ACCESSIBILITY TO URBAN PARKS AND HEALTH OUTCOMES AT THE NEIGHBORHOOD LEVEL

A Thesis
Presented to
The Academic Faculty

by

Nicholas Boyd

In Partial Fulfillment
of the Requirements for the Degree
Master of Science in the
School of Civil and Environmental Engineering

Georgia Institute of Technology
AUGUST 2018

ACCESSIBILITY TO URBAN PARKS AND HEALTH OUTCOMES AT THE NEIGHBORHOOD LEVEL

Approved by:

Dr. Kari Watkins, Advisor
School of Civil and Environmental Engineering
Georgia Institute of Technology

Dr. John E. Taylor
School of Civil and Environmental Engineering
Georgia Institute of Technology

Dr. Gregory S. Macfarlane
Transport Foundry

Date Approved: July 27, 2018

To Emma

ACKNOWLEDGEMENTS

I would like to express my deepest thanks to Dr. Kari Watkins, who encouraged me to take the idea that became this thesis and run with it, supporting me in every step along the way. I owe quite a debt of gratitude to Dr. Gregory Macfarlane, without whom this thesis would not have been possible. Dr. Macfarlane's expertise in spatial econometrics was my north star while working on this research and I offer him my thanks for his bottomless patience and willingness to help. My thanks also go to Dr. John Taylor, who provided me with thoughtful feedback and data to see this project through to completion.

Developing any research project can be a stressful endeavor and, as such, I owe my thanks to the people who supported me on a more personal level throughout this process. I extend my thanks to my parents and to Midge, John & Chrise for taking an interest in my work and always being willing to listen to me talk about it, despite their lack of familiarity with the technical details involved. Lastly, I thank Emma, who was my emotional rock throughout this process and has always been there for me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

This research identifies the correlation between access to urban parks and physical activity and obesity outcomes at the neighborhood level. Using data for New York City, we created a new measure for access to parks called 'park choice accessibility.' Park choice accessibility uses a destination choice-derived framework to interact distance to parks and the quality of those parks as defined by their size and other potential amenities. A small park very close to a neighborhood can have an impact on health outcomes, but a larger park at a similar distance may have an even larger impact. Similarly, a large park can be further away and still have an impact on health outcomes. We assess whether park choice accessibility is associated with increased physical activity or decreased prevalence of obesity at the neighborhood level, controlling for spatially correlated and endogenous effects in addition to socioeconomic covariates such as age, marital status, income, and educational attainment.

Our results suggest that there is no statistically significant relationship between park access and obesity prevalence. However, better park access is associated with a marginal increase in physical activity, suggesting that improving park access throughout cities may serve as a pathway toward achieving physical activity benchmarks. Importantly, this study is subject to several limitations. Public health data were obtained at the census tract level, which may reveal nuances that are not visible at larger spatial scales such as the city or county levels but may obscure relationships that are visible at more disaggregated scales. Additionally, distances from census tracts to parks were calculated using centroids, which may not represent the true distance people would have to travel to access a park from a given census tract. Using network distance and population weighted centroids could provide more accurate estimates of the real distance people

would have to travel to get to parks in the network. Moreover, calculating the distance to park entrances instead of park centroids would further improve distance estimates.

CHAPTER 1. INTRODUCTION

The United States is currently facing an epidemic of obesity and chronic diseases, which are non-communicable diseases of long duration and typically slow development, including cardiovascular diseases, chronic respiratory diseases, diabetes, stroke, joint and bone diseases, and cancer (World Health Organization 2014). Recent statistics suggest that approximately 300,000 premature deaths each year can be attributed to chronic diseases (World Health Organization 2014). According to the Trust for Public Land, "On average, an obese American spends nearly $1,500 more per year in health care costs than an American of normal weight, for a national total of $147 billion in direct medical expenses" (Harnik & Welle 2011, p.5). As obesity and chronic disease have become rampant, it is no surprise that healthcare costs have risen to nearly one-fifth of the United States' gross domestic product (Harnik & Welle 2011).

While a moderate amount of regular physical activity has been established as an effective strategy for reducing and managing obesity and many of the aforementioned chronic diseases (Office of Disease Prevention and Health Promotion 2008; Centers for Disease Control and Prevention 2009a; Durstine et al. 2013), Wolf notes that "more than 50 percent of U.S. adults do not get enough physical activity to provide health benefits; 24 percent are not active at all in their leisure time. Activity decreases with age and sufficient activity is less common among women than men, and among those with lower incomes and less education" (2008, p. 22). Similarly, the United States Department of Health and Human Services notes that less than fifty-percent of Americans meet established recommendations for moderate to vigorous physical activity, or MVPA (U.S. DHHS 2010). As the epidemiological transition from infectious to chronic diseases is now complete in the developed world, increasing physical activity has become a vital public health task in the 21st century.

Research has shown that the design of the built environment influences a range of behaviors, including those related to physical activity. The presence of trees and other vegetation in outdoor environments have been shown to be positively associated with physical activity (Pretty et al 2005); one finding suggests that after sidewalks and trails have been constructed, the introduction of natural elements positively impacts motivation to engage in physical activity (Suminski et al. 2005). Additional evidence indicates that commonly vegetated areas, such as parks and open space, support outdoor physical activity (Giles-Corti et al. 2005; Wells et al. 2007). Perhaps most telling is the finding that "people in large cities perceive themselves to be generally healthier if a greater percentage of the living environment is greenspace, are inclined to be more active, and claim the ability to relax faster" (Wolf 2008, p. 24). Thus, by providing space for active recreation, public parks and greenspaces may increase the likelihood of engaging in more physical activity. As such, public investment in parks can be thought of as a public health intervention for chronic diseases and conditions, as well as general population health.

To date, the literature exploring the relationship between parks and health outcomes, specifically those related to physical activity and obesity, has yielded mixed results. Importantly, there is considerable variation in the design of past studies, including the spatial scale of analysis, the population of interest, and the measure of proximity and/or accessibility. While studies of neighborhood level health impacts do exist, they typically focus on discrete populations of interest within individual parks instead of examining the impacts of larger park networks (Bancroft et al. 2015; Coutts 2008; Aspinall et al. 2010; Rigolon and Nemeth 2016; Sallis et al. 2016). Other studies have used the city and metropolitan statistical area as the spatial scale of analysis, which in some cases has resulted in the discovery of positive associations between increased park space and positive health impacts but may conceal more nuanced relationships that exist at smaller spatial scales (Larson, Jennings & Cloutier 2016). Evaluating the health impacts of parks and greenspaces using a smaller unit of analysis is an important research gap to fill, especially considering that park use and physical activity within parks varies considerably according to residential proximity to parks and park facilities as well as a number of sociodemographic factors (Kaczynski et al. 2014). There are important urban design considerations here; is it better to build a single large park with many different amenities, or to build a series of smaller parks nearer people's homes. Similarly, do parks placed near low-income households affect the behavior of those households? This study attempts to fill that gap using New York City as a case study example. Using a new measure of park access called 'Park Choice Accessibility,' we find that that there is no statistically significant relationship between park access and obesity prevalence but having better access to parks is associated with a modest increase in physical activity participation.

CHAPTER 2. RESEARCH BACKGROUND

2.1 Review of Urbanization & Public Health Outcomes

Across the world, the population of cities is growing at an unprecedented rate. With a current estimated global urban population of over 3.5 billion people, the United Nations Population Fund estimates that by 2030, 5 billion people will inhabit cities worldwide. By 2050, an additional 3 billion people will live in cities – a sixty percent increase over twenty years – with much of this growth expected to take place in the developing world (United Nations Population Fund, 2016). Despite the greater expected concentration of urban growth in other parts of the globe, the trend of urbanization is highly visible in the United States as well. According to the U.S. Census Bureau, "The nation's urban population increased by 12.1 percent from 2000 to 2010, outpacing the nation's overall growth rate of 9.7 percent for the same period" (Ratcliffe 2012, p. 1). Nowak and Walton estimate that urban land area as a percentage of total land area in the U.S. will increase from 3.1% in 2000 to approximately 8.1% by the year 2050, with urbanized places collectively comprising a land area larger than the state of Montana (2005). More than eighty percent of the U.S. population now lives in urban areas, compared to sixty-four percent in 1950 (United States Census Bureau 2007).

The global trend of urbanization has profound implications for population health, both positive and negative. Because urbanization corresponds with an increase in population density compared to rural and suburban settlement patterns, residents of dense urban areas throughout the world have better access on average to many health services and health-promoting amenities simply because of their closer proximity to such resources (Larson, Jennings & Cloutier 2016). However, despite this benefit of increased density and the ability of cities to provide many opportunities for innovation, economic growth and social progress, historical and current evidence suggests that rapid urban growth often leads to congestion and numerous negative environmental and human health outcomes. Interactions between growing urban populations and their environment, marked by intensive resource consumption and ecologically harmful patterns of development, lead to numerous undesired consequences including pollution and sanitation issues as well as racial and socioeconomic disparities (Larson, Jennings & Cloutier 2016). These urbanization-induced stressors can heighten the susceptibility of populations to a wide range of health problems (Larson, Jennings & Cloutier 2016).

Because of the considerable growth of urbanization throughout the world, many city planning and public health professionals have begun to pay more attention to the role of the built environment in promoting or discouraging healthy behaviors. Until recently, most large-scale health promotion efforts focused on individual-level interventions intended to educate people about healthy lifestyles and behaviors, touching on topics including diet and exercise. According to Coutts (2008) however, this trend is shifting as professionals "have begun adopting an ecological paradigm, accepting that both individual and environmental determinants play a role in health behavior. This new, arguably revisited, public-health paradigm accounts not only for the compositional (who you are) but also for the contextual (where you are) influences on physical activity" (p. 552). As professionals begin to operate from the assumption that the design and configuration of the built environment can facilitate or inhibit physical activity, they are increasingly looking to public spaces like parks and greenways as key elements of the built environment that can support exercise (Coutts 2007; Bedimo-Rung, Mowen & Cohen 2005).

2.2 Review of the Health Benefits of Parks

A large body of literature exists documenting the many benefits of urban parks and greenspaces to human health and wellbeing. Importantly, these benefits stem from the provision of ecosystem services, which occur when the natural environment supplies something that people demand, improving quality of life and well-being (Larson, Jennings & Cloutier 2016). These services can include the provisioning of goods such as fresh water and agricultural products; regulatory functions including protection of drinking water quality, heat mitigation, air purification, and stormwater management; and cultural functions, such as improving aesthetics, providing opportunities for recreation, tourism, and physical and mental health, and promoting biodiversity (McDonald 2015). According to Wolch, Byrne and Newell, parks and greenspaces provide "…a wide range of ecosystem services that could help combat many urban ills and improve life for city dwellers—especially their health…Ecosystem services provided by urban greenspace not only support the ecological integrity of cities, but can also protect the public health of urban populations" (2014, p. 234). Table 1 summarizes commonly cited benefits of parks and greenspaces.

Table 1 - Commonly Cited Benefits of Parks and Greenspaces

Category	Summary of Benefits
Physical Health	Provision of clean drinking water, fostering increased physical activity, promoting faster healing in hospitals, reduction of heat-related mortality, reduced incidence of cardiovascular-related mortality, improved air quality and related reductions in respiratory-related mortality, reduced incidence of childhood asthma, and improved birth outcomes (Benedict and McMahon 2006; Cotrone 2015; Akbari, Pomerantz, and Taha 2001; Beattie, Kollin, and Moll 2000; Nowak 2002; Lovasi et al. 2008; Wolf 2008; Mitchell and Popham 2008; Donovan et al. 2013; National Urban and Community Forestry Advisory Council 2015; Stone and Norman 2006)
Mental Health	Reduced stress and mental fatigue, reduced aggression, enhanced emotional and cognitive development, improved behavioral outcomes in youth (Benedict and McMahon 2006; Wolch, Byrne, and Newell 2014; Kuo and Sullivan 2001a; NUFAC 2015; Ernston 2013; Ulrich 1981; Ulrich et al. 1991; Lee and Maheswaran 2010)
Social Health	Enhanced community aesthetics, crime reduction, increased social interaction (Benedict and McMahon 2007; Kuo and Sullivan 2001b; Kuo 2003; Wolfe and Mennis 2012; Sullivan, Kuo, and DePooter 2004)
Economic Health	Provision of ecosystem services, increased residential property values and municipal property tax revenues, attraction of more shoppers and increased economic activity to commercial districts (American Forests 1997; Benedict and McMahon 2006; Coder 1996; McDonald 2015; Lerner and Poole 1999; Anderson and Cordell 1988; Seila and Anderson 1982, 1984; Donovan and Butry 2010; Schwab 2009; Wolf 1999).

Given the general benefits of parks and greenspaces in Table 1, several studies have attempted to quantify the impacts of these spaces on different facets of health and wellbeing across cities, yielding mixed results. Larson, Jennings and Cloutier (2016) used self-reported scores on the Gallup-Healthways Wellbeing Index to evaluate the relationship between different areas of wellbeing, including physical health, and park quantity, quality, and accessibility in 44 U.S. cities. The authors found that “Park quantity (measured as the percentage of city area covered by public parks) was among the strongest predictors of overall wellbeing, and the strength of this relationship appeared to be driven by parks’ contributions to physical and community wellbeing” (Larson, Jennings & Cloutier 2016, p. 1). While the authors found positive associations between wellbeing and park quality and accessibility, these relationships were not statistically significant.

Additionally, the authors note that income may be a poor predictor of wellbeing and suggest that the relationship between income and wellbeing may be moderated by other factors, including park access (Larson, Jennings & Cloutier 2016).

A study by West, Shores and Mudd (2012) used park data from the Trust for Public Land's 2010 City Park Facts and public health data from the Behavioral Risk Factor Surveillance System (BRFSS) to examine relationships between the density of parkland, parkland per capita, and levels of physical activity and obesity for 67 metropolitan statistical areas in the U.S. The study found a significant, positive association between park density and physical activity and a significant, negative association between park density and obesity. In a study of New York City, Stark et al. (2014) found that the "proportion of neighborhoods that was large or small park space and park cleanliness were associated with lower BMI among NYC adults after adjusting for other neighborhood features such as homicides and walkability, characteristics that could influence park usage" (p. 2).

Interestingly, in a study by Richardson et al. (2012) that examined the relationship between urban greenspace and selected mortality rates, the authors did not find a statistically significant relationship between the quantity of urban greenspace and mortality caused by lung cancer, diabetes, heart disease or car accidents, but found that all-cause mortality was significantly higher in greener cities. The authors speculate that this is because greener cities are often sprawling cities, suggesting that the negative health effects of urban sprawl may outweigh the positive health effects of having more greenspace (Richardson et al. 2012). In a meta-analysis of 20 peer reviewed journal articles exploring the relationship between parks and objectively measured physical activity, Bancroft et al. (2015) found that five studies reported a significant positive association between the two, six studies produced mixed results, and nine studies found no association at all.

While several of the aforementioned studies yielded results consistent with the hypothesis that park access and use confer health benefits stemming from increased physical activity and attendant decreases in obesity, all of the studies – except some that were included in the meta-analysis by Bancroft et al. (2015) – were conducted at the level of the city or metropolitan statistical area, potentially concealing variation in health outcomes at more fine-grained levels of analysis (Larson, Jennings & Cloutier 2016).

2.3 Review of Measuring Accessibility

To evaluate the impacts of parks and greenspaces on health outcomes at any scale, one of the fundamental tasks involved is the selection of a measure of accessibility. In a comprehensive meta-analysis of published studies that measure active accessibility (accessibility using non-motorized travel modes including walking and cycling), Vale, Saraiva and Pereira (2014) identified four broad categories of studies based on the active accessibility metric employed: "distance-based, gravity-based or potential, topological or infrastructure-based, and walkability and walk score-type measures" (p. 209). Distance-based measures account only for the Euclidean distance between origins and destinations, while infrastructure-based measures explicitly incorporate relevant transportation networks like roads and sidewalks to more accurately measure travel time and distance. Gravity-based measures incorporate cost measures to model accessibility as a function of a destination's attractiveness (i.e., size, commercial activity, etc.) and the cost of traveling to that destination from a given origin (i.e., travel time or distance) (Vale, Saraiva & Pereira 2014).

Importantly, Vale, Saraiva and Pereira acknowledge that there is not yet a consensus on the most appropriate accessibility metric to use in a given setting, noting that "ways to measure active accessibility are as varied as the number of scholars that measure them" (2014, p. 227). However, their meta-analysis highlights some of the limitations associated with each of the types of metrics described. For example, in describing distance-based accessibility measures, the authors note that such measures are "extremely sensitive to the way in which travel impedance (i.e., distance) is measured. Accordingly, four types of distance can be identified: Euclidean distance, Manhattan distance, shortest network distance, and shortest network time" (Vale, Saraiva & Pereira 2014, p. 216). The appropriateness of one distance-based measure over another can vary significantly depending on the topography of the environment and the travel mode that is being employed. In describing gravity-based accessibility measures, which "assume that travel is a derived demand and there is a tradeoff between the benefit of the opportunity and the cost to reach it from a given origin," the authors note that such measures do not always explicitly account for land use characteristics near origins and destinations, which may impact that true accessibility of those places (Vale, Saraiva & Pereira 2014, p. 219).

In evaluating methodologies of measuring access to urban services, including parks and greenspaces, Logan et al. (2017) note that existing and commonly used approaches "often simplify their measure of proximity by using large areal units and by imposing arbitrary distance thresholds," which often results in access-poor populations being overlooked (p. 1). The authors concede that many existing approaches have long been necessary because the computational power required to use higher-resolution analytical techniques was unavailable. However, due to recent advances in computational power and the advent of municipal open-data policies, Logan et al. (2017) recommend that future analyses of accessibility disaggregate population data to the building or parcel level and use network distance instead of Euclidean distance to measure proximity.

CHAPTER 3. METHODOLOGY

This study makes two important contributions to the study of relationships between park access and health outcomes. The first pertains to how access to parks is measured. Whereas past studies have used purely distance-based areal population metrics (i.e., buffers) and/or the land area devoted to parks to operationalize park access, this study borrows methods from transportation engineering and analysis, specifically the modeling of transportation destination choice and trip generation and attraction, to develop a new measure of accessibility. This new metric is a highly adaptable measure of park access that is capable of incorporating multiple park system attributes simultaneously to generate a weighted measure of accessibility across an entire park network. The second contribution pertains to the application of spatial econometric methods to estimate the impacts of park access on obesity and physical activity outcomes. Past studies have failed to account for the presence of spatial dependence in the variables included in regression models used to estimate relationships of interest, resulting in estimates that may be biased and/or inconsistent. The econometric methods used in this study correct for spatial dependence and produce unbiased estimates as a result. A discussion of the methodologies employed to realize these two contributions follows, beginning with a description of the data preprocessing that was conducted prior to beginning the analysis.

3.1 Data Preprocessing

Development of the measure of park accessibility used in this study and subsequent regression analysis to evaluate relationships of interest was done using a geographic information system (GIS) and two spatial datasets: a shapefile of public parks and greenspaces within New York City's municipal boundaries and a shapefile of census tract boundaries containing several sociodemographic variables of interest. Prior to calculating the park access metric and developing regression models, the parks shapefile was checked for data accuracy to ensure that all features within the shapefile were, in fact, parks. Upon inspection, several features were identified for removal from the data sets, including public sports facilities such as Yankee Stadium and its surrounding parking lots. After removing these features, the size, in acres, of all remaining parks was calculated using ESRI's ArcGIS software and appended to the attribute table of the parks shapefile. The census tract shapefile was examined visually to identify outlying census tracts that might skew subsequent analysis, either because they were considerably isolated from other tracts by distance or were significantly different in size compared to other census tracts. All outlying census tracts were removed.

Several sociodemographic variables have been identified in the literature as important correlates of health outcomes, including race, income, educational attainment, and others (Larson, Jennings & Cloutier 2016). Relevant sociodemographic data were obtained at the census tract level from the American Community Survey's 2011-2015 5-year estimates (ACS). Public health data were obtained at the census tract level from the Centers for Disease Control and Prevention's 500 Cities Project (CDC), which publishes key public health statistics such as obesity and physical activity participation rates for the largest 500 cities in the United States (CDC 2018). After cleaning the data and appending public health and sociodemographic variables to the census tracts shapefile, the parks and census tracts data sets were imported into the R statistical software package for analysis. A total of 2,195 census tracts and 12,491 parks were included in the shapefiles. Due to missing data, 2,115 census tracts were used for subsequent regression analysis. Table 2 presents the variables that were included in the initial regression models.

Table 2 - Public Health and Sociodemographic Variables

Name	Type	Description	Source
Obesity	Dependent	The % of Behavioral Risk Factor Surveillance System (BRFSS) respondents aged ≥ 18 years with a body mass index ≥ 30 kg/m2 based on self-reported height and weight.	CDC
Physical Inactivity		The % of Behavioral Risk Factor Surveillance System (BRFSS) respondents aged ≥ 18 years who reported no leisure-time physical activity in the preceding two weeks.	
Income	Independent	The % of residents within a certain income range, segmented into 10 categories, ranging from ≤ $10,000 to ≥ $200,000.	ACS 2011-2015 5-Year Estimates
Population Density		Logarithm of the number of residents per acre	
Fulltime Work		The % of residents who worked 50 to 52 weeks in the preceding 12 months.	
College Degree		The % of residents with a bachelor's degree or higher.	
Single		The % of residents aged 15 years and over that is divorced, separated or never married.	
Age 0-17		The % of residents age 0-17.	
Age 18-29		The % of residents age 18-29.	
Age 30-64		The % of residents age 30-64.	
Age 65+		The % of residents age 65+.	
Race White		The % of white residents.	
Race Black		The % of black residents.	
Race Native American		The % of Native American residents.	

Table 2 continued

Name	Type	Description	Source
Race Asian	Independent	The % of Asian residents.	
Race Pacific Islander		The % of Pacific Islander residents.	
Race Other		The % of residents who identify as a race other than those listed above.	
Hispanic		The % of residents who are Hispanic.	

3.2 Evaluating Access to Parks

In a typical park destination choice model, the probability of a person residing in census tract i choosing park *j* from the set of all parks *P* is:

$$P(j) = \frac{\exp(U_{ij})}{\sum_{p \in P} \exp\,(U_{ip})}$$

where the empirical utility of each park U_{ij} is a function of the travel costs d from tract i to park j, and the amenities A at *j*

$$U_{ij} = \beta \quad A_j + \beta_d d_{ij}$$

A key purpose in applying this framework is a result from theoretical economics, which holds that the logarithm of the denominator in equation 1 – called the *log-sum* – represents the *consumer surplus* for the choice maker, or the total value of all the choices in the choice set (Train 2003).

$$CS_i = \log(\sum_{p \in P} \exp\left(U_{ip}\right))$$

In other words, CS_i represents the total accessibility to all the parks in the region, weighted for the amenities of the parks and their distance from the residents' homes. In principle, A may be any linear-in-parameters function of the attributes of the park, including its size, the presence of sports fields or playgrounds, the access fee, etc. This represents an improvement over classifying access in terms of the binary proximity (e.g., a park within ½ mile). The natural logarithm of the denominator is a measurement of the weighted total accessibility of the choice set from census tract j and is what is used as the measure of park accessibility in this study. Intuitively, census tracts with better access to parks will have higher log-sums than census tracts with inferior access.

For this study, the only attribute of the park we have access to is the parks' size in acres, and the Euclidean distance between the tracts and parks in miles, calculated respectively with GIS software in ESRI ArcMap and R (Bivand & Lewin-Koh 2017). In standard practice, the β parameter coefficients on size and distance would be estimated from a survey of park use. As this was unavailable, we undertook a manual calibration process. Calibration of the β parameters for the size and distance terms was done by mapping different specifications of the park access metric and visually inspecting the resulting distribution of park access throughout the city to check for reasonableness. We assert that the coefficient on distance is negative, and the coefficient on size is positive; thus, larger and nearer parks contribute more to accessibility than smaller more distant ones.

Figure 1 shows two iterations of the β calibration process for NYC that are clearly not reflective of the true distribution of park access throughout the city. The first iteration shown in

Figure 1 uses a β size parameter of 0.00001 and a β distance parameter of -0.04. The second iteration uses a β size parameter of 0.01 and a β distance parameter of -0.02. Figure 2 shows a map of the distribution of park access in NYC using the final calibrated size and distance β parameters of 0.00001 and -5, respectively.

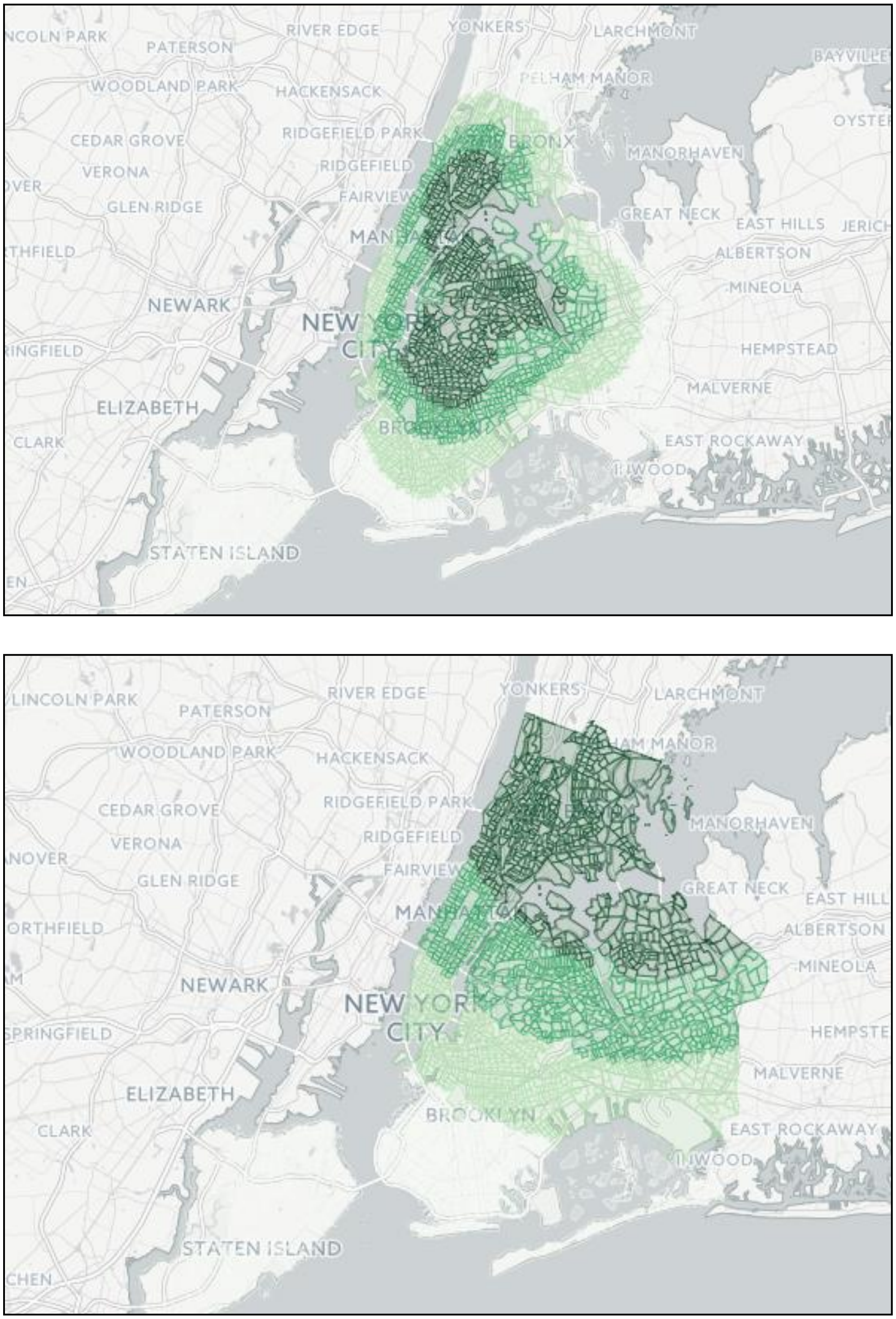

Figure 1 - Mis-specified Iterations of Size and Distance β Calibration in NYC(Top: Size 0.00001, Distance -0.4; Bottom: Size 0.01, Distance -0.02)

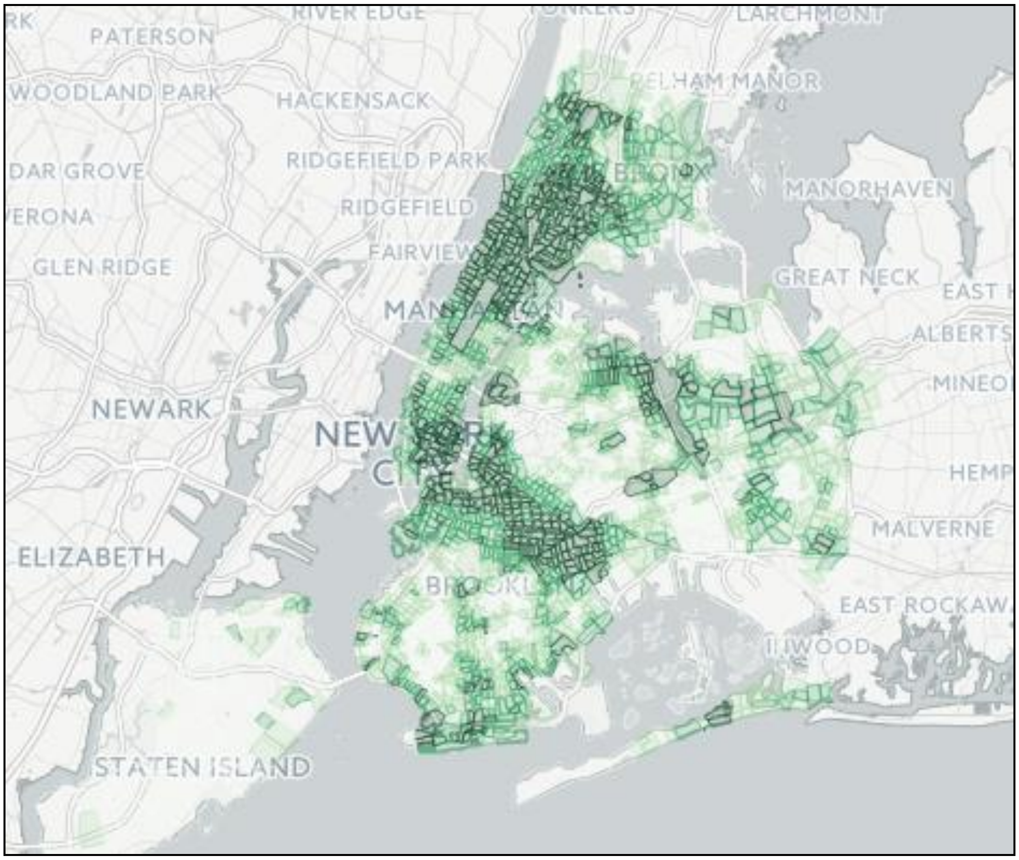

Figure 2 - Map of NYC Park Access Using Final Calibrated Size & Distance β Parameters (0.00001, -5)

An advantage of using the log-sum framework described above to calculate park accessibility is that this framework is highly flexible and can incorporate many different attributes of a park network, with data availability serving as the only limitation. For example, where travel time estimates from census tracts to parks are available, these estimates could be substituted for, or used in conjunction with, the distance measurements that are already included in the model presented here. Additionally, metrics that capture the presence or lack of park amenities that may be conducive to physical activity and positive health outcomes could also be included using the log-sum framework.

In addition to size and distance attributes, we were interested in incorporating some measure of park usage into the park access metric. While survey data would have been ideal for this purpose, none was available. However, we were able to acquire data for the number of Tweets recorded through the Twitter social media platform within each park for the month of September 2014 and used this data to estimate park usage. After developing the park access log-sum framework using size and distance parameters, data became available for the number of Tweets recorded through the Twitter social media platform within each park for the month of September 2014. In theory, parks with more Tweets are the sites of more activity and greater visitation than parks with fewer Tweets. Based on this assumption, the natural logarithm of the number of Tweets recorded in each park in September 2014 was added to the log-sum framework as a third parameter in the calculation of park access, such that $U_i = \beta_s S_i + \beta_d D_i + \beta_t T_i$, where T_i is the natural logarithm of the number of Tweets recorded in each park and β_t is the corresponding coefficient. A Tweet β parameter of 0.001 was used in the final calibrated models.

3.3 Applying Spatial Econometric Methods to Model Relationships of Interest

Multivariate regression is a fundamental econometric tool that is used for the purposes of identifying causal relationships between phenomena and for prediction and forecasting of future trends (Angrist & Pischke 2014). One of the most commonly used multivariate regression tools is ordinary least squares (OLS) linear regression, in which "it is assumed that the values of the coefficients of the independent (explanatory) variables are constant across the spatial extent of the study" (Srinivasan 2016, p. 1). The formula for the standard linear model using OLS is:

$$Y = \alpha + X\beta + e,$$

where Y denotes the dependent variable, α denotes the intercept, X denotes a matrix of exogenous explanatory variables with corresponding coefficients β, and e is a vector of error terms that are assumed to be independent of one another (Elhorst 2010).

Within the realm of regional science, regression modeling is often done using data that is inherently spatial in nature, such as neighborhood demographic characteristics, real estate prices, migration flows between regions, and movement between origins and destinations using various components of the transportation network (LeSage 2014). Importantly, classic regression techniques like OLS make several key assumptions, including normality of the dependent variable, a lack of strong correlations between the independent variables, and independence of observations, among others. According to LeSage (2014), spatial data "typically violates the assumption that each observation is independent of other observations made by ordinary regression methods," making techniques like OLS inappropriate for modeling many spatial data sets (p. 1).

When working with spatial data, modeling techniques like OLS can sometimes produce misleading results because of a phenomenon known as spatial dependence, or spatial autocorrelation, "which is a property of data that arises whenever there is a spatial pattern in the values, as opposed to a random pattern that indicates no spatial autocorrelation" (Srinivasan 2016, p. 1). This phenomenon was most famously described by geographer Waldo Tobler through what he called the First Law of Geography: "everything is related to everything else, but near things are more related than distant things" (Tobler 1970, p. 236). Tobler famously applied this law to a simulation of urban population growth in Detroit, MI (Tolber 1970).

Several global and local measures of spatial autocorrelation have been developed to test for independence of observations across the spatial extent of a study, including Moran's I, which is perhaps the most commonly used metric of spatial autocorrelation, as well as Geary's C and Anselin's Local Moran's I, which reveals local patterns of clustering (Srinivasan 2016). Where spatial autocorrelation is indicated by the results of one or more of the aforementioned statistical tests, classic regression techniques like OLS, which do not control for the effects of spatial dependence, should be discarded in favor of methods that explicitly account for and control spatial effects (LeSage 2014, Srinivasan 2016, Elhorst 2010, LeSage and Pace 2009).

A number of spatial econometric techniques have been developed over time to correct and/or control for the presence of spatial dependence, resulting in more reliable estimates of spatial relationships of interest. These techniques incorporate a spatial weights matrix, *W*, into the formula of the regression equation being used, which specifies the relationship between neighboring spatial units. Two commonly used neighbor relationships are rook contiguity, in which a spatial unit's neighbors are defined as those that are directly above, below or to either side, and queen contiguity, in which a spatial unit's neighbors can lie in any direction, including along diagonals (Golgher & Voss 2015; LeSage & Pace 2009).

Importantly, there are three different kinds of spatial interaction effects that help to explain the occurrence of spatial dependence: endogenous interaction effects, exogenous interaction effects, and correlated effects (Elhorst 2010, Manski 1993). Each type of interaction effect has different implications for the type of spatial econometric technique that should be used to control for underlying spatial dependence. In this study, four types of spatial econometric techniques were used to model the relationship between park access and obesity and physical activity outcomes: the spatial lag model (or spatial autoregressive model), the spatially-lagged error model, the spatial Durbin error model, and the spatial Durbin model. The specification of each model is described in turn below, as well as the motivation for each and the types of interaction effects that each controls for. This section ends with a discussion of how the most appropriate models for this study were selected.

3.3.1 Spatial Autoregressive Model (SAR)

The SAR model is motivated by endogenous interaction effects between the value of the dependent variable in a target unit of analysis and the average value of the dependent variable in neighboring units, generating "a process of 'global spillover' indicating that changes in an independent variable anywhere in the study domain will affect the value of the dependent variable everywhere, even when the declaration of neighborhood influences implicit in the matrix W represents simple 1st-order contiguity" (Golgher & Voss 2015, p. 180). The formula for the SAR model is similar to that of the classic OLS model but includes an average of neighboring values of the dependent variable, such that:

$$Y = \alpha + \rho W y + X\beta + e,$$

where ρ is the coefficient of the spatially lagged dependent variable, Wy (Golgher & Voss 2015, Elhorst 2010, LeSage & Pace 2009).

3.3.2 Spatial Error Model (SEM)

The SEM model is motivated by correlated effects, "where similar unobserved environmental characteristics result in similar behavior" (Elhorst 2010, p. 11). To determine whether the SEM model is appropriate for a given data set, the residuals of an OLS regression must be examined. If there is strong spatial autocorrelation in the residuals (error terms), a SEM model may be the correct econometric method to apply.

Like the SAR model, the formula for the SEM model is quite similar to that of the classic OLS model but differs in its inclusion of a spatially autocorrelated error term, such that:

$$Y = \alpha + X\beta + u, \quad u = \lambda W u + e,$$

where u is the spatially autocorrelated error term, "λ is the coefficient expressing the average strength of spatial correlation among the errors (conditional on W) and W is the weight matrix representing the spatial structure of neighbor influences among the residuals" (Golgher & Voss 2015, p. 179).

3.3.3 Spatial Durbin Error Model (SDEM)

The SDEM model is motivated by both correlated effects and exogenous interaction effects between the values of the independent variables in neighboring spatial units and the value of the dependent variable in a target spatial unit (Golgher & Voss 2015). The formula for the SDEM model differs from the classic OLS model through inclusion of a spatially autocorrelated error term and spatially-lagged independent variables, such that:

$$Y = \alpha + X\beta_1 + WX\beta_2 + u, \quad u = \lambda W u + e,$$

where *WX* is a matrix of spatially lagged independent variables with coefficients β_2 and u is the spatially autocorrelated error term containing coefficient λ (LeSage 2014; Elhorst 2010). Elhorst (2010) notes that use of the SDEM model is fairly uncommon in the literature.

3.3.4 Spatial Durbin Model (SDM)

The SDM model is motivated by endogenous interaction effects between the value of the dependent variable in a target spatial unit and the value of the dependent variable in neighboring spatial units as well as exogenous interaction effects between the values of the independent variables in neighboring spatial units and the value of the dependent variable in a target spatial unit (Elhorst 2010). First introduced by Anselin in 1988, the SDM model features a spatially lagged dependent variable and spatially lagged independent variables, such that:

$$Y = \alpha + \rho Wy + X\beta_1 + WX\beta_2 + e,$$

where ρ is the coefficient of the spatially lagged dependent variable, *Wy*, and WX is a matrix of spatially lagged independent variables with coefficients β_2 (LeSage & Pace 2009; Elhorst 2010; Golgher & Voss 2015). The SDM model does not feature a spatially autocorrelated error term.

3.3.5 Selecting the Most Appropriate Spatial Model

The literature reveals a difference in opinion regarding how to compare spatial econometric models and select the most appropriate one for a given data set. According to Elhorst (2010), the first step in model comparison is to estimate an OLS regression and use the Lagrange multiplier (LM) test to evaluate whether the SAR or SEM model is more appropriate to describe the observations. If the OLS model is rejected in favor of one or both of the spatial models, the SDM model should then be estimated. A likelihood ratio (LR) test can then be used to examine whether the SDM model can be simplified to a SAR or SEM model. If simplification is not possible, then the SDM model best describes the data. Otherwise, one of the two simplified models should be selected according to the results of the LR test.

In contrast to the process Elhorst (2010) describes, LeSage (2014) argues that in practice, the only two spatial econometric models that are ever needed are the SDM and SDEM models. Selection between these is dependent on the type of spillover process being modeled. According to LeSage, “A (spatial) spillover arises when a causal relationship between the *r*th characteristic/action of the *i*th entity/agent located at position *i* in space exerts a significant influence on the outcomes/decisions/actions (y_j) of an agent/entity located at position j” (2014, p. 3). LeSage distinguishes between two kinds of spatial spillovers: local and global. A defining feature of local spillovers is that there are no feedback effects or endogenous interactions between spatial units. Global spillovers, on the other hand, are characterized by the presence of both. If the process being modeled is one whose underlying data generation process consists of local spillover effects, then the SDEM model is the more appropriate one to use. However, if the underlying data generation process consists of global spillovers, the SDM model is preferred. LeSage notes that “Despite the fact that global spillover situations are likely rare, the spatial regression specifications most commonly used in applied regional science literature are those associated with global spillovers, not local” (2014, p. 5).

CHAPTER 4. RESULTS

Results are presented below for models pertaining to obesity and physical activity in turn. For each outcome of interest, model coefficients and goodness-of-fit measures are shown for the spatial autoregressive model, the spatial error model, the spatial Durbin error model, and the spatial Durbin model discussed in Chapter 3 above. Three models of each type are shown – the first does not include park access as an independent variable, while the second includes park access as a function of only size and distance parameters and the third adds the natural logarithm of the number of Tweets in each park to the calculation of the park access measure.

Importantly, not all of the independent variables listed in Table 2 are included in the models below. Early iterations of each model revealed income to be highly insignificant, resulting in its removal from subsequent iterations. Additionally, early model iterations used the raw value of the park access measure as an independent variable. However, using the natural logarithm of park access instead of the raw value resulted in a considerably better model fit across all model types. As such, the natural logarithm of park access is used in the regressions detailed below. The following sections present an overview of how the optimal model was selected for each dependent variable. Each subsection concludes with a comprehensive presentation of the results for the optimal models.

4.1 Modeling the Impact of Park Access on Obesity Prevalence

Table 3 below presents the model coefficients for all the aforementioned spatial models using obesity prevalence (see Table 2) as the dependent variable. For all model types, the inclusion of the park access measure as an independent variable improves goodness-of-fit compared to the models without park access, as evidenced by the lower log likelihood value for Model 2 of each model type compared to Model 1. Adding Tweets into the park access measure further improves goodness-of-fit for all model types, as evidenced by the even lower log likelihood value for Model 3 of each type. Including Tweets in the formulation of park access improves goodness-of-fit by a much larger margin than simply adding park access as a function of size and distance to a model with no park access measure.

The SAR models were discarded because the most likely SAR model (Obesity SAR 3) did not fit the data as well as the most likely SEM, SDM and SDEM models. The SEM models were also discarded because the SDM and SDEM models showed superior goodness-of-fit. While the most likely SDM and SDEM models (Obesity SDM 3 & SDEM 3) were within approximately 10 points of one another on the log likelihood scale, Obesity SDM 3 proved superior and was selected as the optimal model to describe the data. Table 4 below presents the model coefficients and the results of impact simulations that were carried out for Obesity SDM 3, which are required to evaluate the effects of the independent variables on the dependent variable in a spatial Durbin model.

According to the impact simulations presented in Table 4, park access does not have a statistically significant impact on obesity prevalence, although the impact coefficients all have a negative sign, indicating that the direction of the relationship is such that increases in park access are expected to correlate with decreases in obesity prevalence. The total impact coefficients show that decreases in population density, the percentage of residents with a college degree, and the percentage of residents in any age and race category are associated with a statistically significant increase in obesity prevalence. Increases in the percentage of single and Hispanic residents are also associated with a statistically significant increase in obesity prevalence.

Several variables show interesting spatial trends in the significance of their impacts. The direct impact coefficients for fulltime employment, the percentage of residents with a college degree, and the percentage of Hispanic residents show that changes in the value of these variables within a target census tract result in a statistically significant increase in obesity prevalence within that same census tract. However, the indirect impact coefficients for these variables are not statistically significant, indicating that changes in the average value of these variables in neighboring census tracts have no statistically significant impact on obesity prevalence in the target census tract. Conversely, the direct impact coefficients for the percentage of white residents and the percentage of residents who identify as a race other than white, black, or Asian are insignificant while the indirect impact coefficients for these variables are highly significant, indicating that changes in the average value of these variables in neighboring census tracts are associated with a statistically significant increase in obesity in a target census tract.

The direct and indirect impact coefficients for the percentage of black residents are significant but have opposing signs, revealing a unique pattern that is unobserved in the other independent variables. The direct impact coefficient indicates that an increase in the percentage of black residents within a target census tract is associated with a statistically significant increase in obesity prevalence within that same tract. However, the indirect impact coefficient indicates that a decrease in the percentage of black residents in neighboring census tracts is associated with a statistically significant increase in obesity prevalence within the target census tract. Essentially, majority black census tracts have a greater prevalence of obesity than majority non-black tracts, but tracts with neighbors that are majority black have a lower prevalence of obesity.

Table 3 - Spatial Models of Obesity Prevalence

	SAR 1	SAR 2	SAR 3	SEM 1	SEM 2	SEM 3	SDM 1	SDM 2	SDM 3	SDEM 1	SDEM 2	SDEM 3
(Intercept)	21.84***	21.96***	30.89***	33.12***	33.38***	32.80***	28.29***	28.54***	32.73***	32.92***	33.64***	34.78***
	(1.82)	(1.78)	(2.33)	(1.68)	(1.68)	(2.09)	(1.71)	(1.78)	(2.01)	(1.94)	(2.07)	(2.14)
log(Pop. Density)	-0.24***	-0.09	-0.12	0.18***	0.19***	0.11	0.18**	0.19***	0.17*	0.07	0.10	0.10
	(0.05)	(0.05)	(0.07)	(0.05)	(0.06)	(0.07)	(0.06)	(0.06)	(0.07)	(0.06)	(0.06)	(0.07)
Fulltime Employment	-0.06***	-0.07***	-0.05***	-0.06***	-0.06***	-0.05***	-0.06***	-0.06***	-0.06***	-0.06***	-0.06***	-0.06***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
College Degree	-0.07***	-0.07***	-0.08***	-0.10***	-0.10***	-0.11***	-0.10***	-0.10***	-0.10***	-0.10***	-0.10***	-0.10***
	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)	(0.01)
Pct. Single	0.07***	0.08***	0.06***	0.05***	0.05***	0.03***	0.06***	0.06***	0.04***	0.06***	0.06***	0.05***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Pct. 18-29	-0.15***	-0.16***	-0.16***	-0.13***	-0.14***	-0.14***	-0.13***	-0.14***	-0.14***	-0.15***	-0.16***	-0.15***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Pct. 30-64	-0.05***	-0.06***	-0.07***	-0.07***	-0.07***	-0.06***	-0.06***	-0.06***	-0.05***	-0.07***	-0.08***	-0.07***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Pct. 65+	-0.08***	-0.08***	-0.09***	-0.09***	-0.09***	-0.09***	-0.08***	-0.08***	-0.08***	-0.09***	-0.09***	-0.09***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Pct. White	0.01	0.01	-0.03	0.01	0.01	0.01	0.03*	0.03	0.01	0.02	0.01	0.01
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
Pct. Black	0.05***	0.05***	0.05*	0.09***	0.09***	0.11***	0.10***	0.09***	0.09***	0.09***	0.08***	0.08***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
Pct. Asian	-0.05**	-0.05**	-0.10***	-0.05***	-0.05***	-0.06**	-0.02	-0.02	-0.05*	-0.04**	-0.04**	-0.06**
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Pct. Other	0.04*	0.04*	0.04	0.03*	0.03*	0.07**	0.05**	0.04**	0.04	0.04*	0.04*	0.04*
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Pct. Hispanic	0.02***	0.03***	0.02***	0.04***	0.04***	0.03***	0.04***	0.04***	0.04***	0.04***	0.04***	0.03***
	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Rho	0.42***	0.43***	0.22***				0.66***	0.65***	0.58***			
	(0.01)	(0.01)	(0.01)				(0.02)	(0.02)	(0.02)			
log(park access w/o Tweets)		-1.04***			-0.17			-0.09			-0.23	
		(0.13)			(0.17)			(0.18)			(0.17)	
log(park access w/Tweets)			-0.14*			-0.07			-0.08			-0.08
			(0.06)			(0.06)			(0.06)			(0.06)
Lambda				0.79***	0.79***	0.69***				0.71***	0.70***	0.62***
				(0.01)	(0.01)	(0.02)				(0.02)	(0.02)	(0.02)
lag.log(Pop. Density)							-0.66***	-0.56***	-0.46***	-0.60***	-0.49***	-0.34**
							(0.07)	(0.08)	(0.08)	(0.11)	(0.11)	(0.12)
lag.Fulltime Employment							0.05***	0.04***	0.04***	0.02	0.01	0.01
							(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)

Table 3 continued

	SAR 1	SAR 2	SAR 3	SEM 1	SEM 2	SEM 3	SDM 1	SDM 2	SDM 3	SDEM 1	SDEM 2	SDEM 3
							(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
lag.College Degree							0.07***	0.07***	0.07***	-0.01	-0.01	-0.00
							(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
lag.Pct. Single							-0.01	0.00	0.00	0.02	0.03*	0.03
							(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
lag.Pct. 18-29							0.02	0.01	0.02	-0.06**	-0.07**	-0.04
							(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
lag.Pct. 30-64							-0.01	-0.01	-0.02	-0.03	-0.03	-0.02
							(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
lag.Pct. 65+							-0.00	-0.00	-0.00	-0.04	-0.03	-0.02
							(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
lag.Pct. White							-0.16***	-0.15***	-0.17***	0.03	0.03	-0.01
							(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)
lag.Pct. Black							-0.21***	-0.19***	-0.21***	0.05**	0.04*	0.01
							(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)
lag.Pct. Asian							-0.14***	-0.13***	-0.15***	-0.01	-0.02	-0.04**
							(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)
lag.Pct. Other							-0.15***	-0.14***	-0.15***	0.08***	0.08***	0.05**
							(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
lag.Pct. Hispanic							-0.03***	-0.02**	-0.02*	-0.00	-0.00	-0.00
							(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
lag.log(park access w/o Tweets)								-0.74**			-0.72*	
								(0.25)			(0.32)	
lag.log(park access w/Tweets)									0.04			-0.05
									(0.09)			(0.11)
Num. obs.	2115	2107	1435	2115	2107	1435	2115	2107	1435	2115	2107	1435
Parameters	15	16	16	15	16	16	27	29	29	27	29	29
Log Likelihood	-4370.51	-4307.17	-3047.49	-4190.58	-4170.24	-2906.72	-4104.13	-4074.80	-2818.85	-4105.42	-4083.93	-2829.15
AIC (Linear model)	9558.21	9464.15	6366.51	9558.21	9464.15	6366.51	9291.58	9150.96	6222.61	9291.58	9150.96	6222.61
AIC (Spatial model)	8771.02	8646.33	6126.98	8411.17	8372.48	5845.43	8262.27	8207.61	5695.70	8264.85	8225.85	5716.29
LR test: statistic	789.19	819.82	241.52	1149.04	1093.67	523.08	1031.31	945.35	528.92	1028.73	927.11	508.32
LR test: p-value	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

***$p < 0.001$, **$p < 0.01$, *$p < 0.05$

Table 4 - Optimal Obesity Model Results

	Coefficient	Direct Impact	Indirect Impact	Total Impact
Intercept	32.73*** (2.01)	-	-	-
Log(Pop. Density)	0.17* (0.07)	0.08104	-0.78295***	-0.70191***
Fulltime Employment	-0.06*** (0.01)	-0.05357***	0.02271	-0.03086
College Degree	-0.10*** (0.01)	-0.09704***	0.02285	-0.07419***
Pct. Single	0.04*** (0.01)	0.04916***	0.05676**	0.10592***
Pct. 18-29	-0.14*** (0.01)	-0.15318***	-0.13753***	-0.29071***
Pct. 30-64	-0.05*** (0.01)	-0.06445***	-0.10195***	-0.16641***
Pct. 65+	-0.08*** (0.01)	-0.09475***	-0.11276***	-0.20752***
Pct. White	0.01 (0.02)	-0.02605	-0.35764***	-0.38369***
Pct. Black	0.09*** (0.02)	0.05125**	-0.33482***	-0.28358***
Pct. Asian	-0.05* (0.02)	-0.08898***	-0.38438***	-0.47336***
Pct. Other	0.04 (0.02)	0.00570	-0.27950***	-0.27380***
Pct. Hispanic	0.04*** (0.01)	0.03748***	0.00167	0.03915**
Log(park access w/Tweets)	-0.08 (0.06)	-0.08098	-0.01640	-0.09738
Rho	0.58*** (0.02)	-	-	-
Lagged Log(Pop. Density)	-0.46*** (0.08)	-	-	-
Lagged Fulltime Employment	0.04*** (0.01)	-	-	-
Lagged College Degree	0.07*** (0.01)	-	-	-
Lagged Pct. Single	0.00 (0.01)	-	-	-
Lagged Pct. 18-29	0.02 (0.02)	-	-	-
Lagged Pct. 30-64	-0.02 (0.02)	-	-	-
Lagged Pct. 65+	-0.00 (0.02)	-	-	-
Lagged Pct. White	-0.17*** (0.01)	-	-	-
Lagged Pct. Black	-0.21*** (0.01)	-	-	-
Lagged Pct. Asian	-0.15*** (0.01)	-	-	-
Lagged Pct. Other	-0.15*** (0.02)	-	-	-
Lagged Pct. Hispanic	-0.02* (0.01)	-	-	-
Lagged Log(park access w/Tweets)	0.04 (0.09)	-	-	-

4.2 Modeling the Impact of Park Access on Physical Activity Participation

Table 5 below presents the model coefficients for the spatial models using physical activity participation (see Table 2) as the dependent variable. Like the obesity models described previously, the inclusion of the park access measure as an independent variable improves goodness-of-fit compared to the models without park access, as evidenced by the considerably lower log likelihood value for Model 2 of each model type compared to Model 1. Adding Tweets into the park access measure further improves goodness-of-fit for all model types, as evidenced by the even lower log likelihood value for Model 3 of each type.

The most likely SAR and SEM models (Model 3) were very similar in terms of their goodness-of-fit but were discarded because they did not fit the data as well as the most likely SDM and SDEM models (also Model 3 of each type). While the most likely SDM and SDEM models were within approximately 4 points of one another on the log likelihood scale, Physical Activity SDEM 3 exhibited slightly superior goodness-of-fit, suggesting that spatial dependence exists only in the error term and not in the dependent variable. However, the results of the Local Moran's I test for spatial dependence and clustering revealed statistically significant patterns of spatial dependence in the physical activity variable. Because of this, a likelihood ratio test was conducted to test the null hypothesis that the SDEM model is a better fit than the SDM model. The results of the test indicated that we cannot reject the null hypothesis, and as such, Physical Activity SDEM Model 3 was selected as the optimal model.

Because interpretation of SDEM models is not dependent on running impact simulations, the model coefficients themselves reveal the impacts of the independent variables on physical activity participation. According to the coefficients, decreases in fulltime employment, the percentage of residents with a college degree or higher, the percentage of residents aged 18-29 and 65 or greater, and the percentage of Asian residents within a census tract are associated with a statistically significant increase in the percentage of BRFSS respondents aged ≥ 18 years who reported no leisure-time physical activity in the preceding two weeks within the same tract. Park access within a census tract is not statistically significant, but the direction of the relationship is such that a decrease in park access within a census tract is expected to be associated with a decrease in physical activity participation within that same tract.

The spatially lagged coefficients reveal two interesting patterns. First, a decrease in the average value of the percentage of Hispanic residents in neighboring census tracts is associated with a statistically significant decrease in physical activity participation in a target census tract. Thus, tracts whose neighbors have a higher proportion of Hispanic residents are more physically active on average than tracts whose neighbors have a lower proportion of Hispanic residents. Additionally, a decrease in the average value of park access in neighboring census tracts is associated with a statistically significant decrease in physical activity participation in a target census tract. Tracts that are surrounded by tracts with good access to parks are more physically active on average than tracts that are surrounded by tracts with inferior park access. Again, while the non-lagged park access variable is not statistically significant, the sign on the coefficient is negative, suggesting that better park access within a tract may be associated with more physical activity within the same tract.

Table 5 - Spatial Models of Physical Activity Participation

	SAR 1	SAR 2	SAR 3	SEM 1	SEM 2	SEM 3	SDM 1	SDM 2	SDM 3	SDEM 1	SDEM 2	SDEM 3
(Intercept)	48.29^{***}	57.24^{***}	61.44^{***}	52.59^{***}	56.76^{***}	61.58^{***}	48.20^{***}	54.63^{***}	54.00^{***}	32.92^{***}	36.34^{***}	41.34^{***}
	(2.36)	(3.33)	(6.29)	(2.30)	(3.36)	(6.42)	(2.42)	(3.28)	(6.86)	(1.94)	(2.61)	(6.69)
log(Pop. Density)	0.75^{***}	0.85^{***}	0.19	0.61^{***}	0.87^{***}	0.27	0.52^{***}	0.88^{***}	0.39	0.07	0.13	-0.23
	(0.07)	(0.11)	(0.27)	(0.07)	(0.12)	(0.27)	(0.08)	(0.13)	(0.27)	(0.06)	(0.10)	(0.26)
Fulltime Employment	-0.15^{***}	-0.16^{***}	-0.12^{***}	-0.15^{***}	-0.16^{***}	-0.12^{***}	-0.14^{***}	-0.16^{***}	-0.13^{***}	-0.06^{***}	-0.06^{***}	-0.06^{**}
	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.02)
College Degree	-0.19^{***}	-0.24^{***}	-0.28^{***}	-0.22^{***}	-0.24^{***}	-0.28^{***}	-0.22^{***}	-0.24^{***}	-0.28^{***}	-0.10^{***}	-0.10^{***}	-0.12^{***}
	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.02)	(0.00)	(0.01)	(0.02)
Pct, Single	0.12^{***}	0.11^{***}	0.06^{*}	0.08^{***}	0.10^{***}	0.06^{*}	0.08^{***}	0.09^{***}	0.06^{*}	0.06^{***}	0.06^{***}	0.03
	(0.01)	(0.01)	(0.03)	(0.01)	(0.01)	(0.03)	(0.01)	(0.01)	(0.03)	(0.01)	(0.01)	(0.03)
Pct. 18 to 29	-0.20^{***}	-0.22^{***}	-0.16^{***}	-0.17^{***}	-0.21^{***}	-0.16^{***}	-0.17^{***}	-0.20^{***}	-0.13^{**}	-0.15^{***}	-0.17^{***}	-0.13^{**}
	(0.01)	(0.02)	(0.04)	(0.01)	(0.02)	(0.04)	(0.01)	(0.02)	(0.04)	(0.01)	(0.02)	(0.04)
Pct. 30 to 64	-0.14^{***}	-0.14^{***}	-0.09^{*}	-0.14^{***}	-0.14^{***}	-0.09^{*}	-0.13^{***}	-0.13^{***}	-0.08^{*}	-0.07^{***}	-0.09^{***}	-0.04
	(0.01)	(0.02)	(0.04)	(0.01)	(0.02)	(0.04)	(0.01)	(0.02)	(0.04)	(0.01)	(0.02)	(0.04)
Pct. 65+	-0.00	-0.01	0.02	0.01	0.01	0.02	0.02	0.02	0.03	-0.09^{***}	-0.10^{***}	-0.15^{***}
	(0.01)	(0.02)	(0.04)	(0.01)	(0.02)	(0.04)	(0.01)	(0.02)	(0.04)	(0.01)	(0.01)	(0.04)
Pct. White	-0.13^{***}	-0.15^{***}	-0.16^{**}	-0.11^{***}	-0.14^{***}	-0.16^{**}	-0.08^{***}	-0.12^{***}	-0.11	0.02	0.00	-0.04
	(0.02)	(0.03)	(0.06)	(0.02)	(0.03)	(0.06)	(0.02)	(0.03)	(0.06)	(0.02)	(0.02)	(0.06)
Pct. Black	-0.12^{***}	-0.14^{***}	-0.16^{**}	-0.08^{***}	-0.12^{***}	-0.16^{**}	-0.05^{*}	-0.11^{**}	-0.09	0.09^{***}	0.07^{**}	0.04
	(0.02)	(0.03)	(0.06)	(0.02)	(0.03)	(0.06)	(0.02)	(0.03)	(0.06)	(0.02)	(0.02)	(0.06)
Pct. Asian	-0.02	-0.04	-0.05	0.01	-0.02	-0.06	0.02	-0.01	-0.02	-0.04^{**}	-0.08^{***}	-0.13^{*}
	(0.02)	(0.03)	(0.06)	(0.02)	(0.03)	(0.06)	(0.02)	(0.03)	(0.06)	(0.02)	(0.03)	(0.06)
Pct. Other	-0.11^{***}	-0.12^{***}	-0.08	-0.06^{**}	-0.10^{**}	-0.08	-0.05^{*}	-0.10^{**}	-0.02	0.04^{*}	0.04	0.08
	(0.02)	(0.03)	(0.07)	(0.02)	(0.03)	(0.07)	(0.02)	(0.03)	(0.07)	(0.02)	(0.03)	(0.07)
Pct. Hispanic	0.04^{***}	0.04^{***}	-0.01	0.06^{***}	0.04^{***}	-0.00	0.07^{***}	0.05^{***}	0.00	0.04^{***}	0.02^{**}	-0.00
	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.03)	(0.01)	(0.01)	(0.02)
Rho	0.18^{***}	0.04^{***}	0.00				0.40^{***}	0.23^{***}	0.02			
	(0.01)	(0.01)	(0.01)				(0.02)	(0.03)	(0.07)			
log(park access w/o Tweets)		-0.08			-0.02			-0.06			-0.09	
		(0.08)			(0.08)			(0.08)			(0.06)	
log(park access w/Tweets)			-0.06			-0.00			0.01			-0.09
			(0.15)			(0.15)			(0.15)			(0.14)
Lambda				0.49^{***}	0.23^{***}	0.13				0.71^{***}	0.49^{***}	0.17^{*}
				(0.03)	(0.04)	(0.08)				(0.02)	(0.03)	(0.08)
lag.log(Pop. Density)							0.02	-0.26	-0.87	-0.60^{***}	-0.61^{***}	0.10
							(0.11)	(0.15)	(0.45)	(0.11)	(0.13)	(0.45)
lag.Fulltime Employment							0.07^{***}	0.04^{**}	-0.00	0.02	0.02	-0.01
							(0.01)	(0.02)	(0.04)	(0.01)	(0.01)	(0.04)

Table 5 continued

	SAR 1	SAR 2	SAR 3	SEM 1	SEM 2	SEM 3	SDM 1	SDM 2	SDM 3	SDEM 1	SDEM 2	SDEM 3
							(0.01)	(0.02)	(0.04)	(0.01)	(0.01)	(0.04)
lag.College Degree							0.08***	0.06***	-0.02	-0.01	0.00	-0.00
							(0.01)	(0.01)	(0.04)	(0.01)	(0.01)	(0.03)
lag.Pct. Single							0.05***	0.03	0.02	0.02	0.03*	0.03
							(0.01)	(0.02)	(0.04)	(0.01)	(0.02)	(0.04)
lag.Pct. 18 to 29							-0.07**	-0.03	0.01	-0.06**	-0.08**	-0.04
							(0.02)	(0.03)	(0.06)	(0.02)	(0.03)	(0.06)
lag.Pct. 30 to 64							-0.04	-0.00	0.03	-0.03	-0.04	0.01
							(0.02)	(0.03)	(0.05)	(0.02)	(0.02)	(0.05)
lag.Pct. 65+							-0.11***	-0.09***	0.09	-0.04	-0.08***	0.02
							(0.02)	(0.02)	(0.06)	(0.02)	(0.02)	(0.06)
lag.Pct. White							-0.13***	-0.08***	0.02	0.03	0.01	-0.01
							(0.02)	(0.02)	(0.05)	(0.02)	(0.02)	(0.04)
lag.Pct. Black							-0.16***	-0.10***	-0.01	0.05**	0.03*	-0.01
							(0.02)	(0.02)	(0.05)	(0.02)	(0.02)	(0.04)
lag.Pct. Asian							-0.13***	-0.09***	0.03	-0.01	0.00	-0.07
							(0.02)	(0.03)	(0.06)	(0.02)	(0.02)	(0.05)
lag.Pct. Other							-0.14***	-0.07*	0.01	0.08***	0.10***	0.09
							(0.02)	(0.03)	(0.07)	(0.02)	(0.02)	(0.06)
lag.Pct. Hispanic							-0.05***	-0.03*	-0.03	-0.00	-0.02	-0.10*
							(0.01)	(0.02)	(0.04)	(0.01)	(0.01)	(0.04)
lag.log(park access w/o Tweets)								-0.28*			-0.31**	
								(0.12)			(0.10)	
lag.log(park access w/Tweets)									-0.61**			-0.46*
									(0.23)			(0.22)
Num. obs.	2115	958	173	2115	958	173	2115	958	173	2115	958	173
Parameters	15	16	16	15	16	16	27	29	29	27	29	29
Log Likelihood	-4864.28	-2235.47	-372.96	-4794.21	-2219.60	-372.12	-4739.83	-2201.07	-359.56	-4105.42	-1971.10	-355.37
AIC (Linear model)	9908.71	4514.80	775.95	9908.71	4514.80	775.95	9779.57	4512.16	775.18	9291.58	4233.38	769.27
AIC (Spatial model)	9758.56	4502.93	777.92	9618.42	4471.20	776.25	9533.66	4460.14	777.13	8264.85	4000.20	768.74
LR test: statistic	152.16	13.87	0.03	292.29	45.61	1.70	247.91	54.03	0.05	1028.73	235.18	2.53
LR test: p-value	0.00	0.00	0.87	0.00	0.00	0.19	0.00	0.00	0.82	0.00	0.00	0.11

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

CHAPTER 5. DISCUSSION

As shown above, park access does not have a statistically significant impact on obesity prevalence at the census tract level, but tracts whose neighbors have good access to parks are more physically active on average than tracts whose neighbors have inferior park access. This may be because physical activity alone is not enough to reduce obesity if minimum physical activity benchmarks are not met and supportive changes in dietary behaviors are not made. Simply put, parks may encourage physical activity and residents of census tracts located in areas with good park access may be more physically active than residents of census tracts with poor access, but this alone is not enough to mitigate obesity.

This finding is corroborated by the maps shown in Figure 3, which depict patterns of spatial dependence and clustering in obesity prevalence and physical activity participation according to the results of the Local Moran's I test.

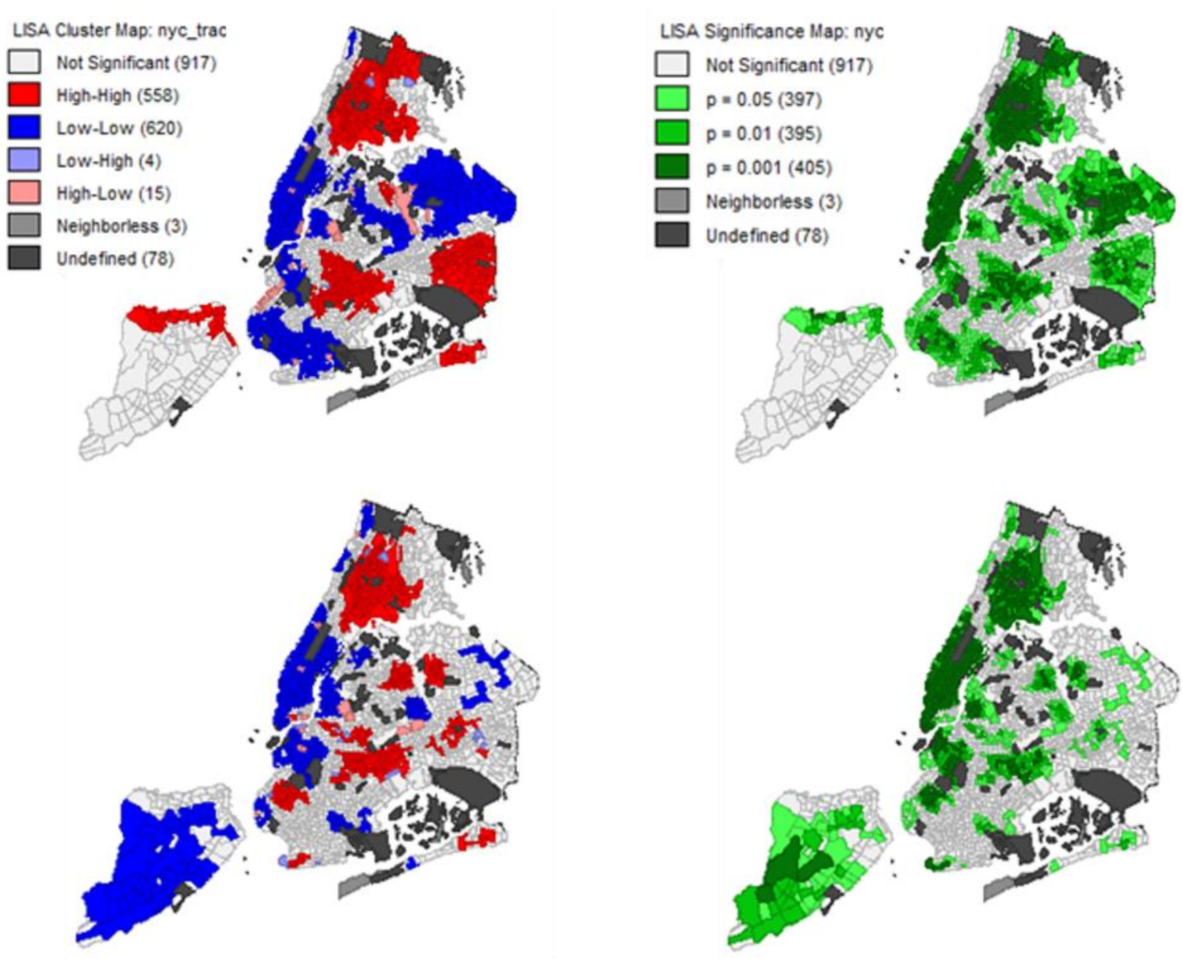

Figure 3 - Local Moran's I Clusters for Obesity Prevalence & Physical Activity Participation (Top: Obesity, Bottom: Physical Activity)

As shown in Figure 3, the places where there is clustering in obesity prevalence are not necessarily the same places where there is clustering in physical activity participation. For example, there are several clusters where physical activity is adequate (Low-Low clusters) but there is no statistically significant clustering in the same area for obesity prevalence. Additionally, there are clusters where obesity prevalence is low (Low-Low clusters) but there is no statistically significant clustering in the same area for physical activity participation. Interestingly, there are also clusters where obesity prevalence is high (High-High clusters) but no clustering exists in physical activity participation. This highlights the fact that there are likely a host of additional factors impacting obesity that physical activity participation alone cannot overcome. Among these factors are the obesity correlates identified in the optimal obesity model that were statistically significant predictors of obesity prevalence, including population density, fulltime employment, educational attainment, marital status, age, and race/ethnicity.

While there is overlap in the statistically significant correlates of obesity prevalence and physical activity participation, obesity prevalence in particular exhibits unique correlations that physical activity participation does not. For example, obesity prevalence is correlated with population density, marital status, the spatial lag of all three age categories, and the spatial lag of all four race categories. Obesity prevalence is also correlated with the non-lagged percentage of black and Hispanic residents. The only correlate that is unique to physical activity participation is the spatial lag of the percentage of Hispanic residents. This suggests that obesity prevalence is impacted by a much greater range of factors than physical activity participation and helps to explain why park access might significantly impact physical activity but not obesity.

Importantly, this study is subject to several limitations which may impact the quality of the analysis and the conclusions we have drawn about the impact of park access on obesity prevalence and physical activity participation. The first pertains to the public health data from which the obesity prevalence and physical activity participation measures were drawn. This data is derived from an annual telephone health surveillance survey and is based on respondent self-reporting, which may introduce nonresponse bias and response bias, respectively (Lim et al. 2013; Rosenman et al. 2011). Nonresponse bias is of particular concern in urban areas, which have seen a sharp decline in telephone survey response rates over time, leading to concerns of under sampling (Lim et al. 2013). Response bias also poses potential problems because respondents are biased toward providing what they view as the 'correct' answer to a question. In health surveillance surveys, this often results in respondents overstating their health and quality of life (e.g., a person may report exercising more frequently than they actually do or may understate their weight) (Rosenman et al. 2011). While these problems could potentially be mitigated by obtaining objectively measured health data, such data were not available to be used for this study.

Another limitation of this study pertains to the incorporation of the number of Tweets recorded within each park in the calculation of the park access metric. Tweets were used as a proxy for activity within parks. Considering this, we assumed that people would be more likely to be physically active in parks with a greater number of Tweets recorded than in parks with less Twitter activity. However, it is possible that the qualities of a park that make it attractive for social media activity make it unattractive for physical activity, in which case the β coefficient for the Tweet parameter would be negative to penalize greater numbers of Tweets. Our assumption about the relationship between social media activity and physical activity within parks could be validated through an objective assessment of physical activity within each park in the system, but this data is not presently available and would be very costly to obtain.

A final limitation pertains to the spatial scale of this analysis. The census tract was used to approximate a neighborhood-level analysis because health data were not available at smaller spatial scales. While evaluating our relationships of interest at the census tract level afforded us the ability to detect nuances that are not visible at larger spatial scales like the city or county levels, it is possible that further disaggregation could reveal additional details that a census tract analysis may obscure. As Logan et al (2017) suggest, disaggregating the data to the parcel or block level could yield even more nuanced results. While this is difficult to do with public health data because of the necessity of protecting individual privacy, obtaining this data at a smaller scale could potentially prove invaluable to this type of analysis. If further disaggregation is not possible, a potential improvement to the methodology employed in this study might come from calculating the distance from tracts to parks using population weighted centroids in order to more accurately capture the distance from where people actually live to the set of parks available to them. Also, using network distance could provide more accurate estimates of the real distance people would have to travel to get to parks in the network. Moreover, calculating the distance to park entrances instead of park centroids would further improve distance estimates.

CHAPTER 6. CONCLUSIONS

City planning and public health professionals are increasingly looking at the improvement of access to public parks and greenspaces as an important pathway toward increasing physical activity participation and decreasing the incidence of obesity and a number of chronic health conditions. Using a new measure of park access called 'Park Choice Accessibility' and several spatial econometric modeling techniques, this study examined the impacts of park access on obesity prevalence and physical activity participation at the census tract level in New York City. The findings suggest that there is no statistically significant relationship between park access and obesity prevalence but having better access to parks is associated with a modest increase in physical activity participation. Importantly, the models presented here indicate that obesity prevalence is impacted by a wider range of factors than physical activity participation, suggesting that increasing physical activity is by itself not enough to mitigate obesity. Future studies would do well to model the relationship between park access and health outcomes at a smaller spatial scale if possible to reveal nuances that may be obscured at the census tract level. Additionally, using park entrances and population weighted centroids or some other measure of residential concentration in the calculation of the distance between parks and residential locations may yield more accurate estimates of park access.

REFERENCES

Akbari, H., M. Pomerantz, and H. Taha. 2001. “Cool Surfaces and Shade Trees to Reduce Energy Use and Improve Air Quality in Urban Areas.” Solar Energy 70 (3): 295–310.

Angrist, J.D., and J.S. Pischke. 2014. *Mastering ’Metrics: The Path from Cause to Effect*. Princeton University Press.

Anderson, L.M., and H.K. Cordell. 1988. “Influence of Trees on Residential Property Values in Athens, Georgia (U.S.A.): A Survey Based on Actual Sales Prices.” Landscape and Urban Planning 15: 153–64.

Aspinall, P.A., C.W. Thompson, S. Alves, T. Sugiyama, R. Brice, and A. Vickers. 2010. “Preference and Relative Importance for Environmental Attributes of Neighbourhood Open Space in Older People.” *Environment and Planning B: Urban Analytics and City Science* 37: 1022–39.

Bancroft, C., S. Joshi, A. Rundle, M. Hutson, C. Chong, C.C. Weiss, J. Genkinger, K. Neckerman, and G. Lovasi. 2015. “Association of Proximity and Density of Parks and Objectively Measured Physical Activity in the United States: A Systematic Review.” *Social Science & Medicine* 138: 22–30.

Beattie, Jeff, Cheryl Kollin, and Gary Moll. 2000. “Trees Tackle Clean Water Regs.” American Forests.

Bedimo-Rung, A.L., A.J. Mowen, and D.A. Cohen. 2005. “The Significance of Parks to Physical Activity and Public Health.” *American Journal of Preventive Medicine* 28 (2S2): 159–68.

Benedict, Mark A., and Edward T. McMahon. 2006. Green Infrastructure: Linking Landscapes and Communities. Washington, D.C.: Island Press.

Bivand, R., Lewin-Koh, N. 2017. Maptools: Tools for Reading and Handling Spatial Objects. R package version 0.8-41. https://CRAN.R-project.org/package=maptools

Centers for Disease Control and Prevention. 2009. “CDC - Healthy Places - Physical Activity.” Healthy Places. http://www.cdc.gov/healthyplaces/healthtopics/physactivity.htm.

Centers for Disease Control and Prevention. 2018. “500 Cities: Local Data for Better Heath.” https://www.cdc.gov/500cities/

Coder, Rim D. 1996. "Identified Benefits of Community Trees and Forests." FOR96-39. Athens, GA: The University of Georgia Cooperative Extension Service Forest Resources Unit.

Cotrone, Vincent. 2015. "The Role of Trees & Forests in Healthy Watersheds: Managing Stormwater, Reducing Flooding, and Improving Water Quality." University Park, PA: Penn State Extension.

Coutts, C. 2008. "Greenway Accessibility and Physical-Activity Behaviour." *Environment and Planning B: Urban Analytics and City Science* 35 (3).

Donovan, Geoffrey H., and David T. Butry. 2010. "Trees in the City: Valuing Street Trees in Portland, Oregon." Landscape and Urban Planning 94: 77–83.

Donovan, Geoffrey H., David T. Butry, Yvonne L. Michael, Jeffrey P. Prestemon, Andrew M. Liebhold, Demetrios Gatziolis, and Megan Y. Mao. 2013. "The Relationship Between Trees and Human Health: Evidence from the Spread of the Emerald Ash Borer." American Journal of Preventive Medicine 44 (2): 139–45.

Durstine, J. Larry, Benjamin Gordon, Zhengzhen Wang, and Xijuan Luo. 2013. "Chronic Disease and the Link to Physical Activity." Journal of Sport and Health Science 2 (1): 3–11.

Ernston, Henrik. 2013. "The Social Production of Ecosystem Services: A Framework for Studying Environmental Justice and Ecological Complexity in Urbanized Landscapes." Landscape and Urban Planning 109: 7–17.

Giles-Corti, B., M.H. Broomhall, M. Knuiman, C. Collins, K. Douglas, K. Ng, A. Lange, and R.J. Donovan. 2005. "Increasing Walking: How Important Is Distance To, Attractiveness, and Size of Public Open Space?" American Journal of Preventive Medicine 28: 169–76.

Golgher, A.B., and P.R. Voss. 2016. "How to Interpret the Coefficients of Spatial Models: Spillovers, Direct and Indirect Effects." *Spatial Demography* 4: 175–205.

Harnik, P., and B. Welle. 2011. "From Fitness Zones to the Medical Mile: How Urban Park Systems Can Best Promote Health and Wellness." Trust for Public Land. https://www.tpl.org/sites/default/files/cloud.tpl.org/pubs/ccpe-health-promoting-parks-rpt.pdf.

Kaczynski, A.T., G.M. Besenyi, S.A. Wilhelm-Stanis, M.J. Koohsari, K.B. Oestman, R. Bergstrom, L.R. Potwarka, and R.S. Reis. 2014. "Are Park Proximity and Park Features Related to Park Use and Park-Based Physical Activity among Adults? Variations by Multiple Socio-Demographic Characteristics." *International Journal of Behavioral Nutrition and Physical Activity* 11 (146).

Kuo, Frances E. 2003. “Social Aspects of Urban Forestry: The Role of Arboriculture in a Healthy Social Ecology.” Journal of Arboriculture 29 (3): 148–55.

Kuo, Frances E., and William C. Sullivan. 2001a. “Aggression and Violence in the Inner City: Effects of Environment via Mental Fatigue.” Environment and Behavior 33 (4): 543–71.

———. 2001b. “Environment and Crime in the Inner City: Does Vegetation Reduce Crime?” Environment and Behavior 33 (3): 343–67.

Larson, L.R., V. Jennings, and S.A. Cloutier. 2016. “Public Parks and Wellbeing in Urban Areas of the United States.” *PLoS ONE* 11 (4).

Lee, A.C.K., and R. Maheswaran. 2010. “The Health Benefits of Urban Green Spaces: A Review of the Evidence.” Journal of Public Health 33 (2): 212–22.

Lerner, and Poole. 1999. “The Economic Benefits of Parks and Open Space.” The Trust for Public Land.

LeSage, J., and R.K. Pace. 2009. *Introduction to Spatial Econometrics*. CRC Press.

LeSage, J.P. 2014. “What Regional Scientists Need to Know about Spatial Econometrics.” *The Review of Regional Studies* 44 (1).

Lim, S., S. Immerwahr, S. Lee, and T.G. Harris. 2013. “Estimating Nonresponse Bias in a Telephone-Based Health Surveillance Survey in New York City.” *American Journal of Epidemiology* 178 (8).

Logan, T.M., T.G. Williams, A.J. Nisbet, K.D. Liberman, C.T. Zuo, and S.D. Guikema. 2017. “Evaluating Urban Accessibility: Leveraging Open-Source Data and Analytics to Overcome Existing Limitations.” *Environment and Planning B: Urban Analytics and City Science* 0 (0): 1–17.

Lovasi, G.S., J.W. Quinn, K.M. Neckerman, M.S. Perzanowski, and A. Rundle. 2008. “Short Report: Children Living in Areas with More Street Trees Have Lower Prevalence of Asthma.” Journal of Epidemiology and Community Health 62: 647–49.

Manski, C.F. 1993. “Identification of Endogenous Social Effects: The Reflection Problem.” *Review of Economic Studies* 60: 531–42.

McDonald, Robert I. 2015. Conservation for Cities: How to Plan & Build Natural Infrastructure. Washington, D.C.: Island Press.

Mitchell, Richard, and Frank Popham. 2008. “Effect of Exposure to Natural Environment on Health Inequalities: An Observational Population Study.” The Lancet 372: 1655–60.

National Urban and Community Forestry Advisory Council. 2015. “Ten-Year Urban Forestry Action Plan: 2016-2026.” University of Virginia Institute for Environmental Negotiation. http://www.urbanforestplan.org/wp-content/uploads/2015/11/FinalActionPlan_Complete_11_17_15.pdf.

Nowak, David J. 2002. “The Effects of Urban Trees on Air Quality.” General Technical Report. Syracuse, NY: USDA Forest Service, Northern Research Station.

Nowak, David J., and Jeffrey T. Walton. 2005. “Projected Urban Growth (2000-2050) and Its Estimated Impact on the US Forest Resource.” Journal of Forestry 103 (8): 383–89.

Office of Disease Prevention and Health Promotion. 2008. “Physical Activity Guidelines for Americans.” U.S. Department of Health and Human Services.

Pretty, J., J. Peacock, M. Sellens, and M. Griffin. 2005. “The Mental and Physical Health Outcomes of Green Exercise.” International Journal of Environmental Health Research 15 (5): 319–37.

Ratcliffe, Michael. 2012. “How Do We Measure Urban Areas?” United States Census Bureau. 2012. https://www.census.gov/newsroom/blogs/random-samplings/2012/04/how-do-we-measure-urban-areas.html.

Richardson, E.A., R. Mitchell, T. Hartig, S.D. Vries, T. Astell-Burt, and H. Frumkin. 2011. “Green Cities and Health: A Question of Scale?” *Journal of Epidemiology and Community Health* 66 (2).

Rigolon, A., and J. Nemeth. 2016. “A QUality INdex of Parks for Youth (QUINPY): Evaluating Urban Parks through Geographic Information Systems.” *Environment and Planning B: Urban Analytics and City Science* 0 (0): 1–20.

Rosenman, R., V. Tennekoon, and L.G. Hill. 2011. “Measuring Bias in Self-Reported Data.” *International Journal of Behavioral and Healthcare Research* 2 (4): 320–32.

Sallis, J.F., E. Cerin, T.L. Conway, M.A. Adams, L.D. Frank, M. Pratt, D. Salvo, et al. 2016. “Physical Activity in Relation to Urban Environments in 14 Cities Worldwide: A Cross-Sectional Study.” *Lancet* 387: 2207–17.

Schwab, James C. 2009. “Planning the Urban Forest: Ecology, Economy, and Community Development.” General Technical Report 555. Planning Advisory Service. Chicago, IL: American Planning Association.

Seila, A.F., and L.M. Anderson. 1982. “Estimating Costs of Tree Preservation on Residential Lots.” Journal of Arboriculture 8: 182–85.

———. 1984. “Estimating Tree Preservation Costs on Urban Residential Lots in Metropolitan Atlanta.” Research Paper 48. Macon, GA: Georgia Forestry Commission.

Stark, J.H., K. Neckerman, G.S. Lovasi, J. Quinn, C.C. Weiss, M.D. Bader, K. Konty, T.G. Harris, and A. Rundle. 2014. “The Impact of Neighborhood Park Access and Quality on Body Mass Index among Adults in New York City.” *Journal of Preventive Medicine* 64: 63–68.

Stone, Brian, and John M. Norman. 2006. “Land Use Planning and Surface Heat Island Formation: A Parcel-Based Radiation Flux Approach.” Atmospheric Environment 40: 3561–73. doi:10.1016/j.atmosenv.2006.01.015.

Sullivan, William C., Frances E. Kuo, and Stephen F. DePooter. 2004. “The Fruit of Urban Nature: Vital Neighborhood Spaces.” Environment and Behavior 36 (5): 678–700.

Suminski, R.R., W.S. Carlos Poston, R.L. Petosa, E. Stevens, and L.M. Katzenmoyer. 2005. “Features of the Neighborhood Environment and Walking by U.S. Adults.” American Journal of Preventive Medicine 28 (2): 149–55.

Tobler, W.R. 1970. “A Computer Movie Simulating Urban Growth in the Detroit Region.” *Economic Geography* 46: 234–40.

Train, K. 2003. *Discrete Choice Methods with Simulation*. Cambridge, UK: Cambridge University Press.

Ulrich, Roger S. 1981. “Natural Versus Urban Scenes: Some Psychophysiological Effects.” Environment and Behavior 13 (5): 523–56.

Ulrich, Roger S., Robert F. Simons, Barbara D. Losito, Evelyn Fiorito, Mark A. Miles, and Michael Zelson. 1991. “Stress Recovery During Exposure to Natural and Urban Environments.” Journal of Environmental Psychology 11: 201–30.

United States Census Bureau. 2007. “Census Atlas of the United States - Chapter 2: Population Distribution.” United States Census Bureau. https://www.census.gov/population/www/cen2000/censusatlas/pdf/2_Population-Distribution.pdf.

United Nations Population Fund. 2018. “Urbanization.” UNFPA. 2018. https://www.unfpa.org/urbanization.

U.S. Department of Health and Human Services. 2010. “The Surgeon General’s Vision for a Healthy and Fit Nation.” U.S. Department of Health and Human Services, Office of the Surgeon General.

Vale, D.S., Saraiva, M. 2016. Active accessibility: A review of operational measures of walking and cycling accessibility. The Journal of Transport and Land Use 9(1): 209-235.

Wells, N.M., S.P. Ashdown, E.H.S. Davies, F.D. Cowett, and Y. Yang. 2007. “Environment, Design, and Obesity: Opportunities for Interdisciplinary Collaborative Research.” Environment and Behavior 39 (1): 6–33.

West, S.T., K.A. Shores, and L.M. Mudd. 2012. “Association of Available Parkland, Physical Activity, and Overweight in America’s Largest Cities.” *Journal of Public Health Management Practice* 18 (5): 423–30.

Wolch, Jennifer R., Jason Byrne, and Joshua P. Newell. 2014. “Urban Green Space, Public Health, and Environmental Justice: The Challenge of Making Cities ‘Just Green Enough.’” Landscape and Urban Planning 125: 234–44.

Wolf, K.L. 1999. “Nature and Commerce: Human Ecology in Business Districts.” In Building Cities of Green: Proceedings of the 9th National Urban Forest Conference, 4. Washington, D.C.

Wolf, Kathleen L. 2008. “City Trees, Nature and Physical Activity: A Research Review.” Arborist News.

Wolfe, Mary K., and Jeremy Mennis. 2012. “Does Vegetation Encourage or Suppress Urban Crime? Evidence from Philadelphia, PA.” Landscape and Urban Planning 108: 112–22.

World Health Organization. 2014. “WHO | Noncommunicable Diseases.” WHO. http://www.who.int/topics/noncommunicable_diseases/