of people that moved from obese to overweight, \((I_2')\) infected people, which are obese population (30≤BMI<40), \((\alpha_2)\) rate of people that moved from obese to extremely obese, \((\beta_3)\) rate of people that moved from extremely obese to obese group, \((I_3')\) infected people, which are extremely obese population (BMI≥40), and \((R)\), which represents recovered people.
The model was performed by using 13 parameters as mentioned before. The available parameters for the study area were, birth rates, death rates, and the number of people for each BMI category (normal weight, overweight, obese, extreme obese). The other parameters rates were using the national trends for the United States that had been used by Thomas et al. (2014). The model was applied at the block group level. The first and last block groups that their obesity prevalence plateaued were selected in order to find the socioeconomic variables and the built environment characteristics of those block groups.

The model was preformed several times with different birth rates to find whether birth rates could change the time, at which the prevalence of obesity would plateau. The three different birth rates were set to be:

1. 20 per 1,000
2. 30 per 1,000, and
3. 40 per 1,000

In addition, the model was applied with the same rates except for the social influence by overweight and obese. The model changed a parameter that was the social influence on overweight and obese, (it was performed with three different social influence rate):

1. Social influence rate on overweight (0.01) and obese (0.02)
2. Social influence rate on overweight (0.20) and obese (0.10)
3. Social influence rate on overweight (0.40) and obese (0.20)

The purpose for changing birth rates and social influence parameters was to figure out which parameter had more influence in affecting and changing the time at which the prevalence of obesity would plateau.
The model that was created by Thomas et al. (2014) was a nationwide model that had not been used for smaller geographic areas, such as counties, census tracts, or census block groups. In this study, the model was extended and it was applied to Summit County, Ohio at the block group level. In order to predict the prevalence of obesity at more detailed geographic levels, this research used the same parameters as used in Thomas’ model but using smaller geographic areas as the unit of analysis. The simulations were applied repeatedly for calculation of obesity prevalence from 2013 to 2023. The results showed the year when the prevalence would plateau in each neighborhood. Using localized values for the model parameters and repeating the model for each small geographic area (Block groups), the simulations in this research would give much more precise and detailed descriptions of predicting the prevalence of obesity. The dynamic model simulation of obesity coding was commissioned to Jing Mao who is a computer science student at Kent State University.

The next chapter presents the findings of this study and how the built environment features and socioeconomic variables were related to obesity, for in the youth population aged 16-24. In addition, it shows the most correlated variables to obesity prevalence in Summit County, OH. Furthermore, the results of applying the dynamic model of obesity prevalence at the block group level in Summit County, OH.
The results discussed in this chapter were calculated by using data from the US Census, Geolytics (http://www.geolytics.com/) and Ohio BMV driver licenses. In this chapter, I present and analyze the results of applying GIS techniques that were discussed in previous chapters.

5.1 Geography of Obesity In Summit County, Ohio

The obesity data used in this study was from the Ohio Bureau driver licenses. The total number of driver licenses was 408,523. By using the street addresses of the driver licenses, the data were geocoded to their geographic locations in the study area. About 98.4% (401,943) of them were geocoded and linked to their physical addresses. As seen in Table 5, the total number of geocoded points was 401,943, which was close to the total number of population in Summit County, OH. Out of the geocoded points, there were 33,484 from the age 16 to 21. Out of the 16 – 21 group, there were 3,833 obese, 7,844 overweight, and 21,373 normal weight (See Table 5). The geocoded records were then aggregated at both the census block groups and tracts. The ratio of obesity was calculated by using the following formula:

\[
\text{Ratio of Obesity} = \frac{\text{Number of obese among people aged (16 – 21)}}{\text{Total number of geocoded points in the census}} \times \text{Tract, Block group among those aged (16 – 21)} \times 1,000
\]

Figure 10. Ratio of Obesity formula

The calculation of obesity ratio was performed using the Field Calculator in ArcGIS. The calculation for obesity ratios was done just for the age group 16 – 21 because of the accuracy of
weight comparing with different driver license ages. People tended not to update their weights when they renewed their licenses. Therefore, only this age group was used in obesity analysis. The ratio of obesity calculation was done for each census tract and block group in the study area in order to see the spatial pattern of obesity whether it was clustered, dispersed, or randomly distributed. Further, to uncover the geographic disparities, the ratios of obesity were visualized at two different geographic scales (Block groups & tracts).

Table 5. Number of geocoded points in Summit County, Ohio

<table>
<thead>
<tr>
<th>BMI Category</th>
<th>All age group</th>
<th>Age 16 – 21</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal Weight (BMI &lt; 24.9)</td>
<td>171,156 (42.5%)</td>
<td>21,373 (64%)</td>
</tr>
<tr>
<td>Overweight (BMI 25 – 29.9)</td>
<td>144,207 (36%)</td>
<td>7,844 (23.5%)</td>
</tr>
<tr>
<td>Obese (BMI 30 – 39.9)</td>
<td>77,400 (19.2%)</td>
<td>3,833 (11.5%)</td>
</tr>
<tr>
<td>Extremely obese (BMI &gt; 40)</td>
<td>9,180 (2.3%)</td>
<td>434 (1%)</td>
</tr>
<tr>
<td>Total number of geocoded points</td>
<td>401,943</td>
<td>33,484</td>
</tr>
</tbody>
</table>

In Summit County, there were 31 cities and township. The largest city was Akron, which covered an area of 62.37 square miles, while the smallest township was Lakemore with an area of 1.67 square miles. Figure 11 represents the average obesity ratios for each city in Summit County. The average obesity ratio for cities in Summit County was calculated from selecting the census block groups in each city and taking the average of their obesity ratios. The obesity ratio was used in order to compare different cities. However, some cities, such as Clinton and Richfield had just one census block group; therefore, the average obesity ratio was not selected. The Clinton city was excluded from the comparison because it had just one census block group and there was no average. For the other cities, the average was calculated from the selected
census block groups that fall inside each city. As mentioned before, the obesity ratio was calculated at the census block groups for all of the study area. The highest top five cities in terms of obesity ratio are Lakemore, Barberton, Akron, Norton, and North Field while lowest obesity ratio is that of Bath Township.

![Average Obesity Ratio by Cities in Summit County, Ohio](image)

*Figure 11. Average obesity ratio for Summit County, Ohio cities*

Figures 13 and 14 show the spatial distribution of obesity ratios by the census tracts and by block groups for the age group 16 to 21. As shown in Figures 13 and 14, the spatial pattern of the ratio of obesity was mostly concentrated at the center of the city of Akron. In Summit County, Ohio, there were 135 census tracts and 452 census block groups. As seen in Figures 13 and 14, darker areas represented areas with high obesity ratios while brighter areas represented areas with lower obesity ratios. At the census tracts level, there were 5 census tracts that had a ratio above 220 per 1,000 people, while at the census block group level there were 38 block...
groups meeting that criterion (Table 6). Table 6 shows the differences between obesity ratios in both the tract and block group levels at five different obesity categories as shown in Figures 13 and 14. As seen in Table 6, applying the same category classification of obesity ratios showed some differences between two geographic scales in terms of the number of geographic units with obese prevalence. The five categories were chosen based on the values of obesity ratios for both tracts and block groups (Figures 13, 14, and Table 6). They were two significant differences between outcomes as applied by census tracts and by census block groups. Applying the same calculation at both different scales might be misleading especially when dealing with larger geographic areas such as tracts.

Table 6. Comparison between obesity ratio between census tract and block group

<table>
<thead>
<tr>
<th>Ratio of Obesity Per, 1000</th>
<th>Tract</th>
<th>Block Group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td># of tracts:</td>
<td># of obese:</td>
</tr>
<tr>
<td>&lt; 100</td>
<td>45</td>
<td>868</td>
</tr>
<tr>
<td>100.1 – 140</td>
<td>33</td>
<td>944</td>
</tr>
<tr>
<td>140.1 – 180</td>
<td>33</td>
<td>1263</td>
</tr>
<tr>
<td>180.1 – 220</td>
<td>20</td>
<td>868</td>
</tr>
<tr>
<td>&gt;200</td>
<td>5</td>
<td>172</td>
</tr>
</tbody>
</table>

Most of the high obesity ratios areas were inside the city of Akron, Ohio. At the block group level, the spatial pattern of obesity ratios was more obvious than that at the tracts level. This might be due to the level of data aggregation at the tract level. However, at the block group level the results showed more details about the locations of high obesity ratios. In addition, the smaller size of the block group gave more details about the spatial distribution of the obesity ratio.
Spatial autocorrelation tool (Morans I) was one of the useful tools that had been applied to analyzing the spatial pattern of obesity ratios. Calculated coefficient values for spatial autocorrelation helped to distinguish the different degrees of clustering in the spatial patterns of obesity. They allowed us to see if obesity ratios were spatially clustered or dispersed. The result from applying this tool for obesity ratio showed a high degree of clustering especially at the census block group (Figure 12). The z-score at the tract level was 19.71 while at the block group level was 21.00. In other words, both geographic levels showed a high degree of spatial clustering in terms of obesity ratios.

Figure 12. Moran’s I Index of Obesity Ratio in Summit County, Ohio
Figure 13. Spatial distribution of obesity ratio in Summit County, Ohio at the block group level
Figure 14. Spatial distribution of obesity ratio in Summit County, Ohio at the tract level
Figures 15 and 16 showed the result of applying hot spot analysis at both census tract and block group levels. Red areas represented the hot spot locations, while blue areas were the locations of cold spots of obesity ratios. Yellow areas were the locations with insignificant confidence levels, which meant that it was a mixture of neighboring high obesity ratio values and low values. In other words, red areas showed the location of high values of obesity ratio that were surrounded by high values of obesity ratio, while blue areas were the locations of the clusters of low values of obesity ratio that are surrounded by low values. As seen in Figures 15 and 16, the hot spot at both scales were located in the city of Akron, while the cold spots were located at the north side of Akron. At the tract level, the distribution of hot spots were obvious by the agglomerated tracts in red and blue. However, at the block group level, hot/cold spots were given with much more details than what we could see by tracts. At the block group level, hot spots were seen in City of Akron and the southwestern parts of Summit County that connected to Akron. The cold spots at both scales were almost the same. However, at the block group there were some cold spots that were located in areas south of Akron, which were not revealed at tract level. The distribution of obesity ratio was mostly located in the city of Akron while other cities show low values of obesity ratio.
Figure 15. Hot spot analysis of obesity ratio at the block group level in Summit County, Ohio
Figure 16. Hot spot analysis of obesity ratio at the tract level in Summit County, Ohio
5.1.1 Ethnicity and Obesity in Summit County, Ohio

In Summit County, the majority of the population were white and they accounted for about 80% of the total population. African Americans were the second largest ethnic group and the largest minority group. They represented about 14.6% of the total population in Summit County. Other ethnic groups, such as Asian, Hispanic, and others together accounted for about 5%. Figure 17 and 18 show the spatial distribution of white and African Americans in Summit County, Ohio at the level of census block groups. The percentages of White and African Americans were calculated by dividing the total number of each ethnic group with the total population in each census block group and multiply it by 100. The following example shows how the percentage calculation was performed:

\[
\frac{\text{Number of African Americans in census block group #391535088001}}{\text{Total population in census block group #391535088001}} \times 100
\]

*Figure 17. African American calculation formula*

As seen from Figure 18, White population had spread over almost all of the Summit County, except in Akron where the majority of African American lived. Census block groups that had a percentage above 70% of White and African Americans were selected in order to see whether high concentration of the two ethnic groups (White & African American) were associated with high obesity ratios. Table 7 shows statistics of those two ethnic groups and obesity ratios:

Table 7. Average obesity ratio and number of obese people for White and African American at the census block group level

<table>
<thead>
<tr>
<th></th>
<th>White &gt; 70%</th>
<th>African American &gt; 70%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ratio of obesity (Average)</td>
<td>144 per 1,000</td>
<td>269.6 per 1,000</td>
</tr>
<tr>
<td>Number of obese population</td>
<td>2,722</td>
<td>484</td>
</tr>
</tbody>
</table>
In Table 7, census block groups with African Americans accounting for more than 70% of the population had higher ratios of obesity than the census block groups with more than 70% of their population being white. The average obesity ratio for high concentration census block groups of African Americans (>70%) is 269.6 per 1,000 while for the average obesity ratio from census block groups that had majority of population being white was 144 per 1,000. The total number of obese people in all White block groups was 2,722 while for all African American block groups the number of obese people was 484.

When looking at obesity ratios, however, African American block groups did have higher obesity ratios than those in white block groups. Even though African Americans were minorities in Summit County, they had a higher obesity ratio than white population. Using the average ratio of obesity was better than just counting the number of obese people in the selected census block groups. For example, using just the number of obese people did not reflect the percentages of obesity in the selected census block groups.

Obesity in general could be related to some traits of ethnic groups than others because of the cultural differences in terms of the types of food each ethnic group consumed. In addition, it could be related to the socioeconomic status of each ethnic group in Summit County. Furthermore, the built environment could be one of the factors that might affect what types of food each group could ate because of the affordable options of food in their neighborhood and physical environment features, such as access to parks, and the availability of sidewalks. All of these factors could help to explain why some ethnic groups had higher obesity ratios than others.
Figure 18. Spatial distribution of White population in Summit County, Ohio by block group
Figure 19. Spatial distribution of African American in Summit County, Ohio by block group.
5.1.2 Socioeconomic Variables, Land Cover, and Obesity in Summit County, Ohio

Socioeconomic variables were among the influential factors that could affect the prevalence of obesity in an area (Gupta et al., 2012). The socioeconomic data was downloaded from American Community Survey that included, education, employment, median family income, and the number of people below and above poverty level. In addition, land cover data, which was categorized as urban versus green space, was also included in the analysis. Figure 20 shows the spatial distribution of urban area versus green space in the study area. The urban area in Summit County covered an area of 116 square miles that accounted for 27% of the total area of Summit County. The green space covered an area of 303 square miles that accounted for 72% of Summit County. The percentages of urban and green space were calculated for both the block group and tract levels.

The ten block groups and ten tracts that had the highest obesity ratios among census block groups and tracts were selected for additional comparison. Table 8 shows the statistics of socioeconomic variables calculated for the selected block groups and tracts. The education variable in Table 8 was the percentage of people that held bachelor degrees or higher. The second variable was the percentage of unemployment. The third variable was the median family income. The fourth and fifth variables were the density per square mile for number of people below and above poverty level. The last two variables were the percentages of urban and green space. The average of each variable was used in order to compare it with the top ten obesity ratios at the both different geographical scales. After the top ten obesity ratio neighborhoods (Block groups & tracts) were identified, the average of the top ten census units were used in order to do the comparison between the two different geographical scales. For the first variable, the education was 7.12% at the block group and 7.6% at the tract level. Both were lower than the
overall average of the entire censuses in the study area. The second variable, which represented unemployment ratios, showed that the top ten block groups and top ten tracts all had much higher obesity ratios than all block groups or all tracts. This suggested that unemployment was a very important factor in affecting obesity ratios. The third variable was median family income, which was an important socioeconomic variable. The average median family income for the top ten block groups was $27,761, which was significantly lower than the overall average of Summit County of, $60,315. At the tract level, median family income at the top ten tracts with the highest obesity ratios was $29,586 while the overall median family income at the tract level was $60,524. As shown in Table 8 median family income was lower in areas with high obesity ratios comparing with the overall average obesity ratio.

Another socioeconomic variable was the densities of people that were below or above poverty level per square mile. At the top ten census block groups, the number of people that were below poverty level was 2,245 per square mile, comparing with, 2,077 people per square mile that were above the poverty level. This indicated that there were more people below poverty level who were obese than those above poverty level who were obese at the top ten block groups with the highest obesity ratios. At the tract level, the results were different. In terms of the number of people per square mile that were below poverty level, there were 1,888 people while the number of people per square mile that were above poverty level was 5,533. The results at the tract level showed that the geographic units of analysis did make a significant difference.

Table 8 also shows the results by urban versus non-urban areas. There were more block groups and tracts inside Akron City than outside the city, 76% at the block group level and 84% at the tract level. The last variable was by the percentages of green spaces. For both block groups and tracts, there were less green space in areas of the highest obesity ratios than the overall green
space, 23% in the selected block groups versus 40% among all block groups and 16% for the selected tracts as opposed to 47% among all tracts. This indicated that the top ten obesity ratio neighborhoods (Block groups & tracts) were more urban and with less green space. The results that were shown in Table 8 indicated that socioeconomic variables at the top ten census units (block groups & tracts) with highest obesity ratios were generally those with lower socioeconomic status (more unemployment rates, less educated, lower income) and more urban with less green space.

Table 8. Socioeconomic variables and land cover percentage for the top ten obesity ratio at the census block group and tract

<table>
<thead>
<tr>
<th></th>
<th>Block group (Top 10)</th>
<th>All block group</th>
<th>Tract (Top 10)</th>
<th>All tract</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Education (%)</strong></td>
<td>7.12</td>
<td>26</td>
<td>7.6</td>
<td>26</td>
</tr>
<tr>
<td><strong>Unemployment (%)</strong></td>
<td>21</td>
<td>11</td>
<td>23.15</td>
<td>11</td>
</tr>
<tr>
<td><strong>Median Family Income</strong></td>
<td>$27,761</td>
<td>$60,315</td>
<td>$29,586</td>
<td>$60,524</td>
</tr>
<tr>
<td><strong>Below Poverty Level (Density Per square mile)</strong></td>
<td>2,245</td>
<td>1,026</td>
<td>1,888</td>
<td>740</td>
</tr>
<tr>
<td><strong>Above Poverty Level (Density Per square mile)</strong></td>
<td>2,077</td>
<td>2,616</td>
<td>5,533</td>
<td>1,814</td>
</tr>
<tr>
<td><strong>Urban (%)</strong></td>
<td>76</td>
<td>56</td>
<td>84</td>
<td>53</td>
</tr>
<tr>
<td><strong>Green spaces</strong></td>
<td>23</td>
<td>40</td>
<td>16</td>
<td>47</td>
</tr>
</tbody>
</table>
Figure 20. Urban versus green spaces in Summit County, Ohio
Figure 21 shows the results of applying Cluster and Outlier analysis on obesity ratios at the block group level. Black areas represented concentrations of high-high values (hot spots), blue areas show the concentration of low-low values (cold spots), yellow areas represented areas with mixed of high-low values, white areas represented areas with low-high values, while gray areas were areas with no significant pattern. The clusters of high values of obesity were located in the center and the southwestern side of the city of Akron. There were 88 census block groups, which could be considered for the category of high-high values. While the clusters of low-low obesity ratios were located at the northeastern and the western sides in Summit County. They accounted for 46 census block groups. The rest of the map shows areas without significant clustering of obesity ratios. They were represented in gray color (Figure 21).

Figure 22 shows the results of applying the Cluster and Outlier analysis at the tract level. As can be seen in Figure 22, the distribution of clusters and outlier as shown by tracts looked to be the same as those at the block group level. However, at the block group level, the spatial pattern was clearer than that at tract level because of the geographic differences between them in terms of the size and shapes of the census units. At the tract level, there were 33 tracts of high values clusters and 21 tracts of low-low values (See Figure 22). The clusters of high and low clusters values of obesity ratio were selected in order to identify their socioeconomic variables.

Table 9 shows the socioeconomic variables of high concentration clusters and low concentrations value clusters of obesity ratio. As seen in this table, the neighborhoods with high-high values of obesity ratio at both geographical scales (Block groups & tracts) had a lower education level comparing with the second category, which has low-low values of obesity ratios. In the education variable, the average education level was 9.18% in the block groups and 9.03% at the tract level. While at the low-low values category, the percentage of education was 52.35%
at the block group and 50.1% at the tract. This indicated that education was inversely associated with obesity ratios. In other words, high concentration of obesity ratios were associated with low education levels.

The second socioeconomic variable, unemployment percentage in the first category (High-high) was 18.01% at the block group and 19.5% at the tract level. While the second category (Low-low), the average unemployment percentage was 5.4% at the block group and 4.30% at the tract. This indicated that high percentages of unemployment were associated with low levels of obesity ratio and lower unemployment percentages were associated with high obesity ratios.

The third socioeconomic variable was median family income. At the high-high clusters category, the median family income at the block group level was $35,297 and $33,075 at the tract level. While at the low-low cluster category, the median family income was $103,812 at the block group level and $102,737 at the tracts. This indicated that high-high clusters of obesity ratios were associated with low median family income.

The last variables were the numbers of people per square mile that were below or above the poverty level. The variation was significant especially for obesity ratios of the people at below the poverty level. There were just 91 such people per square mile at the low-low clusters of obesity ratios comparing with 1,942 per square mile at the high-high values clusters of obesity ratios. The same results also were true at the tract level. There were 1,540 such people per square mile at the high-high clusters category while at the low-low clusters there were only 212 such people per square mile.

For the last socioeconomic variable, the results showed that, at the block group level, there were 3,072 people per square mile who were above the poverty level in areas of high
concentration of obesity ratios. This could be compared with 1,623 people per square mile at the low-low concentration of obesity ratios. While at the tract level, the results were opposite. At the high-high clusters of obesity ratios, there were 3,664 people per square mile while at the low-low clusters of obesity ratios there were 4,429 people per square mile. This could be related to the level of aggregation of at the tract level, which could change the results.

The percentage of urban areas at the high-high category at the block group level was 82%, comparing with 79% at the tract level. At the low-low category the percentage of urban areas at the block group level was 31% while at the tract level it was 24%. The last variable was the percentages of green space. At the block group level, the percentage of green space at the high-high category was 18%, comparing with 21% at the tract level. At the low-low category, the results showed higher percentages of green space, which was 69% at the block group level and 75.5% at the tract level. This indicated that high-high concentrations of obesity ratios were associated with high percentages of urban areas in the neighborhoods and low percentages of green space. On the other hand, concentrations of low-low values of obesity ratio were associated with low urban areas and high percentages of green space in the neighborhoods.

Socioeconomic variables were one of the factors that could be associated with obesity ratios. In addition, the percentages of urban and green space in association with obesity ratios in terms of high-high and low-low concentrations of obesity ratios. The next section discusses the built environment, the physical environment features, and how they were related to obesity quantitatively.
Table 9. Socioeconomic variables and land cover of the high and low clusters of obesity ratio at the block group and tract level

<table>
<thead>
<tr>
<th></th>
<th>Block group (High-High)</th>
<th>Block group (Low-Low)</th>
<th>Tract (High-High)</th>
<th>Tract (Low-Low)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Education (%)</strong></td>
<td>9.18</td>
<td>52.35</td>
<td>9.03</td>
<td>50.1</td>
</tr>
<tr>
<td><strong>Unemployment (%)</strong></td>
<td>18.01</td>
<td>5.40</td>
<td>19.5</td>
<td>4.30</td>
</tr>
<tr>
<td><strong>Median Family Income</strong></td>
<td>$35,297</td>
<td>$103,812</td>
<td>$33,075</td>
<td>$102,737</td>
</tr>
<tr>
<td><strong>Below Poverty Level</strong></td>
<td>1,942</td>
<td>91</td>
<td>1,540</td>
<td>212</td>
</tr>
<tr>
<td>(Density per square mile)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Above Poverty Level</strong></td>
<td>3,072</td>
<td>1,623</td>
<td>3,664</td>
<td>4,429</td>
</tr>
<tr>
<td>(Density per square mile)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Urban (%)</strong></td>
<td>82</td>
<td>31</td>
<td>79</td>
<td>24</td>
</tr>
<tr>
<td><strong>Green Spaces (%)</strong></td>
<td>18</td>
<td>69</td>
<td>21</td>
<td>75.5</td>
</tr>
</tbody>
</table>
Figure 21. Cluster and outlier analysis of obesity ratio at the block group level
Figure 22. Cluster and outlier analysis of obesity ratio at the tract level
5.2 Built Environment Features and Obesity in Summit County, Ohio

Built environment was one of the associated factors that were related to obesity ratios. Built environment included, access to food outlets (Fresh vs non-fresh), parks, open green spaces, fitness centers, bus stops, street connectivity, and other features. In this study, data for several of physical environment features were available. Different tools in ArcGIS were used in the built environment features and obesity ratio:

5.2.1 Network Analyst (Service Area)

This tool was used to measure the walkability index. Service area tool calculated the distance around the feature such as food outlet by using the road network. This tool was applied around several physical environment features such as, grocery stores, parks, bus stops, and service plazas. As mentioned in the methodology section, the service area distances around the four features (grocery stores, parks, bus stops, and service plazas) were set to be 0.5 mile inside the urban areas such as Akron, Ohio, and 1.5 miles in sub-urban cities. This analysis was carried out at the census block groups level. The census block groups were selected based on whether the service area around the four features covered the centroid of the block group that it fell in. In other words, if the centroid of the block group fell inside the service area then the entire block group was considered as having access to the features (grocery stores, parks, bus stops, service plazas). Figure 23 shows the results of applying service area tool and the walkability index. The index in the map ranged from zero to four. Areas where low walkability index values were presented in yellow and darker areas were those with high walkability index values. Akron had just a few areas with high walkability index values comparing with the surrounding sub-urban areas, such as Cuyahoga Falls at the north east of Akron, Fairlawn at the north west of Akron,
Barberton at the west south of Akron, and also the east side of Akron where Tallmadge is located. Table 10 shows the statistics of the average obesity ratio, number of obese people, and walkability index scores.

Table 10. Walkability index score, average obesity ratio, and number of obese people by census block group in Summit County, Ohio

<table>
<thead>
<tr>
<th>Walkability score</th>
<th>Average obesity ratio</th>
<th>Number of obese people</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 0.10</td>
<td>112.18</td>
<td>1,286</td>
</tr>
<tr>
<td>0.10 – 0.25</td>
<td>139.43</td>
<td>1,036</td>
</tr>
<tr>
<td>0.26 – 0.50</td>
<td>144.86</td>
<td>849</td>
</tr>
<tr>
<td>0.51 – 0.75</td>
<td>128.67</td>
<td>342</td>
</tr>
<tr>
<td>&gt; 0.75</td>
<td>126.06</td>
<td>594</td>
</tr>
</tbody>
</table>

As seen in Table 10, the lowest average obesity ratio was in sub-urban areas, which was 112.18 per 1,000. That obesity ratio occurred in a neighborhood with a walkability score of less than 0.10. While the highest average obesity ratio was at the walkability score of 0.5 (in the category of 0.26 – 0.50). That neighborhood had an obesity ratio of 144.86 per 1,000. For the number of obese people in neighborhood, the highest numbers of obese people were at the lowest walkability score, which was 1,286. They were shown in Figure 22 in the bright yellow color. The smallest number of obese people in a neighborhood was 342. That was with walkability score of 0.75 (0.51 – 0.75). What we could conclude from Table 10 was that neighborhoods with lower walkability scores tended to have higher obesity ratios and with more obese population. In addition, neighborhoods, with higher scores for walkability tended to have fewer obese people and lower obesity ratios.
Figure 23. Results of Walkability index in Summit County, Ohio
5.2.2 Network Analyst (Closest Facility)

Closest Facility tool in ArcGIS/Network Analyst was used to calculate the distances from census block group centroids to all of the facilities included in the study. These facilities were the available built environment features in the study area such as, food outlets, parks, fitness center, and bus stops. In order to see whether these features were associated with obesity ratios in different neighborhoods.

The Service Area tool calculated distances to neighborhoods (block groups) around each included facility. As opposed to that, the Closest Facility tool calculated the shortest distance between two features. Both of the two tools used the road network to calculate the distances. The centroids of block groups were used as the starting points when calculating the distances from each block group to the nearest physical environment features. This tool was used at just the block group level.

After calculating the distances between each block group centroid to different physical environment features, each block group was assigned a distance to the closest physical environment features. The calculated distances were used to assess whether the distances affected the obesity ratios. Specifically, shorter such distances implied more opportunities for physical in the neighborhoods and consequently obesity ratios.

In addition, to figure out the correlated features that were related to obesity ratios. Figure 24 shows the spatial distribution of grocery outlets, which represented the fresh food outlets. The grocery stores represented in the map were all the grocery stores, which included small and large chain supermarkets. In the study area, there were 110 grocery stores. Most of them were located in the central parts of Akron and in the north side of Summit County, OH. In Figure 24, dark
brown census block groups represented longer distances from their centroids to grocery stores while bright yellow areas represented closer distances to grocery stores.

Figure 25 shows the spatial distribution of fast food outlets and the distances from centroids of block groups to the closest outlets. According to the downloaded data from A to Z database website (http://www.atozdatabases.com), there were 334 fast food outlets in the study area. Fast food outlets were located in most of the urban area in Summit County, especially in the city of Akron (Figure 25). Comparing the distances between Figures 24 and 25, it was obvious that there were more neighborhoods that were closer to fast food outlets than to grocery stores.

The third physical environment feature was parks. In the study area, there were 130 parks located in Akron, north and south side of Summit County, (Figure 26). It was clear that there were enough access from different neighborhoods to public parks. Figure 26 shows the distribution of public parks and the distances to each neighborhood. Few census block groups were within 2 miles to any parks. Figure 27 shows the spatial distribution of fitness centers and the distances to them from each block group. In the study area, there were 103 fitness centers and most of them were located in Akron and the north and south sides of Summit County. The last physical environment feature examined was bus stops. In the study area, there were 121 bus stops and most of them were in the City of Akron and the cities surrounded of Akron (Figure 28).

Using the calculated distances from each census block group to its nearest physical environment features, the top ten census block group of obesity ratio were selected in order to see whether the distances between block groups to physical environment features are correlated with obesity ratios in the block groups. Table 11 shows the results of this correlation. According
to Table 11, the closest physical environment feature in the ten block groups (with the highest obesity ratios) was bus stops. The average distance from centroids of block groups to bus stops was 0.56 mile in the top ten block groups with the highest obesity ratios. Fast food outlets were found to be in shorter distances to centroids of the block groups with the highest obesity ratios than other physical environment features. This indicated that bus stops and fast food outlets were closer to neighborhood (block groups) centroids that have high obesity ratios.

Table 11. Average distance between top ten obesity ratio and grocery stores, fast food, parks, fitness centers, and bus stops

<table>
<thead>
<tr>
<th>Facility</th>
<th>Distance</th>
<th>Maximum</th>
<th>Minimum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grocery stores</td>
<td>0.61</td>
<td>1.68</td>
<td>0.11</td>
</tr>
<tr>
<td>Fast Food</td>
<td>0.58</td>
<td>1.19</td>
<td>0.11</td>
</tr>
<tr>
<td>Parks</td>
<td>0.77</td>
<td>1.33</td>
<td>0.23</td>
</tr>
<tr>
<td>Fitness centers</td>
<td>1.09</td>
<td>2.4</td>
<td>0.20</td>
</tr>
<tr>
<td>Bus stops</td>
<td>0.56</td>
<td>1.29</td>
<td>0.16</td>
</tr>
</tbody>
</table>

Closest Facility tool was one of the most useful tools in GIS technology because it could measure the geographical access by calculating the distances along streets in a road network rather than just calculating the distance between two features using an Euclidean straight line. Closest Facility tool provided the distances from each block group to their closest physical environment features. Such calculated distances were one of the explanatory variables in the geographically weighted regression models discussed in the next section.
Figure 24. Distance to grocery stores in Summit County, Ohio at the census block group level.
Distance to Fast Food In Summit County, OH At the Block Group level

Figure 25. Distance to fast food in Summit County, Ohio at the census block group level
Figure 26. Distance to parks in Summit County, Ohio at the census block group level.
Figure 27. Distance to fitness centers in Summit County, Ohio at the census block group level
Figure 28. Distance to bus stops in Summit County, Ohio at the census block group level
5.3 Results of Applying Geographically Weighted Regression

GWR as a tool is a useful one for investigating the spatial variation between dependent and explanatory variables. This tool shows the locations of the strongest or weakest relationship between variables over space. Table 12 shows the results of adjusted $R^2$ for each independent variable and the only dependent variable, which was obesity ratios. As can be seen from Table 12, the adjusted $R^2$ at census tracts level were higher than those of the block groups. This was probably due to the aggregation of smaller geographic units (block groups) to larger units (tracts), which caused the detailed variations among smaller units to be lost. This issue had been well documented in geography literature as the modifiable unit problem, or MAUP.

The highest adjusted $R^2$ at the tract level was at the education attainment variable, which was 0.62 and 0.35 at the block group level. The lowest adjusted $R^2$ value was in the number of vehicles, which was 0.45 at the tract level and 0.29 at the block group level. Applying GWR for each independent variable, the results helped to figure out the independent variables that had the strongest associations with obesity ratios in the study area, and where they are located.

At the tract level, the correlation was obvious from some of the variables because there were some results that were above 0.50, which showed how strong the correlation between the dependent and independent variables was. Another important thing after distinguishing the strength of the correlation between each independent variable and the dependent variable was the direction of the correlation; whether it was a positive or a negative correlation. Positive correlation implied that once the values in the dependent variable increased, the values in the independent variable increased at the same time. Negative correlation implied that values in dependent and independent variables change in opposite directions. The correlation between the
dependent variable could be strong whether it was negative or positive based on the results of correlation analysis.

Table 12. Results of adjusted $R^2$ and each independent variable

<table>
<thead>
<tr>
<th>Independent variables</th>
<th>Tract Adjusted $R^2$</th>
<th>Block group Adjusted $R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Education attainment</td>
<td>0.62</td>
<td>0.35</td>
</tr>
<tr>
<td>Unemployment</td>
<td>0.57</td>
<td>0.31</td>
</tr>
<tr>
<td>Median household income</td>
<td>0.61</td>
<td>0.35</td>
</tr>
<tr>
<td>Density of people below poverty level per square mile</td>
<td>0.52</td>
<td>0.32</td>
</tr>
<tr>
<td>Density of people above poverty level per square mile</td>
<td>0.51</td>
<td>0.30</td>
</tr>
<tr>
<td>Road density</td>
<td>0.52</td>
<td>0.30</td>
</tr>
<tr>
<td>Distance to grocery stores</td>
<td>0.50</td>
<td>0.29</td>
</tr>
<tr>
<td>Distance to fast food stores</td>
<td>0.48</td>
<td>0.31</td>
</tr>
<tr>
<td>Distance to all type of restaurants</td>
<td>0.55</td>
<td>0.32</td>
</tr>
<tr>
<td>Distance to parks</td>
<td>0.45</td>
<td>0.30</td>
</tr>
<tr>
<td>Distance to fitness centers</td>
<td>0.45</td>
<td>0.31</td>
</tr>
<tr>
<td>Distance to bus stops</td>
<td>0.50</td>
<td>0.31</td>
</tr>
<tr>
<td>Percentage of urban</td>
<td>0.50</td>
<td>0.30</td>
</tr>
<tr>
<td>Percentage of green spaces</td>
<td>0.50</td>
<td>0.32</td>
</tr>
<tr>
<td>Number of vehicles</td>
<td>0.45</td>
<td>0.29</td>
</tr>
<tr>
<td>Time to work (Below 30 min)</td>
<td>0.45</td>
<td>0.29</td>
</tr>
<tr>
<td>Time to work (Above 30 min)</td>
<td>0.45</td>
<td>0.34</td>
</tr>
</tbody>
</table>

Figure 29 shows the direction of the correlation between obesity ratios and education attainment at tract and at block group levels. At the tract level, the direction of the correlation was more obvious and clearer than that in block group level. However, both of them had the same direction of correlation, which was a negative correlation. That implied, the higher obesity
ratios were, the lower education attainment would be. High education attainment value was correlated with low obesity ratio.

Figure 30 shows the same negative correlation between obesity ratios and median family income. That suggested that the higher the median family income was, the lower obesity ratio could be. Areas with high obesity ratios typically had lower median family income. At the census tract level, the correlation was obvious and stronger than block group.

Figure 31 shows the relationship between obesity ratios and unemployment percentages at the tract and block group levels. The direction of the correlation was positive. The relationship showed that the neighborhoods with high obesity ratio were associated with high percentage of unemployment, while neighborhoods with low obesity ratio were associated with lower percentage of unemployment. The relationship was also stronger at the tract level than at block group level. The correlations between obesity ratios and socioeconomic variables, such as unemployment, education, median family income, and unemployment were stronger comparing with those with other variables. Exploring the direction and correlation between dependent variable and independent variables by using the graphs that shows the direction of the correlation was an important step before applying the GWR tool.
Figure 29. Relationship between ratio of obesity and education attainment at tract and block group level
Figure 30. Relationship between ratio of obesity and median family income at tract and block group level
Figure 31. Relationship between ratio of obesity and Unemployment at tract and block group level
After applying GWR for each independent variable, it was used in two additional models. The first model included socioeconomic variables and the second model included the built environment features, both used obesity ratios as the dependent variable. The socioeconomic variables were based on the available data that were gathered from different sources, which included education attainment, median family income, percentage of unemployment, and the density per square mile for people that were below or above the poverty level. For the second model, the built environment variables were based on the distances calculated around food outlets, parks, road density, and percentages of urban and green space. The model were applied at two different scales (Block groups & tracts) with the same variables to find the geographic disparities at the two different scales. Table 13, explains what were the selected independent variables for each model.

Table 13. Independent variables for the two GWR models

<table>
<thead>
<tr>
<th>Socioeconomic variables</th>
<th>Built environment variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Education attainment</td>
<td>- Distance to all food outlets</td>
</tr>
<tr>
<td>- Median Family income</td>
<td>- Distance to parks</td>
</tr>
<tr>
<td>- Unemployment</td>
<td>- Road density</td>
</tr>
<tr>
<td>- Density of people below poverty level per square mile</td>
<td>- Percentage of urban</td>
</tr>
<tr>
<td>- Density of people above poverty level per square mile</td>
<td>- Percentage of green spaces</td>
</tr>
</tbody>
</table>

To assess the outcomes from two GWR models (1), the spatial distributions of the residuals from the two models were checked to see if there was any spatial dependency. The spatial distribution of the residuals were randomly distributed for both models except in the
second model at the block groups level where the spatial distribution of residuals were clustered (Figure 32, 33, 34, and 35).

Figure 32. Spatial distribution of residuals for GWR model (1) at the tract level
Figure 33. Spatial distribution of residuals for GWR model (1) at the block group level

Given the z-score of -0.411933755385, the pattern does not appear to be significantly different than random.
Given the z-score of -0.138992750894, the pattern does not appear to be significantly different than random.
Figure 34. Spatial distribution of residuals for GWR model (2) at the tract level

Given the z-score of 2.26102875451, there is a less than 5% likelihood that this clustered pattern could be the result of random chance.

Figure 35. Spatial distribution of residuals for GWR model (2) at the block group level
Figures 36 and 37 show the results of applying GWR in the first model (with the socioeconomic variables and obesity ratios) at the tracts and block groups levels. For the first model, the adjusted $R^2$ at the tract level was 0.67, which was higher than in block groups 0.41. The adjusted $R^2$ suggested that 67% of the variation in the dependent variable could be explained by the variations in the independent variable at the tract level. As seen in Figures 36 and 37, blue areas represented the locations with low localized $R^2$ values while red areas represented with-high values of $R^2$. In other words, red areas showed stronger associations between the dependent variable, which was obesity ratio, and the independent variables, which were socioeconomic variables.

In the first model and at the tract level, some areas, such as the central and south eastern parts of Akron show only weak associations between the dependent and independent variables when compared with that of the northeastern part of Akron. At the tract level, the spatial pattern of localized $R^2$ showed more areas with low correlation especially in the central part of Akron city and the southern part of Summit County. While northeastern side of Summit County showed a stronger association between the variables. In the first model, the adjusted $R^2$ at the tract level was higher than that at the block group level. However, at the block group level the pattern of localized $R^2$ showed more details than that at the tract level because of the geographic sizes differences between the two different geographical scales (Figures 36 & 37).
In the second model, the adjusted $R^2$ at both tract and block group levels were lower than those in the first model. They were 0.55 at the tract level and 0.23 at the block group level. Figure 38 shows the distribution of localized $R^2$ values at the tract level. Almost the entire city of Akron had shown a weaker association between the dependent variable and independent variables. While tracts in south side of Summit County, had the strongest association when compared with those of tracts in other areas. At the block group level the spatial distribution of localized $R^2$ areas in the southeast and west of Akron, had weaker association when compared with the east side of Akron, west and south side of Summit County, where there were strong associations between the variables (Figure 39). However, the results in the second model at the block group level could not be trusted because the spatial distribution of the residuals was clustered. From the results of model (1) and (2), it was clear that the socioeconomic variables were more associated and related with obesity ratios in the first model, when compared with the results from the second model.
Figure 36. GWR results of model (1) at the census tract level
Results of GWR (Model 1) At The Bgroup level
Adjusted $R^2 = 0.40$

Figure 37. GWR results of model (1) at the block group level
Figure 38. GWR results of model (2) at the census tract level
Figure 39. GWR results of model (2) at the census block group level
5.4 Extension of Thomas Dynamic Model

The dynamic model that was created by Thomas et al. (2014) was applied at each census block group individually in the study area. The extended model was based on the available data such as the prevalence of obesity (i.e., the ratios of population classified as normal weight, overweight, obese, and extremely obese) in each block group. The other variables such as the birth and death rate were obtained from Ohio Department of Health. The birth rate in Summit County was 11.3 per 1,000 while the death rate was 5.16 per 1,000. The other model parameters values were set based on the assumptions (listed below) as used in Thomas model et al. (2014). The dynamic model was extended and applied for each census block group, to find the disparities of obesity ratios among block groups and to find when the prevalence of obesity would plateau in these block groups.

For each census block group, the model was performed using several parameters as shown in Figure 40:

(1) Probability of being born in obeseogenic environment, which was based on the percentage of reproductive age female who were overweight or obese.

(2) Death rate of obese and extremely obese people, which were set to be higher than the death rate of people of normal weight.

(3) Social influence of overweight with other BMI categories.

(4) Social influence on obese, which was lower than the rate of social influence of overweight.

(5) Rate of weight gain at the exposed, overweight, obese, and extremely obese categories.

(6) Rate of weight recovered from different BMI classes, obese to overweight and overweight to normal weight.
(7) Birth and death rates.

All the parameters rates used were the same as Thomas dynamic model except, the birth rates, death rates, and the number of normal weight, obese, extreme obese people, which was available at the block group level. As shown in Figure 40, parameters that were highlighted in red, were those that had changed.

The extended dynamic model was applied from the year 2013 to 2023 to see what areas based on the available data would plateau first in terms of obesity prevalence. Figure 41 shows the results of applying the dynamic model at the census block group level. Red areas represented the neighborhoods where their obesity prevalence had plateaued. Blue areas show the neighborhoods where obesity prevalence would occur later in time. By year 2015 (using 2013 data), some of the census block groups had their obesity prevalence plateaued (shown in red). By 2016, there would be more block groups having their obesity prevalence plateaued, especially at the north and east north, and the south side of Summit County. By 2017, more than half of the census block groups would have turned red, indicating that their obesity prevalence would plateau. By 2018 to 2021, the entire census block groups would turned red, emphasizing that by 2021, the entire Summit County would have their obesity prevalence plateaued at the block group level.
Figure 40. Screenshot of the dynamic model parameters and software interface
Figure 41: Results of applying the dynamic model at the census block group in Summit County, Ohio
After running the extended dynamic model, the first and last block groups to have their obesity prevalence plateaued were selected. Figure 42 shows the locations of the block groups where that their obesity prevalence were first and last plateaued. The first and last block groups were selected in order to find out the conditions of their socioeconomic status, such as, education attainment, median family income, and percentage of unemployment. Furthermore, other variables were examined, including the percentages of urban and green space, and distances to parks and food outlets.

Table 14, compares the first block groups and last census block groups that reached the plateau of obesity prevalence the earliest and latest. The average education attainment in the first groups was higher than that of the last block groups. For the first block groups, the education attainment was 32.15% and it was 7.42% for the last block groups. The median family income was also higher than in the first group than that in the last block groups. In the first group, the median family income was $65,095 and at the last block groups had $33,296. The final socioeconomic variable, which was unemployment, shows a higher percentage in the last block group (15.11%) than that in the first block group (6.40%). In terms of the percentage of urban areas, the percentage of urban areas in the first block groups (61.18%) was higher than that in the last block groups (81.36%). Finally, the percentage of green spaces in the first block groups was 39% and 18.62% for the last block groups.

In terms of the distances to grocery stores, fast food, and parks, the first block groups had a shorter distance if compared with that in the last block groups. The comparison between the first and the last block groups about when their obesity prevalence would turned plateau indicated that the first block group had a higher socioeconomic status but shorter distances to parks, and food outlets than those of the last block groups.
Table 14. Comparison between the first and last block groups that their obesity prevalence become plateau.

<table>
<thead>
<tr>
<th></th>
<th>First block group</th>
<th>Last block group</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Education attainment (%)</strong></td>
<td>32.15</td>
<td>7.42</td>
</tr>
<tr>
<td><strong>Median Family Income</strong></td>
<td>$65,095</td>
<td>$33,296</td>
</tr>
<tr>
<td><strong>Unemployment (%)</strong></td>
<td>6.40%</td>
<td>13.25</td>
</tr>
<tr>
<td><strong>Urban (%)</strong></td>
<td>61.18</td>
<td>81.36</td>
</tr>
<tr>
<td><strong>Green spaces (%)</strong></td>
<td>39</td>
<td>18.62</td>
</tr>
<tr>
<td><strong>Distance to grocery stores (miles)</strong></td>
<td>1.20</td>
<td>0.80</td>
</tr>
<tr>
<td><strong>Distance to fast food (miles)</strong></td>
<td>0.85</td>
<td>0.50</td>
</tr>
<tr>
<td><strong>Distance to parks (miles)</strong></td>
<td>1.03</td>
<td>0.75</td>
</tr>
</tbody>
</table>

The extended model was used with the same parametric values as those in Thomas et al. (2014), except for the probability of being born in obesogenic environment and the birth/death rates to find what census block groups would plateau first in terms of obesity prevalence. This approach was used to see what changes would occur if those two parameters were changed. The extended dynamic model was performed with three different birth rates 0.20, 0.30, and 0.40 per 1,000 with a probability of being born in obesogenic environment of 1.
Figure 42. First and last block groups that plateaued in Summit County, Ohio
Figure 43 shows the percentage of census block group that their obesity prevalence plateaued for the three different birth rates (0.20, 0.30, and 0.40). With the first birth rate (0.20 per 1,000) in the year 2015, the obesity prevalence began to plateau with of 37% of the block groups. As can be seen from Figure 42, using different birth rates did not change too much in terms of the time at which the prevalence of obesity would plateau in Summit County. However, increasing the birth rate to 0.30 and to 0.40 per 1,000 would increase the percentage of block groups that plateaued in 2017 (See Figure 43). This indicated that birth rates and the probability of being born in obesogenic environment increased the time by which obesity prevalence would plateau. The dynamic model was also applied with the same parametric rates, expect for the rate of the social influence by overweight and obese individuals.

The extended model was also used to find out whether the influence of overweight and obese individuals increase or decrease the time, by which the prevalence of obesity in Summit County would plateau. For social influence:

1. the influence rate by overweight and obese were set as 0.01 and 0.02
2. the rate by overweight and obese were set as 0.20 and 0.10
3. the rate were set as 0.40 for overweight and 0.20 for obese (Figure 43).

For social influence rate (1) the obesity prevalence would not plateau until 2023. With the second social influence rate (2), the obesity prevalence was growing faster because of the influence of overweight and obese. The social influence rate (3) increased, the obesity prevalence more than the first and second social influence rates. In the third social influence rate, the obesity prevalence would plateau by 2019 or 2020 (Figure 44). This indicated that the social influence rates on overweight and obese were more effectible factors than the birth rates in terms of the time at which the prevalence of obesity would plateau.
Figure 43. Time at which the obesity prevalence would plateau with different birth rates

Figure 44. Time at which the obesity prevalence would plateau at different social influence rates
5.5 Discussion

The literature suggested that socioeconomic variables were associated with obesity prevalence (Giske et al., 2011). This concept was validated in Summit County, where high obesity ratios were associated with low educational level, low median family income, and high unemployment rate. The literature also suggested that built environment and physical environment features affected the prevalence of obesity (Gordon-Larsen et al., 2006).

Network analyst tool (closest facility) was used to calculate the distances between physical environment features and centroids of each census block group along streets in the road network. The finding indicated that for the block groups with the top ten obesity ratios, bus stops and fast food outlets had the shortest distances when compared with other physical environment features. This could explain that fast food outlets had the shortest distances to the centroids of block groups with the top ten obesity ratios.

Consequently, those block groups had the highest obesity ratios. Consuming fast food regularly with a lower physical activity increased the chances for being obese. In terms of ethnicity, as mentioned in the analysis section, dominated the neighborhoods of white population had a lower obesity ratios compared with dominated African Americans neighborhoods, even though African Americans were minority in the study area. The findings showed that neighborhoods with predominantly white population (>70%) had a lower average obesity ratio (144 per 1,000 population) and neighborhoods with predominantly African American (>70%) had a higher average obesity ratio (269.6 per 1,000 population). Census tracts and block groups were used in order to do comparison between different geographic scales. Census blocks were not used; however, it is the smallest geographic unit. The purpose of not using census block was the confidentiality of people’s information. A study on obesity prevalence in Summit County,
OH found that census block groups gives more details than using census block groups because of the level of aggregation (Lee et al., 2014). Therefore, the reason why using tracts and block groups was for comparison purposes. Moreover, block groups defines the neighborhood better than tracts.

Built environment was one of the features that could affect the prevalence of obesity. At the top ten neighborhoods with the highest obesity ratios, there were more urban neighborhoods than rural ones and there was less green space in those neighborhoods than others. This could explain that obesity was highly associated with urban neighborhoods with less green space. The results in the second model of GWR (built environment model) did not reflect the idea that built environment features affected obesity prevalence. In contrast, the socioeconomic variables, which showed stronger relationship with obesity ratios were influential at both tract and block group levels.

Applying GWR with each independent variable individually helped to distinguish the independent variables that were the most influential with obesity ratios. Socioeconomic variables, such as the education attainment, and unemployment rate were the most influential variables on obesity ratios in neighborhoods in Summit County. GWR was performed at two different geographical scales, census tracts and block groups. At the census tract level, the results of applying GWR in the first model (socioeconomic variables) showed that a strong association between dependent and independent variables because of the high value of adjusted $R^2$ (0.67%) in tracts but a slightly weaker association in block groups (0.41%). The second model (Built environment features) showed only weak associations at both scales because of the low values of adjusted $R^2$. 

109
The dynamic model that had been used by Thomas et al. (2014) was performed at each census block group in order to find the time when the obesity prevalence in Summit County would plateau. The results showed that the block groups that plateaued the earliest had higher values in the socioeconomic variables than those of the block groups that plateaued last. This indicated that block groups with higher socioeconomic status had lower obesity prevalence that would plateau first. However, the distances to parks, grocery stores, and fast food were lower in the block groups that plateaued the last. The extended dynamic model was also applied with different birth rates, 20, 30, and 40 (per 1,000 population). From the results, the time at which obesity ratios would plateau did not differ among different birth rates. However, the social influence rates by overweight and obese had more influence than the birth rates.

5.5.1 Advantage / Disadvantage of Census Tracts and Block Groups:

This study used two different geographic scales for the analysis (Block groups & tracts) in order to find how much difference would occur when applying two different geographic scales. In the United States, the standard hierarchy of census geographic boundaries was organized in the following order:

- nation,
- regions,
- states,
- counties,
- census tracts
- census block groups and,
- census blocks,
The data for this study was available at the tracts and block groups level. In terms of the advantage of using tracts level, tracts were used in various studies as unit of analysis and to define the relationship between neighborhood and food environment because of the availability of the data at the tracts level (Fan et al., 2014). According to the US Census, tracts have a population, ranging from 1,200 to 8,000 people. While census block groups have a range of 600 and 3,000 people each. The main advantage of using block groups over tracts is to visualize the statistical data geographically. In general, different studies used census tracts in their analysis because the availability of the data at the tract geographical scale while at the block groups the data are rarely available especially health data. However, there is a disadvantage of using tracts because the generalization and the margin of error at the tracts level are higher than those of block groups. For example, one census tract could be larger than multiple census block groups. In terms of the spatial size of the census boundaries, the smaller geographic boundaries the more details provided. In the study area, there are 135 census tracts and 452 census block groups, which provided more details and information than census tracts. The spatial size of tracts varied between different areas in the study area (urban verses rural); however, at the block group level the spatial size is smaller, which was more precise than tracts.

Using census tracts and block groups for analysis were useful especially to define the neighborhood level; however, at the tract level, the data were more available than block group. In terms of the disadvantages, the boundary of the tract could overlay with other neighborhood or city boundaries. Therefore, block groups were useful in representing the neighborhood better than tracts.
CHAPTER 6

CONCLUSION

Obesity is one of the main cause of death in the United States and many other developed countries. Many chronic diseases were related to obesity, such as diabetes, heart attack, hypertension, and others. The obesity ratio for adult population (age 16 – 21) in Summit County, Ohio, was 123.26 per 1,000, which was a high number. Recent studies on obesity prevalence focused on the factors that were associated with obesity, such as socioeconomic variables and built environment features. One of the purposes of this study was to find the associated variables that were related to obesity ratios in Summit County, at the two-neighborhood levels, census tracts and census block groups. GIS techniques were used in order to associate the physical environment features to obesity ratios in the study area. GIS techniques were also used to find out the association between the socioeconomic variables and built environment features and obesity ratio.

Finding out the factors associated with obesity prevalence could help policy makers to design and apply strategies that could reduce prevalence of obesity. Another purpose of this study was to extend and apply the dynamic model that was proposed by Thomas et al. (2014) to the neighborhood level. In this study, neighborhoods were defined as census block groups, although census tracts were used for comparison purposes. Thomas’ model was extended and applied to the neighborhood level to find out when the prevalence of obesity in each neighborhood would plateau first.
According to the available data, such the populations of normal weight, overweight, obese, extremely obese, birth, and death rate, the obesity prevalence in Summit County, would plateau by 2021. The results from applying the extended dynamic model indicated that the social influence for overweight subpopulation and obese was more influential than changes in birth rates. Once the social influence for overweight and obese subpopulation were defined as three different rates, the numbers of census block groups that plateaued were different from the results of applying the extended dynamic model at different birth rates. In the birth rates model, there was not a big difference between the three different birth rates in terms of the time at which obesity prevalence would plateau.

In conclusion, GIS techniques such as geographically weight regression, network analyst, Cluster and Outlier analysis, and spatial statistics, were all useful in distinguishing and assessing the association between the relevant variables with obesity ratios in the study area. In addition, applying the extended dynamic model at the census block group level helped to visualize the neighborhoods that would have their obesity prevalence plateaued. The extended model allowed simulations to be carried out to see the different paces individual neighborhoods having their obesity prevalence plateaued. Outcomes from the extended model could be visualized to assist the understanding of spatio-temporal trends of obesity prevalence at the block group level over time between 2013 and 2023. Understanding the degree to which influential factors were associated with obesity ratios, such as built environment or socioeconomic variables were essential in order to design and implement strategic actions to reduce obesity prevalence in neighborhood level. Such strategic actions could include increasing the awareness of the health effects of being obese and developing the built environment in a way that encouraged physical activity to reduce obesity prevalence. Furthermore, applying the extended dynamic model at the
neighborhood level helped to identify the locations of where the obesity prevalence would plateau first.

6.1 Limitations

One of the limitations in this study was the availability of the data that could be linked with GIS. There were many factors that could be related to obesity, such as crime rates at the block group level. Unfortunately, few of such data were available to this research. However, that could be a direction for future research.

Another limitation was the BMI derived from the Ohio Bureau of Motor Vehicle. Some people tended to provide inaccurate information or failed to update information once they had their driver’s licenses. Such information included heights and weights as reported by license holders. This potential inaccuracy was especially the case about their weights. In addition, the obesity data in Summit County, Ohio was just for those who had driver licenses issued. These data did not include those who chose not to have or failed to have driver’s licenses.

Furthermore, the data was gathered from different sources with different dates. For example, the demographics of the census was issued in 2010, driver licenses records were from 2008 until 2012, and the food outlets were gathered based on the year 2014. That was one of the limitations with using different data sources; however, this was the best data I could assemble for this study.

One limitation of using BMI data was that it was not always correct. BMI values were not simple to interpret or compare, especially for people with BMI greater than 30 being defined as obese. Because there might be some people that are physically fit but have high BMI values. For example, a BMI value that was considered high implying a person was overweight or obese could in fact be physically fit (Figure 44). However, the BMI values were the most widely used
measurement for obesity because a BMI value it used the ratio of the height and weight. At this time, there was not any better indicator than BMI to use for defining overweight, obese, or extremely obese.

The last limitation was with the extended dynamic model that had been applied at the census block group level. It was difficult or even impossible to predict how individuals moved between different BMI categories overtime. Such movements could only be assumed for the purpose of simulations. Therefore, it was based on assumptions and national trends of the United States. However, the only available data that I could use in the extended dynamic model were the number of normal weight, overweight, obese, and extremely obese as calculated from the self-reported heights and weights taken from driver license data of different years. The other parameters that were used in the model were based on assumptions for how they would change in the future.

*Figure 45. The same BMI score with a different body shape*
6.2 Future directions

The future directions of this research in this study are to further investigate the obesity prevalence as a public health issue and to fit with more associated factors that are related to obesity. One of the important directions is to focus on built environment with more research in neighborhoods where obesity prevalence are high. In addition, the dynamic model of obesity prevalence is based on several assumptions whose validity greatly depended on whoever defining them. Therefore, finding the precise rates or best estimates for the parameters in the extended dynamic model would be critical for a successful study in the future. Another important future direction of this study is to find a solution to prevent obesity and reduce the prevalence of obesity by starting from the neighborhood level. Solutions could be done by increasing the awareness of the side effects of obesity and how important to increase physical activity at the neighborhood level in order to reduce the chances for being obese.
REFERENCES


Giles-Corti, B., & Donovan, R. J. (2002). Socioeconomic status differences in recreational physical activity levels and real and perceived access to a supportive physical environment. *Preventive medicine*, 35(6), 601-611.


local food with diet quality – a comparison of assessments based on surveys and geographic information systems: the multi-ethnic study of atherosclerosis. *American Journal of Epidemiology.* 167(8), 917-24


Thomas, D. M., Weedermann, M., Fuemmeler, B. F., Martin, C. K., Dhurandhar, N. V., Bredlau,


## APPENDIX

Description for each equation (Thomas et al. 2014)

<table>
<thead>
<tr>
<th>Subpopulations</th>
<th>Description</th>
<th>Formulas</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Susceptible (BMI &lt; 25)</strong></td>
<td>Proportion of birth entering a non-obesogenic environment BMI &lt; 25</td>
<td>$\mu N(1 - p)$</td>
</tr>
<tr>
<td><strong>S(t)</strong></td>
<td>Fraction of the population dies (normal death rate)</td>
<td>$-\mu S$</td>
</tr>
<tr>
<td></td>
<td>Fraction of the normal weight become exposed</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Contact with overweight individuals (25≤BMI≤30)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Contact with obese individuals (30≤BMI≤40)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Increase in weight not related to social contact with overweight or obese</td>
<td>$-\alpha S$</td>
</tr>
<tr>
<td><strong>Exposed E(t)</strong></td>
<td>Proportion of births born in obesogenic environment</td>
<td>$p(\mu N)$</td>
</tr>
<tr>
<td></td>
<td>Fraction of the population dies (normal weight)</td>
<td>$-DE$</td>
</tr>
<tr>
<td></td>
<td>Fraction of the population becomes overweight (25≤BMI≤30)</td>
<td>$-\alpha E$</td>
</tr>
<tr>
<td></td>
<td>Fraction of the recovered individuals becomes susceptible to re-infection</td>
<td>$\rho R$</td>
</tr>
<tr>
<td></td>
<td>Fraction of the normal weight become exposed</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Contact with overweight individuals (25≤BMI≤30)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Contact with obese individuals (30≤BMI≤40)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Increase in weight not related to social contact with overweight or obese</td>
<td>$-\alpha S$</td>
</tr>
<tr>
<td><strong>Infected (Overweight 25≤BMI≤30)</strong></td>
<td>Fraction of the population dies (normal death rate)</td>
<td>$-DI_1$</td>
</tr>
<tr>
<td></td>
<td>Fraction of exposed individuals become infected</td>
<td>$\alpha E$</td>
</tr>
<tr>
<td></td>
<td>Fraction of the infected (overweight) move to obese category (30≤BMI≤40)</td>
<td>$-\alpha_4 I_1$</td>
</tr>
<tr>
<td></td>
<td>Fraction of overweight recover and return to normal weight</td>
<td>$-\rho I_1$</td>
</tr>
<tr>
<td></td>
<td>Fraction of obese individuals return to overweight category</td>
<td>$\beta_2 I_2$</td>
</tr>
<tr>
<td>Category</td>
<td>Formula</td>
<td></td>
</tr>
<tr>
<td>--------------------------------</td>
<td>-------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Infected (Obese 30≤BMI≤40) Iₜ(ᵗ)</td>
<td>Fraction of the population dies (different death rate than normal weight population) $-D₀I₂$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fraction of overweight become obese $α₁I₁$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fraction of obese individuals move to extremely obese category (BMI≥40) $-α₂I₂$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fraction of obese move to overweight category $-β₂I₂$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fraction of extremely obese individuals move to obese category $β₃I₃$</td>
<td></td>
</tr>
<tr>
<td>Infected (Extremely obese BMI≥40) Iₜ(ᵗ)</td>
<td>Fraction of the population dies (different death rate than normal weight population) $-D₀I₃$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fraction of infected obese individuals move to extreme obese category $α₃I₁$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fraction of extreme obese individuals move to obese category $-β₃I₃$</td>
<td></td>
</tr>
<tr>
<td>Recovered (BMI≤25)</td>
<td>Fraction of the population dies (normal death rate) $-D₀R$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fraction of infected (overweight) recovered $ρ₁I₁$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fraction of the recovered returns to exposed category $ρₘR$</td>
<td></td>
</tr>
</tbody>
</table>