2014

Built Environment and Risk of Obesity in the United States: A Multilevel Modeling Approach

Yanqing Xu
Louisiana State University and Agricultural and Mechanical College, yxugeo@gmail.com

BUILT ENVIRONMENT AND RISK OF OBESITY IN THE UNITED STATES: A MULTILEVEL MODELING APPROACH

A Dissertation

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

in

The Department of Geography and Anthropology

by
Yanqing Xu
B.E., China University of Geosciences, 2008
M.S., Wuhan University, 2010
December 2014

Acknowledgements

Completion of this doctoral dissertation would not have been possible without the help and support of several people. First of all, I would like to extend my deepest gratitude to my academic advisor, Professor Fahui Wang, for his motivation, enthusiasm, patience and tremendous knowledge. His valuable guidance and unconditional support helped me from the very beginning throughout my work on this dissertation. Thank you for everything you have done for me, Professor Wang! I would also like to thank Professor Lei Wang for his guidance, patience, understanding, and most importantly, his friendship during my graduate study. He always helped me solve the technical obstacles and encouraged me to continue the research. My sincere thanks also go to Professor Eugene Kennedy, who tirelessly advised me on the statistical models and scientific writing. I really enjoyed the every Friday meeting with Professor Kennedy, and we had endless topics to discuss. I also thank the Graduate School Dean's representative Professor Haosheng Huang for his advice during my proposal defense and final defense.

I am grateful to the encouragement and support of many friends, especially Bedoor Mahammad, Quan Tang and Myrtho Joseph who have always been supportive of me through these difficult years. I will cherish forever the friendship with several visiting scholars from China, with whom I have shared the office. They are easy-going, humble, considerate, and self-motivated, and have been a constant force exerting a positive influence on me.

I also thank Ms. Dana Sanders for her patience and kindness in processing all my paperwork. She has always made herself available to clarify my questions. I thank Professors Barry Keim, Xuelian Meng and Heather McKillop for their help. Mr. Luke E Driskell helped me with software installation and updates, data acquisition and computer hardware.

I extend my biggest appreciation to the love, care, and support of my family. I would like to thank my husband, Cong Fu, for his encouragement, support, and everything. His tolerance of my occasional gloomy moods is a pledge in itself of his strong devotion and love. I thank my parents for their endless love and faith in me. Also, I thank Cong's parents for their consistent encouragement and support.

Thanks to all those people who have believed in me. Because of you, my academic experience has been one that I will cherish forever.

Table of Contents

List of Tables

List of Figures

Abstract

This dissertation provides a series of exploratory analyses of the relationship between built environment and obesity by using multiple data sets and employing the state-of-art Geographic Information Systems methods. Several built environment factors including street connectivity, walkability and food environment, are for the first time measured across 48 contiguous states of the U.S., built from a fine geographic scale such as the census tract level.

Based on the nationwide BRFSS data, the first study used the Geographically Weighted Regression (GWR) model to analyze the obesity rates at the county level. The model results reveal that overall obesity rates are negatively related to walk score and street connectivity, but positively related to poverty rates and metro classification, while the effect of fast-food-to-full-service restaurant ratio is not evident. The strength of each variable's effect also varies significantly across the country.

To mitigate the ecological fallacy, the second study used a multi-level modeling (MLM) approach by accounting for individual attributes such as demographic, socioeconomic and behavior variables. Furthermore, models for areas of different urbanicity levels were tested. The national study found that obesity risk initially increases with the urbanicity level and then drops, resembling an inverted-V shape. The results lend support to the role of built environment in influencing people's health behavior and outcome, and promote public policies that need to be sensitive to the diversity of demographic groups and geographically adaptable.

Defining neighborhoods at the county level may be problematic in the previous MLM study since people's activity space is seldom countywide. The third study added another level (zip code area) to the MLM analysis of the BRFSS data in Utah. The results showed that at the zip code level, poverty rate and distance to parks are significant and negative covariates of the

odds of overweight and obesity; and at the county level, food environment is the sole significant factor with a stronger fast food presence linked to higher odds of overweight and obesity. These findings suggested that obesity risk factors lie in multiple neighborhood levels and built environment need to be defined at a neighborhood size relevant to residents' activity space.

Chapter 1 Introduction

Obesity is a major risk factor for heart disease, diabetes, stroke, depression, sleep apnea, osteoarthritis, and some cancers. The obesogenic environment thesis suggests that disparities of obesity prevalence are attributable to differentiated exposure to a healthy food environment that promotes healthier dietary choices and built environments that encourage physical activities. For example, easy access to fast food is likely to promote more meals or increase consumption of high fat meals, leading to higher caloric intake and propensity of weight gain. Regular physical activity can help control weight and improve health. Broadly speaking, built environment includes not only human-made resources and infrastructure designed to support human activity, but also food environment such as restaurants and grocery stores, as compared with the natural environment.

Although the association between built environment and obesity has been studied extensively, several challenges prevent us from obtaining a comprehensive understanding of how various obesogenic built environment factors affect weight status. For instance, lack of quality data and complexity of computation, accurate quantitative measures of built environment has previously been difficult or infeasible, particularly at a large scale such as a national scope. Most of the existing studies are localized and yield often conflicting results. The majority of studies focus on one or a very few built environment factors. To fill these gaps, the focus of this dissertation is on new and creative measures of built environment in the U.S. The research also seeks to reveal possibly different effects of built environment in various geographic settings.

After the detailed introduction in the measurement of those built environment variables, a global model was used to analyze the overall relationship and GWR model was used to identify regional differences. The regression model has found that the walk score and street connectivity are negatively relatedly to obesity, and that poverty rate and metro are positively related to obesity, while the fast-food-to-full-service restaurant ratio is not significant. These findings were translated to qualitative inferences that could help policy making. This analyses was based on aggregated data which ignoring individual variability. In fact, Individual behaviors such as eating habit and physical activity do not occur itself; rather, they are influenced by socio-environmental factors including built environment. In order to overcome this possible ecological fallacy, where relationships observed in groups are assumed to hold for individuals (Freedman 1999), multilevel models were then used to analyze the influence of built environment on obesity by incorporating individual-level risk factors.

Based on the global multilevel models using samples in the nationwide area, county-level socio-demographic structure such as a lower racial-ethnic heterogeneity index or a higher poverty rate is linked to a higher obesity risk. Among the built environment variables, a poorer street connectivity and a more prominent presence of fast-food restaurants are associated with a higher obesity risk. While the effect of walk score is not evident in influencing obesity risk, a higher walk score is indeed linked to a lower rate of physical inactivity. Overall, obesity risk initially increases with the urbanicity level and then drops, resembling an inverted-V shape. These results led to the examination of possible variability of association between built environment and obesity across different urbanization levels which is another important highlight of the research.

The issue of appropriate area unit for defining the neighborhood effect is another concern in public health. Therefore, this research continues to examine the neighborhood effects at different levels on association of built environment factors with individual obesity. Due to the data limitation, the study area of Utah was used in both the zip-code and county level. The results suggest that observed built environmental influences on overweight and obesity are sensitive to these different spatial units. Net of individual controls and place-based poverty prevalence, distance to parks seems to be the only significant built environmental variable that is consistent with the hypothesis, that is, the longer distance to parks, the less spatial park accessibility, the higher odds of overweight and obesity. The results on the food environment are inconsistent across zip code and county level analyses. Walk score and street connectivity, measures of neighborhood walkability, are not significantly linked to odds of individuals' excessive body weight in this sample. These findings suggest that the contextual variables need to be defined in a way that reflects human mobility pertaining to the specific trip purposes.

The remainder of the dissertation is structured as follows. Chapter 2 is the review of the literature for the research. Chapter 3 introduces the neighborhood variables, including the social-demographic variables, built environment variables and different urbanization levels. The built environment measurements include street connectivity, walk score, food environment, and accessibility to parks. As a baseline study, Chapter 4 provides an ecological analysis of association of neighborhood variables with obesity rates at the county level by using both global and local regression models. Chapter 5 examines how built environment variables (measured at the county level) affect individual's physical activity behavior and body weight by using multilevel modeling to control for the effects of individual attributes. Studies reported in Chapters 4 and 5 have a national scope (covering the 48 contiguous states of the U.S.). Chapter 6 further advances the research by focusing on the State of Utah with the built environment variables measured at both zip code and county levels. It uses three-level models to detect whether built environment factors measured at different neighborhood sizes exert different influences on individuals' body weight, and thus shreds light on possibly appropriate neighborhood sizes for measuring particular built environment factors. Chapter 7 summarizes the results and conclusions from the preceding chapters and discusses future work.

The data used in this research were from different sources. In Chapter 4, the individual data was from the BRFSS carried out in 2012 in the conterminous United States area which is the newest published dataset. Chapter 5 used the Utah BRFSS dataset in the year of 2007, 2009 and 2011. Census 2010 data was used to capture the social-demographic variables, including race heterogeneity and poverty rate. Street dataset in 2005 was used in Chapter 4, while Chapter 5 used the newest street dataset in 2009 from ESRI data 2012. Walk score was captured by the Walk Score API which take advantage of the real-time traffic data and it was collected in the year of 2012. Restaurant data was from the County Business Patterns (CBP) which is an annual series providing subnational economic data by industry. In the dataset, restaurants are classified into fast food and full service by their type.

Chapter 2 Literature review

During the past twenty-five years, the United States has experienced an unparalleled rise in its residents' body weights. Overweight and obesity are the two factors that reflect ranges of weight that are greater than what is commonly considered to be healthy for a given height. Body Mass Index (BMI; weight in kilograms/ (height in centimeters/100)^2) is generally used for people to measure their health condition and it provides a reliable indicator of body fatness for most people and is used to screen for weight categories that may lead to health problems (Doyle et al. 2006). Overweight was defined as a BMI of 25.0 to 29.9 and obesity was defined as a BMI of 30.0 or higher (Flegal et al. 2010). The current obesity epidemic, the main topic in this research, has become a significant contributing factor of several leading causes of morbidity and mortality, including heart disease, stroke, diabetes and some cancers (Zhang, Lu and Holt 2011). The U.S. Department of Health and Human Services (HHS) awarded more than $119 million to states and U.S. territories to support public health efforts to reduce obesity and increase physical activity (Wakefield 2004). If the prevalence of obesity continues, 13 states could have adult obesity rates over 60 percent, 39 states could have rates above 50 percent, and all 50 states could have obesity rates above 44 percent by 2030 (Gates 2012).

Obesity can be caused by many factors. Although prevention strategies for obesity and its related risk may be obtained etiologically, the influence of built environment is an emerging interest. According to the Centers for Disease Control and Prevention (CDC), obesity trends vary geographically: only Colorado and the District of Columbia had a prevalence of obesity less than 23% in 2010; over thirty-three states had a prevalence equal or greater than 25%; eleven of these states (Mississippi, Alabama, West Virginia, South Carolina, Kentucky, Tennessee, Texas, Louisiana, Michigan, Missouri, and Oklahoma) had a prevalence of obesity equal or greater than

31%. The variations across smaller geographic areas such as counties and census tracts are even greater. In searching for factors driving the obesity epidemic and its geographic variations, the importance of neighborhood socio-demographic and built environment characteristics were highlighted among researchers in the recent decade. The built environment refers to human-made resources and infrastructure designed to support human activity, such as buildings, roads, parks, restaurants, grocery stores and other amenities, as compared with the natural environment (Pierce, Ernest and Ashworth 2012). Exposure to different built environments may contribute to the geographical variation of obesity trend.

The built environment variables are mostly defined in three domains: physical activity, land use and transportation, and food environments (Feng et al. 2010). To define the environment variables related to physical activity, greenness and access to recreational facilities are the two most popular measurements. Bell (2008) examined associations among age- and gender-specific BMI z-scores in a satellite-derived measure of greenness. They measured greenness by using Normalized Difference Vegetation Index (NDVI) which was derived by converting pixel values in satellite images to continuous measurements than ranges from -1 to +1 (Grigsby, Chi and Fiese 2011). Results showed that greenness is inversely associated with the BMI z-scores of children and youth at 2 years. Casey (2008) evaluated physical activity environment by designing a survey to get answers from responses. A mean physical activity "access" variable was created of all answered responses to actively place, walk to different destinations, sidewalks present and shoulders of roads safe for walking and community pleasant for physical activity. Comparing with the subjective measurement, Miles (2008) used objective measures of the physical activity environment by counting the number and location of recreational facilities and destinations within the neighborhoods, the location of sidewalks, and the layout of streets. Different conclusions were made from these researchers by checking the relationship between physical activity environment and weight status. Gordon (2006) used an 8-km radius around the residence as neighborhood and arrived at the conclusion that odds of overweight declined with increasing number of physical activity facilities per census block group. Nevertheless, Burdette (2004) examined the relationship between overweight in preschool children and environmental factors and found that there is no association between proximity to playgrounds.

Land use/transportation environment contains the 3D (density, diversity and design), connectivity, walkability and sprawl. Population density, residential density and employment density are the most common factors to define density (Smith et al. 2008b); land use mix and entropy index are popular ways to define diversity (Bodea, Garrow and Meyer 2008); bus stop density and subway stop density are the main factors for design (Rundle and Freeman 2007). According to Li, each unit increase in land-use mix was associated with a 25% reduction in the prevalence of overweight/obesity. Rutt and Cleman (2005) got the opposite conclusion that living in areas with greater mixed land use was associated with higher BMI values. Until now, there is no specific definition for walkability. In ESRI's News, "Walkability is a measure of the effectiveness of community design in promoting walking and bicycling as alternatives to driving cars to reach shopping, schools, and other common destinations" (Rattan, Campese and Eden 2012). The Centers for Disease Control and Prevention (CDC), the World Health Organization (WHO) and other health organizations emphasize the importance of walkability to prevent obesity. Both Saelens (2003) and Doyle (2006) got the same conclusion that there is a statistically significant inverse association between walkability and BMI.

Ewing (2003) developed a county sprawl index using the similar process as metropolitan sprawl index from Smart Growth America. The sprawl index is based on four factors: residential density, neighborhood mix of uses, strength of activity centers and downtown, and accessibility of the street network (Ewing, Pendall and Chen 2002). They found that the county sprawl index had small but significant associations with obesity. In another of Ewing's work (2006), they did both cross-sectional and longitudinal study on US adolescent participants. Cross-sectional analysis demonstrated overweight is significantly related to urban sprawl and longitudinal analyses showed no statistically significant relationships between urban sprawl and changes in BMI over time. Lopez (2004) examined the association between urban sprawl and the risk for being overweight or obese among US adults and found that urban sprawl is significantly associated with overweight and obesity. In his research, urban sprawl was based on density and compactness and the sprawl index is defined as $SI_i = 50((S\%_i – D\%_i) + 1)$, where $S\%_i$ = % population in low density census tracts and $D\%_i$ = % population in high-density census tracts. Kelly-Schwartz (2004) used the same sprawl index but got a different conclusion: there is no association between the metropolitan area-level sprawl index and BMI.

Food environment is an important factor related to people's weight status. There are many ways to identify food environment variables: fast food restaurant density, population per fast-food restaurant, fast-food restaurant proximity, average food pricing, distance to usual grocery store and so on (Papas et al. 2007). Metha and Chang (2008) measured restaurant density as the number of restaurants per 10,000 individuals. Restaurant were classified into fast-food or full-service categories (Chou, Grossman and Saffer 2004, Morland et al. 2002). They concluded that fast-food restaurant density and a higher ratio of fast-food to full-service restaurants were associated with higher risk of being obese. Wang (2007) calculated food environment as store proximity and count of stores per square mile and got the conclusion that higher neighborhood density of small grocery stores and closer proximity to chain supermarkets were associated with higher BMI among women. Maddock (2004) examined the relationship between obesity and prevalence of fast food restaurant at state level and concluded that decreasing numbers of square miles and increasing population per fast-food restaurant were significantly associated with an increasing statewide trend of obesity. Except these associations, Burdette and Whitaker (2004) found that there is no association between proximity to fast-food restaurants. Although different conclusions were obtained from these reviews, most of them found that there is association between food environment and obesity.

It is rare that obesity related research using spatial methods which can identify clusters of individuals exhibiting similar health behaviors or patterns (Schuurman, Peters and Oliver 2009). The Centers for Disease Control defines a cluster as “an unusual aggregation, real or perceived, of health events that are grouped together in time and space and that are reported to a health agency” (MMWR 1990). It is common to find out that the risk of obesity trends are clustered with links to built environment. Mobley et al. (2004), for example, found evidence of a correlation between high-BMI clusters and low socioeconomic status of the surrounding community. Monda and Popkin (2005) found moderately and highly active youth had significantly decreased odds of overweight in both cross-sectional and longitudinal designs by using spatial analysis. According to Schlundt et al. (2006), obesity, diabetes, and hypertension were found to be clustered based on census tract variables. Vanasse et al. (2006) indicated that the prevalence of obesity varied adequately between regions with higher values being associated with low leisure-time physical activity and low fruit and vegetable consumption. In general, spatial analysis is a useful tool for researchers to effectively direct scarce public health resources on vulnerable regions (Pouliou and Elliott 2009)

Multilevel models are common in public health which comes from socio-ecological theories that emphasize the importance of social and environmental factors in determining human behavior and health outcomes (Huang et al. 2009). Individual behaviors such as eating habit and physical activity do not occur itself; rather, they are influenced by socio-environmental factors. In this case, multilevel models are necessary for illustrating individual behaviors including individuals' body weight (Papas et al. 2007). Multilevel models can be called in different ways, including hierarchical linear model (Raudenbush and Bryk 2001), random coefficient models (Bonoit 2009), mixed-effects models (Pinheiro and Bates 2000), covariance structure models (Muthen 1994) and growth-curve models (McArdle and Epstein 1987).

A notable research by Wen and Kowaleski (2012) used multilevel modeling to explore whether neighborhood built environment attributes are significant correlates of obesity risk and mediators of obesity disparities by race–ethnicity. They run the models by different genders and got the conclusion that built environment is a significant correlate of obesity risk but is not much of a mediator of obesity disparities by race-ethnicity. Kim et al. (2006) explored the relations between social capital measured at the US state and county levels and individual obesity and leisure-time physical inactivity by using multilevel logistic models. Different levels including individual level, county level and state level were conducted and the results indicated that little support was found for mediation by social capital for the associations of urban sprawl and income inequality with obesity. Other articles (Boardman et al. 2005, Monteiro et al. 2004, Sundquist, Malmstrom and Johansson 1999, Malmstrom, Sundquist and Johansson 1999, Kennedy et al. 1998, Fraser et al. 2010) used multilevel modeling also provided us the different associations of obesity related to socio- and built environment.

Chapter 3 Choice of neighborhood variables

3.1 Social-demographic variables

Analysis of built environment should also include its interaction with people's wellbeing status and race/ethnic composition (Kirby et al. 2012). (Zhu and Lee 2008) studied the socio-economic status associated with the built environments. They found that given the similar built environment, economic and ethnic disparities exist in the environmental support for walking. Similarly, (Li, Wen and Henry 2014) concluded that built environments and socio-economic conditions were integrated and that both played important roles on obesity prevention. Therefore, in addition to the built environment variables, two socio-demographic variables were selected: poverty rate (estimated percent of people of all ages in poverty) and ethnic heterogeneity derived from the Census 2010 data. Ethnic heterogeneity reflects the racial-ethnic composition which is defined as $1- \sum p_i^2$, where p_i is the fraction of the population in a given group (Sampson 1989). The ethnic heterogeneity index ranges between 0 and 1. If the value equals to 0, it means that there is only one racial/ethnic group in the unit; while a value approaching 1 reflects a maximum heterogeneity. These two socio-demographic variables were also suggested by the experts of the public health studies (personal communications).

3.2 Physical activity environment

3.2.1 Street Connectivity

Intersections are identified from the street centerline data and connectivity is based upon the number of nodes from the streets at each intersection. Intersections with a starting or ending note of an edge or an intersection of 3-way or more edges are included in the connectivity index calculation (Wang, Wen and Xu 2013). Otherwise, starting or ending of an edge and 2-way connections are not included in our research. Intersection density is used based on the number of intersections within the area to measure street connectivity.

$$SC_i = \# \text{ of intersections} / \text{area} \tag{1}$$

Intersection density corresponds closely to block size -- the greater the intersection density, the smaller the blocks. Small blocks make a neighborhood walkable. Street network density and intersection density are highly and positively correlated with each other (Aurbach 2010). Different areas have different patterns of intersection density, and the differences will become larger when street network density decrease from urban to suburban and then rural areas. The measurement is based on a census tract level and then aggregated to county adjusted by population.

3.2.2 Walk score

Walk score (http://www.walkscore.com/) is a measure based on the distances from a point of interest to nearby amenities. The walk score algorithm has been used in many public health studies (Brewster et al. 2009, Cortright 2009, Duncan et al. 2011, Jones 2010, Kirby et al. 2012, Kumar 2009, Li et al. 2014, Rauterkus, Thrall and Hangen 2010, Zhu and Lee 2008). (Brewster et al. 2009) showed that neighborhood walk score was related with the level of physical exercise, and hence could predict the levels of obesity, hypertension, and diabetes. (Jones 2010) studied the walk score and its association with activity levels. They found that the walk score is correlated with the GIS-derived walkability index ($r = 0.63$ $p < 0.0001$). Duncan et al. (2011) concluded that the walk score algorithm could produce valid measure of walkability, particularly at the 1600-meter buffer. They suggested that the walk score could be used across multiple scales.

The walk score algorithm requires user input for location of the amenities such as food, retail, education, recreation, and entertainment, which in this research are sourced from public domain map providers - Google, Education.com, Open Street Map, and Localeze. The algorithm calculates a linear combination of the Euclidean distances from point of interest to the amenities. The weights in the linear combination are determined by facility type priority and a distance decay function (Front Seat 2013). Walk score ranges from 0 (the lowest) to 100 (the highest). 0-49 is defined as car-dependent, while 0-24 means almost all errands require a car and 25-49 means a few amenities within walking distance; 50-69 is defined as somewhat walkable (some amenities with walking distance); 70-89 is defined as very walkable (most errands can be accomplished on foot) and the number between 90-100 is walker's paradise where daily errands do not require a car (Front Seat 2013).

(Front Seat 2013) provides an application programming interface (API) to query the Walk Score database through URL calls, eliminating the need for manually working with the website interface (Front Seat 2013). A Python program was developed to automatically request walk scores from the server through the Walk Score API. In order to avoid the bias caused by concentration of population in limited space within a large area, population-weighted centroids of census tracts are used instead of simple geographic centroids (Wang and Luo 2005). The population data is from the census 2010 at the census block level. The walk scores by census tracts are then aggregated to the county level so to match the scale of the obesity data. The aggregation takes the population as the weight term. The weight is determined by the ratio between the population of a lower level unit (e.g. census tract) and a higher level unit (e.g. county).

$$W_k = \sum_{i=1}^{n_k} \mathrm{Pop_i} * \mathrm{W_i} / \mathrm{Pop_k} \tag{2}$$

where W_k is the walkability for the aggregated geographic unit (k) to walk in the streets, n_k is the number of lower level units to be aggregated in a higher level unit k, Pop_i is the population of the ith lower level unit, and Pop_k is the total population of the higher level unit k.

3.3 Food Environment

This research hypothesized that accessibility to fast food restaurant is an indicator of extra calorie intake per population because food consumption relying fast food restaurants may promote more meals or may increase consumption of high fat meals, leading to higher calorie intake (Lopez 2007). *Food environment* was captured by fast-food restaurant presence. Food consumption relying on fast food restaurants is likely to promote more meals or increase consumption of high fat meals, leading to higher caloric intake (Lopez 2007). The restaurant data was from the U.S. Economic Census (http://www.census.gov/econ/) and County Business Patterns (CBP) (http://www.census.gov/econ/cbp/). In these studies, we used the most recent data available in 2007 and 2012, with restaurants classified into fast-food and full-service. At the county level, food environment was measured as the *ratio of fast-food and full-service restaurants*. In the study area of Utah, many of the zip code areas did not have any restaurants, and calibrating such a ratio would be infeasible. Therefore, at the zip code level, we used the *fast-food accessibility* to capture the food environment. The accessibility measure follows the widely adopted accessibility index such as

$$A_i = \sum_{j=1}^{n} [S_j f(d_{ij}) / (\sum_{k=1}^{m} P_k f(d_{kj}))] \tag{3}$$

where P_k is population at location (i.e., zip code) k, and S_j is the number of fast food restaurants at location j, d is the travel time between them, and the common gravity model (i.e., power function) is adopted to define the distance decay function $f(d)$ (Wang 2012).

3.4 Urbanicity

According to Lopez (2006), the relation between built environments and obesity are different between inner city neighborhoods and sub-urban ones. Neighborhoods in the inner city tend to have greater street connectivity, higher walk score and more sidewalks but still have higher obesity rate since inner cities usually have less attractive and less safe environments deterring physical activity (Weir, Etelson and Brand 2006). The previous findings suggest that it is necessary to examine the possible variation of built environments' effects on obesity by an area's urbanicity. The 2006 and 2013 NCHS (National Center for Health Statistics) Urban-Rural Classification Scheme for counties were used for the classification of urbanicity (NCHS). There are six-level NCHS urban-rural categories, including large central metro, large fringe metro, medium metro, small metro, micropolitan and noncore.

In order to more accurately capture urbanicity, this research also uses another definition based on the 2010 Census Urban and Rural Classification (Census 2013). The Census defines an urban area with minimal criteria of population and population density using much smaller geographic units such as census tracts and census block. For each county, its urbanicity is defined as a continuous urbanization ratio, i.e., urban population in urban areas over the total population in the county (Wang et al. 2013).

Chapter 4 Geographical regression analysis of the build environment and obesity in the U.S.

4.1 Introduction

Regression models were used to study the relationship between obesity and the environmental factors such as fast food density (Rose et al. 2009), land use pattern (Heath et al. 2006, Duncan et al. 2010, Yamada et al. 2012), poverty (Maroko et al. 2009), and walkability (Casagrande et al. 2011b). However, it should be noted that in the spatial scale of administrative level public health studies, regression models could be spatially non-stationary, namely, that the coefficients of the regression model are spatially variable (Brunsdon, Fotheringham and Charlton 1998). In such a case, local regression models such as the Geographically Weighted Regression (GWR, Fotheringham, Brunsdon and Charlton 2002) could avoid the 'ecological fallacy' problem (Holt et al. 1996), and better explain the variability of obesity. In addition, we could gain better understanding of the phenomenon by interpreting the spatial pattern of the coefficients (Brunsdon et al. 1998). (Maroko et al. 2009) examined the relationship between park accessibility and social economic status characteristics such as poverty, language barrier, population density and percent of minority ethnic groups in the New York City by using the global and GWR regression models. They found only a weak relationship of the accessibility of parks and the physical activity variables with the obesity rate. Their results suggested there existed spatial non-stationarity in the regression models. GWR has been demonstrated to be an effective tool to analyze obesity in a geographical context (Chalkias et al. 2013, Chen and Truong 2012, Chi et al. 2013, Dijkstra et al. 2013, Edwards et al. 2010, Fraser et al. 2012, Wen, Chen and Tsai 2010). However, only very few have studied the state-wide obesity problem in the U.S. continental area. (Chi et al. 2013) used GWR and a k-mean clustering analysis method to examine the association of the food environment and some other socio-economic variables with obesity in the U.S. Their work set a basis of a new analysis framework, i.e. using agglomerates to explain the spatial patterns of the regression coefficients. The built environment factors, however, were not their focus.

4.2 Statistics and spatial pattern of input data

The mean value of the obesity rate among adults is 27.39% (Table 1). The distribution patterns of the variables are shown in Figure 1. Higher overall obesity is clustered within the east south central areas. The highest obesity rate is 42.10% in Holmes County, Mississippi. In the race heterogeneity map of Figure 1, race heterogeneity is higher in the south area of the United States, whereas the Queens County in New York has the highest race heterogeneity, which means that there is a prominent diversity of ethnic groups in that area. The average poverty rate is 15.44% and it is much lower in the northeast areas. The areas of high street connectivity and walk score are the highly-urbanized northeast and west coast. The ratio of fast-food to full-service restaurants is low in the Midwest areas. Among the 3109 counties, 1086 of them (about 35%) are metro while the rest are non-metro.

Table 1 Summary values of dependent and independent variables used in OLS and GWR

Variables	Min	Max	Mean	SD	No. Observations
Dependent variable					
Obesity Rate	12.40	42.10	27.39	3.62	3109
Independent variables					
Race Hetero	0.01	0.78	0.27	0.20	3109
Poverty Rate	2.50	48.50	15.44	6.22	3109
Street Connectivity	0.62	336.08	30.42	37.55	3109
Walk Score	0.00	84.75	11.17	14.99	3109
Ratio of fast-food-to--full-service restaurant	0.00	18	2.06	1.38	3109
Urbanization					
Metro					1086
Non-metro					2023

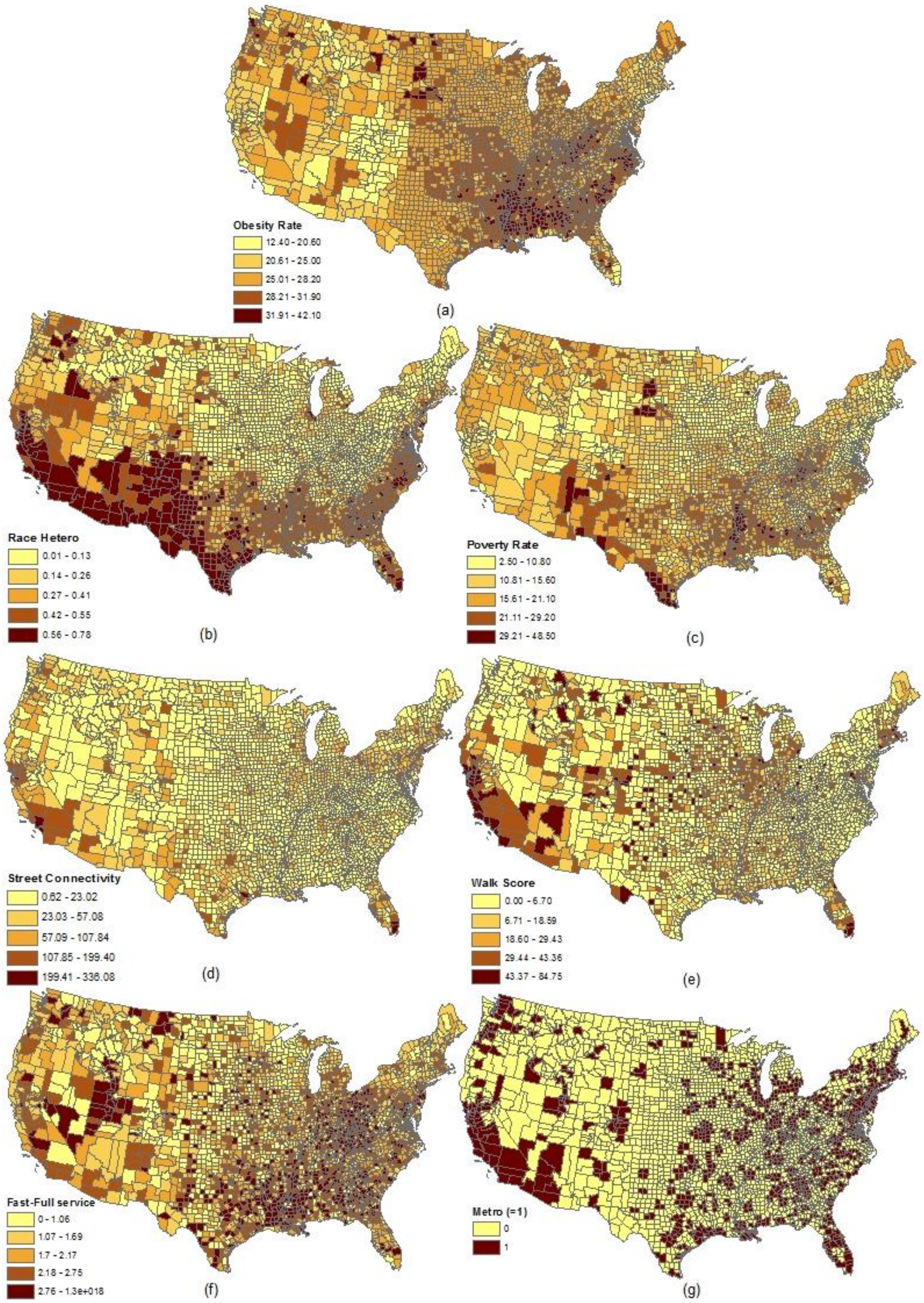

Figure 1 Distribution patterns of the variables: (a) Obesity Rate; (b) Race Heterogeneity; (c) Poverty Rate; (d) Street Connectivity; (e) Walk Score; (f) Fast Food Ratio; (g) Urbanicity

4.3 Methods

4.3.1 Ordinary Least Squares Regression

The regression takes the age-adjusted rates of obesity among adults as the dependent variable; and the independent variables are race heterogeneity, poverty rate, ratio of fast-food to full-service restaurants, street connectivity, walk score and urbanicity. The relationship was examined on a county-wide basis with cross-sectional analysis by using an Ordinary Least Squares (OLS) regression. The purpose is to test the significance of the variables and potential multicollinearity problems among the variables. The model is set as:

$$\mathrm{OB} = \beta_0 + \beta_1 \mathrm{RaceHetero} + \beta_2 \mathrm{Poverty} + \beta_3 \mathrm{Ratio} + \beta_4 \mathrm{SC} + \beta_5 \mathrm{WS} + \beta_6 \mathrm{Metro} + \varepsilon \qquad (4)$$

where OB stands for obesity, β_0, β_1, β_2, β_3, β_4, β_5 and β_6 are the regression coefficients, and ε is the random error in the two models.

Moran's *I* is used to test the spatial autocorrelation of the residuals from the regression model:

$$I = \frac{n}{\sum_{i=1}^{n}\sum_{j=1}^{n} Wji} \times \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} Wij(x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \qquad (5)$$

where *n* is the total number of counties in the area, *i* and *j* represented different counties, x_i is the residual of *i*, and $\bar{x}$ is the mean of residuals. W_{ij} is a measure of spatial proximity pairs of *i* and *j* (Wong and Lee 2005). The values of Moran's I would be between -1 and +1. -1 means negative autocorrelation which implied nearby locations tended to have dissimilar values; +1 means positive autocorrelation which indicated that similar values tended to occur in adjacent areas. Along with the index, Z-scores are usually reported for the statistical significance test. If Z is out of the range of ± 1.96, the non-hypothesis of the randomness test is rejected at the 95% level, which means the pattern is spatially auto-correlated. Otherwise, the spatial arrangement would be regarded as completely random (Lin and Wen 2011, Goodchild 1986).

The OLS regression result shows that the significant variables for obesity are race heterogeneity, poverty rate, street connectivity and walk score (Table 2). The ratio of fast-food to full-service restaurants is not significant. The poverty variable has a positive coefficient (0.31), indicating that the relationship is positive, or in other words obesity prevalence is higher in areas with high poverty rate. In addition, the positive sign of the urbanization variable indicates that residents living in more urbanized areas are more likely at a higher risk of obesity. This confirms the previous findings that that urban areas usually have more disadvantaged populations (i.e., low socioeconomic status or minorities) and less safe environments for people to take physical activities (Doyle et al. 2006, Weir et al. 2006). The negative sign of the race heterogeneity variable suggests that it is more common for the minorities to get obese. The coefficient for street connectivity and walk score is negative and significant, confirming that higher street connectivity and walk score are related to lower obesity rate. The VIF values in the table do not suggest any multicollinearity among the independent variables. The coefficient of determination r2 for obesity is 0.30, where there was a significant amount of variance unexplained. The residual maps (Figure 2) show some spatial autocorrelation in the residuals. The Moran's I of the residuals is 0.31 ($p < 0.01$). The spatial autocorrelation in the residuals suggests there is some spatially correlated variability unexplained by the global OLS model. Instead of the global model, we shall use the local regression model, which allows the regression coefficients to vary over the spatial domain.

4.3.2 Geographically Weighted Regression

GWR is a localized regression model that allows the parameters of a regression estimation to vary over the spatial domain (Lin and Wen 2011). The model can be expressed as:

$$\mathrm{OB_i} = \beta_{0i} + \beta_{1i}\mathrm{RaceHetero} + \beta_{2i}\mathrm{Poverty} + \beta_{3i}\mathrm{Ratio} + \beta_{4i}\mathrm{SC} + \beta_{5i}\mathrm{WS} + \beta_{6i}\mathrm{Metro} + \varepsilon_i \quad (6)$$

Table 2 Ordinary Least Squares result

Variable	Coefficient	StdError	p-value	VIF
Intercept	23.46	0.17		
Race Hetero	-1.69	-5.09	0.00	1.45
Poverty Rate	0.31	0.01	0.00	1.42
Street Connectivity	-0.02	0.002	0.00	1.54
Walk Score	-0.01	0.004	0.00	1.08
Ratio of fast-food-to-full-service	-0.00	0.00	0.28	1.01
Metro	1.13	0.00	0.00	1.42
Moran's *I*	0.31			
Adjusted R^2	0.30			
AICc	15,701			

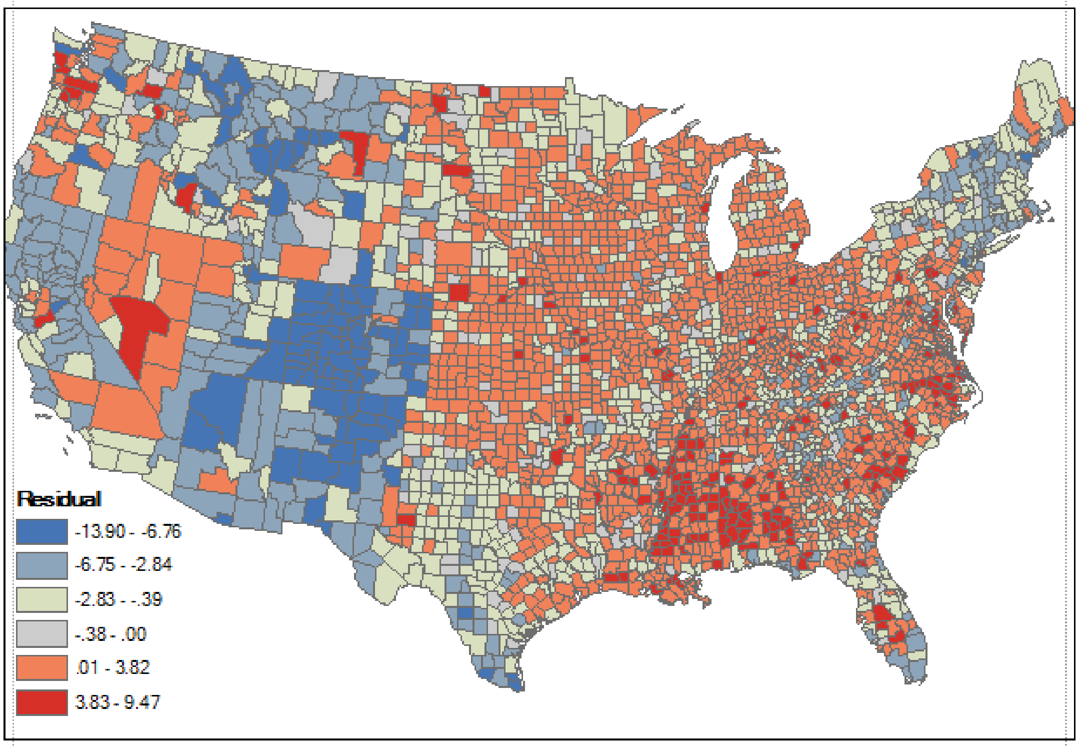

Figure 2 Residuals from the OLS regression. The map shows strong spatial autocorrelation.

where β_{ni} refers to the estimated regression coefficients at county *i*. The spatial variability of an estimated local regression coefficient was examined to determine whether the underlying process exhibited spatial heterogeneity (Fotheringham, Brunsdon and Charlton 2000). The optimal solution of the regression equation in GWR is constrained by a geographically weighted matrix W_i (Fotheringham et al. 2002):

$$\beta_i = (X^T W_i X)^{-1} X W_i Y \tag{7}$$

where W_i is defined by the spatial neighboring relations between points:

$$W_i = \begin{pmatrix} W_{i1} & 0 & 0 & \cdots & 0 \\ 0 & W_{i2} & 0 & \ldots & 0 \\ 0 & 0 & W_{i3} & \ldots & 0 \\ . & . & . & \ldots & . \\ 0 & 0 & 0 & \ldots & W_{in} \end{pmatrix} \tag{8}$$

where W_{ij} is the impact between location *i* and location *j* (*i and j* = 1…n) defined by the distance between them and a kernel function. The closer the data points are, and the stronger impact they have on each other, and therefore a large W_{ij}. The kernel function is usually a Gaussian function with a band width. The adaptive kernel band width calibrated by minimizing the Akaike information criterion (AIC) value of the regression model was used.

The analyses were done in ERSI ArcGIS 10.1 and GWR 4 software packages. The results from the GWR model (Table 3) show significant improvement over the OLS model. The model returns an overall r^2 of 0.72, much better than the OLS model (r^2 = 0.30). And the lower AIC value indicates the GWR model is better than OLS. Figure 3 shows the maps of the locally weighted r^2 between the observed and fitted values. Furthermore, the residuals of the GWR model only have a slight level of spatial autocorrelation (Moran's I = 0.01).

The spatial distribution of r^2 is not even over the study area (Figure 3). Some counties have high r^2 up to 0.85 and some are very low. Generally, the counties in most areas of the north central states and the states of Mississippi, Alabama and Florida have better regression results than others. Figure 4 and Figure 5 shows the maps for coefficients of intercept, race heterogeneity, poverty rate, street connectivity, walk score, the ratio of fast-food-to-full-service restaurants and urban-rural classification, and the *t* values representing the fitting level for each specific variable in GWR. The cartographic method by Mennis (2006) is adopted to map coefficient values and their significance simultaneously.

Table 3 Geographically Weighted Regression results

	Min	25% quartile	50% quartile	75% quartile	Max
Intercept	13.66	23.09	24.96	26.29	31.05
Race Hetero	-14.57	-1.30	1.34	3.75	11.43
Poverty Rate	-0.35	0.10	0.18	0.27	0.52
Street Connectivity	-0.05	-0.02	-0.01	-0.004	0.04
Walk Score	-0.13	-0.02	-0.01	0.004	0.07
Ratio of fast-food- to-full-service	-0.30	0.00	0.00	0.15	1.75
Metro	-11.55	-0.33	0.09	0.55	2.68
Moran's *I*	0.01				
Adjusted R^2	0.72				
AICc	13,215				

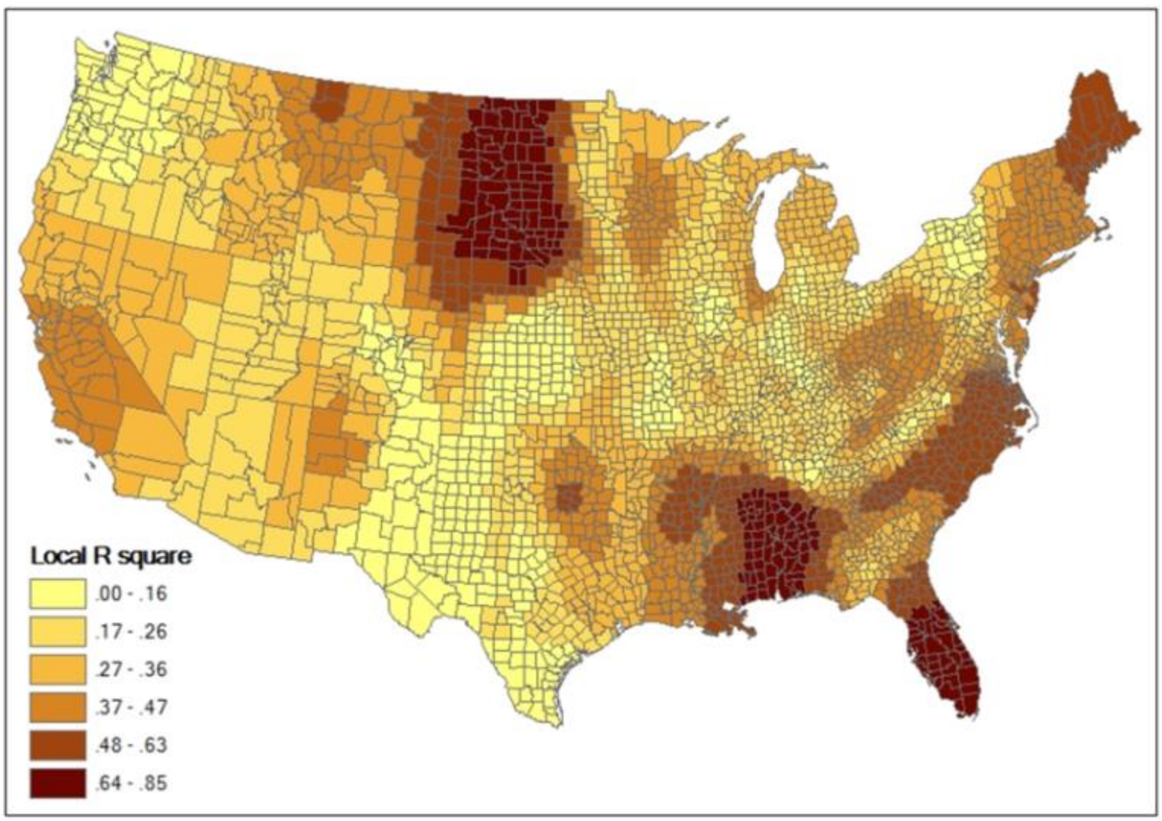

Figure 3 Coefficients of determination (R^2) from the GWR model

Figures 4 shows the spatial patterns of the GWR model coefficients. The intercepts are lower in the mountain areas and the northeast counties, indicating generally lower obesity in those areas (Figure 4a). Figure 4b shows along the west coast and some areas inwards there are significantly positive coefficients of the racial disparity. Poverty rate is strongly associated with obesity in most of the counties, except for some counties in Colorado (Figure 4c). Further investigation into the areas with negative coefficients might be interesting, which however goes beyond the scope of this research. The consistency in the poverty coefficients leads to the general consensus that socio-economic disadvantage/poverty might be the prevalent factor of the obesity problem in the U.S. counties. The relationship between the street connectivity and obesity is negative in most counties, with outliers of slightly positive values in the mountain areas (Figures 4d and 4e). The outliers are mainly in low population density areas. It suggests that in areas of low population density, increase in street connectivity or walkability may not reduce obesity. The ratio of fast-food to full-service restaurants is strongly and positively related to obesity rate in the northeast areas and some counties from the state of Washington (Figure 4f). However, this variable and the Urbanicity variable (Figure 4g) do not relate much to the obesity problem in most area of the country. Therefore, they are not included in the discussion of spatial clusters in the following section.

4.3.3 Regionalization

The coefficients maps have strong spatial correlation due to the use of local samples in the GWR model. The spatial pattern of coefficients and their *t* values reflect some underlying physical or social-cultural mechanisms. For example, in Figure 4b we can observe clusters of strong positive coefficient values of racial heterogeneity in the north-west and the west coast area, covering the major areas of California, Nevada, Idaho, Montana, North Dakota, and a part of

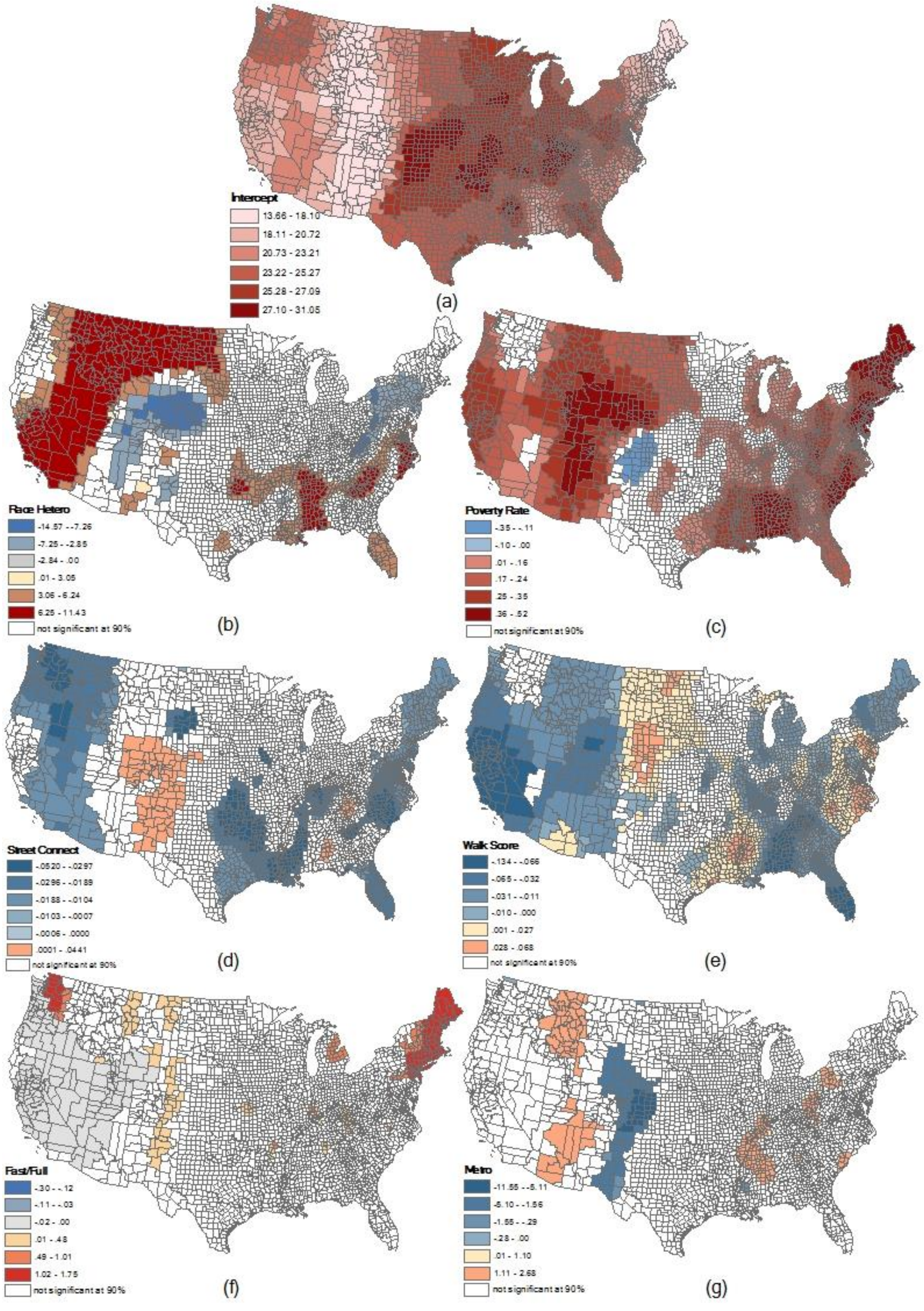

Figure 4 GWR coefficients: (a) Intercept; (b) Race Heterogeneity; (c) Poverty Rate; (d) Street Connectivity; (e) Walk Score; (f) Fast Food Ratio; (g) Urbanicity

South Dakota. If any future research is to be conducted on the obesity problem of different racial groups, these areas would be interesting. Therefore, regionalization of coefficients and their significance level can define geographical regions of high and low significant and non-significant coefficients. By creating those regions, it could better reveal the heterogeneity among the U.S. counties in their obesity problem. Three variable coefficients: race heterogeneity, poverty rate and street connectivity are used in the regionalization analysis. Other variables in the regression are omitted because their significant levels are generally low (Figure 4). The procedure to delineate the regions is as following:

1. For each variable, the counties are classified to three codes based on the sign of their coefficient and significance at 95%: 1 – significantly positive, 2- significantly negative, and 3 – not significant at 95%.
2. Use the "dissolve" algorithm in the GIS to eliminate the boundaries of counties in the same class and spatially adjacent to each other. This will generate regions representing homogeneous area of each variable, e.g. the coefficient values in the region are all positive or negative or non-significant.
3. Generalize the region maps by eliminating those smaller ones. The goal of the generalization is to avoid the regionalization being too fragmented. Remaining are 3-4 large regions for each coefficient after the generalization.
4. Intersect the three region maps to create the final regionalization map (Figure 5).

The regionalization map shows the U.S. counties are grouped as 16 regions in 7 classes. The classification is based on the signs of the coefficients of the selected three variables – poverty, racial heterogeneity, and street connectivity, as summarized in Table 4. Class 1 includes the states of New York, Connecticut, Pennsylvania, and Maryland. The two significant variables in this area are poverty (+) and race heterogeneity (-). The symbols + or – in the parenthesis represent a positive or negative sign of the coefficient. Class 2 includes multiple clusters scattered in the map, including the eastern part of the Gulf Coast, the south west mountain areas of Utah, Arizona, and part of Colorado and New Mexico, the Great Lakes area and its basin, and the area around Memphis, Tennessee. In these areas, the only significant variable is poverty rate. Class 3 includes major areas of California, Montana, Wyoming, North Dakota, the west Utah and small part of the south areas. None of the three variables is significant in these areas, suggesting that the regression model could not explain much of the variability of the obesity problem. Class 4 is located at the northeast and northwest corners of the map, as well as the border area between Texas and Louisiana. The two significant variables in this area are poverty (+) and street connectivity (-). It suggests that in this area, policies that help the poor or promote walkable environment would help in reducing obesity. Class 5 includes the coast of Virginia and North Carolina - or so called "the Dominion of Atlantic", Nevada, east part of California, Oregon and Washington, and most areas of Idaho. All three variables are significant in this class: Race (+), Poverty (+), Street Connectivity (-). Policies related to these variables would all be effective. Class 6 includes the adjacent areas of Utah, Colorado, Nebraska and Wyoming. Class 6 has a positive sign of street connectivity and a negative sign of race heterogeneity, both of which are against the general hypothesis of the regression model. The population density is generally low in this region. Class 7 is the central zone of the U.S., from Texas all the way up to the north border of the country. None of the three variables is significant in these areas. In other words, regressions cannot explain much of the variability of the obesity problem there. More variables should be included to study the obesity problem in these areas.

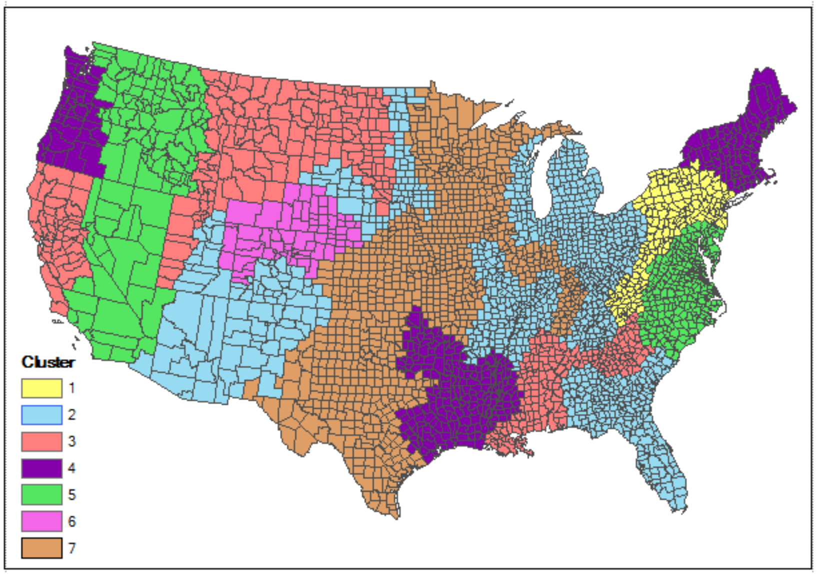

Figure 5 Geographical regions created from the GWR coefficients

Table 4 Classification of the regions based on the coefficient values

Classes	Race Heterogeneity	Poverty Rate	Street Connectivity
1	-	+	0
2	0	+	0
3	+	+	0
4	0	+	-
5	+	+	-
6	-	+	+
7	0	0	0

("+" means positive significant, "-"means negative significant, "0" means not significant)

4.4 Discussion

As the first attempt to use the Walk Score and street connectivity at the county level concerning public health, this research confirms the previous findings about the role of walkability in reducing obesity at the community level (Frank, Andresen and Schmid 2004). Both the global OLS model and local regression model have showed that Walk Score is a significant factor to explain variability of obesity in the U.S. The aggregated Walk Score at the county level from the Front Seat algorithm is proven significant in modeling obesity by regressions. In other words, a feasible way of measuring walkability was demonstrated at the county scale to be used in research of public health. The consistency between the global and local regression models suggests the generality of this approach to measuring walkability. Nonetheless, some outliers in the local regression model can be observed in Figure 4e. These counties, mainly distributed in the central zone of the country either have positive coefficients (the higher the walk score, the higher the obesity) or non-significant (walk score does not matter). These outliers might be caused by the inconsistency of the data used in the Walk Score algorithm. The use of centroid of census tracts to approximate the population centers in the algorithm might be one of the reasons.

While the global OLS regression model can measure the relationship between the obesity rate and the six explanatory variables: race heterogeneity, poverty rate, street connectivity, walk score, ratio of fast-food- to-full-service and urban-rural classification, the local regression model, GWR has its strength in finding geographical heterogeneity among the counties by the clustered spatial pattern of their coefficients. In fact, the spatial patterns of the coefficients are more favorable than the regression itself to a geographical analysis. General statistic methods used in Human Geography have been criticized for the attempt to generalize human objects and neglect the spatial structure of the society. The use of the localized regression model (GWR) compensates the weakness of the statistic models that neglect spatial heterogeneity. It turns out the GWR is more powerful in explaining the variability of obesity in use with the selected independent variables: race heterogeneity, poverty rate, street connectivity, walk score, ratio of fast-food- to-full-service and urban-rural classification. The spatial pattern of the coefficients is actually more interesting to Human Geographers than the regression itself. In each of the coefficient maps (Figure 4), one can visually identify distinguishable areas and clusters. It is evident that the public health policies cannot depend upon a global model. For example, poverty rate is identified as a significant positive contributing factor of obesity by the global model; but from the local model areas with negative or non-significant coefficients were able to identified, indicating that the global model's conclusion does not apply to these regions (Figure 4c). Public policies should be flexible in accordance to the unique characteristics of each region.

Furthermore, I would like to comment on the methodologies used in our research and by others in Human Geography studies. Our regionalization analysis partitions the entire study area into multiple patches that have unique characteristics regarding to their coefficient values and significance levels from the local regression model. The outcome is similar to what has been used by the Regional Geography paradigm. Regional Geography studies the unique combination of characteristics in an area (Peet 1998). Despite the similarity in their form and descriptive nature, our approach is fundamentally different from that of the traditional Regional Geography that has been criticized of its lack of scientific justifications, and of that the regions defined from the traditional approach are subjective and unpredictable. In contrast, our approach is based on the quantitative information from the regression models – i.e., the region divisions are empirically defined. The sign of the coefficients and their significant level (95%) were used as a threshold to define different regions. Therefore the regions created from our regionalization analysis are predictable and scientifically justifiable, and essentially it is GIS that makes such an approach possible. Hence, one of the possible purposes of this paper is to illustrate and promote the use of GIS spatial analysis and statistics on public health studies.

The unique characters of the classes defined in table 4 for the regions in Figure 5 could improve the policy-making procedure of the obesity problem. It helps answering two types of questions: "What measures could be taken to reduce the obesity risk in area X?" and "What are the areas that measure Y could reduce obesity rate?" To answer the first question, one shall read the class of area X from the map of Figure 5 and find whatever variables that are significantly defined for that class from Table 4. To answer the second question, one just needs to look up from Table 4 the classes that marked as significant measure of Y and then refer to the Figure 5 map to find the counties in those classes. In this way, policy-making personals with no expertise in GIS and quantitative methods would be able to dissect such a report.

Although the ultimate goal of public health research is to thoroughly understand the obesity problem related to the physical and socio-economic conditions, this research only focused on several built environment variables and social status variables. The variables were selected according to the major hypotheses about obesity (Rundle, Roux and Freeman 2007). Even for built environment, there are many other variables that were not selected, such as land use mix, access to park and neighborhood crime rates, which were mentioned in previous research (Talen and Anselin 1998). Furthermore, individual's socioeconomic status such as age, gender, income, marital status, education level and employ status was not taken into account. To improve the understanding of obesity and built environment associations, it is possible to adopt some space-time analysis framework, such as stratifying different years instead of analyzing one epidemic year. By doing so it could provide more detailed patterns of spatial autocorrelation changes of obesity-built environment relationship. Moreover, weather was not include as a variable because it was not a common practice in previous research; but as suggested from our GWR model analysis of the clustering pattern of the counties, weather might be an explanatory factor that results in such a spatial pattern. At last, linear regression cannot handle non-linear relationships. Certain transformation will be necessary if non-linear terms are identified.

Although we do not observe any non-linearity in the variables used in this research, cares should be taken if more variables are included in the future work.

4.5 Summary

To summarize, in this research the obesity problem and related built environment factors were analyzed over the counties in lower 48 states and DC by using the regression models with a GIS. A global model was used to analyze the overall relationship and GWR model to identify regional differences. The agreement among most counties about the poverty rate, street connectivity and walk score was found in relation to obesity; I also found different model coefficients among the counties about race heterogeneity, food environment and urban-rural classification. These findings were translated to qualitative inferences that could help policy making. GIS made the local regression and regionalization possible and converted the quantitative statistics to a geographical analysis problem. Such data analysis methodology and framework could enhance our understanding of the obesity problem over the U.S. I expect similar approaches are to be applied to other public health problems in the U.S. or other countries.

Chapter 5 Built Environment and Obesity by Urbanicity in the U.S.

5.1 Introduction

Obesity is a major risk factor for heart disease, diabetes, stroke, depression, sleep apnea, osteoarthritis, and some cancers (Ahima and Lazar 2013). Regular leisure time physical activity can help control weight and improve health. However, less than half (48.4%) of adults of 18 years of age and over meet the Physical Activity Guidelines for aerobic physical activity in 2011 (National Center for Health Statistics 2013), and more than one-third (34.9%) adults were obese in 2011-2012 (Ogden et al. 2013). The medical costs for obese people were $1,429 higher than those of normal weight in 2008 (Finkelstein et al. 2009). Obesity prevalence rates vary a great deal across states from 20.5% in Colorado to 34.7% in Louisiana in 2012 (CDC 2012), and even more among smaller geographic areas such as counties.

The cause of obesity arises from a positive energy balance over time. Energy intake is basically from food and drink, and energy consumption is related to individual's physical activity. An individual with a high level of consumption of fast foods and sugar-sweetened beverages (Pereira et al. 2005, Schulze et al. 2004) and a low level of physical activity (Koh-Banerjee et al. 2003) has a high risk of obesity. The obesogenic environment thesis suggests that disparities of obesity prevalence are attributable to differentiated exposure to a healthy food environment that promotes healthier dietary choices and built environments that encourage physical activities (Swinburn, Egger, and Raza 1999; (Powell, Spears and Rebori 2010). Built environment refers to human-made resources and infrastructure designed to support human activity, such as buildings, roads, parks, restaurants, grocery stores and other amenities, as compared with the natural environment (Pierce et al. 2012).

There is a large body of literature examining the relationship between built environment (including factors such as access to healthy food, distance to nearby amenities, walkable urban form and neighborhood safety) and obesity (Feng et al. 2010, Papas et al. 2007, O Ferdinand et al. 2012, Durand et al. 2011). However, due to challenges of data requirements and computation complexity for measuring obesogenic built environments, few studies are on a national scale until very recently. Among the recent national studies, Wen & Kowaleski-Jones (2012) and Wen et al. (2013) considered two major built environment factors such as distance to the nearest parks and street connectivity, and Wang et al. (2013) focused on the role of population-adjusted street connectivity. The present nationwide analysis considers two built environment factors that have not been included in previous studies of such a scale, namely walk score and the ratio of fast-food to full-service restaurants.

Furthermore, recent literature suggests that the linkage between built environment and physical activity (and thus obesity) vary in different geographic settings such as urban versus rural areas (Monnat and Pickett 2011, Ding and Gebel 2012, Ewing et al. 2014). Urban neighborhoods have more sidewalks, mixed land uses, better street connectivity and more playgrounds than rural areas (Lopez and Hyness 2006). Within urban area, children in inner city neighborhoods are engaged in less physical activity than those in suburban areas (Weir, Etelson and Brand (2006). More anxiety about neighborhood safety may deter physical activity and help explain a higher obesity rate in inner city areas (Felton et al. 2002, Wilson et al. 2004). A recent study shows that better street connectivity reduces obesity risk only in suburbia of large metropolitan areas, not central city areas or smaller metropolitan or rural areas (Wang et al. 2013). This research examines the association between built environment and obesity with an emphasis on the likely variability across different levels of urbanicity.

On the methodological front, multilevel models are common in public health research. Individual behaviors such as eating habit and physical activity do not occur itself; rather, they are influenced by socio-environmental factors including built environment (Huang et al. 2009). This study uses the multilevel modeling approach to analyze the influence of built environment on adult physical inactivity and obesity in the U.S. while controlling for individual attributes (e.g., race, age, gender, marital status, education attainment, employment status, income, and whether an individual smokes). The next section explains data processing and definition of variables. Section 5.3 presents the multilevel models and related results. Section5.4 discusses the results and highlight findings. The section is concluded with a brief summary and discussion of future research.

5.2 Data Sources and Variable Definitions

5.2.1 Individual Variables from BRFSS

The Behavioral Risk Factor Surveillance System (BRFSS) is an annual health-related telephone survey system for tracking risk behaviors, health conditions, and use of preventive services in the U.S. since 1984. Since 2011, the survey data added cell phone only respondents to landline respondents that were covered by the survey data for 1984-2010. We used the 2012 BRFSS data set (http://www.cdc.gov/brfss/annual_data/annual_2012.html), the most recent one available at the time of this research being conducted. The data set contains a large volume of individual data geocoded to county. After eliminating the records with missing values for variables used in this study, the study area includes 328,156 observations from the BRFSS in the 48 conterminous states and Washington D.C.

The BRFSS data contains two dependent variables used in this research: *physical inactivity* and *obesity*. Physical inactivity refers to no leisure-time physical activity or exercise in the last month as reported. Individuals with BMI > =30 were considered as obese. They are coded as binary, i.e., 1 for no physical activity and 0 otherwise, 1 for being obese and 0 otherwise.

Individual independent variables are also from the BRFSS data set (Table 1). In addition to *age* (18+), "age squared" is added to check the curvilinear impact of age in the multilevel models in the next section. *Race-ethnicity* is categorical including non-Hispanic Black, Hispanic, and others with non-Hispanic White as the reference category. Binary variables include *sex* (female as the reference category), *employment status* (not employed as the reference category), *marital status* (currently not married as the reference category), and *smoker* (non-smoker as the reference category). *Education* and *income* are numerical such as: education level = 1-4 (1 for "did not graduate high school", 2 for "graduated high school", 3 for "attended college or technical school", 4 for "graduated from college or technical school"), income level = 1-5 (1 for "less than $15,000", 2 for "15,000 to less than $25,000", 3 for "$25,000 to less than $35,000", 4 for "35,000 to less than $50,000", 5 for "$50,000 or more."

5.2.2 Rates of Physical Inactivity and Obesity for Various Socio-Demographic Groups

Table 5 summarizes the sample distributions across the individual socio-demographic variables reported in the 2012 BRFSS. The overall physical inactivity rate is 23.49%, and the overall obesity rate is 29.25%. Among the four major racial-ethnic groups, non-Hispanic whites account for the vast majority (80%) and can be considered as the reference category, both physical inactivity rate (PIR) and obesity rate (OBR) for non-Hispanic Blacks or Hispanics are higher than the averages and more so for non-Hispanic Blacks, and the PIR for others is slightly higher than the average but the OBR for others is slightly lower than the average. The PIR increases with age, so does the OBR till the 54-65 age group but drops in the 65+ age group. The

Table 5 BRFSS Individual Variables and Distributions

Demographic Variables		Sample Size	Physical inactivity rate (PIR, %)	Obesity rate (OBR, %)
All		*328,156*	*23.49*	*29.25*
Race-ethnicity	Non-Hispanic Whites*	262,745	22.29	27.69
	Non-Hispanic Blacks	29,697	30.82	42.67
	Hispanics	20,154	27.80	31.79
	Others	15,560	24.22	26.68
Age	18-29	27,817	13.78	20.41
	30-41	47,614	16.87	29.82
	42-53	69,925	21.08	31.96
	54-65	90,479	24.47	33.25
	65+	92,321	30.70	25.65
Gender	Men	139,697	21.07	29.26
	Women*	188,459	25.29	29.24
Married	Yes	175,530	19.73	28.58
	No*	152,626	27.82	30.02
Education	Did not graduate high school (1)	25,139	44.17	36.28
	Graduated high school (2)	92,497	32.23	32.81
	Attend college or technical school (3)	89,963	22.82	31.60
	Graduate from college or technical school (4)	120,557	12.98	23.29
Employed	Yes	144,165	17.65	29.42
	No*	183,991	28.07	29.11
Income	Less than $15,000 (1)	38,300	40.22	35.89
	$15,000 to less than$25,000 (2)	58,007	34.72	33.01
	$25,000 to less than $35,000 (3)	37,480	28.30	30.33
	$35,000 to less than$50,000 (4)	48,081	22.63	29.95
	$50,000 or more (5)	146,288	13.71	25.52
Smoker	Yes	55,530	33.46	26.29
	No*	272,626	21.46	29.85

Note: * indicates the reference category in the group.

latter suggests a curvilinear association of the variable “age” with obesity. The PIR for women is higher than men, but their OBRs are about the same. The married has a lower PIR and a lower OBR than their married counterparts. Both the PIR and OBR drop with increasing educational attainment. The employed has a lower PIR than the unemployed, but their OBRs are very close. Like “educational attainment”, both the PIR and OBR drop with increasing income. Smokers have a higher PIR but a lower OBR. For the most part, the trend for the PIR is consistent with that of OBR. However, they also differ in several cases such as the minor discrepancy in their associations with age, gaps in their associations with marital status and employment status, and the major contrast in the associations with smokers/non-smokers. The above observations do not consider the joint effects of multiple variables let alone the neighborhood effects, and thus are preliminary.

5.2.3 Neighborhood Variables at the County Level from Census and Other Sources

All neighborhood variables are defined at the county level as county is the smallest geographic unit identified in the BRFSS dataset. Guided by the literature, two social-demographic variables are included: poverty rate and race heterogeneity, both derived from the Census 2010 data. *Poverty rate* is the estimated percent of people of all ages in poverty. *Racial-ethnic heterogeneity* reflects the racial-ethnic composition defined as $1- \sum p_i^2$, where p_i is the fraction of the population in a given group (Sampson 1989). This study includes six racial-ethnic groups (Non-Hispanic Whites, Blacks, Asians/Pacific Islander, Hispanics, American Indians/Alaska Natives, and others) for calculating the index in a county. The heterogeneity index ranges between 0 and 1. If the value equals to 0, it means that there is only one racial/ethnic group in the unit; while a value approaching 1 reflects a maximum heterogeneity.

The built environment is also measured at the county level, and includes street connectivity, walkability and food environment. Intersection density (i.e., number of intersections per km^2) is the most commonly used index to measure *street connectivity*. Ball et al. (2012) concluded that street connectivity is not significantly associated with either adult BMI or BMI categories. Wang et al. (2013) argued that intersection density varies a great deal within a large geography area such as county, and the conventional measure of street connectivity can be biased for a county with the majority of population concentrated in limited urban area. Therefore, "population-adjusted street connectivity" is a preferred choice. In implementation, intersection density is calculated at the census tract level and then aggregated to the county level by computed a weighted average value (using population as weight). *Walkability* is measured by the Walk Score (http://www.walkscore.com/) based on the algorithm developed by the Front Seat Management (http://www.frontseat.org/). It calculates the Euclidean distances from a point of interest to nearby amenities such as food, retail, education, recreation, and entertainment, and then integrates them by a linear combination of these distances with weights that account for facility type priority and a distance decay function (Front Seat 2013). Similarly, walk score is first obtained at the census tract level and then aggregated to the county level by computing the population-weighted averages. Food consumption relying on fast food restaurants is likely to promote more meals or increase consumption of high fat meals, leading to higher caloric intake (Michimi and Wimberly 2010). Some studies used the number of fast-food restaurants per capita to measure the food environment (Wang et al. 2007, Jay 2004, Lamichhane et al. 2013). Such an approach does not account for the availability of choices between healthy and unhealthy food by consumers. This research uses the *ratio of fast-food to full-service restaurant numbers* at the county level to measure food environment. The restaurant data is extracted from the 2012

County Business Patterns (CBP), an annual series providing subnational economic data by industry (http://www.census.gov/econ/cbp/). In the dataset, restaurants are classified into fast food and full service. To our knowledge, walk score and food environment are for the first time used in a national study of built environment for obesity risk.

For *urbanicity*, we first use the 2013 NCHS Urban-Rural Classification Scheme for Counties prepared by the National Center for Health Statistics (NCHS), in accordance with the 2010 OMB (Office of Management and Budget) standards for defining metropolitan and micropolitan areas (Ingram DD and Franco SJ 2014). There are six urban-rural categories such as large central metro, large fringe metro, medium metro, small metro, micropolitan and noncore, where noncore is used as the reference category for coding. In order to more accurately capture urbanicity, this research also uses another definition based on the 2010 Census Urban and Rural Classification (Census, 2014). The Census defines an urban area with minimal criteria of population and population density using much smaller geographic units such as census tracts and census block. For each county, its urbanicity is defined as a continuous urbanization ratio, i.e., urban population in urban areas over the total population in the county ((Wang et al. 2013)).

5.2.4 Variability of County-Level Variables across NCHS Urban-Rural Categories

Figure 6a-f show the spatial patterns of the aforementioned county-level variables. Given the emphasis of examining the association of built environment and obesity by urbanicity, it is valuable to examine the variability of each county-level variable across the urbanicity categories (here based on the NCHS classifications as an example). In addition to the two social-demographic variables and three built environment variables, we also calculate the average physical inactivity and obesity rates in the counties.

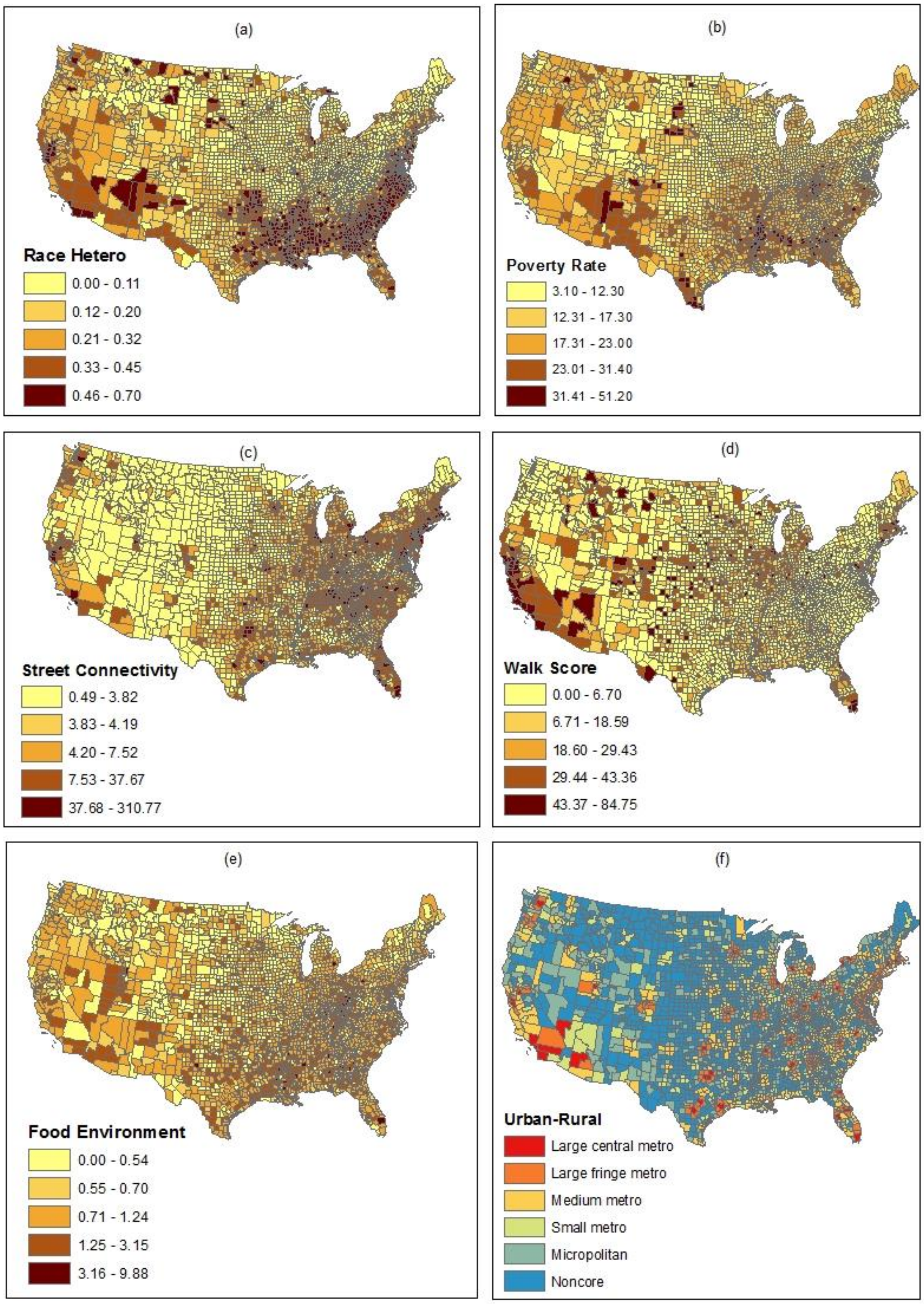

Figure 6 County-level variables: (a) racial-ethnic heterogeneity; (b) poverty rate; (c) street connectivity; (d) walk score; (e) food environment; (f) urbanicity

As the urbanicity decreases from large central metro to noncore counties, (1) both the average physical inactivity rate (PAR) and obesity rate (OB) increase from the lowest to the highest, so does the average, as shown in Figure 7a; (2) both the average racial-ethnic heterogeneity index and walk score decrease in general (only slightly higher in medium metro than in large fringe metro), as shown in Figure 7b and 7e; and (3) both the average street connectivity and fast-food to full-service restaurants ratio decrease, as shown in Figure 7d and 7f. For the average poverty rate, the order is noncore > micropolitan > large central metro > small metro > medium metro > large fringe metro, as shown in Figure 7c. In other words, the poverty rate is the highest at the two ends of urbanicity (rural counties such as in noncore or micropolitan and urban core such as in large central metro) and declines toward the middle with the lowest poverty rate in suburbia (fringe) of large metro.

Are differences in the average values statistically significant across the urban-rural classifications? This may be answered by conducting the ANOVA (analysis of variance) test. Here a regression model is introduced for the same purpose for its simplicity and easy interpretation (Wang et al. 2014). Five dummy variables can be used to code the six urbanicity categories. The noncore counties are selected as the reference type and coded as $X_1=X_2=X_3=X_4=X_5=0$. Counties of any other type are coded by assigning a value "1" to one of the dummy variables and "0" to the rest four (e.g., $X_1=1$ and $X_2=X_3=X_4=X_5=0$ for large central metro counties; $X_2=1$ and $X_1=X_3=X_4=X_5=0$ for large fringe metro counties; and so on). Denoting the variable of interest (say, "obesity rate (OBR)") as the dependent variable Y, the model is written as

$$Y = b_0 + b_1 X_1 + b_2 X_2 + b_3 X_3 + b_4 X_4 + b_5 X_5 \tag{9}$$

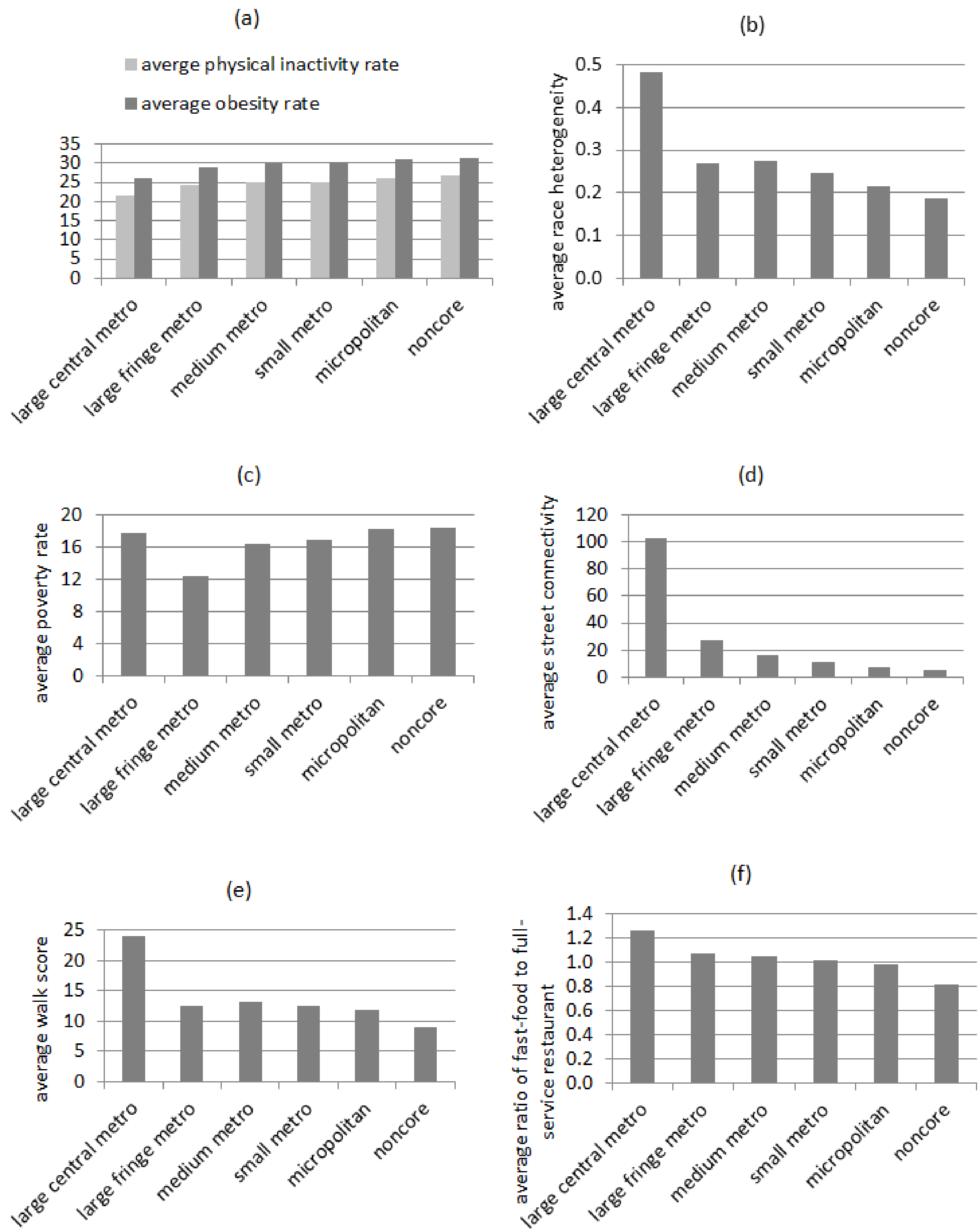

Figure 7 Averages of county-level variables by urbanicity: (a) physical inactivity rate and obesity rate, b) racial-ethnic heterogeneity, (c) poverty rate, (d) street connectivity, (e) walk score; (f) food environment

In Equation (9), the intercept b_0 is the average value of the variable for the reference category (here noncore counties), coefficient b_i is the difference in averages of the variable between the counties of a category coded X_i =1 and the counties of the reference category, and corresponding t-values indicate whether the differences are statistically significant. The results are reported in Table 6. For example, for obesity rate, the average for noncore counties is 31.259, for large central metro counties is 31.259-5.359=25.900, for large fringe metro counties is 31.259-2.261=28.998 and so on. The results reported in Table 2 are consistent with Figures 7a-7f. Moreover, the corresponding t-values indicate that the differences between the reference category (noncore counties) and any other types of counties are statistically significant in most cases.

Once again, the above discussion is based on analysis of aggregated data for a single variable at a time and has limited value. The actual effect of county-level variables needs to be examined in a multilevel modeling schema.

5.3 Multilevel Modeling

5.3.1 Overall Models

Multilevel modeling (MLM) examines the risk of individual health behavior (i.e., physical inactivity) or outcome (i.e., obesity) by considering both individual and neighborhood-level (county) variables. Tables 7 and 8 present the results, i.e., odds ratio of multilevel logistic models for the study area. There are four models for each, labeled “PI” and “OB” for physical inactivity and obesity, respectively. In Table 7, model 1 is the unconditional model with only individual-level predictors, model 2 adds the county-level variables. In Table 8, in order to capture the effect of urbanicity, model 3 adds five dummy variables to code the six NCHS classifications (noncore county as the reference category), and model 4 uses the continuous

Table 6 Regressions for testing variability of county-level variables across urban-rural classifications

	Physical inactivity rate	Obesity rate	Racial-ethnic heterogeneity	Poverty rate	Street connectivity	Walk score	Food environment
Large central metro	-5.254*** (-8.55)	-5.359*** (-9.96)	0.296*** (15.19)	-0.753 (-0.96)	97.903*** (44.37)	15.137*** (8.24)	0.440*** (5.11)
Large fringe metro	-2.765*** (-9.49)	-2.261*** (-8.86)	0.080*** (8.66)	-6.085*** (-16.41)	22.407*** (21.41)	3.628*** (4.16)	0.257*** (6.29)
Medium metro	-1.882*** (-6.47)	-1.175** (-4.61)	0.087*** (9.45)	-2.062*** (-5.57)	11.312*** (10.83)	4.262*** (4.90)	0.235*** (5.76)
Small metro	-1.741*** (-5.89)	-0.907*** (-3.51)	0.057*** (6.14)	-1.487*** (-3.96)	6.274*** (5.91)	3.613*** (4.09)	0.197*** (4.76)
Micropolitan	-0.766*** (-3.21)	-0.260 (-1.24)	0.028*** (3.77)	-0.242 (-0.80)	2.277** (2.66)	3.047*** (4.27)	0.166*** (4.97)
Non-core	*26.908**** *(197.26)*	*31.259* *(261.63)*	*0.188**** *(43.40)*	*18.437**** *(106.22)*	*4.976**** *(10.16)*	*8.869**** *(21.74)*	*0.817**** *(42.70)*
R^2	0.051	0.051	0.095	0.085	0.428	0.029	0.025

***$p \leq 0.001$, **$p \leq 0.01$, *$p \leq 0.05$

Table 7 Odds ratios of multilevel logistic models for physical inactivity (PI) and Obesity (OB)

	Model PI1	Model PI2	Model OB1	Model OB2
Individual variables				
Non-Hispanic Black	1.205***	1.196***	1.666***	1.671***
Hispanic	1.217***	1.219***	1.055**	1.059***
Other race/ethnicity	1.173***	1.171***	0.922***	0.925***
Age (18+)	1.037***	1.036***	1.125***	1.124***
Age^2	0.999***	0.999***	0.999***	0.999***
Male	0.867***	0.868***	1.060***	1.062***
Married	1.002	0.998	0.952***	0.947***
Education (1-4)	0.712***	0.714***	0.833***	0.834***
Employed	0.843***	0.844***	0.781***	0.782***
Income (1-5)	0.818***	0.820***	0.919***	0.922***
Smoker	1.571***	1.570***	0.612***	0.611***
County variables				
Racial-ethnic Heterog.		0.985		0.885*
Poverty		1.008***		1.010***
Street connectivity		0.999*		0.998***
Walk Score		0.998***		0.999
Fast food ratio		1.052***		1.049***
AIC	*334287.3*	*334158.8*	*390893.6*	*390705.0*

***$p \leq 0.001$, **$p \leq 0.01$, *$p \leq 0.05$

variable "urban ratio" and its square term. See the previous section on the definitions of reference categories for several categorical individual variables such as race-ethnicity, sex, marital status, employment status, and smoker or nonsmoker.

Based on Tables 7 and 8, the effects of individual variables *largely* confirm the preliminary observations from Table 5 on the distributions of PI and OB rates by various socio-demographic groups, but some details are new. Even when the findings may appear consistent from the two tables, the MLM results have more clarity for the statistical significance associated with each variable and are also more reliable because the effects of neighborhood variables are controlled for. The differences are highlighted here. Note that the findings are also consistent across the four PI models and across the four OB models.

Table 8 Odds ratios of multilevel logistic models with urbanicity for physical inactivity (PI) and Obesity (OB)

	Model PI3	Model PI4	Model OB3	Model OB4
Individual variables				
Non-Hispanic Black	1.194***	1.183***	1.675***	1.681***
Hispanic	1.222***	1.211**	1.059***	1.064***
Other race/ethnicity	1.170***	0.150***	0.926***	0.930***
Age (18+)	1.036***	1.035***	1.124***	1.122***
Age^2	0.999***	0.999***	0.999***	0.999***
Male	0.868***	0.868***	1.062***	1.061***
Married	0.997	0.997	0.947***	0.945***
Education (1-4)	0.714***	0.715***	0.834***	0.836***
Employed	0.844***	0.845***	0.783***	0.782***
Income (1-5)	0.820***	0.819***	0.922***	0.920***
Smoker	1.570***	1.568***	0.611***	0.607***
County variables				
Racial-ethnic Heterog.	1.008	1.105	0.847***	0.897*
Poverty	1.008***	1.007***	1.011***	1.007***
Street connectivity	0.999	1.000	0.998***	0.999***
Walk Score	0.998***	0.998***	0.999	0.999
Fast food ratio	1.053**	1.062***	1.042***	1.040***
Large central metro	0.864***		1.003	
Large fringe metro	0.973		1.088***	
Medium metro	0.928***		1.063**	
Small metro	0.921***		1.057*	
Micropolitan	0.959*		1.051*	
Urban ratio		1.048		1.301***
Urban ratio squared		0.795*		0.702***
AIC	*334145.2*	*325877.5*	*390696.7*	*380893.1*

****p*≤ 0.001, ***p*≤ 0.01, **p*≤ 0.05

Non-Hispanic Black and Hispanic have higher risks of physical inactivity and obesity than their non-Hispanic white counterparts; and between the two major minority groups, the odd ratio of obesity is even higher for non-Hispanic Black than for Hispanic, but the odd ratio of physical inactivity is reversed (i.e., higher for Hispanic than for non-Hispanic Black). The latter finding (the reversed gaps in PI and OB between the two groups) is new from MLM. Both risks of physical inactivity and obesity increase initially with age and then drops after passing a certain age. The curvilinear effect of age is present in both PI and OB here, but absent for PI from Table 5. Males tend to be more physically active, but bear a higher risk of obesity. The latter finding is also new from MLM (certainly much stronger and more evident). In the MLMs, marital status is not significant for physical inactivity, but being married is negatively associated with the risk of obesity. This suggests that the large gap in PI between the married (19.73%) and the unmarried (27.82%) from Table 5 may be caused by other confounding factors (age and others), and does not necessarily imply that the marital status is a factor in influencing physical activity. The lower obesity ratio for the married (also from Table 5) remains after other variables are controlled for. Higher education, being employed and higher income are all associated with lower risks of physical inactivity and obesity. Smokers have a higher risk of being inactive but a lower risk of obesity.

There are several discrepancies in an individual variable's associations with PI and OB risks. It is understandable that smokers may tend to be more physical inactive while maintaining lower body weight since nicotine consumption increases energy expenditure and could suppress appetite (Chiolero et al. 2008). It is rather puzzling in others (e.g., the reversed gaps between Non-Hispanic Black and Hispanic, lower PI risk but higher OB risk for males, indifferent for PI but lower OB risk for the married). Why is the effect on PI not transferred to the same one (or even the opposite one) on OB for the above population groups? Unless there is evidence of different behavior in food and beverage intakes or different metabolism, one may question the reliability of PI (a subjective assessment loosely defined) in comparison to OB (a rather more objective measure based on BMI) (Wang et al. 2013: 10-11). We will keep this in mind, and hereafter focus more on the MLM results on obesity.

Net of individual controls, models PI2 and OB2 in Table 7 add two socio-demographic variables and three built environment measures at the county level, and models PI3, PI4, OB3 and OB4 in Table 8 add the effect of urbanicity. Declining AIC values from model PI1 to PI2 to PI3 to PI4 and also from OB1 to OB2 to OB3 to OB4 confirm the value and validity of MLMs, particularly models PI4 and OB4 with urbanicity defined by urban ratio. Racial-ethnic heterogeneity is not significantly associated with physical inactivity but negatively associated with obesity. Poverty rate is positively associated with both physical inactivity and obesity risks. Among the built environment variables, the ratio of fast-food-to-full-service restaurants is positively associated with physical inactivity and obesity risks in all models. Street connectivity is negatively associated with obesity (but not significant with physical inactivity), and walk score is negatively associated with physical inactivity (but not significant with obesity). Physical inactivity largely decreases with the level of urbanicity (measured in either NCHS classifications or urban ratio), which is consistent with the preliminary observation from Table 5. However, based on model OB3, obesity risk is the lowest in noncore and large central metro counties (with no significantly statistical difference between them), and increases gradually in the order of micropolitan, small metro, medium metro, and large fringe metro. That is to say, with the exception of large central metro with the highest urbanicity, obesity risk climbs up with increasing urbanicity. It is captured by the curvilinear effect of urban ratio in model OB4, i.e., obesity risk increases with urban ratio and comes down after a certain urban ratio. This finding on obesity risk from Table 8 is different from the preliminary reading from Table 5. Again, one possible reason for the deviation between PI and OB models is the gap in measurement reliability between the two.

5.3.2 Models by Urbanicity Levels

In order to test the complexity of urbanicity's impact, we extract the subsets of data by urban-rural classifications. In other words, we are interested in examining whether the effects of individual and county-level variables are consistent in various geographic settings, here in different urbanicity levels. For the aforementioned reason, this subsection only presents the results on obesity. Tables 9 and 10 present the MLM results by the six NCHS urban-rural county categories and by the urban ratio ranges, respectively. Here we highlight the differences from those based on all samples in the study area. Among the individual variables, the effects of most of the variables (e.g., non-Hispanic Black, age, education, employment, income and smoker) remain consistent across all six NCHS categories (Table 9) or across the five urban ratio ranges (Table 10), but others vary. For example, both Tables 9 and 10 show that the higher obesity risk for Hispanic is no longer significant in the less urbanized areas, and other race/ethnicity tends to have lower obesity risk in the more urbanized areas but higher obesity risk in the less urbanized areas. The higher obesity risk for male suggested previously is only present in the middle pack of urbanicity levels (not significant in large central metro or noncore areas from Table 9, even a lower risk in completely urban areas not significant in completely rural areas from Table 10). While overall the married tends to have a lower risk of obesity, such an effect is not significant particularly in less urbanized areas (i.e., noncore or micropolitan from Table 9, completely rural from Table 10).

Among the county-level variables, racial-ethnic heterogeneity is now not significant in all areas, and poverty is no longer significant in moderately urbanized areas (i.e., small metro or micropolitan from Table 9, marginally or mostly urban areas from Table 10). Among the built environment variables, the relationship between street connectivity and obesity becomes

Table 9 Odds ratios of multilevel logistic models for obesity by NCHS urban-rural classification

	Large central metro (*N*=37,354)	Large fringe metro (*N*=50,806)	Medium metro (*N*=53,722)	Small metro (*N*=35,562)	Micropolitan (*N*=40,551)	Noncore (*N*=15,617)
Individual variables						
Non-Hispanic Black	1.866***	1.513***	1.689***	1.552***	1.606***	1.684***
Hispanic	1.121**	1.048	1.073*	1.080	1.010	0.973
Other race/ethnicity	0.824***	0.786***	0.941	1.027	1.138**	1.164*
Age (18+)	1.128***	1.122***	1.123**	1.126***	1.118***	1.108***
Age squared	0.999***	0.999***	0.999***	0.999***	0.999***	0.999***
Male	1.004	1.148***	1.056***	1.073***	1.047*	1.028
Married	0.893***	0.919***	0.964	0.950*	0.982	0.997
Education (1-4)	0.800***	0.825***	0.832***	0.845***	0.860***	0.894***
Employed	0.779***	0.790***	0.778***	0.750***	0.782***	0.828***
Income (1-5)	0.959***	0.928***	0.916***	0.900***	0.899***	0.914***
Smoker	0.660***	0.623***	0.604***	0.607***	0.546***	0.587***
County variables						
Race Heterogeneity	0.911	0.931	0.864	0.974	0.981	0.839
Poverty	1.019***	1.018***	1.007*	0.994	1.000	1.009*
Street connectivity	0.999***	0.998***	0.998	0.998	1.002	1.003
Walk Score	0.998***	0.999	0.999	1.000	1.001	1.001
Fast food ratio	1.006	0.999	1.152***	1.145**	1.198***	1.032
AIC	*61864.7*	*78092.6*	*88965.2*	*51172.4*	*63099.2*	*37431.2*

***$p \leq 0.001$, **$p \leq 0.01$, *$p \leq 0.05$

Table 10 Odds ratios of multilevel logistic models for obesity by urban ratio ranges

	Completely urban (0.99, 1.00]	Highly urban (0.90, 0.99]	Mostly urban (0.50, 0.90]	Marginally urban (0.01, 0.05]	Completely rural [0, 0.01]
Individual variables					
Non-Hispanic Black	1.907***	1.662***	1.605***	1.622***	1.681***
Hispanic	1.043	1.129***	1.084**	0.997	0.949
Other race/ethnicity	0.829*	0.809***	0.874***	1.099**	1.224**
Age (18+)	1.127***	1.125***	1.126***	1.117***	1.112***
Age squared	0.999***	0.999***	0.999***	0.999***	0.999***
Male	0.900**	1.082***	1.090***	1.061***	1.044
Married	0.853***	0.924***	0.946***	0.967*	0.990
Education (1-4)	0.797***	0.804***	0.836***	0.857***	0.872***
Employed	0.753***	0.797***	0.774***	0.779***	0.813***
Income (1-5)	0.966*	0.941***	0.914***	0.906***	0.914***
Smoker	0.657***	0.665***	0.616***	0.565***	0.563***
County variables					
Race Heterogeneity	1.029	0.731	0.881	0.865	0.981
Poverty	1.020***	1.014***	1.004	1.002	1.011***
Street connectivity	0.998***	1.000	0.998	0.998	1.004
Walk Score	0.999	0.998*	0.999	1.000	1.000
Fast food ratio	0.963	1.003	1.203***	1.135***	1.025
AIC	*21854.4*	*70995.4*	*138288.8*	*110971*	*38490.1*

***$p \leq 0.001$, **$p \leq 0.01$, *$p \leq 0.05$

insignificant in less urbanized areas, better walk score is only linked to reduced risk of obesity in large central metro areas (Table 9) or highly urban areas (Table 10), a higher ratio of fast-food to full-service restaurants is associated with a higher risk of obesity in moderately urbanized areas (i.e., medium metro, small metro and micropolitan from Table 9, marginally or mostly urban areas from Table 10). In other words, the positive effects of better street connectivity and walk score on lowering obesity risk are present in highly urbanized areas, and the food environment is in play more so in areas of middle-range urbanicity.

The consistency in results from Tables 9 and 10 validates the two systems of urban-rural classifications. Some of the effects of independent variables derived from the "global" model are altered in the "urbanicity-specific" models, more so for the neighborhood variables than individual variables. This implies that some obesity risk factors are sensitive to variation of geographic settings.

5.4 Discussion and Concluding Comments

The objective of this study is to explore the role of contextual attributes such as built environment in contributing to physical inactivity and obesity risks. While the measurements of neighborhood built environment are similar to those commonly investigated in the literature, the implementations at the national level, particularly walk score and food environment, are new. There is a significant association between built environment variables and physical inactivity/obesity, net of individual attributes and neighborhood socio-demographic characteristics. Another important highlight is the examination of possible variability of association between built environment and obesity across different urbanization levels.

Based on the BRFSS data, results from the multilevel models show that individual variables such as age, education level, employment status and income are consistent between their impacts on physical inactivity and obesity. There are some disconnections between impacts of other individual attributes (e.g., race-ethnicity, sex, marital status and smoking behavior) on physical inactivity and those on obesity. Barring distinctive behaviors of food-beverage intake or metabolisms among the various socio-demographic groups, one may suspect a possible reliability gap in the measurement of these two dependent variables.

Based on the global models using samples in the whole study area, county-level socio-demographic structure such as a lower racial-ethnic heterogeneity index or a higher poverty rate is linked to a higher obesity risk. Among the built environment variables, a poorer street connectivity and a more prominent presence of fast-food restaurants are associated with a higher obesity risk. While the effect of walk score is not evident in influencing obesity risk, a higher walk score is indeed linked to a lower rate of physical inactivity. Overall, obesity risk initially increases with the urbanicity level and then drops, resembling an inverted-V shape. The results lend support to the relevance of built environment in potentially influencing people's health behavior and outcome.

Finally, the analysis on data subsets reveals the variability of effects of both individual and county-level variables in areas of different urbanicity levels. For instance, with comparison to the findings from the global models, the higher obesity risk for Hispanic is no longer significant in the less urbanized areas, neither is the higher obesity risk for male in areas of the highest/lowest urbanicity, nor is the lower risk for the married in rural areas. These findings on the individual attributes call for more in-depth studies that may uncover possibly distinctive behavior of these demographic groups in different geographic environments. Similarly, for county-level built environment variables, better street connectivity and walk score lowers obesity risk only in the highly urbanized areas, and food environment seems to be more of a factor in areas of middle urbanicity levels. Both street connectivity and walk score reflect walkability, whose variability is most likely to play a role in people's health behavior across large cities but to a less extent in small-medium cities or rural areas. The prominent influence of food environment in areas of moderate urbanicity is interesting. One plausible theory may be that due to the ubiquity of fast-food restaurants in U.S., accessibility of fast food is fairly uniform in large cities or countryside and only exhibits a certain variability in areas between the two. Testing this theory or ones on built environment begs for data with finer geographic resolutions than the county level available to this study.

Several limitations of the study need to be acknowledged. The first issue concerns the data. Both measures of physical inactivity and obesity rely on the survey data from BRFSS. As pointed out previously, physical activity is loosely defined as “leisure-time physical activity in the last 30 days” reported by oneself, and raises the concern of reliability. In addition, county is the smallest geographic unit geocoded by the BRFSS data. A finer geography resolution would help us define built environment at a spatial scale that is more relevant to people’s activity space such as zip code area or census tracts (Krieger et al. 2003, Sturm and Datar 2005). The average size of the counties in the study area is 2,502.11 km^2. Urban planners assume that one quarter mile (0.4 km) is a comfortable range for pedestrians (Rundle et al. 2007). Secondly, the measurements of built environment can be more comprehensive in future work. Limited by data availability and time, this study does not include variables such as accessibility of recreational facilities (e.g., parks, gyms), presence of mixed land use, climate change and others that have been suggested to affect health behavior and outcome. Lastly, this study is cross-sectional without considering any temporal changes. The built environment defined is the present state of environment for an individual. A person’s BMI reflects the accumulated effect of one’s living environment and behavior, both of which may have changed. The research may establish the link between an environment factor and obesity, but cannot tell whether the neighborhood factor causes residents to live healthy or whether healthy individuals choose to live in neighborhood with such an environment.

Chapter 6 Multilevel Built Environment Features and Individual Odds of Overweight and Obesity in Utah

6.1 Introduction

Although Utah is among the states with the lowest obesity rates in the U.S., the estimated prevalence of overweight and obesity is over 60% according to the BeeWell Utah (http://home.utah.edu/~u0145007/Bee%20Well%20Utah/facts.html).

Multilevel modeling is commonly used in research on obesity etiology by incorporating both individual-level risk factors and neighborhood characteristics (Wen and Maloney 2011, Wang et al. 2013). Individual variables are usually obtained directly from surveys. Built environment factors are often measured and constructed at some neighborhood level(s) from various data sources. One challenge is to determine what constitutes an appropriate neighborhood scale or size in defining built environment. For example, in analyzing overweight risks, Gordon-Larsen et al.(2006) used an 8-km radius around one's residence as a reasonable range to define available physical activity facilities. In a study on overweight risks in preschool children, Burdette and Whitaker (2004) defined relevant environment as distances from a child's residence to the nearest public playground and fast food restaurant. Rutt and Coleman (2005) defined neighborhood as a 0.25–mile radius around each person's residence to examine the association between mixed land use and BMI. In examining the impact of urban sprawl index on obesity rate, Ewing et al. (2003) used the county level and Kelly-Schwartz et al. (2004) chose primary metropolitan statistical areas (PMSA). Other studies in this field employed smaller area units such as census tracts (Wen and Maloney 2011) and zip code areas (Wang et al. 2012) to define neighborhoods, mainly depending on what geographic identifiers are available in the data used in research. The wide variability in neighborhood size without a fair justification of its choice may lead to questions of stability and reliability of research results, an issue related to the modifiable areal unit problem (MAUP) (Fotheringham and Wong, 1991).

More recently, several MLM-based studies examined the issue of appropriate area unit(s) for defining the neighborhood effect in public health. It is widely acknowledged that effective interventions on health behaviors and outcomes occur on multiple levels (Nader et al. (2008). Mobley et al. (2008) examined how contextual variables in four types of geographic areas (post code areas, primary care service areas, medical service study areas, and county) affected the use of mammography service, and found inconsistent results across the four levels. Another study offered some insights speculating that small local areas might reflect social support while a large area unit might reflect geo-political units and minorities' political influence (Kuo, Mobley and Anselin 2011). Wang et al. (2012) constructed a new level of geographic areas from zip code areas with comparable population size to examine the neighborhood effect when neighborhoods are defined in different sizes. Kwan (2012b) used a term "the uncertain geographic context problem (UGCoP)" to refer to unstable results derived from different delineations of contextual units, and went on to suggest that contextual units should be defined in a way that captures people's actual or potential activity spaces (Kwan 2012a).

The current research continues this line of work to examine the neighborhood effects at both zip code and county levels on association of several built environment factors with individual odds of overweight and obesity. We seek to explore appropriate neighborhood units for a particular built environment factor in a representative sample of state of Utah. The results show that empirical results of built environmental influences differ across these two contextual levels.

In this research, improved measures of built environment were used: street connectivity, walk score, park accessibility and food environment at different contextual units. Street connectivity is adjusted by population size in order to capture the unevenness in geographic distribution of population groups. Walk score is promoted by Front Seat in recent years to capture walkability which is seldom used in the obesity study. Park accessibility is aggregated from a small area unit to zip code and county.

6.2 Data and variable definitions

Individual-level data used in this study are from the Utah Behavioral Risk Factor Surveillance Survey (BRFSS) collected in 2007, 2009 and 2011, which is an ongoing telephone (landline or cellular phones) survey by the Utah Department of Health in conjunction with the CDC for assessing health conditions and risk in the non-institutionalized Utah adult population (18 years and older). The 2007 and 2009 BRFSS were based on landline telephone numbers only and the 2011 BRFSS was based on both landline and cell phone numbers when recruiting subjects and collecting data. The 2011 BRFSS data reflects a change in weighting methodology (raking) and the addition of cell phone only respondents while the 2007 and 2009 BRFSS were solely based on landline subject recruiting and data collection (http://www.cdc.gov/brfss/annual_data/annual_2011.htm). The BRFSS data (http://health.utah.gov/opha/OPHA_BRFSS.htm) contains rich information on individual socio-demographic characteristics, behavioral factors and health conditions with zip code provided for each respondent. After deleting a small amount of missing data, 21,961 observations are used in the research. Among these records, there are 9,962 men and 11,999 women. Some zip code boundaries have changed over time, and a few zip codes are points. By checking the postal service website and other online sources, a unified GIS layer of 299 zip codes in 29 counties was able to be constructed as shown in Figure 8.

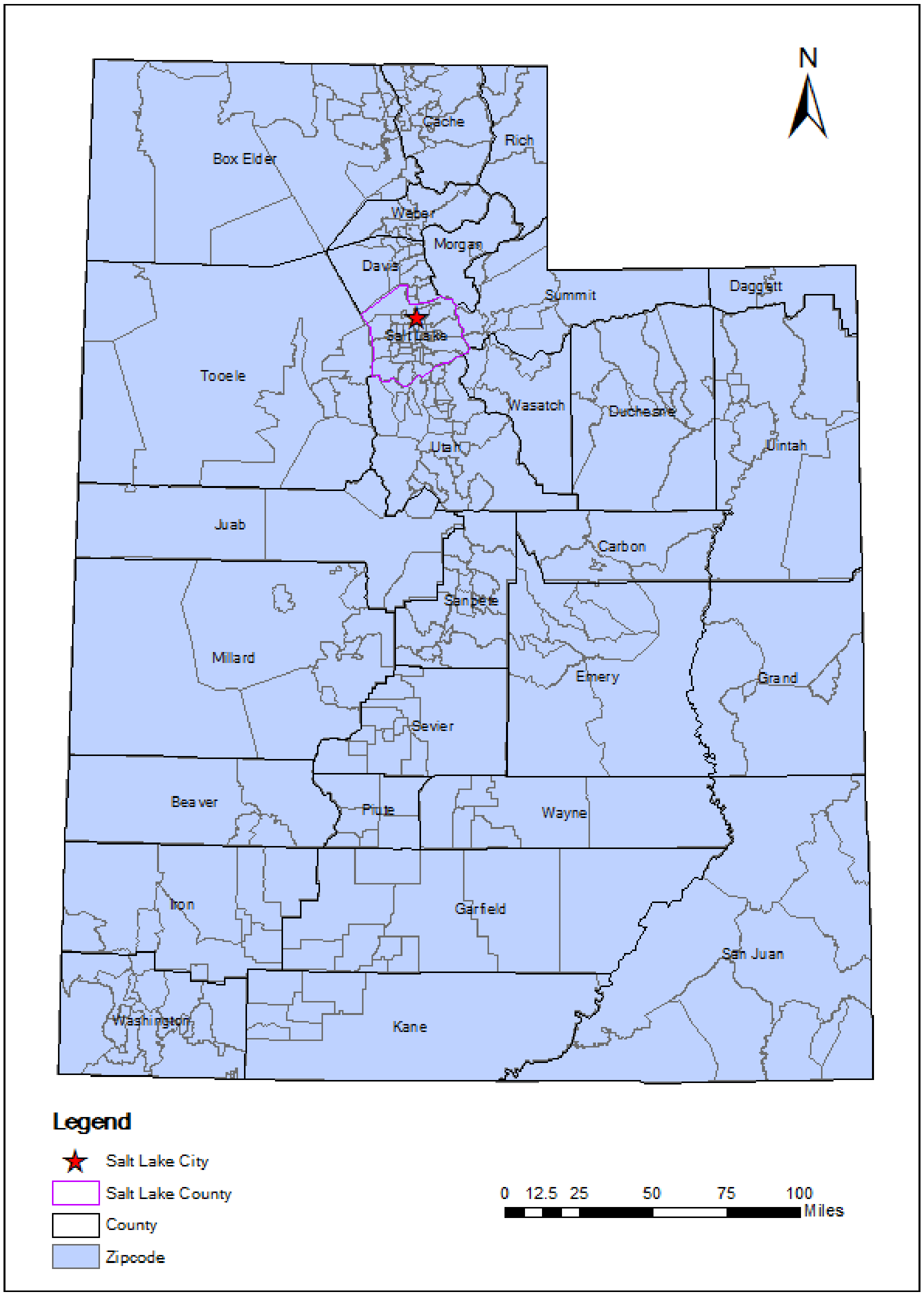

Figure 8 Boundaries of the zip code and county in Utah

Descriptive statistics for the Utah residents in the study sample are shown in Table 11. More than 60% of the study participants are either overweight or obese and the prevalence of obesity in this sample is 24.2%. The majority of the residents are white. About 70% of sample received college degree or above.

Table 11 Individual variables from the BRFSS (2007, 2009, 2011; *n* = 21,961)

Variables	Sample size	Sample %
Female	11,999	54.6
Non-Hispanic Whites	20,505	93.3
College degree or above	15,433	70.3
Currently married	15,255	69.5
Current smokers	6,229	28.4
Employed for wages	10,616	48.3
Self-employed	2,289	10.4
Out of work for more than 1 year	438	2.0
Out of work for less than 1 year	550	2.5
Homemaker	2,750	12.5
Student	489	2.2
Retired	4,829	22.0
Obese (BMI 30.0 and above)	5,315	24.2
Overweight and obese (BMI 25.0 and above)	13,281	60.5

Body Mass Index (BMI) was calculated based on self-reported height and weight: BMI = mass (kg)/ (height (m)) ^{2}. According to the CDC, an adult who has a BMI between 25 and 29.9 is considered overweight, while BMI of 30 or higher is obese (http://www.cdc.gov/obesity/adult/defining.html). Two levels of excessive weight were examined in this study, obesity (BMI>=30) and overweight plus obesity (BMI>=25). Socio-demographic variables including age (continuously measured), gender, race (whites versus non-whites), employment status (categorical), education level (college graduates versus below bachelor's degree), marital status (currently married or not) and smoking status(currently smoking or not) were controlled for in the analysis following previous work (Wen and Kowaleski-Jones 2012). Age squared was added to further control for potential nonlinear age effect. Race/ethnicity was dichotomously measured into whites versus non-whites given the vast majority of the respondents were white. Employment status was characterized into several groups including "employed for wages" (as the reference category), "self-employed", "out of work for more than one year", "out of work for less than one year", "homemaker", "student", and "retired." Education was dichotomously measured given the threshold effect of college credentials on obesity prevention (Wen and Kowaleski-Jones 2012).

Except the variables mentioned in the above chapter, the distance to the nearest *park* was constructed from the 2008 park dataset, also from the aforementioned ESRI Data DVD. National, state and local parks and forests are included in the dataset. There were 275 public parks and forest units in Utah, and 24 of them with areas smaller than 4,000 square feet were not included in this study. For better accuracy, distance to the nearest park was calculated from each census block centroid (Zhang et al. 2011) and then aggregated to zip code and county levels as street connectivity and walk score.

Table 12 reports mean, median, and ranges of neighborhood variables for the zip code areas and counties. We are aware of the gaps in dates among the data sources for the variables: BRFSS data 2007-2011, census data for poverty in 2010, street connectivity and distance from park in 2008, food environment in 2007 and walk score derived from the contemporary sources in 2013 (when most the data extraction and processing were conducted). It is considered acceptable given the limitation of data availability.

6.3 MLM Analysis

After eliminating cases with missing data for BMI or demographic characteristics at the individual level, the analysis included 21,961 individuals nested within 299 zip codes that were nested within 29 counties. In other words, the hierarchical structure of the data has three levels: individuals (level 1) in zip codes (level 2) in county (level 3). Individuals living in the same zip code area or the same county share the same environmental characteristics at the corresponding level. That is to say, the neighborhood contextual variables are defined at two levels (zip code

Table 12 Variables at the zip code and county levels

Data source (Year)	Neighborhood Characteristics	Mean		Median		Range	
		Zip Code	County	Zip Code	County	Zip Code	County
Decennial Census (2010)	% Poverty	0.74	11.73	0.09	11.20	0.00-48.86	4.80-25.80
ESRI Data DVD (2008)	Street Connectivity	8.45	29.13	1.13	13.15	0.02-83.79	0.91-173.46
Online (2013)	Walk score	10.25	6.20	0.00	0.00	0.00-92.00	0.00-32.84
ESRI Data DVD (2008)	Distance to park	12.00	13.04	10.49	9.67	0.38-2.17	0.80-46.06
Economic Census (2007)	Food environment [1]	84.64	2.87	35.84	3.10	2.17-958.49	0.00-5.33

[1] Food environment means fast-food restaurant accessibility at zip code level and the ratio of fast-food restaurant to full-service restaurant at county level

and county). Three-level random intercept logistic regression analyses were performed using SAS ProcGlimmix (Gibbs 2008). Model 1 tested the effect of individual and zip code variables. Model 2 added county-level factors to Model 1. Model 3 was the final model including all significant place-based contextual variables in previous models. Akaike Information Criterion (AIC) value for each model was also reported to gauge a model's balance between its fitness of power and degrees of freedom.

Table 13 presents the odds ratios of multilevel logistic models for the risk of obesity (BMI >=30). The effects of all the individual variables are fairly consistent across all models. White is not significant in any models. Female gender, college education, self-employment, homemaker, married and smoking are negatively associated with the odds of obesity. Age is positively associated with the odds of obesity, but the negative and significant coefficient for the "age squared" variable suggests this trend is reversed after reaching a certain age. Zip code level poverty prevalence (Models 1, 2 and 3) and county level ratio of fast-food to full-service restaurants (Models 2 and 3) are the only two place-based covariates exhibiting significant and positive associations with individual-level odds of obesity. Based on the AIC values, Model 3 is preferred.

Table 14 presents the results for overweight and obesity. Currently married is not significant anymore and student becomes negatively significant in Model 1. Other individual variables have the same effects as Table 13. In Model 1, fast food restaurant accessibility is negatively associated with the odds of overweight and obesity. Poverty prevalence (Models 1 and 2) and distance to the closest parks (Model 2) are positive covariates at zip code level but the effect of poverty is rendered insignificant in Model 3. At the county level, only the ratio of fast-food to full-service restaurants is a significant covariate positively associated with the odds of overweight or obesity (i.e., BMI>=25) (Models 2 and 3). Based on the AIC values, Model 3 is preferred.

Table 13 Adjusted Odd Ratios (95% Confidence Interval) of the Multilevel Logistic Models for Odds of Obesity (BMI>=30)

	Model 1	Model 2	Model 3
Individual-level variables			
Age (18+)	1.133***	1.133***	1.133***
Age^2	0.999***	0.999***	0.999***
Female	0.845***	0.846***	0.845***
White	1.063	1.063	1.059
Married	0.886**	0.885**	0.887**
College	0.834***	0.835***	0.827***
Self-employed	0.748***	0.749***	0.752***
Out of work for more than 1 year	1.142	1.144	1.129
Out of work for less than 1 year	1.119	1.123	1.113
Homemaker	0.829***	0.828***	0.826**
Student	0.879	0.876	0.838
Retired	1.054	1.055	1.050
Smoker	0.930*	0.931*	0.933*
Zip code-level variables			
Poverty	3.149**	3.686**	3.471**
Street connectivity	1.002	1.002	
Walk Score	0.999	1.000	
Distance to park	1.007	1.011	
Fast food accessibility	1.000	1.000	
Metro	1.037	1.025	
County-level variables			
Poverty		0.996	
Street connectivity		1.000	
Walk Score		1.004	
Distance to park		0.991	
Ratio of fast-food to full-service		1.172***	1.160***
Metro		0.875	
AIC	23599.08	23595.30	23581.16

Sample size: 21,961 individuals living in 299 zip codes, 29 counties.
***$p \leq 0.001$, **$p \leq 0.01$, *$p \leq 0.05$ (two-tailed tests)

6.4 Discussion

A unique feature of the current study is that it fit three-level multilevel models to simultaneously examine several built environmental features in their associations with odds of excessive body weight at two geographic aggregation levels: zip code and county. Two different levels of excessive body weight, overweight plus obesity and obesity alone were examined. The results suggest that observed built environmental influences on overweight and obesity are sensitive to these nuances. Net of individual controls and place-based poverty prevalence, distance to parks seems to be the only significant built environmental variable that is consistent

Table 14 Adjusted Odd Ratios (95% Confidence Interval) of the Multilevel Logistic Models for Odds of Overweight and Obesity (BMI>=25)

	Model 1	Model 2	Model 3
Individual-level variables			
Age (18+)	1.135***	1.136***	1.136***
Age^2	0.999***	0.999***	0.999***
Female	0.475***	0.475***	0.475***
White	1.058	1.058	1.054
Married	1.039	1.039	1.040
College	0.823***	0.824***	0.820***
Self-employed	0.820***	0.821***	0.821***
Out of work for more than 1 year	0.964	0.964	0.962
Out of work for less than 1 year	0.967	0.970	0.969
Homemaker	0.734***	0.734***	0.734***
Student	0.861*	0.859	0.858
Retired	0.941	0.941	0.942
Smoker	0.945*	0.945*	1.768*
Zip code-level variables			
Poverty	2.104**	2.376*	1.768
Street connectivity	1.000	1.000	
Walk Score	1.000	1.000	
Distance to park	1.009	1.014*	1.012***
Fast food accessibility	0.999*	0.999	
Metro	1.003	0.975	
County-level variables			
Poverty		0.997	
Street connectivity		1.000	
Walk Score		1.005	
Distance to park		0.991	
Ratio of fast-food to full-service		1.128***	1.120***
Metro		0.926	
AIC	27604.79	27599.70	27585.17

Sample size: 21,961 individuals living in 299 zip codes, 29 counties.
***$p \leq 0.001$, **$p \leq 0.01$, *$p \leq 0.05$ (two-tailed tests)

with our hypothesis, that is, the longer distance to parks, the less spatial park accessibility, the higher odds of overweight and obesity. However, this effect is only manifested for the odds of being overweight or obese rather than being obese alone. Meanwhile, the results on the food environment are inconsistent across zip code and county level analyses. In addition, walk score and street connectivity, measures of neighborhood walkability, are not significantly linked to odds of individuals' excessive body weight in this sample.

Poverty rate is the only placed-based socio-demographic variable included in the analyses as a control variable. Both zip code and county level poverty rates were examined. It turns out the zip code-level poverty effect is more stable across the model configurations and body weight outcomes compared to built environment features. By contrast, county-level poverty was never significant in the presence of zip code-level poverty. This finding suggests that socioeconomic status, captured by poverty rate, should play a more important role at smaller geographic unit. County-level poverty has a weaker influence on the individual compared to zip code-level poverty as the latter captures socioeconomic contexts of more immediate social surroundings.

Three types of built environment features including walkability, park accessibility and food environment were examined. Unexpectedly, none of the two walkability measures, namely street connectivity and walk score, were significant. Both variables were objectively measured and theoretically expected to be conductive to leisurely or non-leisurely walking and thus help with prevention against excessive weight gain. The empirical discrepancies are intriguing but not without antecedent (Berke et al. 2007). Several reasons are possible for this result. Our measures of walkability are not precise enough and the exposure misspecification may partly explain the null finding. Lacking information on individual address, geographic centroids of each zip code area as the focal point were used to measure street connectivity and walk score. Within-area variations can not be captured in this way. In addition, there may be interaction effects between walkability and other neighborhood factors such as socioeconomic status and ethnic composition. A recent study conducted in Baltimore found that walkability was only negatively linked to lower odds of obesity among individuals living in predominantly white and high-SES neighborhoods whereas the association between walkability and obesity among individuals living in low-SES neighborhoods was not significant after accounting for the confounders (Casagrande et al. 2011a). Other interaction effects may also exist. It is also possible that walkability effects are simply just weaker compared to other built environment features like food environments and park accessibility in Utah. However, population-based studies also conducted in Utah (Smith et al. 2008a, Zick et al. 2013) used different walkability indicators and examined the walkability and obesity link reporting that increasing levels of walkability decrease the risks of excess weight. Perhaps empirical results of the walkability and excessive weight link are to some extent to the specific walkable-environment measures used in the analysis.

Distance to parks captures spatial inaccessibility to local parks representing one type of neighborhood activity-promoting public amenities. A significant and positive effect of this variable was found at the zip code level but not at the county level. This is consistent with previous findings that the association between neighborhood environments and health outcomes are stronger for smaller units such as zip code and census tracts (Krieger et al. 2003, Sturm and Datar 2005). The result also makes intuitive sense, that is, individuals' exercise levels are likely to be more responsive to parks nearby rather than those located distantly. Compare to walkability, presence of local parks is a stronger built environment factor of individuals' odds of excessive weight in our analysis.

While walkability and park accessibility are both hypothesized to be environmental factors promoting physical activity, the food environment is supposed to affect the other key energy balance factor, dietary intake. There are many ways to capture the food environment and calculating the number of fast food restaurant per capita is a common method in many researches (Wang et al. 2007, Jay 2004). In this study, density of BMI-unhealthy food outlets was captured by focusing on per-capita exposure to fast food. Instead of using the conventional method, the presence and density of fast food outlets were operationalized differently for the two spatial units, zip code areas and counties. Fast-food restaurant accessibility was defined at the zip code level and the ratio of fast-food outlets to full-service outlets was used at the county level. Results show that there is slightly negatively association between fast food accessibility and risk of overweight and obesity at the zip code level. Although the association at the zip code level in Model 1 is counterintuitive, it is no longer significant after adding the county-level variables. For fast food ratio at the county level, it is strongly positively associated with the risk of unhealthy outcome and obesity ($p \leq 0.001$). The explanation is that full-service restaurants are typically providing healthy food, while fast-food restaurants are typically main source of unhealthy, energy dense processed foods (Michimi and Wimberly 2010). This is the only variable that is significant at the county level. Since people normally drive to buy fast food beyond the zip code they live, perhaps the adequate scale for defining food environment need to be expanded beyond zip code areas.

6.5 Concluding Remarks

Based on the BRFSS data in Utah, this research examines the associations between neighborhood built environments and individual odds of overweight and obesity after controlling for individual risk factors. Four neighborhood built environment factors measured at both zip code and county levels are street connectivity, walk score, distance to parks, and food environment. Two additional neighborhood variables, namely the poverty rate and urbanicity, are also included as control variables.

Several study limitations should be kept in mind when interpreting study findings. First of all, this study is cross-sectional without taking the time effects. The built environment variables describe an individual's location at a specific time which does not account for how long the residents have lived in that address. For example, people with high BMI may reflect years of accumulation but only live in that area while doing the survey. The cross-sectional analysis cannot tell whether neighborhood environment factors cause individuals to live health or whether health individuals choose to live in neighborhood with good environment characteristics. To better sort of selection versus causation, longitudinal analyses should be conducted in the future. Second, the measurement of overweight/obesity was relied on self-reported weight and height. Under reporting may occur if individuals who are older or heavier. Lastly, there are omitted built environment factors that are important but not examined in this study. For example, the mixed land use may increase people's physical activities and reduce obesity. Highly mixed commercial and residential land uses can provide goods and services within individuals' walking or bicycling distances.

Despite the limitations, several strengths of this study are noteworthy. A key contribution of the current study is its simultaneously examining both physical activity and food environments at two different geographic units. To the best of our knowledge, this is the first 3-level study examining contextual effects of the built environments on individuals' odds of excessive weight. The MLM results show that among the four built environment variables, (1) at the zip code level, distance to parks is the only significant (and negative) covariate of the odds of overweight and obesity; and (2) at the county level, food environment is the sole significant factor with stronger fast food presence linked to higher odds of overweight and obesity. As residents normally walk to parks for recreational activities but drive to restaurants for food, the relevant built environments vary in spatial range. The findings suggest that obesity risk factors lie in multiple neighborhood levels and built environment need to be defined at a neighborhood size relevant to residents' activity space. This raises the issue of "uncertain geographic context problem (UGCoP)" and suggests that the contextual variables need to be defined in a way that reflects human mobility pertaining to the specific trip purposes.

Chapter 7 Conclusion and Future Work

This chapter summarizes the results and discussions of the previous chapters. Both Ordinary Least Squares and Geographically Weighted Regression were used to test the relationship between built environment and obesity rate by using the aggregated dataset. This approach may lead to ecological fallacy, where relationships observed in groups are assumed to hold for individuals. Besides the aggregate regression models, the need to consider environmental and contextual variables in the social and behavioral sciences has taken into account. Multilevel models have grown in popularity in large part because they provide a means to explicitly model the influence of context on many individual level processes. However, in applications of these and other statistical models that incorporate context into the analysis, rarely is physical location or distance between entities considered. In order to obtain a comprehensive understanding of how environmental attributes affect people's behavior, this dissertation examines the relationship between built environment and obesity by using different models and sources of data.

The quantitative measured of the built environment variables by using GIS techniques and the nationwide of the study area are the most important merits in this research. Multilevel models which have the ability to model contextual questions were then used to study the contextual and organization effects on people's weight status. There are three main parts of this dissertation. The first one and the second one were focused on county-level analysis with the study area of the conterminous United States. The third part simultaneously examined both physical activity and food environments at two different geographic units: county and zip-code in the state of Utah. The results suggest that obesity risk factors lie in multiple neighborhood levels and built environment need to be defined at a neighborhood size relevant to residents' activity space.

While the measurements of neighborhood built environment are similar to those commonly investigated in the literature, the implementations at the national level, particularly walk score and food environment, are new. Furthermore, regionalization analysis was applied in order to identify the attributes of the areas with higher rates of obesity which will be a useful tool for public health researchers and policy makers to effectively optimize scarce public health resources on disadvantage regions. Multilevel models including both two-level and three-level models were performed to predict the risk of obesity based on a function of predictor variables at more than one level. It contributes to a better understanding of the specific individual, socio- and built environment variables that are associated with obesity which may provide insight into potentially risk factors to the current obesity epidemic.

7.1 Summary of the results and conclusions

(1) By reviewing the aggregate level, the regression model has found that the walk score and street connectivity are negatively relatedly to obesity, and that poverty rate and metro are positively related to obesity, while the fast-food-to-full-service restaurant ratio is not significant. While the global OLS regression model can measure the relationship between the obesity rate and the explanatory variables, GWR has its strength in finding geographical heterogeneity among the counties by the clustered spatial pattern of their coefficients. A regionalization method was used to group the U.S. counties to regions based on their GWR coefficients. Qualitative inferences of policies are made available with the regions to facilitate our better understanding of the obesity problem associated with the built environment.

(2) Multilevel modeling is used to control for the effects of individual socio-demographic characteristics such as race-ethnicity, age, sex, marital status, education attainment, employment status, income level, and whether an individual smokes. Neighborhood variables include built environment, socio-demographic factors and urbanicity level at the county level. The relationship between built environment and obesity was checked by urbanicity in the conterminous United States area. County-level socio-demographic structure such as a lower racial-ethnic heterogeneity index or a higher poverty rate is linked to a higher obesity risk. Among the built environment variables, a poorer street connectivity and a more prominent presence of fast-food restaurants are associated with a higher obesity risk. While the effect of walk score is not evident in influencing obesity risk, a higher walk score is indeed linked to a lower rate of physical inactivity. Overall, obesity risk initially increases with the urbanicity level and then drops, resembling an inverted-V shape. The results lend support to the relevance of built environment in potentially influencing people's health behavior and outcome. Urbanization level differences are found for these associations by analyzing the data subsets. The influences of poverty, street connectivity and walk score on obesity are stronger in the urban areas. The positive association between the food environment and physical inactivity/obesity is stronger among non-metro areas. The results demonstrate that different geographic settings should be taken into account among the obesity research.

(3) The Utah BRFSS data include information on 21,961 individuals geocoded to zip code areas. Individual variables include BMI (body mass index) and socio-demographic attributes such as age, gender, race, marital status, education attainment, employment status, and whether an individual smokes. Neighborhood built environment factors measured at both zip code and county levels include street connectivity, walk score, distance to parks, and food

environment. Two additional neighborhood variables, namely the poverty rate and urbanicity, are also included as control variables. Multilevel modeling results show that at the zip code level, poverty rate and distance to parks are significant and negative covariates of the odds of overweight and obesity; and at the county level, food environment is the sole significant factor with stronger fast food presence linked to higher odds of overweight and obesity. These findings suggest that obesity risk factors lie in multiple neighborhood levels and built environment need to be defined at a neighborhood size relevant to residents' activity space. A key contribution of this study is its simultaneously examining both physical activity and food environments at two different geographic units. To the best of our knowledge, this is the first 3-level study examining contextual effects of the built environments on individuals' odds of excessive weight.

7.2 Suggestions for future work

Although the ultimate goal of public health research is to thoroughly understand the obesity problem related to the physical and socio-economic conditions, this research only focused on several built environment variables and social status variables. We selected the variables according to the major hypotheses about obesity. In addition to street connectivity, walk score, park accessibility and food environment, other build environment variables such as land use mix, neighborhood crime rates, and greenness could be included as input data to predict people's health status. Except the commonly used built environment variables, some physical environment variables including weather (temperature, precipitation, or disaster etc.) will be considered in the future study.

This study was mostly focus on county level as neighborhood, only a small study area Utah was checked at the zip code level. For some variables, county or zip code may not be suitable to describe people's activity space. Therefore, smaller geographic unit, such as census tract will be used in the future study. Also, small population problem will be taken into account since it may cause unstable rate estimates and suppress data in sparsely populated areas. Regionalization which is to combine small units into large areas to ensure population is comparable across areas will be used in the future study.

This whole study is cross-sectional without considering any temporal changes. The built environment defined is the present state of environment for an individual. A person's BMI reflects the accumulated effect of one's living environment and behavior, both of which may have changed. The research may establish the link between an environment factor and obesity, but cannot tell whether the neighborhood factor causes residents to live healthy or whether healthy individuals choose to live in neighborhood with such an environment.

References

Ahima, R. S. & M. A. Lazar (2013) The Health Risk of Obesity—Better Metrics Imperative. *Science,* 341**,** 856-858.

Aurbach, L. 2010. The Power of Intersection Density. In *Thoroughfares and Transit*.

Ball, K., K. Lamb, N. Travaglini & A. Ellaway (2012) Street connectivity and obesity in Glasgow, Scotland: Impact of age, sex and socioeconomic position. *Health & Place,* 18**,** 1307-1313.

Bell, J. F. & J. S. Wilson (2008) Neighborhood Greenness and 2-Year Changes in Body Mass Index of Children and Youth. *American Journal of Preventive Medicine,* 35**,** 547-553.

Berke, E. M., T. D. Koepsell, A. V. Moudon, R. E. Hoskins & E. B. Larson (2007) Association of the Built Environment With Physical Activity and Obesity in Older Persons. *Am J Public Health,* 97**,** 486-492.

Boardman, J. D., J. M. Saint Onge, R. G. Rogers & J. T. Denney (2005) Race Differentials in Obesity: The Impact of Place. *Journal of Health and Social Behavior,* 46**,** 229-243.

Bodea, T., L. Garrow & M. Meyer (2008) Explaining obesity with urban form: a cautionary tale. *Transportation,* 35**,** 179-199.

Bonoit, K. 2009. Random-coefficient models. In *Introduction to Multilevel Models*.

Brewster, M., D. Hurtado, S. Olson & J. Yen. 2009. Walkscore.com: A New Methodology to Explore Associations Between Neighborhood Resources, Race, and Health.

Brunsdon, C., S. Fotheringham & M. Charlton (1998) Geographically weighted regression - modelling spatial non-stationarity. *Journal of the Royal Statistical Society Series D-the Statistician,* 47**,** 431-443.

Burdette, H. L. & R. C. Whitaker (2004) Neighborhood playgrounds, fast food restaurants, and crime: relationships to overweight in low-income preschool children. *Preventative Medicine,* 38**,** 57-63.

Casagrande, S. S., J. Gittelsohn, A. Zonderman, M. Evans & T. Gary-Webb (2011a) Association of Walkability with Obesity in Baltimore City, Maryland. *Am J Public Health,* 101**,** S318-S324.

Casagrande, S. S., J. Gittelsohn, A. B. Zonderman, M. K. Evans & T. L. Gary-Webb (2011b) Association of walkability with obesity in Baltimore City, Maryland. *Am J Public Health,* 101 Suppl 1**,** S318-24.

Casey, A. A., M. Elliott & K. Glanz (2008) Impact of the food environment and physical activity environment on behaviors and weight status in rural U.S. communities. *Preventative Medicine,* 47**,** 600-604.

CDC. 2012. Overweight and Obesity.

Census. 2013. 2010 Census Urban and Rural Classification and Urban Area Criteria. ed. U. S. C. Bureau.

Chalkias, C., A. G. Papadopoulos, K. Kalogeropoulos, K. Tambalis, G. Psarra & L. Sidossis (2013) Geographical heterogeneity of the relationship between childhood obesity and socio-environmental status: Empirical evidence from Athens, Greece. *Applied Geography,* 37**,** 34-43.

Chen, D.-R. & K. Truong (2012) Using multilevel modeling and geographically weighted regression to identify spatial variations in the relationship between place-level disadvantages and obesity in Taiwan. *Applied Geography,* 32**,** 737-745.

Chi, S.-H., D. S. Grigsby, N. Bradford & J. Choi (2013) Can Geographically Weighted Regression improve our contextual understanding of obesity in the US? Findings from the USDA Food Atlas. *Applied Geography,* 44**,** 134-142.

Chiolero, A., D. Faeh, F. Paccaud & J. Cornuz (2008) Consequences of smoking for body weight, body fat distribution, and insulin resistance. *The American Journal of CLINICAL NUTRITION,* 87**,** 801-809.

Chou, S.-Y., M. Grossman & H. Saffer (2004) An economic analysis of adult obesity: results from the Behavioral Risk Factor Surveillance System. *Journal of Health Economics,* 23**,** 565-587.

Cortright, J. 2009. Walking the Walk. In *How Walkability Raises Home Values in U.S. Cities*. Impresa, Inc.

Dijkstra, A., F. Janssen, M. D. Bakker, J. Bos, R. Lub, L. J. G. V. Wissen & H. Eelko (2013) Using Spatial Analysis to Predict Health Care Use at the Local Level: A Case Study of Type 2 Diabetes Medication Use and Its Association with Demographic Change and Socioeconomic Status. *PLoS ONE,* 8.

Ding, D. & K. Gebel (2012) Built environment, physicalactivity, and obesity: What have we learned from reviewing the literature? *Health & Place,* 18**,** 100-105.

Doyle, S., A. Kelly-Schwartz, M. Schlossberg & J. Stockard (2006) Active Community Environments and Health: The Relationship of Walkable and Safe Communities to Individual Health. *Journal of the American Planning Association,* 72**,** 19-31.

Duncan, D. T., J. Aldstadt, J. Whalen, S. J. Melly & S. L. Gortmaker (2011) Validation of Walk Score® for Estimation Neighborhood Walkability: An Analysis of Four US Metropolitan Areas. *International Journal of Environmental Research and Public Health,* 8**,** 4160-4179.

Duncan, M. J., E. Winkler, T. Sugiyama, E. Cerin, L. duToin, E. Leslie & N. Own (2010) Relationships of Land Use Mix with Walking for Transport: Do Land Uses and Geographical Scale Matter? *J Urban Health,* 87**,** 782-95.

Durand, C., M. Andalib, G. Dunton, J. Wolch & M. Pentz (2011) A systematic review of built environment factors related to physical activity and obesity risk: implications for smart growth urban planning. *Obes Rev.,* 2011**,** 5.

Edwards, K. L., G. P. Clarke, J. K. Ransley & J. Cade (2010) The neighborhood matters: studying exposures relevant to childhood obesity and the policy implications in Leeds, UK. *J Epidemiol Community Health,* 64**,** 194-201.

Ewing, R., R. C. Brownson & D. Berrigan (2006) Relationship Between Urban Sprawl and Weight of United States Youth. *American Journal of Preventive Medicine,* 31**,** 464-474.

Ewing, R., G. Meakins, S. Hamidi & A. C. Nelson (2014) Relationship between urban sprawl and physical activity, obesity, and morbidity-Update and refinement. *Health & Place,* 26**,** 118-126.

Ewing, R., R. Pendall & D. Chen. 2002. Measuring sprawl and its impact.

Ewing, R. & T. Schmid (2003) Relationship Between Urban Sprawl and Physical Activity, Obesity, and Morbidity. *American Journal of Health Promotion,* 18**,** 47-57.

Ewing, R., T. Schmid, R. Killingsworth, A. Zlot & S. Raudenbush (2003) Relationship Between Urban Sprawl and Physical Activity, Obesity, and Morbidity. *American Journal of Health Promotion,* 18**,** 47-57.

Felton, G., M. Dowda, D. Ward, R. Dishman, S. Trost, R. Saunders & R. Pate (2002) Differences in physical activity between black and white girls living in rural and urban areas. *J Sch Health,* 72**,** 250-5.

Feng, J., T. A. Glass, F. C. Curriero, W. F. Stewart & B. S. Schwartz (2010) The built environment and obesity: A systematic review of the epidemiologic evidence. *Health & Place,* 16**,** 175 -190.

Finkelstein, E. A., J. G. Trogdon, J. W. Cohen & W. Dietz (2009) Annual Medical Spending Attributable To Obesity: Payer-And Service-Specific Estimates. *Health Affairs,* 28**,** w822-w831.

Flegal, K. M., M. D. Carroll, C. L. Ogden & L. R. Curtin (2010) Prevalence and Trends in Obesity Among US Adults, 1999-2008. *The Journal of the American Medical Association,* 303**,** 235 - 241.

Fotheringham, A. S., C. Brunsdon & M. Charlton. 2000. *Quantitative Geography: Perspectives on Spatial Data Analysis*. Newbury Park, CA: Sage Publications Ltd.

---. 2002. *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. New York, NY, USA: Wiley.

Frank, L. D., M. A. Andresen & T. L. Schmid (2004) Obesity Relatinships with Community Design, Physical Activity, and Time Spent in Cars. *American Journal of Preventive Medicine,* 27**,** 87-96.

Fraser, A., T. Kate, C. Macdonald-Wallis & N. Satter (2010) Association of maternal weight gain in pregnancy with offspring obesity and metabolic and vascular traits in childhood. *Circulation,* 121**,** 2557-64.

Fraser, L. K., MBChB, G. P. Clarke, J. E. Cade & K. L. Edwards (2012) Fast Food and Obesity A spatial Analysis in a Large United Kingdom Population of Children Aged 13-15. *American Journal of Preventive Medicine,* 42**,** e77-e85.

Freedman, D. A. 1999. Ecological Inference and the Ecological Fallacy. Berkeley: University of California.

Front Seat. 2013. Walk Score API.

Gates, S. 2012. American Obesity In 2030: Most U.S. Residents Will Be Obese Within Next 2 Decades.

Gibbs, P. 2008. An Introduction to Generalized Linear Mixed Models Using SAS RPOC GLIMMIX. SAS Technical Support.

Goodchild, M. F. 1986. *Spatial autocorrelation*. Geo Books.

Gordon, P., M. Nelson & P. Rage (2006) Inequality in the Built Environment Underlies Key Health Disparities in Physical Activity and Obesity. *Pediatrics,* 117**,** 41--24.

Grigsby, D. S., S.-H. Chi & B. H. Fiese (2011) Where they live, how they play: Neighborhood greenness and outdoor physical activity among preschoolers. *International Journal of Health Geographics,* 10.

Heath, G. W., R. C. Brownson, J. Kruger, R. Miles, K. E. Powell & L. T. Ramsey (2006) The Effectiveness of Urban Design and Land Use and Transport Policies and Practices to Increase Physical Actity: A Systematic Review. *Journal of Physical Activity & Health,* 3**,** S55 - S76.

Holt, D., D. G. Steel, M. Tranmer & N. Wrigley (1996) Aggregation and Ecological Effects in Geographically Based Data. *Geographical Analysis,* 28**,** 244-261.

Huang, T. T., A. Drewnowski, S. K. Kumanyika & T. A. Glass (2009) A Systems-Oriented Multilevel Framework for Addressing Obesity in the 21st Century. *Prev Chronic Dis,* 6.

Ingram DD & Franco SJ. 2014. 2013 NCHS urban-rural classification scheme for counties. ed. N. C. f. H. Statistics. Vital Health Stat.

Jay, M. (2004) The Relationship Between Obesity and the Prevalence of Fast Food Restaurants: State-Level Analysis. *American Journal of Health Promotion,* 19**,** 137-143.

Jones, L. I. 2010. Investigation neighborhood walkability and its accossication with physical activity levels and body composition of a sample of Maryland adolescent girls. In *Department of Epidemiology and Biostatistics*. College Park: University of Maryland.

Kelly-Schwartz, A., J. Stockard, S. Doyle & M. Schlossberg (2004) Is Sprawl Unhealthy? : A Multilevel Analysis of the Relationship of Metropolitan Sprawl to the Health of Individuals. *Journal of Planning Education and Research,* 24**,** 184-96.

Kennedy, B. P., I. Kawachi, R. Glass & D. Prothrow-Stith (1998) Income distribution, socioeconomic status, and self rated health in the United States: multilevel analysis. *BMJ,* 317**,** 917-921.

Kim, D., S. V. Subramanian, S. L. Gortmaker & I. Kawachi (2006) US state- and county-level social capital in relation to obesity and physical inactivity: A multilevel, multivariable analysis. *Social Science & Medicine,* 63**,** 1045-1059.

Kirby, J., L. Liang, H. Chen & Y. Wang (2012) Race, place, and obesity: the complex relationships among community racial/ethnic composition, individual race/ethnicity, and obesity in the United States. *American Journal of Public Health,* 102**,** 1572-8.

Krieger, N., J. Chen, P. Waterman, M. Soobader, S. V. Subramanian & R. Carson (2003) Choosing area based socioeconomic measures to monitor social inequalities in low birth weight and childhood lead poisoning: The Public Health Disparities Geocoding Project (US). *J Epidemiol Community Health,* 57**,** 186-99.

Kumar, R. 2009. Walkability of Neighborhoods: A Critical Analysis of Zoning Codes. In *School of Planning*. Cincinnati: University of Cincinnati.

Kuo, T.-M., L. R. Mobley & L. Anselin (2011) Geographic disparities in late-stage breast cancer diagnosis in California. *Health & Place,* 17.

Kwan, M.-P. (2012a) How GIS can help address the uncertain geographic context problem in social science research. *Annals of GIS,* 18**,** 245-255.

--- (2012b) The Uncertain Geographic Context Problem. *Annals of the Association of American Geographers,* 102**,** 958-968.

Lamichhane, A. P., J. Warren, R. Putee, D. E. Porter, M. Bottai, E. J. Mayer-Davis & A. D. Liese (2013) Spatial patterning of supermarkets and fast food outlets with respect to neighborhood characteristics. *Health & Place,* 23**,** 157-164.

Li, K., M. Wen & K. A. Henry (2014) Residential Racial Composition and Black-White Obesity Risks: Differential Effects of Neighborhood Social and Built Environment. *International Journal of Environmental Research and Public Health,* 11**,** 626-642.

Lin, C.-H. & T.-H. Wen (2011) Using Geographically Weighted Regression (GWR) to Explore Spatial Varying Relationships of Immature Mosquitoes and Human Densities withe the Incidence of Dengue. *International Journal of Environmental Research and Public Health,* 8**,** 2798-2815.

Lopez, R. (2004) Urban Sprawl and Risk for Being Overweight or Obese. *American Journal of Public Health,* 94**,** 1574-1579.

Lopez, R. P. (2007) Neighborhood Risk Factors for Obesity. *Risk Factors and Chronic Disease,* 15**,** 2111-2119.

Lopez, R. P. & H. P. Hyness (2006) Obesity, physical activity, and the urban environment: public health research needs. *ENVIRONMENTAL HEALTH,* 5.

Malmstrom, M., J. Sundquist & S.-E. Johansson (1999) Neighborhood Environment and Self-Reported Health Status: A Multilevel Analysis. *American Journal of Public Health,* 89**,** 1181-1186.

Maroko, A. R., J. A. Maantay, N. L. Sohler, K. L. Grady & P. S. Arno (2009) The complexities of measuring access to parks and physical activity sites in New York City: a quantitative and qualitative approach. *International Journal of Health Geographics,* 8.

McArdle, J. & D. Epstein (1987) Latent growth curves within developmental structural equation models. *Child Development,* 58**,** 110-33.

Mehta, N. K. & V. W. Chang (2008) Weight Status and Restaurant Availability: A Multilevel Analysis. *American Journal of Preventive Medicine,* 34**,** 127-133.

Mennis, J. (2006) Mapping the Results of Geographically Weighted Regression. *The Carographic Journal,* 43**,** 171-179.

Michimi, A. & M. C. Wimberly (2010) Associations of supermarket accessibility with obesity and fruit and vegetable consumption in the conterminous United States. *International Journal of Health Geographics,* 9**,** 94-112.

Miles, R., L. B. Panton, M. Jang & E. M. Haymes (2008) Residential context, walking and obesity: Two African-American neighborhoods compared. *Health & Place,* 14**,** 275-286.

MMWR. 1990. Guidelines for Investigating Clusters of Health Events. In *Morbidity and Mortality Weekly Report,* 1-16. Atlanta: Centers for Disease Control and Prevention.

Mobley, L. R., E. A. Finkelstein, O. A. Knavjou & J. C. Will (2004) Spatial Analysis of Body Mass Index and Smoking Behavior among WISEWOMAN Participants. *Journal of Women's Health,* 13**,** 519-528.

Mobley, L. R., T.-M. Kuo & L. Andrews (2008) How Sensitive Are Multilevel Regression Findings to Defined Area of Context? A Case Study of Mammography Use in California. *Medical Research and Review,* 65**,** 315-337.

Monda, K. & B. Popkin (2005) Cluster analysis methods help to clarify the activity-BMI relationship of Chinese youth. *Obes Res,* 13**,** 1042-51.

Monnat, S. M. & C. B. Pickett (2011) Rural/urban differences in self-rated health: Examining the roles of county size and metropolitan adjacency. *Health & Place,* 17**,** 311-319.

Monteiro, C., W. Conde, B. Lu & B. Popkin (2004) Obesity and inequities in health in the developing world. *International Journal of Obesity,* 28**,** 1181-1186.

Morland, K., S. Wing, A. D. Roux & C. Poole (2002) Neighborhood Characteristics Associated with the Location of Food Stores and Food Service Places. *American Journal of Preventive Medicine,* 22**,** 23-9.

Muthen, B. O. (1994) Multilevel Covariance Structure Analysis. *Sociological Methods & Research,* 22**,** 376-398.

Nader, P. R., R. H. Bradley, R. M. Houts, S. L. McRitchie & M. O'Brien (2008) Moderate-to-Vigorous Physical Activity From Ages 9 to 15 Years. *JAMA,* 300**,** 295-305.

National Center for Health Statistics. 2013. Health, United States, 2012: With Special Feature on Emergency Care. Hyattsville, MD.

NCHS. 2006. NCHS Urban-Rural Classification Scheme for Counties.

O Ferdinand, A., B. Sen, S. Rahurkar, S. Engler & N. Menachemi (2012) The relationship between built environments and physical activity: a systematic review. *American Journal of Public Health,* 102**,** e7-e13.

Ogden, C. L., M. D. Carroll, B. K. Kit & K. M. Flegal. 2013. Prevalence of Obesity Among Adults: United States, 2011-2012. In *NCHS Data Brief*. Hyattsville, MD: National Center for Health Statistics.

Papas, M. A., A. J. Alberg, R. Ewing, K. J. Helzlsouer, T. L. Gary & A. C. Klassen (2007) The Built Environment and Obesity. *Epidemiologic Reviews,* 29**,** 129-143.

Peet, R. 1998. *Modern Geographical Thought*. Oxford: Blackwell.

Pereira, M., A. Kartashov, C. Ebbeling, L. Van Horn, M. Slattery, D. J. Jocobs & D. S. Ludwig (2005) Fast-food habits, weight gain, and insulin resistance (the CARDIA study): 15-year prospective analysis. *Lancet,* 365**,** 36-42.

Pierce, I., V. Ernest & J. Ashworth. 2012. Tool for Health and Resilience in Vulnerable Environments. Health Department and Community Health Center.

Pinheiro, J. C. & D. M. Bates. 2000. *Mixed-Effects Models in S and S-PLUS*. Springer.

Pouliou, T. & S. J. Elliott (2009) An exploratory spatial analysis of overweight and obesity in Canada. *Preventative Medicine,* 48**,** 362-367.

Powell, P., K. Spears & M. Rebori. 2010. What is Obesogenic Environment? Reno: University of Nevada Cooperative Extension.

Rattan, A., A. Campese & C. Eden. 2012. Modeling Walkability. ESRI.

Raudenbush, S. W. & A. S. Bryk. 2001. *Hierarchical Linear Models: Applications and Data Analysis Methods*. SAGE Publications.

Rauterkus, S. Y., G. I. Thrall & E. Hangen (2010) Location Efficiency and Mortgage Default. *JOSRE,* 2.

Rose, D., P. L. Hutchinson, N. Bodor, C. Swalm, T. Farley, D. Cohen & J. Rice (2009) Neighborhood Food Environments and Body Mass Index: the importance of in-store contents. *American Journal of Preventive Medicine,* 37**,** 214-219.

Rundle, A. & L. M. Freeman (2007) The urban built environment and obesity in New York City: A multilevel analysis. *American Journal of Health Promotion,* 21**,** 326-34.

Rundle, A., A. Roux & L. M. Freeman (2007) The Urban Built Environment and Obesity in New York City: A Multilevel Analysis. *American Journal of Health Promotion,* 21**,** 326-34.

Rutt, C. D. & K. J. Coleman (2005) Examining the relationships among built environment, physical activity, and body mass index in El Paso, TX. *Preventative Medicine,* 40**,** 831-841.

Saelens, B. E., J. F. Sallis, J. B. Black & D. Chen (2003) Neighborhood-Based Differences in Physical Activity: An Environment Scale Evaluation. *American Journal of Public Health,* 93**,** 1552-8.

Sampson, R. J. (1989) Community Structure and Crime: Testing Social-Disorganization Theory. *American Journal of Sociology,* 94**,** 774-802.

Schlundt, D., M. Hargreaves & L. McClellan (2006) Geographic clustering of obesity, diabetes, and hypertension in Nashville, Tennessee. *J Ambul Care Manage,* 29**,** 125-32.

Schulze, M. B., J. E. Manson, D. S. Ludwig, G. Colditz, M. J. Stampfer & W. C. Willett (2004) Sugar-sweetened beverages, weight gain, and incidence of type 2 diabetes in young and middle-aged women. *Journal of the American Medical Association,* 292**,** 927-934.

Schuurman, N., P. A. Peters & L. N. Oliver (2009) Are Obesity and Physical Activity Clustered? A Spatial Aanlysis Linked to Residential Density. *Epidemiology,* 17**,** 2202-9.

Smith, K. R., B. B. Brown, I. Yamada, L. Kowaleski-Jones, C. D. Zick & J. X. Fan (2008a) Walkability and Body Mass Index. *American Journal of Preventive Medicine,* 35**,** 237-244.

--- (2008b) Walkability and Body Mass Index: Density, Design, and New Diversity Measures. *American Journal of Preventive Medicine,* 35**,** 237-244.

Sturm, R. & A. Datar (2005) Body mass index in elementary school children, metropolitan area food prices and food outlet density. *Public Health,* 119**,** 1059-68.

Sundquist, J., M. Malmstrom & S.-E. Johansson (1999) Cardiovascular risk factors and the neighborhood environment: a multilevel analysis. *International Journal of Epidemiology,* 28**,** 841-845.

Talen, E. & L. Anselin (1998) Assessing spatial equity: an evaluation of measures of accessibility to public playgrounds. *Environment and Planning,* 30**,** 595-613.

Vanasse, A., M. Demers, A. Hemiari & J. Courteau (2006) Obesity in Canada: where and how many? *International Journal of Obesity,* 30**,** 677-83.

Wakefield, J. (2004) Fighting Obesity Through the Built Environment. *Environmental Health Perspectives*, 112**,** A616 - A618.

Wang, F. & W. Luo (2005) Assessing spatial and nonspatial factors for healthcare access: towards an integrated approach to defining health professional shortage areas. *Health & Place,* 11**,** 131-146.

Wang, F., M. Wen & Y. Xu (2013) Population-Adjusted Street Connectivity, Urbanicity and Risk of Obesity in the U.S. *Applied Geography,* 41**,** 1-14.

Wang, M. C., S. Kim, A. A. Gonzalez, K. E. MacLeod & M. A. Winkleby (2007) Socioeconomic and food-related physical characteristics of the neighbourhood environment are associated with body mass index. *J Epidemiol Community Health,* 61**,** 491-498.

Weir, L. A., D. Etelson & D. A. Brand (2006) Parents' perceptions of neighborhood safety and children's physical activity. *Preventive Medicine,* 43**,** 212-217.

Wen, M. & L. Kowaleski-Jones (2012) The built environment and risk of obesity in the United States: Racial–ethnic disparities. *Health & Place,* 18**,** 1314-1322.

Wen, M. & T. N. Maloney (2011) Latino Residential Isolation and the Risk of Obesity in Utah: The Role of Neighborhood Socioeconomic, Built-Environmental, and Subcultural Context. *J Immigr Minor Health,* 13**,** 1131-1141.

Wen, T.-H., D.-R. Chen & M.-J. Tsai (2010) Identifying geographical variations in poverty-obesity relationships: empirical evidence from Taiwan. *Geospatial Health,* 4**,** 257-265.

Wilson, D., K. Kirtland, B. Ainsworth & C. Addy (2004) Socioeconomic status and perceptions of access and safety for physical activity. *Ann Behav Med,* 28**,** 20-8.

Wong, D. W.-S. & J. Lee. 2005. *Statistical anaylysis of geographic information with ArcView GIS and ArcGIS*. John Wiley & Sons, Inc.

Yamada, I., B. B. Brown, K. R. Smith, C. D. Zick, L. Kowaleski-Jones & J. X. Fan (2012) Mixed land use and obesity: an empirical comparison of alternative land use measures and geographic scales. *The Professional Geographer,* 64**,** 157-177.

Zhang, X., H. Lu & J. B. Holt (2011) Modeling spatial accessibility to parks: a national study. *International Journal of Health Geographics,* 10**,** 1 - 14.

Zhu, X. & C. Lee (2008) Walkability and safety around elementary schools economic and ethnic disparities. *American Journal of Preventive Medicine,* 34**,** 282-90.

Zick, C. D., H. Hanson, J. X. Fan, K. R. Smith, L. Kowaleski-Jones, B. B. Brown & I. Yamada (2013) Re-visiting the relationship between neighbourhood environment and BMI: an instrumental variables approach to correcting for residential selection bias. *International Journal of Behavioral Nutrition and Physical Activity,* 10.

Vita

Yanqing Xu was born in Jingmen City, Hubei Province, China. She went to college in 2004 and received her Bachelor of Engineering degree in Geographic Information Systems (GIS) at the China University of Geosciences, Wuhan, China, in 2008. In the meantime, she also obtained a Bachelor of Arts degree in English at the Huazhong University of Science and Technology, Wuhan, China. Since the fall of 2008, she began her graduate study and earned the Master of Science degree in Cartography and GIS in the Department of Remote Sensing at Wuhan University in 2010. She also held an internship in the Bureau of Mapping and Surveying of Hubei Province from 2009 to 2010. In the fall of 2010, she has been a Ph.D. student in the Department of Geography and Anthropology, Louisiana State University, supported by a teaching assistantship and other scholarships. Her research focuses on human geography, public health, GIS and spatial analysis.