Georgia State University
Digital Archive @ GSU

Sociology Dissertations | Department of Sociology

11-17-2011

HOV to the MD? A Multilevel Analysis of Urban Sprawl and the Risk for Negative Health Outcomes

William Mark Sweatman Ph.D.
Georgia State University, wmsweatman@yahoo.com

Follow this and additional works at: http://digitalarchive.gsu.edu/sociology_diss

Recommended Citation
Sweatman, William Mark Ph.D., "HOV to the MD? A Multilevel Analysis of Urban Sprawl and the Risk for Negative Health Outcomes" (2011). *Sociology Dissertations.* Paper 61.

This Dissertation is brought to you for free and open access by the Department of Sociology at Digital Archive @ GSU. It has been accepted for inclusion in Sociology Dissertations by an authorized administrator of Digital Archive @ GSU. For more information, please contact digitalarchive@gsu.edu.

HOV TO THE MD?

A MULTILEVEL ANALYSIS OF URBAN SPRAWL AND THE RISK FOR NEGATIVE HEALTH OUTCOMES

by

WILLIAM MARK SWEATMAN

Under the direction of Dawn Baunach, Ph.D.

ABSTRACT

Urban sprawl often has a negative connotation, used as a derogatory label for certain forms and consequences of land development that are seen as environmentally and socially unpleasant. Although sprawl may be seen as offensive, there may be other, far greater and more harmful consequences of sprawl. The literature indicates that rates of negative health outcomes, such as obesity, tend to be higher in more developed areas. However, aside from a few studies, little empirical research looks specifically at the influence of sprawl when it comes to individual health. This research project focuses on sprawl and examines the relationships it has with health behaviors and health outcomes. By analyzing data from the CDC's 2003 Behavioral Risk Factor Surveillance System (BRFSS), an annual telephone survey of adults that include more than two-hundred self-

reported and calculated variables, I investigate the associations between sprawl, physical activity, body weight, and health outcomes using Structural Equation Modeling (SEM). By employing SEM, my research differs from previous research in this field by adding not only additional layers to the evaluation of sprawl and health outcomes, but also allows for the evaluation of associations through various "paths" instead of looking at variables within simpler hierarchical regression models. In addition to direct effects, it also allows for the determination of indirect, or mediated, effects between variables within a path model. Even though no direct relationship between sprawl and health outcomes was revealed, sprawl did show to have a statistically significant indirect effect on health outcomes mediated by physical activity and body weight. Physical activity is also shown to mediate the relationship between sprawl and body weight. Additionally, physical activity reveals both a direct and indirect effect on health outcomes, with its indirect effect being mediated by body weight. Finally, physical activity and body weight are both shown to have statistically significant direct effects on health outcomes. In the concluding chapter I propose a new path model in light of the results of the analyses of data in order to represent the associations between sprawl, physical activity, body weight, and health outcomes more accurately.

INDEX WORDS: Urban Sprawl, Health, Health behaviors, Health outcomes, Diabetes, Heart attack, Heart disease, High cholesterol, Hypertension, Stroke, Multilevel analysis

HOV TO THE MD?

A MULTILEVEL ANALYSIS OF URBAN SPRAWL

AND THE RISK FOR NEGATIVE HEALTH OUTCOMES

by

WILLIAM MARK SWEATMAN

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

in the College of Arts and Sciences

Georgia State University

2011

Copyright by
William Mark Sweatman
2011

HOV TO THE MD?

A MULTILEVEL ANALYSIS OF URBAN SPRAWL

AND THE RISK FOR NEGATIVE HEALTH OUTCOMES

by

WILLIAM MARK SWEATMAN

Committee Chair: Dawn Baunach, Ph.D.

Committee: Charlie Jaret, Ph.D.

Mary Ball, Ph.D.

Electronic Version Approved:

Office of Graduate Studies

College of Arts and Sciences

Georgia State University

December 2011

DEDICATION

To my parents, thank you so much for the social and financial capital you have given me. Without the resources you blessed me with I would never have been able to achieve the dream of obtaining my doctorate. And even more so, thank you so much for the love you have given me throughout the years. I love you both so very much. Thank you!

ACKNOWLEDGEMENTS

I would like to thank the chair of my committee, Dr. Dawn Baunach, for all of her advice and support. Her knowledge of quantitative sociology provided me with great guidance during my dissertation journey. I also would like to thank my committee members, Dr. Charlie Jaret and Dr. Mary Ball, for their constructive criticism and knowledge of sociological matters. My dissertation is well-rounded and impressive due to the guidance and advice from my entire dissertation committee.

I would like to thank Dr. Molly Perkins and Dr. Mignon Montpetit for their expert knowledge of multilevel statistical analysis and their unwavering support and encouragement.

I would also like to thank my colleagues for their encouragement and support during graduate school, especially Pam Regus, Josie Parker, Evelina Sterling, Linda Danavall, and Emmie Cochran-Jackson. Each of you has made the journey through the doctorate program not only endurable, but fun!

TABLE OF CONTENTS

ACKNOWLEDGEMENTS ..... **V**

LIST OF TABLES ..... **IX**

LIST OF FIGURES ..... **X**

CHAPTER 1

INTRODUCTION **1**

CHAPTER 2

REVIEW OF THE LITERATURE **6**

PHYSICAL ACTIVITY AND HOW IT RELATES TO WEIGHT AND HEALTH 6

SUBURBANIZATION AND HOW IT RELATES TO PHYSICAL ACTIVITY 8

STUDIES LOOKING AT SPRAWL AND HEALTH 12

CHAPTER 3

CONCEPTUAL FRAMEWORK **15**

HYPOTHESES 18

ADDITIONAL CONSIDERATIONS 20

CHAPTER 4

MEASURES OF URBAN SPRAWL **22**

STUDIES FOCUSING ON SPRAWL AND HEALTH OUTCOMES 23

CONCLUSION 27

CHAPTER 5

METHODOLOGY **29**

DEPENDENT VARIABLES 32

INDEPENDENT VARIABLES 35

Control Variables 37

Sociodemographic Control Variables 37

Personal Health Control Variable 41

Regional Control Variable 43

Analytic Technique 43

Model Analyses 44

Similarities and Difference between this Study and Ewing et al. Study 49

Limitations 50

Findings 52

CHAPTER 6

SPRAWL AND NEGATIVE HEALTH OUTCOMES 53

Prediction of Negative Health Outcomes with Sprawl 54

Age, Education, Income, and Gender as Predictors of Negative Health Outcomes 59

Race and Ethnicity as Predictors of Negative Health Outcomes 62

Marital Status as Predictors of Negative Health Outcomes 64

Personal Health as Predictors of Negative Health Outcomes 65

Region as a Predictor of Negative Health Outcomes 66

Fit Statistics 67

Summary 69

CHAPTER 7

SPRAWL AND NEGATIVE HEALTH OUTCOMES WITH PHYSICAL ACTIVITY AND BMI AS MEDIATING VARIABLES INCLUDING ALL HYPOTHESIZED PATHS 71

Prediction of Physical Activity 72

Prediction of Body Mass Index (BMI) 75

Physical Activity and BMI as Independent Variables 77

Strongest Variables in the Prediction of Negative Health Outcomes 78

Changes in Prediction of Negative Health Outcomes from Previous Models 81

Indirect Effect of Sprawl through Physical Activity on Negative Health Outcomes, Paths H1 and H6 ... 82

Indirect Effect of Sprawl through BMI on Negative Health Outcomes, Paths H4 and H3 ... 83

Indirect Effect of Sprawl through Physical Activity and BMI on Negative Health Outcomes, Paths H1, H2, and H3 ... 84

Indirect Effect of Physical Activity through BMI on Negative Health Outcomes, Paths H2 and H3 ... 85

Indirect Effect of Sprawl through Physical Activity on BMI, Paths H1 and H2 ... 85

CHAPTER 8

DISCUSSION AND CONCLUSION ... 86

Validity of Hypotheses ... 86

Associations between Variables in the Conceptual Model ... 92

Best Fitting Analytical Model ... 94

Proposal of a New Conceptual Model ... 95

Contributions to the Literature ... 97

Implications at an Applied Level ... 98

Possibilities for Future Research ... 99

Conclusion ... 101

REFERENCES ... 105

APPENDIX A ... 112

LIST OF COUNTY-LEVEL SPRAWL INDEX, FIPS, AND REGION CLASSIFICATION ... 112

APPENDIX B ... 125

COUNTY SPRAWL INDEX VARIABLES ... 125

APPENDIX C ... 126

STANDARDIZED REGRESSION COEFFICIENTS ... 126

LIST OF TABLES

Table 2.1 Central City Population and Job Loss 9

Table 5.1 Difference in Means Significance Testing 31

Table 5.2 Descriptive Statistics 32

Table 6.1 Zero-order Relationship between Sprawl and Negative Health Outcomes 56

Table 6.2 Baseline Model SEM Regression Results for Negative Health Outcomes. 57

Table 7.1 Model 2 SEM Regression Results for Negative Health Outcomes 73

LIST OF FIGURES

Figure 3.1 Established (Solid) and Speculative (Dashed) Relationships 17

Figure 3.2 Conceptual Model Used in this Dissertation 18

Figure 5.1 Analytical Model 46

Figure 5.2 Independent, Dependent, and Mediating Variables 47

Figure 6.1 Baseline Analytical Path Model 54

Figure 7.1 Analytical Path Model Two 71

Figure 8.1 New Conceptual Model 96

CHAPTER 1

INTRODUCTION

It is not uncommon for many people to associate urban sprawl with a negative connotation, using the term as a derogatory label for certain forms and consequences of land development that are seen as environmentally and socially unpleasant (Freeman 2001; Razin 1998; Wassmer 2002). Although many people may view sprawl as offensive, there may be other, far greater and more harmful consequences of sprawl. The literature indicates that rates of negative health outcomes, such as obesity, tend to be higher in areas that are more sprawling (McCann and Ewing 2003; Kelly-Schwartz et al. 2004; Ewing et al. 2003). A recent study supports this association indicating a relationship between urban sprawl and the risk for being overweight or obese (Lopez 2004). Aside from a few studies, little empirical research looks specifically at the influence of sprawl when it comes to the health of individuals. However, urban sprawl is receiving growing attention, and the empirical evidence that does exist generally supports links between built environmental conditions and health outcomes (Giles-Corti and Donovan 2002).

Associations between the built environment and health are not new. In the late 19th century, public health practitioners realized the effects of the built environment on the public; how the very place where people lived and worked affected their health. Unsanitary sewage and water conditions, dark airless tenement housing, and toxic industrial wastes all contributed to the spread of disease. In response to such conditions, planners advocated public infrastructure, such as water and sewer lines, building codes, and zoning plans to separate people from toxins and reduce population concentrations. Today, urban areas are much more sanitary places, where citizens suffer less from infectious diseases. However, the built environment continues to

influence public health. People who reside in sprawling suburban areas are now suffering from chronic health conditions that earlier generations did not. The lack of physical activity, poor diets, air pollutants, and environmental toxins may cause many people to suffer from chronic health problems such as heart disease, asthma, and diabetes at rates previous generations did not (Perdue 2004). It is just as important today, as it was over one hundred years ago, to understand associations between the built environment and public health.

It has long been documented that body weight is associated with negative health outcomes, such as diabetes, cardiovascular disease, and cancer, and that the risk of mortality increases with the severity of obesity (Calle et al. 1999). Unfortunately, rates of obesity are increasing rapidly in the United States (Flegal et al. 1998; Kuczmarski et al.1994; Bianchini, Kaaks, and Vainio 2002). Obesity, and the plethora of diseases that come along with it, used to be blamed on the fact that Americans eat too much fattening food. Although this may be true, researchers now are focusing on another component of the crisis, that of low levels of physical activity. It is well agreed upon that Americans are too sedentary and weigh too much (McCann and Ewing 2003).

Decreasing body mass and maintaining a proper weight is a means to reduce the risks for deadly diseases such as diabetes and heart disease, and one very important way to maintain a healthy weight is to be physically active (Flegal, Carroll, Kuczmarski, and Johnson 1994). Medical research has shown that walking and similar forms of moderate physical activity help maintain a healthy weight. However, the lack of physical activity, as well as being overweight, factor into more than 200,000 premature deaths annually. Therefore, it is very important for individuals to be physically active in order to reduce their risk for obesity and its associated health conditions (Mokdad, Bowman, Ford, Vinicor, Marks, and Koplan 2001). However, being

physically active on a regular basis may pose a significant problem for those who reside in sprawling areas.

Community form has a strong relationship to one's health. The design of communities influences health by encouraging or discouraging routine physical activity involved in daily life. People living in sprawling areas, for example, may miss out on significant health benefits available, such as walking to the store, to work, or other places as part of a daily routine (McCann and Ewing 2003). Patterns of streets within neighborhoods, such as those found in many suburban subdivisions, affect how people use their cars and their propensity to walk. Metropolitan areas with high levels of urban sprawl tend to have higher per capita vehicle miles traveled daily, even after controlling for factors such as income, size of metropolitan area, and location within the nation. This suggests that people in high-sprawl areas drive more, quite possibly at the expense of daily physical activity (Lopez 2004). This association between development patterns and health outcomes can be seen as an indirect effect mediated by physical activity and body weight.

Even though some research linking sprawl with health status exists, there is the need for more empirical research on this relationship. This research project helps fill the need for additional empirical measures by looking for associations between sprawl, physical activity, weight, and disease. It specifically follows up on and refines the approach and results of Ewing et al. (2003). It tests the relationship between all of these variables by focusing on a conceptual model linking each variable via different hypotheses. It not only addresses the issue of how sprawl relates to health outcomes, but also addresses the issue of how sprawl relates to physical activity and weight. Various "paths" through the conceptual model allow me to determine which variables have the strongest association on the outcome variables – diabetes, heart attack, heart

disease, high cholesterol, hypertension, and stroke. These variables are outcomes in my conceptual model because previous research shows they have a high association with a lack of regular physical activity and body weight. Therefore, based on previous studies it is reasonable to hypothesize that individuals who reside in more sprawling areas are more prone to have a higher risk for such conditions because they are less likely to obtain the beneficial aspects of regular daily physical activity due to the design of their built environment.

There are many questions that arise when looking for associations between sprawl and health outcomes, such as: How does sprawl specifically affect health outcomes? Does county-level sprawl have a direct association with individual-level health outcomes? Most likely not, but it may have a direct effect on levels of individual physical activity. It is apparent from many studies that weight gain does have a causal effect on health outcomes and that less physical activity does affect weight (Calle et al. 1999; Flegal et al. 1998; Kuczmarski et al. 1994; Bianchini et al. 2002; McCann and Ewing 2003; Mokdad et al. 2001), but what association, if any, does a higher level variable like sprawl have on individual-level health?

This research project explores the questions listed above by focusing on sprawl and determining relationships it has with physical activity, body weight, and specific health outcomes. By analyzing data from the CDC's Behavioral Risk Factor Surveillance System (BRFSS) (Centers for Disease Control and Prevention 2003), an annual telephone survey of adults that includes more than two hundred self-reported and calculated variables, I test for any significant associations between sprawl, physical activity, body weight, and health outcomes that exist by using a statistical technique called Structural Equation Modeling (SEM). By employing SEM, my research differs from previous research in this field by adding not only additional layers to the evaluation of sprawl and health outcomes, but also allows for the evaluation of

associations through various “paths” instead of looking at variables within simpler hierarchical regression models.

CHAPTER 2

REVIEW OF THE LITERATURE

This section explores the theoretical connections between the various parts of the conceptual model briefly described above. It first discusses the variables most proximal to health outcomes, that of physical activity and how it relates to body weight, and then proceeds through the model to the most distal variable, that of sprawl. Finally, this section concludes with previous studies that focus on sprawl and health outcomes. By understanding the connections between each part of the model, as well as the strengths and weaknesses of previous studies involving sprawl and health, I am able to propose statistical models to best test possible associations between sprawl, physical activity, body weight, and health outcomes.

Physical Activity and How it Relates to Weight and Health

Conditions such as obesity, diabetes, and hypertension have reached epidemic levels in the United States. Previously, these conditions were blamed on Americans' overconsumption of fattening foods. Although this may be true, there is another component to the crisis – physical inactivity. An increasing body of evidence suggests that moderate forms of regular physical activity, such as walking, can have beneficial effects on public health. Regular exercise allows individuals to maintain a healthy weight, as well as bestow other health benefits (McCann and Ewing 2003; Ewing et al.2003).

Unfortunately, physical inactivity is now a major health problem in the United States. Compelling evidence suggests physical inactivity is a significant contributing factor in several chronic diseases and conditions, and the hazards of a sedentary lifestyle have led numerous

activist groups to promote public health recommendations in support of physical activity (Blair, LaMonte, and Nichaman 2004). One suggestion to achieve greater physical activity is to reduce sedentary behavior by incorporating more incidental activity into daily routines. Many experts agree that the only way to maintain a healthy weight and lifestyle is to be physically active (Saris et al. 2003; McCann and Ewing 2003; MacLennan 2004).

Even though research demonstrates the benefits of moderate physical activity in maintaining a healthy weight, physical activity remains low in the United States and such inactivity is blamed for more than 200,000 premature deaths each year. Physical inactivity may soon overtake tobacco as the nation's predominant health risk (McCann and Ewing 2003; MacLennan 2004). Another staggering statistic is the fact that the majority of Americans report not obtaining enough exercise to meet the recommended minimum of twenty minutes of strenuous activity three days per week or thirty minutes of moderate activity five days per week. Even more shocking is the fact that one in four Americans remain completely inactive during their leisure time (McCann and Ewing 2003).

For those who are overweight, achieving and maintaining a healthy weight requires more than the amount of daily physical activity recommended for the average person. Compelling evidence suggests that obese and formerly obese individuals require sixty to ninety minutes (as opposed to thirty minutes for the average individual) of moderate physical activity five days per week in order to maintain a healthy weight (Saris et al. 2003). The prospect for many Americans achieving a healthy weight seems bleak, as the majority of Americans need to increase their physical activity and those who are already overweight have a much harder battle to fight in order to achieve their ideal weight and health.

One thing is certain, health experts believe most Americans are too sedentary and weigh too much. As a consequence, conditions associated with inactivity, such as obesity, diabetes, and heart disease, have reached epidemic levels. A major question guiding this debate is whether the design of communities makes it difficult for people to obtain physical activity in order to maintain a healthy weight (McCann and Ewing 2003; Ewing et al. 2003), again an association suggesting an indirect effect between sprawl and health status.

Independent of how much people walk in their leisure time, body mass index and obesity levels are found to be higher for individuals who reside in more sprawling areas. Urban form may have a stronger relationship to one's body weight than does walking for leisure, suggesting that people living in sprawling areas miss out on significant health benefits available by walking, biking, climbing stairs, and other types of physical activity as part of their daily lives. Communities designed for walking, such as those found in more dense urban areas, seem to encourage an extra 15 to 30 minutes of walking per week. This extra 15 to 30 minutes of walking per week could translate into a 150-pould person losing and/or keeping off one to two pounds each year. This estimated extra walking and reduction in weight is in addition to the recommended minimum weekly amounts of physical activity discussed earlier (McCann and Ewing 2003).

Suburbanization and how it Relates to Physical Activity

In the 1970s, 69 percent of the American population lived in what is classified as metropolitan statistical areas (MSAs). In the 1980s that figure rose to 75 percent. By the 1990s, 77 percent of the nation's population lived in a metropolitan area. Although a greater proportion of the population now resides in metropolitan areas, only a small percentage of such population

Table 2.1 Central City Population and Job Loss

Decade	Metropolitan Area Jobs Located in the Central City	Metropolitan Area Population Residing in Central City
1950s*	70%	57%
1960s*	63%	49%
1970s*	55%	43%
1980s*	50%	40%
1990s*	45%	37%

Sources: Mieszkowski and Mills (1993)

lives and works in the central city (Mieszkowski and Mills 1993). Table 2.1 demonstrates how central cities in metropolitan areas lost percentages of both jobs and population over the decades.

It is clear from the figures in Table 2.1 that suburbanization has changed the American landscape, bringing residents out of the center of metropolitan areas. This change in place of residence and lifestyle in modern American society has led to a reliance on automobiles to access jobs that are now located mostly in suburban areas throughout metropolitan regions. Even though numerous jobs are now located in the suburbs where many people live and work, automobiles are still necessary to access most employment locations due to the lack of adequate public transportation in the suburbs. Therefore, today the automobile is not only a staple in the lives of Americans, but a necessity, due to low-density development (Mieszkowski and Mills 1993). In fact, residential density is significantly related to the degree in which residents rely on the automobile (Freeman 2001). Gone are the days of walking to and from places of daily activities for many individuals.

The essential nature of this form of mobility allows jobs and homes to be miles apart. As development continues outward from the central cities, housing and services grow farther apart. This development pattern makes it just about impossible for pedestrian mobility. Therefore, Americans are driving more and more each year due to the increasingly spread-out nature of

metropolitan areas. Additionally, the average time per trip is increasing, including the commute to work, which society now accepts as commonplace (U.S. Department of Transportation 1999; 2004). It can be argued that sprawl is a direct result of a society centered on the automobile (Glaeser and Kahn 2003).

It can be seen that sprawl has influenced community design, which in turn has created environments that may significantly affect regular physical activity. In fact, some argue that today's built environment designs regular physical activity out of everyday life. Many times, the quickest, if not only, way to get to places for normal daily activities, such as work, school, and stores, is to drive. This dependence on automobiles has created communities where behaviors beneficial to health, such as walking, are basically non-existent. Even more pervasive are drive-through services, which allow individuals to bank, pick up dry cleaning, order food, and ironically, get their medication, conveniently without ever taking a step. Even sidewalks are designed "out" of many suburban areas making it impossible for individuals to safely walk from place to place (MacLennan 2004).

Ewing, Pendall, and Chen (2002a) found that people living in sprawling areas tend to drive more, own more cars, breathe more polluted air, face a greater risk of traffic fatalities, walk less, and use public transit less than those who reside in areas that are not as sprawling. Such factors lead people to weigh more and suffer from hypertension and other negative health conditions. In contrast, those who reside in compact areas, such as New York City, tend to drive less and walk more. These findings hold true even after controlling for sociodemographic factors such as age, education, gender, and race. In fact, sprawl and its component factors (dependence on automobiles, less physical activity, etc.) are found to be greater predictors of health than demographic control variables (Ewing et al. 2002a).

The most likely way community design influences weight and health is by either encouraging or discouraging routine physical activity in daily life. For most people, this means walking to the store, to work, or other such places as a part of their daily routine. When it comes to whether or not people get regular exercise in their leisure time, such as running, working out, gardening, etc., the degree of sprawl seems to have very little influence, as people in both sprawling and compact areas are equally likely to report they exercise in some fashion (McCann and Ewing 2003). However, the degree of sprawl does make a difference in how people engage in the most common, accessible, and free form of exercise – that of walking. Individuals in more sprawling areas report less time walking in their leisure time than those who reside in compact locales. For every 50-point increase in their sprawl scale, which ranges from 63.12 to 352.07, McCann and Ewing (2003) find that people are likely to walk fourteen minutes less per month for exercise, even when controlling for gender, age, education, ethnicity, and other factors. Their study also shows that routine physical activity is a significant factor in lower BMI (Body Mass Index) of individuals residing in more compact communities.

Do McCann and Ewing's (2003) associations between urban form and physical activity always hold true? What about a different side of this argument? Is it possible that those residing in compact urban areas could walk less because they do not have as far to walk? Maybe they take elevators more than those in sprawling areas because the buildings are tall, and taking twenty or thirty flights of stairs is just not practical. On the other hand, those who reside in sprawling areas may get more daily walking than those in dense areas because they have large parking lots to traverse and sprawling buildings to walk around in instead of more compact, taller buildings.

Studies Looking at Sprawl and Health

There are a few studies that focus on measuring the health effects of sprawl, including McCann and Ewing (2003), which is a follow-up study of research conducted by Ewing et al. (2003). Both of these studies look at how sprawl affects physical activity, obesity, and chronic disease, specifically diabetes, coronary heart disease, and hypertension. Their results show that individuals in sprawling areas are more likely to have a higher BMI. McCann and Ewing found that a 50-point increase in the degree of sprawl (a scale developed by Ewing et al. and measured at the county level ranging from 63.12, indicative of more sprawl, to 352.07, indicative of a compact area, with an average score of 100) relates to an increase in BMI by 0.17 points, which translates to just over one pound for the average person. They also found that sprawl has a small but significant effect on minutes walked per month and individuals in sprawling areas are more likely to have hypertension. However, sprawl did not return any significant effects on diabetes or coronary heart disease.

Another study also investigated the connection between sprawl and weight. Conducted by Lopez (2004) and titled "Urban Sprawl and the Risk for Being Overweight or Obese," an association between levels of obesity in the South (20%) and urban sprawl is discussed. Lopez contends that many of the metropolitan areas with the highest levels of urban sprawl are in the southern United States. From this association, Lopez set forth to quantitatively test for statistical significance between urban sprawl and the risk for being overweight or obese. Lopez's analysis shows that sprawl is associated with both an increased risk for being overweight (0.2% for each 1 point increase in his sprawl index, which is measured on a scale from 0 to 100, with 0 being the most compact) and an increased risk for being obese (0.5% for each 1-point increase in his

sprawl index). He found the risk for being obese is greater than the risk for being merely overweight when it comes to sprawl's effect.

All three of these studies have some shortcomings though. One issue is the validity of height and weight reported by participants (Lopez 2004; O'Toole 2002a; Bowlin et al. 1993; Jackson et al. 1992). Critics such as O'Toole (2002a) point to the definition of BMI itself, which is a standard measure of weight-to-height used to determine whether or not people are overweight or obese, as being problematic because it does not distinguish between body types or consider the overall health of an individual. There is also the issue of whether or not individuals are being honest when reporting their height and weight.

Even though critics point to the fact that height and weight may be over- and understated, respectively, in the BRFSS survey, it does not mean the data are completely flawed. Most people will estimate their weight downward, providing a systematic error. As far as height, this is more of a random error, since some people will provide a height that is too high and other will provide a height that is too low. This is less problematic than the systematic error found with weight. Every survey has its shortcomings and is plagued with the fact that some people will not be honest (Rowland 1990). However, measures of BRFSS data have been determined to have high reliability and validity, including height and weight (Nelson et al. 2001; Nelson et al. 2003).

Another issue with these studies is the fact that the data represent only one point in time (Lopez 2004; O'Toole 2002a; Bowlin et al. 1993; Jackson et al. 1992). Hence, there is no way of measuring whether or not respondents became overweight or obese while residing in the metropolitan area in which they were surveyed. It may be the case that some became overweight while residing in a very dense, less sprawling metropolitan area then moved to a very sprawling, less dense metropolitan area before participating in the survey, or vice versa. However,

according to Lopez, most people tend not to change metropolitan areas; instead remaining in one metropolitan area for long periods of time. Therefore, results should be a fairly appropriate reflection of individuals' exposure to sprawl. However, there is the issue of people making intra-metropolitan moves, such as from the city center to a suburban area, or vice versa. In moves like these, individuals are more likely to be exposed to varying degrees of sprawl.

This issue of varying degrees of sprawl leads to a final issue with Lopez's (2004) study, which is its significant ecological bias in that it does not reflect how sprawl varies within metropolitan areas. Metropolitan areas are not homogenous, but rather differ from inner city to suburbs, a factor not controlled for in Lopez's study. Ewing et al. (2003) and McCann and Ewing (2003) did, however, improve this issue that Lopez failed to address by employing a sprawl index measured at the county level. This allows for a more precise measure of how sprawl affects individual-level health without assuming that metropolitan-level sprawl is the same within a region or affects individuals residing in very different parts of the same metropolitan area in the same manner. A summary and critique of the different measures of sprawl, along with additional ways of quantifying sprawl is discussed in detail in Chapter 4.

The urbanized environment has consequences for health, but exactly how does urban sprawl relate to health outcomes? Is there a clear association? There may be many influences on one's health, and sprawl may only have a small and/or indirect effect. On the other hand, there may be a strong association between sprawl and health outcomes. It is important now, more than ever, to understand just how critical the decisions to plan, regulate, zone, and build the American landscape are to individual-level health. The better the effects of these decisions are understood, the better choices Americans can make in how to maintain their best health in the sprawling American landscape.

CHAPTER 3

CONCEPTUAL FRAMEWORK

Many different ideas have emerged as to the origins of urban sprawl. Theories range from anti-urban attitudes, racism, and increased affluence to economics and government policies. The latter scenario theorizes that policies at the local, state, and national level, such as homeowner subsidies, highway programs, infrastructure subsidies, and federal income tax deductions, have fostered sprawl. These policies, according to proponents of this theory, encourage city dwellers to move to the suburbs in favor of single-family home ownership instead of apartment living in crowded, often dirty, cities. Such suburban home ownership was made possible by government sponsorship of superhighways, suburban infrastructure, long-term amortized mortgages, and federal mortgage insurance, so the theory goes (Bruegmann 2005; Glaeser and Kahn 2003; Brueckner and Fansler 1983; Lindstrom and Bartling 2003).

Another theory favors that of technology as the cause of sprawl. As new communication and transportation options were made available, growth was able to disperse from city centers. According to this theory, in the past two centuries the railroad tended to concentrate growth and population within cities. The automobile, however, made it possible to disperse this concentration. This theory of sprawl advocates that before enhanced transportation options, cities were dense, but this density yielded to highly dispersed growth with the introduction of the automobile. This introduction of a personal means of long-range transportation, replacing private horse-drawn carriages, made it possible for long-distance commutes to and from the suburbs (Bruegmann 2005; Glaeser and Kahn 2003; Lindstrom and Bartling 2003).

Of course, it takes proper highway infrastructure to support commutes via automobile. Some critics blame postwar freeways as the cause of sprawl. However, it is not logical to place

complete blame on the highway system for the cause of sprawl. As ironic as it seems, these roads were heavily supported by central-city interests. It was believed that a highway system would reinforce the centrality of downtowns and make it easier for people from throughout the region to get there, much like the railroads did in the past. Such roads did make getting downtown much quicker, but they made it just as easy to leave the city as well. Even with such a connection, there is no concrete evidence to prove that decentralization of cities and subsequent sprawl throughout metropolitan areas was caused solely by postwar freeways. In fact, there is the argument that the decentralization caused by the highway system was no different than the decentralization caused by its predecessor, the railroad. Both have caused some dispersal, as well as centralization. The amount of each depends on many factors, including individual choices (Bruegmann 2005; Glaeser and Kahn 2003).

No matter how suburban or rural an area of a metropolitan region may be, most residents are closely tied economically and socially to the urban world, and therefore are dependent on that world (Bruegmann 2005; Glaeser and Kahn 2003; Lindstrom and Bartling 2003). This dependence on the urban environment makes certain technologies, like automobiles and vast road systems, necessary for mobility within metropolitan regions, many times at the expense of pedestrian-friendly street networks. No matter what caused sprawl and continues to fuel it, critics contend that it designs regular, beneficial, physical activity out of everyday life through the patterns of mobility it promotes, which leads to a path model developed by Ewing et al.

Figure 3.1 shows the model Ewing et al. (2003) developed from their analyses looking at the relationships between sprawl, physical activity, obesity, and disease. They determined that established relationships (represented by solid lines) exist between physical activity, obesity, and

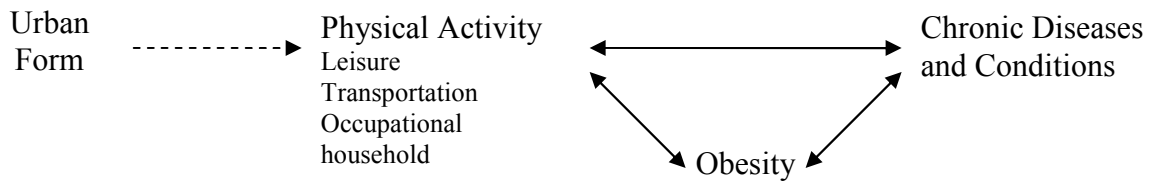

Figure 3.1 Established (Solid) and Speculative (Dashed) Relationships, from Ewing et al. (2003)

chronic diseases and conditions. Analyses also point to a speculative relationship between urban form and physical activity. Ewing et all. (2003) developed their conceptual model (Figure 3.1) after they performed their analyses of the data, but they did not conduct analyses to determine if any indirect effects exist between the variables. One of my goals and contributions in this dissertation research is to investigate possible indirect effects between sprawl, physical activity, and body weight in regards to their relationship to health outcomes (depicted in Figure 3.2)
As it can be seen in my conceptual model (Figure 3.2), hypothesized relationships between the variables depict direct and indirect effects. This conceptual model is the basis for my research. I use it not only to look for associations between sprawl, physical activity, weight, and health outcomes, but also text for indirect effects. I also look to answer questions, such as: 1) Is there a significant association between sprawl and physical activity and if so, in what direction does sprawl affect physical activity? Do individuals residing in less sprawling areas get significantly more or less physical activity on average than those in more sprawling areas? and 2) Does physical activity have a significant association with health outcomes by way of individual-level weight? It is important to find out which, if any, associations between sprawl, physical activity, weight, and health outcomes are the strongest in order to proceed with additional, more refined research testing the causal relationships depicted in this model. Variables controlling for

individual-level characteristics, such as sociodemographics, personal health, and region, are also included in statistical models in order to test each hypothesis.

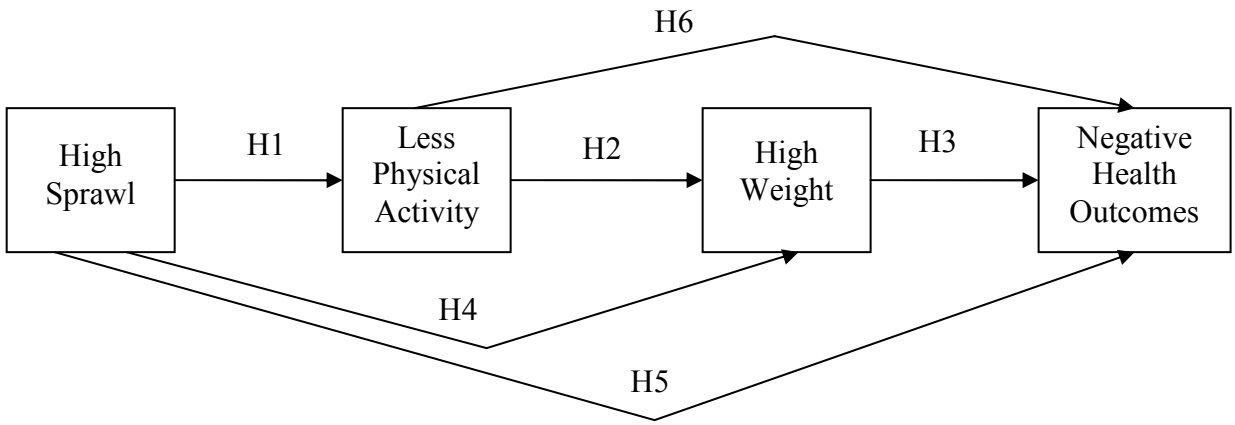

Figure 3.2 Conceptual Model Used in this Dissertation

Hypotheses

Taking direction from previous studies and the conceptual model outlined above, I hypothesize that there may be associations between sprawl, levels of physical activity, body weight, and health outcomes. Therefore, I hypothesize the following:

Hypothesis 1 Sprawl is negatively related to physical activity. As sprawl increases, individual levels of physical activity decrease.

I suspect that built environmental form has a strong relationship to one's health and that the design of communities influences health by encouraging or discouraging routine physical activity. Previous research suggests that people living in sprawling areas miss out on significant health benefits available, such as walking to the store, to work, or other places as part of a daily routine (McCann and Ewing 2003). This lack of beneficial regular physical activity may be explained by over-reliance on automobiles and/or by patterns of streets within neighborhoods,

such as those found in many suburban subdivisions, as they seem to affect people's propensity to walk (Lopez 2004).

Hypothesis 2 Physical activity is negatively related to weight. As individual levels of physical activity decrease, individual weight increases.

Obesity has reached epidemic levels in the United States and physical inactivity is now implicated as one of the causes of this condition. An increasing body of evidence suggests that moderate forms of regular physical activity, such as walking, can have beneficial effects in maintaining a healthy weight (McCann and Ewing 2003; Ewing et al. 2003). Unfortunately, the majority of Americans report not obtaining enough exercise to meet the recommended weekly minimum, and many Americans remain completely inactive during their leisure time (McCann and Ewing 2003).

Hypothesis 3 Weight is positively related to negative health outcomes. As individual weight increases, so does the chance for individual-level negative health outcomes, such as diabetes, heart attack, heart disease, high cholesterol, hypertension, and stroke.

Health experts believe most Americans are overweight and conditions such as diabetes and heart disease have reached epidemic levels as a consequence of such excessive weight (McCann and Ewing 2003; Ewing et al. 2003). Such connections between body weight and health conditions have long been documented, as well as an increase in the risk of mortality coinciding with the severity of obesity (Calle et al. 1999).

Hypothesis 4 Sprawl is positively related to weight. As sprawl increases, so does individual-level weight.

This hypothesis seeks to determine if there is a direct effect between sprawl and weight. It is possible that no direct effect exists, but in order to be certain I test this hypothesis in order to

determine how much variance in the individual-level variable of weight can be explained by the macro-level variable of sprawl, controlling for other factors.

Hypothesis 5 Sprawl has a positive relationship to individual-level negative health outcomes. As sprawl increases, so does the chance for negative health outcomes, such as diabetes, heart attack, heart disease, high cholesterol, hypertension, and stroke, especially when mediated by physical activity and BMI.

Based on existing research, I hypothesize a positive zero-order correlation between sprawl and negative health outcomes. As sprawl increases I expect a positive relationship between sprawl and health problems to be revealed, especially when controlling for other variables such as sociodemographics, personal health, and region. However, with the addition of the mediating variables of physical activity and BMI, I hypothesize that the relationship between sprawl and negative health outcomes will shrink and become statistically insignificant.

Hypothesis 6 Physical activity is negatively related to negative health outcomes. As individual levels of physical activity decrease, the chance for negative individual-level health outcomes, such as diabetes, heart attack, heart disease, high cholesterol, hypertension, and stroke increase.

Physical activity has been shown to have beneficial health outcomes (Perdue 2004; McCann and Ewing 2003; Ewing et al. 2003; Ewing et al. 2002a); therefore, it is theoretically logical to test for the variance explained by the relationship between these two individual-level variables.

Additional Considerations

There are multiple additional variables besides sprawl, levels of physical activity, and weight that need to be accounted for when statistically modeling for associations between sprawl and health outcomes, variables which many previous studies fail to address. Epidemiologists

have termed this multi-variable association the 'web of causation' to refer to the fact that health and disease are not explained by simple bivariate relationships (Krieger 1994). Instead, health and disease are explained by a complex network of numerous interconnected risks and factors, or multiple causations. Factors that may be important in explaining associations for health outcomes have been absent in previous studies that examine sprawl and health outcomes at the individual- and metropolitan-level. When looking at sprawl and health outcomes, other mediating factors may be involved. Therefore, it is important to statistically model other independent and control variables, such as race, age, gender, education, and region. In addition, a variable that measures whether or not an individual is consciously increasing his or her level of physical activity in order to control their weight is examined as well.

CHAPTER 4

MEASURES OF URBAN SPRAWL

Many different expressions have been used to describe urban sprawl. The uncontrolled, unplanned spread of urban development into areas adjoining the edge of a city and a continuous network of low density urban communities are just two descriptions of sprawl found in the literature (Galster et al. 2001; Razin 1998). Complicating the situation of defining sprawl is the fact that many times it is expressed as a noun, where it describes a condition characterizing all or part of an urban area at a particular point in time. Other times it is used as a verb, describing the process of converting land over a period of time from non-urban to urban uses, or as changes in the extent or intensity of urbanization, particularly at urban fringes. By defining sprawl as either a condition (a noun) or a process (a verb), ambiguity and idiosyncrasy set in, making it impossible to know with confidence the causes, consequences, or effects of sprawl, as well as the effectiveness of policies designed to control it (Wolman et al. 2005; Galster et al. 2001; Fulton et al. 2001; Ewing et al. 2002a). Ambiguity in defining sprawl creates problems regarding its measurement.

Measures of sprawl vary as much as the definitions for sprawl itself and range from the simple to the complex. Such varied ways of measurement make it difficult to pinpoint what exactly is meant by sprawl, how it should be measured, and the geographical areas and types of land that should be considered (Wolman et al. 2005). In order to track sprawl, scholars, as well as planners and policy makers, need a way to define and measure it, as well as be able to demonstrate how and to what degree sprawl has genuine implications (Wolman et al. 2005; Galster et al. 2001; Fulton et al. 2001; Ewing et al. 2002a). Fortunately, measurements and

components of sprawl are found throughout the literature, and several recent studies address how to define and empirically operationalize the concept.

This section discusses recent studies that focus on sprawl and health and describe their measures of sprawl. It concludes with a discussion of the measure of sprawl in which I use for analyzing my conceptual model.

Studies Focusing on Sprawl and Health Outcomes

There are two recent studies focusing on sprawl and its association with health. In a study looking at the measurement, distribution, and trends of sprawl in the 1990s, Lopez and Hynes (2003) define sprawl as a process where the overall pattern of metropolitan land development consists of populations residing in lower-density developments. In Lopez and Hynes' sprawl index, metropolitan areas with much of their population concentrated in certain areas are considered less sprawling than metropolitan areas with a population that is evenly distributed across the entire region. Even though their sprawl index is fairly simple, it is a useful measure of sprawl and is based on accessible public data. Lopez and Hynes contend that concentration (the distribution of density) is an important factor in measuring sprawl and state that focusing on density computations alone across metropolitan areas will result in an index that is misleading, because sprawl is also a function of how density is distributed. Unfortunately Lopez and Hynes' scale fails to reflect the spatial positions of low- and high-density tracts. Additionally, their sprawl index is based on subjective cut-off points for low- and high- density tracts. If their cut-off points were changed, so would their sprawl index calculations (Jaret et al. 2009). The sprawl index used in the Lopez and Hynes study is the same one Lopez (2004) used

in his study looking at the relationship between sprawl and the risk for being overweight or obese.

Another important study of sprawl that measured and computed it for U. S. metropolitan areas was done by Ewing et al. (2002a, b). This was a landmark study conducted by Rutgers and Cornell Universities for Smart Growth America, a national public interest group that promotes smart growth policies. In this study, the researchers define sprawl as a process where development across the landscape far outpaces population growth and provides people with poor accessibility. Ewing et al. assert that a sprawling landscape consists of: (1) a population that is dispersed in extensive low-density development; (2) homes, shops, and workplaces that are rigidly separated from one another; (3) a network of roads that consist of large blocks with poor access; and (4) a lack of well-defined activity centers, such as downtowns. These authors state that other features of sprawl, such as the lack of transportation choices and difficulty walking, are results of these four unique dimensions of sprawl.

Ewing et al. set forth to create a sprawl index that can be measured and analyzed based on the four dimensions of sprawl they identify. Each dimension comprises several measurable components that were tested to ensure they added a unique perspective to the overall representation of sprawl. For example, residential density includes the proportion of residents living in very spread-out areas, the proportion of residents living close together, and overall density, as well as other measures. Ewing et al. argue that they have created the most comprehensive attempt to define and quantify sprawl in the United States. A list of the four factors measuring sprawl and the sources for data can be found in their study.

Ewing et al. (2002a, b) computed their four dimensions of sprawl by performing factor analysis on the numerous variables representing each given dimension they identified. Based on

the factor loadings, they created an index score for each dimension, as well as a fifth composite score, for 83 metropolitan areas. Their resulting indices are one of the most comprehensive attempts to define and quantify sprawl in the United States. However, their scores are based on 1990 Census data, as there was little data available for the year 2000 when they created their index. Therefore, their index is not as up to date as other sprawl indices and leaves researchers who desire to use the Ewing et al.'s sprawl method the task of computing their own, more up to date index. Since metropolitan areas are dynamic and increase in size and area over time, it is imperative that the most up-to-date index on sprawl be utilized for any type of analysis and/or comparison (Jaret et al. 2009).

Another study that focuses on sprawl and health was conducted by McCann and Ewing (2003), which was derived from the landmark study by Ewing et al. (2003). The McCann and Ewing research was a follow up study utilizing a county-level sprawl index developed for the Ewing et al. research. In addition to their metropolitan-level sprawl index, Ewing et al. also developed a county-level sprawl index, using a very similar measure to their metropolitan-level index. While looking at measuring the health effects of sprawl as it relates to physical activity, obesity and chronic disease, McCann and Ewing contend that although the metropolitan-level sprawl index developed by Ewing et al. is an extremely comprehensive means of measuring sprawl, they needed a finer degree of information for their study. So, they turned to the county-level sprawl index developed by Ewing et al. This research conducted by McCann and Ewing is very similar to my dissertation. The main differences are the fact that I also test for indirect effects among the variables and employ additional health outcomes.

Ewing et al.'s (2003) county-level sprawl index used relevant data from the Ewing et al. metropolitan sprawl study to create a county-level index that scores 448 counties utilizing six

variables from the U.S. Census and the Department of Agriculture's Natural Resources Inventory measure on residential density and street network connectivity. Ewing et al. conducted factor analysis to derive their sprawl index from their data sources. Even though fewer data are available at the county level and their index is less comprehensive than the metropolitan-level sprawl index, it is still the most complete measurement of sprawl available at the county level.

Ewing et al. (2003) developed their county-level sprawl index based on variables that reflect two dimensions of sprawl – residential density and street network connectivity, two of the original four dimensions employed by Ewing et al. (2002a, b). Overall, they utilized six variables to develop their sprawl index, which include: (1) population density per square mile; (2) percentage of population living at densities less than 1,500 per square mile; (3) percentage of population living at densities greater than 12,500 per square mile; (4) net population density of urban lands (excludes lands not directly related to the figure, such as green spaces and roads); (5) average block size in square miles; and (6) percentage of small blocks (≤ 0.01 square mile). The technique utilized by Ewing et al. provided the researchers with a sprawl index that is a comprehensive measurement of sprawl at a finer, more appropriate level. A list of the factors, variables, and sources Ewing et al. utilized for their county-level sprawl index is listed in Appendix B.

The sprawl index that Ewing et al. (2003) developed is somewhat counterintuitive in that high scores indicate low levels of sprawl, whereas low scores represent high levels of sprawl. For their county-level sprawl index, scores range from 63.12 (high sprawl) for Geauga County (which is a mostly rural county) in the Cleveland, OH metropolitan area to 352.07 (low sprawl) for the very compact New York County (Manhattan) in the New York, NY metropolitan area. Both of these two cases are considered outliers with most counties clustered around an average

score of 100. It should be mentioned that counties on the lower end of the scale (those indicative of high sprawl by this county-level index) are actually more exurban than suburban and therefore do not necessarily accurately represent sprawl for certain counties, like that of for Geauga County, OH.

Conclusion

When determining an appropriate method for quantifying sprawl, it is important to keep in mind that sprawl is very subjective in nature. It is also imperative to capture as many vital aspects of sprawl as possible, as well as to determine the correct geographical area and level at which to calculate it. Variables that accurately represent the theoretical abstraction of sprawl must be considered in order to objectively measure it. Since operational variables are seldom complete and accurate representations of underlying constructs and they are subject to measurement and sampling errors, it is important to use multiple variables in order to capture the essence of the construct. One or two variables cannot adequately capture and truly represent the inherent complexity of sprawl. Therefore, multiple variables are needed to represent its various dimensions (Ewing et al. 2002a). Unfortunately, many studies fall short in determining the vital aspects of sprawl and measuring it adequately.

Keeping in mind that sprawl is a construct that must be operationalized properly, I utilize the Ewing et al. (2003) county-level sprawl index due to its multiple-variable representations of two dimensions of sprawl. The Ewing et al. county-level sprawl index is an excellent choice for my research because spatial form is inherently multidimensional and does not stem from one single process. Rather, it is a complex phenomenon that interconnects many different social and economic processes (Timms 1971; Massey and Denton 1988). Even though no absolute definition of sprawl exits, it tends to be defined by its characteristics, such as those identified by

Ewing et al. (2002a. b. 2003) and McCann and Ewing (2003). For this reason, I utilize the Ewing et al. index (the same one utilized in the McCann and Ewing study) for my research concerning the connection between sprawl, physical activity, weight, and health outcomes.

Since certain data were not available for all counties, Ewing et al. (2003) lost a total of 73 counties that were included in their metropolitan-level sprawl index. However, they gained 90 additional counties by incorporating other metropolitan areas, such as Chattanooga, TN-GA, Mobile, AL, and Augusta-Aiken, GA-SC in their study. Additionally, the BRFSS does not report data for certain counties that are included in the Ewing et al. (2003) sprawl index; therefore, I lose a certain amount of cases from the BRFSS dataset because they do not have a corresponding county-level sprawl index.

I also have the issue with more rural and/or exurban counties like Geauga, OH. I have kept counties that are more rural or exurban in my analyses for several reasons. Counties like Geauga, OH are metropolitan counties, as defined by the U.S. Census Bureau, for a reason. They are tied to their respective metropolitan core both socially and economically and are not recent additions to metropolitan areas. They have been a metropolitan county for at least ten years. Additionally, exurbanites often have to travel farther than their suburban counterparts to get to places such as stores, schools, etc. because commercial development is often farther apart in the exurbs.

Sprawl index scores for all 448 counties included in the Ewing et al. (2003) and McCann and Ewing (2003) studies can be found in Appendix A.

CHAPTER 5

METHODOLOGY

Most of the data for analyzing my conceptual model comes from the Behavioral Risk Factor Surveillance System's (BRFSS) 2003 Chronic Disease and the Environment dataset (Centers for Disease Control and Prevention 2003). The BRFSS is an annual telephone survey of adults that includes more than two hundred self-reported and calculated variables. It is an excellent source of information concerning the health status and habits for the U.S. population. This is the same source for data in the Ewing et al. (2003) and McCann and Ewing (2003) research. Instead of 2003 data, those studies incorporated data from the years 1998 to 2000.

The Behavioral Risk Factor Surveillance System is an annual telephone survey that utilizes a questionnaire distributed to each of the 50 states. It is developed jointly by the CDC's Behavioral Surveillance Branch (BSB) and the States. The questionnaire is constructed at the BRFSS Working Group annual meeting in February of each year. Representatives from the National Center for Chronic Disease Prevention and Health Promotion (NCCDPHP) and other parts of the CDC propose BRFSS questions for consideration, along with input and feedback each state provides on the proposed content. After their annual meeting the BSB designs the core components, as well as optional modules, and sends them to the states, where each state may add questions they have designed or acquired for their own purposes for health surveillance.

After the questionnaire has been designed and distributed to each of the 50 states, the BSB provides each state samples of telephone numbers. Each state must then review its sampling methodology with a state statistician and BSB to make sure data collection procedures are in place and that they follow correct methodology. States then conduct interviews each month using the prescribed protocol and enter the data into a computer-assisted telephone

interviewing (CATI) computer file, then edit and correct completed interviews. States then submit their data to the BSB where it is weighted according to state-specific population estimates and distributes the information accordingly.

The BRFSS survey protocol requires all states to ask the core component of questions without modification; although states may choose to add any of the optional modules, as well as state-added questions after the core component. Electronic monitoring is a routine and integral part of the monthly survey procedures for all interviewers. If electronic monitoring is not used, then a 5% random sample of each month's interviewees must be called back to verify the quality of selected responses.

The BSB states that an eligible household for surveying is a housing unit that has a separate entrance where occupants eat separately from other persons on the same property and such household is occupied by its members as a principal or secondary place of residence. Eligible household members include those who are 18 years and older, related or unrelated, roommates, and domestic workers who consider the dwelling their home. Completed interviews must include age, race, and gender. If such values are not entered, imputed values are generated and used only to assign post-stratification weights. The average time to complete an interview for the 2003 BRFSS annual survey was 20.8 minutes and the response rate (defined as completed interviews plus partially completed interviews, divided by all eligible interviews) was 48.3% (Centers for Disease Control and Prevention 2003).

It should be noted that not all cases included in the 2003 BRFSS dataset were utilized due to reconciliation with Ewing et al.'s (2003) sprawl index. After reconciliation, I lost 115 counties from their sprawl index because they were not included in the BRFSS dataset. Additionally, I lost 474 counties from the BRFSS dataset because they were not included in

Ewing et al. sprawl index. However, after reconciliation between the Ewing et al. county-level sprawl index and the 2003 BRFSS dataset, there were cases from 326 counties across 109 metropolitan areas in 40 states included in analyses. Every effort was made to ensure the BRFSS dataset was reconciled with the Ewing et al. sprawl index accurately. This was done by matching each case in the BRFSS individual-level database with its correct county-level sprawl index score. Matching was completed using the FIPS code, a code employed by the U.S. government to assign each county in the nation a unique number. The BRFSS includes these codes and lists them for each case. By looking up the FIPS code for each county in Ewing et al.'s sprawl index, I was able to match each case with its correct sprawl index score.

Since I was unable to use certain cases from the BRFSS dataset in analyses, I compared the means of certain variables to determine if cases used in analyses differ systematically from the cases not used. Table 5.1 shows the results of this comparison. It details the difference in means between used and unused cases. T-tests return highly significant results stating that unused cases do in fact differ systematically from those that were used in analyses.

Table 5.1 Difference in Means Significance Testing

	Used Cases	**Unused Cases**
Variable	**% / Mean**	**% / Mean**
Dependent		
Diabetes	7.4%	8.7%***
Heart Attack	3.7%	5.1%***
Heart Disease	4.4%	5.8%***
High Cholesterol	34.2%	35.6%***
Hypertension	26.9%	30.4%***
Stroke	2.2%	3.0%***
Independent/Mediating		
Physical Activity (hours/week)	1.33	0.67***
Body Mass Index	26.63	27.06***
Control		
Increased Physical Activity	70.0%	67.0%***

*p < 0.050, **p < 0.010, ***p < 0.001

In order to better understand the data and variables in this study, dependent, independent, and control variables are discussed below. Table 5.2 lists descriptive statistics for all variables included in analyses.

Dependent Variables

There are six different dependent variables. They include diabetes, heart attack, heart disease, high cholesterol, hypertension, and stroke. All data for the dependent variables were obtained from the BRFSS dataset.

Diabetes. This variable shows whether or not individuals in the sample have diabetes. It is operationalized as being either *yes*, *no*, or *yes – pregnancy-related diabetes*. I coded all cases that were *yes – pregnancy related diabetes*, which was 0.90% of the sample, as *no* because it is a temporary condition related to some pregnancies and is not a normal day-to-day condition as with other chronic cases of diabetes. In my final dataset, 7.4% of the sample has diabetes.

Heart Attack. This variable shows whether or not individuals in the sample have ever had a heart attack. In my final dataset, 3.7% of individuals in the sample have had a heart attack. Heart attack was not defined in the survey; rather interviewers relied on whether or not interviewees were ever told by a medical professional that they have had a heart attack.

Table 5.2 Descriptive Statistics

Variable	Range	% Mean/SD
Dependent		
Diabetes	–	7.4%
Heart Attack	–	3.7%
Heart Disease	–	4.4%
High Cholesterol	–	34.2%
Hypertension		26.9%
Stroke	–	2.2%

Table 5.2 continued

Variable	Range	% Mean/SD
Independent and Mediating		
Sprawl Index (reversed)	0 – 282.90	237.66/28.71
Physical Activity (hours/week)	0 – 19.97	1.33/1.59
Body Mass Index	8.38 – 99.98	26.63/5.52
Control		
Age	18 – 99	48.25/16.95
Education		
Less than 9th Grade	–	2.6%
Grades 9 – 12	–	5.6%
High School Graduate (or GED)	–	26.2%
Some College or Technical School	–	27.0%
College Graduate	–	38.6%
Income		
Less than $10,000	–	4.7%
$10,000 - $14,999	–	4.7%
$15,000 - $19,999	–	6.7%
$20,000 - $24,999	–	8.7%
$25,000 - $34,999	–	12.8%
$35,000 - $49,999	–	17.0%
$50,000 - $74,999	–	18.1%
$75,000 or more	–	27.3%
Gender		
Male	–	39.0%
Female	–	61.0%
Race/Ethnicity		
White, not Hispanic	–	75.9%
Black, not Hispanic	–	9.9%
Asian, not Hispanic	–	3.1%
Native Hawaiian or other Pacific Islander, not Hispanic	–	0.4%
American Indian or Alaskan Native, not Hispanic	–	0.8%
Other race, not Hispanic	–	0.8%
Multi-racial, not Hispanic	–	1.6%
Hispanic	–	7.7%
Marital Status		
Married	–	52.8%
Member of unmarried couple	–	3.1%
Widowed	–	10.0%
Separated	–	2.7%
Divorced	–	13.7%
Never married	–	17.7%
Region		
Northeast	–	30.5%
South	–	29.6%
Midwest	–	18.5%
West	–	21.4%
Increased Physical Activity	–	70.0%

Heart Disease. This variable shows whether or not individuals in the sample have heart disease. In my final dataset, 4.4% of individuals in the sample have heart disease. Heart disease was not defined in the survey; rather interviewers relied on whether or not interviewees were ever told by a medical professional that they have heart disease.

High Cholesterol. For those who have had their cholesterol checked, this variable shows whether or not individuals in the sample have high cholesterol. In my final dataset, of those who had their cholesterol checked, 34.2% of them have high cholesterol. Like that of heart attacks and heart disease, high cholesterol was not defined in the survey; rather interviewers relied on whether or not interviewees were ever told by a medical professional that they have high cholesterol.

Hypertension. This variable shows the prevalence of high blood pressure in the sample. It consists of whether or not an individual was told on two or more different visits to a physician or other health professional that they have high blood pressure. As with the variable for diabetes, hypertension is operationalized as being either *yes*, *no*, or *yes – pregnancy-related hypertension*. I coded all cases that were *yes – pregnancy related hypertension*, which is 1.10% of the sample, as *no*. As with the reasoning for coding those with pregnancy related diabetes as *no*, I coded those with pregnancy related hypertension as *no* because this is a temporary condition and not a normal day-to-day condition as with other chronic cases of hypertension. In my final dataset, 26.9% of the sample has hypertension. Hypertension was not specifically defined in the survey; rather interviewers relied on whether or not interviewees were ever told by a medical professional that they have hypertension.

Stroke. This variable shows whether or not an individual in the sample has ever had a stroke. In my final dataset, 2.2% of the sample has had a stroke.

Independent Variables

Three different independent variables are incorporated into my research. They include sprawl, physical activity, and BMI. Data for independent variables are available from the Ewing et al. (2003) study concerning sprawl and health, as well as the BRFSS dataset. Below are detailed descriptions of each independent variable.

Sprawl. As mentioned previously, I employ the Ewing et al. (2003) county-level sprawl index. This index allows me to analyze for any statistical connections between sprawl, physical activity, weight, and health outcomes at the county and individual level. The Ewing et al. county-level sprawl index is preferable because it addresses the fact that spatial form does not stem from one single process; that it consists of more than one variable. It is a complex phenomenon that interconnects many different social and economic processes that occur over time (Timms 1971; Massey and Denton 1988).

The ordering of the Ewing et al. sprawl index is rather counter-intuitive, with the least sprawling county having the highest score. In order to have a more intuitive sprawl index, I reversed their sprawl index so it would start at zero. I accomplished this with a simple mathematical calculation. I subtracted each county's sprawl score from 352.07, allowing me to have New York County, NY (Manhattan) start the sprawl scale at 0 and Geauga County, OH end it with 288.95. However, no individuals residing in the least sprawling county in the Ewing et al. sprawl index (Geauga County, OH) were included in the BRFSS dataset. Therefore, in my final dataset, I had a range of 0 (Manhattan) to 282.90 (Yadkin county, NC) and respondents' average sprawl index score was 237.66 with a standard deviation of 28.71. It should be noted that using this sprawl index resulted in low multiple regression coefficient values (many so low that

computer output indicated values of only 0.000); therefore, I divided my reversed sprawl index by 10 in order to obtain readable coefficient values in my regression analyses.

Physical Activity. Physical activity is shown to have beneficial health outcomes, especially when it relates to conditions such as diabetes, heart disease, high cholesterol, and stroke. In addition to the positive health effects of physical activity, previous research shows that the lack of physical activity may have a correlation to sprawl (Perdue 2004; McCann and Ewing 2003; Ewing et al. 2003; Ewing et al. 2002a). Therefore, physical activity is an important independent variable. The BRFSS uses two variables to determine levels of physical activity. They include: (1) the average amount of minutes per day and the number of days per week individuals receive moderate physical activity; as well as (2) the average amount of minutes per day and days per week individuals receive vigorous physical activity. Vigorous physical activity is defined by the BRFSS as activities that cause large increases in breathing or heart rate, while moderate activities cause only small increases in breathing or heart rate. The BRFSS calculates the number of minutes individuals receive both moderate and vigorous physical activity per week. I took both of these calculated variables and added them together to obtain the total minutes of physical activity each case in the sample receives per week. I then divided that total by 60 in order to determine how many hours of physical activity each individual in the sample gets per week. In my final dataset, individuals have a mean of 1.33 hours of physical activity per week, with a standard deviation of 1.59. Physical activity ranges from 0 to 19.97 hours per week.

Body Mass Index (BMI). BMI is the construct I use to represent weight. BMI is a standard measurement of weight to height that reliably determines overweight and obesity in individuals (McCann and Ewing 2003; Guo et al. 2002; National Institute of Health 1998). BMI

is calculated by dividing an individual's weight by his or her height in inches squared, then multiplying by 703 ([weight in pounds/height in inches^2] x 703). Individuals with BMI of 25 or higher are considered overweight and those with BMI of 30 or more are considered obese (McCann and Ewing 2003; Guo et al. 2002; National Institute of Health 1998). The 2003 BRFSS calculates BMI for each case within the dataset and uses a slightly different formula. The 2003 BRFSS transforms reported weight and height from pounds and inches to kilograms and meters, respectively, then performs the following mathematical calculation: *weight in kilograms / (height in meters * height in meters)*. The BRFSS does not utilize decimal places for this variable and notes that two decimal places are implied. Therefore, for my analyses I divided this variable by 100 to obtain the correct BMI for each case. In my final dataset, the average BMI for individuals in the sample is 26.63, with a standard deviation of 5.52. BMI ranges from 8.38 to 99.98. BMI values greater than 50.00 are considered super morbidly obese and there are 50 cases in the sample that have a BMI over 50.00 (Dresel, Kuhn, and McCarty 2004).

Control Variables

Control variables consist of sociodemographics, a personal health variable, and a variable that determines region within the nation. These control variables are explained below, with detailed information for each.

Sociodemographic Control Variables

Sociodemographic variables, such as age, gender, race, educational attainment, income, and marital status are shown in previous research to be important predictors of health outcomes (Lopez and Hynes 2003; Paeratakul et al. 2002; Kelly-Schwartz et al. 2004; McNeill, Kreuter,

and Subramanian 2006). Therefore, they are included in statistical models as controls for my research. Each control variable is measured at the individual level. All data for sociodemographic control variables are included in the BRFSS dataset. More information concerning each sociodemographic control variable is detailed below.

Age. Studies show that healthy behaviors, such as physical activity, decline with age (McCann and Ewing 2003; Ewing et al. 2003; Egede and Zheng 2002; Frank and Engelke 2001; Giles-Corti and Donovan 2002). Therefore, many of the health outcomes being investigated in this dissertation could be influenced by age. In light of this possibility, age is a control variable for determining relationships between sprawl and health outcomes. Age is measured in number of years and is a continuous variable. The mean age within the sample is 48.25 with a standard deviation of 16.95. Age ranges from 18 to 99 years of age.

Education. Previous research shows that those with higher educational attainment are more likely to engage in healthy behaviors, such as regular physical activity, commitment to proper diets, and reduced tobacco use (McCann and Ewing 2003; Ewing et al. 2003; Giles-Corti and Donovan 2002; Ford et al. 1991; Jeffery et al. 1991). However, there is some research that shows negative health outcomes, such as obesity, are associated with higher levels of education (Mokdad et al. 1999). Therefore, educational attainment is controlled for in analyses. Education consists of an ordinal variable that measures: (1) those who never attended school or only kindergarten (0.1%); (2) those who attended grades 1 – 8 (2.5%); (3) those who attended grades 9 – 12 (5.6%); (4) those who are high school graduates (or have a GED) (26.2%); (5) those who have some college or technical school (27.0%); and (6) those who are college graduates (38.6%). Since those who never attended school or only kindergarten and those who attended grades 1 – 8 are so low in numbers, I combined these two categories for analyses. Also, in order to test for

non-linear effects, I tested separate analyses with education as a set of dummy variables with those who have less than a high school education as the reference category. I found that education is very linear in its effects on health outcomes and that each higher level category of education increases its effect and significance on the outcome variables when run as a set of dummy variables.

Income. Research continually shows a clear association between negative health outcomes and low income. This is in part due to the fact that low income people do not have access to proper nutrition and health care coverage (Day 2006; Strunk and Reschovsky 2004; Subramanian and Kawachi 2006; Gorin and Moniz 2004; Eberstadt and Satel 2004). In order to determine how earnings affect health outcomes, an ordinal variable for annual household income, with eight categories ranging from: (1) less than $10,000 (4.7%); (2) $10,000 to less than $15,000 (4.7%); (3) $15,000 to less than $20,000 (6.7%); (4) $20,000 to less than $25,000 (8.7%); (5) $25,000 to less than $35,000 (12.8%); (6) $35,000 to less than $50,000 (17.0%); (7) $50,000 to less than $75,000 (18.1%); to (8) $75,000 or more (27.3%), were included in analyses. Like that of education, I also ran analyses with income as a set of dummy variables with the lowest annual income level as the reference category to test for non-linear effects. As with education, I found income to be very linear in its effects on health outcomes and that each higher level category of income increases its effect and significance on the outcome variables when tested as a set of dummy variables.

Gender. Research shows associations between women and less physical activity, which ultimately leads to negative health outcomes (Mokdad et al. 1999; Frank and Engelke 2001; Giles-Corti and Donovan 2002). Therefore, gender is included in analyses as a control. The

reference category for the gender dummy variable is males, with 61% of the sample being comprised of females.

Race and Ethnicity. Previous research shows that Blacks and Latinos are more likely to have greater negative health outcomes, such as obesity and diabetes, and get less than the recommended amount of physical activity, when compared to other racial and ethnic groups (Day 2006; McCann and Ewing 2003; Egede and Zheng 2002; Mokdad et al. 1999). Therefore, race is included in analyses as controls. Racial categories that the BRFSS places participants into include: (1) White only; (2) Black or African American only; (3) Asian only; (4) Native Hawaiian or other Pacific Islander only; (5) American Indian or Alaskan Native only; (6) Other race only, and (7) Multiracial. The BRFSS also asks participants if they are of Hispanic or Latino origin. From these racial and ethnic categories, the BRFSS recodes each case into the following categories: (1) White, not Hispanic (75.9%); (2) Black, not Hispanic (9.9%); (3) Asian, not Hispanic (3.1%); (4) Native Hawaiian or other Pacific Islander, not Hispanic (0.4%); (5) American Indian or Alaskan Native, not Hispanic (0.8%); (6) Other race, not Hispanic (0.8%); (7) Multi racial, not Hispanic (1.6%); and (8) Hispanic (7.7%). The reference category for the race/ethnicity dummy variables is those who are non-Hispanic White.

Marital Status. Research has found that marital status plays an important role in mortality rates (Umberson 1992; Gove 1973; Stroebe and Stroebe 1983). Married couples have lower rates of mortality than do their unmarried counterparts. This advantageous effect seems to have more benefit for men than it does for women, as wives often take on the responsibility of controlling the health of their husbands (Umberson 1992). In addition, a transition from married to unmarried status is associated with an increase in negative health behavior, especially for widows who are unable to cope with the death of their spouse (Umberson 1992; Gove 1973;

Stroebe and Stroebe 1983). Therefore, marital status is included in analyses as a set of dummy control variables. Those who are: (1) widowed (10.0%); (2) separated (2.7%)/divorced (13.7%); and (3) never married (17.7%); were compared to those who are married (52.8%)/member of an unmarried couple (3.1%).

Marital Status * Sex Interaction Effect. Since marital status has been shown to be more beneficial for married men than for married women regarding health, an interaction effect for married women is examined to determine any significant statistical effects for this category of women when it comes to my conceptual model. Thirty percent of the sample is comprised of married females.

Personal Health Control Variable

Techniques employed by previous studies to determine relationships between sprawl and negative health outcomes fail to account for what epidemiologists term as the 'web of causation,' the conviction that health and disease are explained by a complex network of numerous interconnected risks and factors, or multiple causations (Krieger 1994). Factors that may be important in explaining associations for health outcomes have been absent in previous studies that examine sprawl and health. Factors related to health, like that of weight control, may also be important in determining associations for health outcomes. Therefore, it is important to statistically model for such a variable when testing my conceptual model. By adding a weight control variable, I am able to determine whether or not there are multiple effects on individual health outcomes, rather than just simple independent and sociodemographic variables. My personal health control variable is measured at the individual level and included in the BRFSS dataset. Below is a description of this variable.

Increased Physical Activity. Physical activity is shown to have beneficial health outcomes, especially when it relates to conditions such as diabetes, heart attack, heart disease, high cholesterol, and stroke. In addition to the positive health effects of physical activity, previous research shows that the lack of physical activity may have a correlation to sprawl (Perdue 2004; McCann and Ewing 2003; Ewing et al. 2003; Ewing et al. 2002a). Along with the mediating variable that measures the total amount of hours of physical activity each respondent receives per week, I also employ a variable that determines whether or not individuals are consciously increasing their amount of physical activity in order to lose or maintain their body weight. In my final dataset, 70% of the cases are actively trying to lose or maintain their body weight by increasing their amounts of physical activity. When examining correlations between total amount of weekly physical activity and increasing physical activity in order to control weight, results clearly show a statistical significance between the two variables, suggesting that those trying to lose or maintain their body weight by increasing their amount of physical activity do in fact tend to get more weekly physical activity than individuals not trying to control their weight by increasing their amount of physical activity. In my final dataset, those who consciously increase their physical activity in order to control their weight have 1.47 mean hours of physical activity per week with a standard deviation of 1.50, compared to those who do not consciously increase their physical activity in order to maintain their weight with an average of 0.95 hours of physical activity per week with a standard deviation of 1.50. So, it can be seen that those who consciously increase their physical activity to control their weight do in fact differ statistically from those who do not increase their physical activity in order to control their weight.

Regional Control Variable

Since Lopez (2004) found an association between levels of obesity in the South and the fact that many metropolitan areas with the highest levels of urban sprawl are in the southern United States, it is important to control for the area of the nation where individuals reside. Below is a description of this variable.

Region. Controlling for region allows me to determine associations between the variables in my conceptual model that may be affected by region. Like the county-level sprawl index I utilize, each case in the sample is matched with its respective region within the United States. These regions include: Northeast (30.5%), South (29.6%), Midwest (18.5%), and West (21.4%), as defined by the 2000 U. S. Census. Region is coded as a set of dummy variables with Northeast as the reference category. A listing of the region each state is included in can be found in Appendix A, next to each state heading.

Analytic Technique

This research project takes a confirmatory, or hypothesis-testing, approach to analyzing a structural theory. There are two important aspects that I propose in this research. First, the causal process I project can be represented by a series of structural (regression) equations. Second, these structural relations can be pictorially modeled (the conceptual model) in order to give a clear conceptualization of the theory under investigation. In order to test all of the variables in my conceptual model simultaneously, I conduct my analyses using a process known as Structural Equation Modeling (SEM). By testing my entire model all together, I am able to determine which of the empirical relationships among variables are consistent with my conceptual model (Byrne 2006; Raykov and Marcoulides 2006). Even though my conceptual

model can be analyzed in its entirety, it is best done in different steps in order to analyze the hypothesized causal processes involved between the variables.

Based on my conceptual model, I theorize "causal" processes that generate observations on multiple variables. SEM provides a method for the quantification and testing of such probable theories. If SEM finds that goodness-of-fit is adequate, then it can be argued that the model is *plausible* for the hypothesized relations among the variables it contains. If goodness-of-fit is inadequate, then the plausibility of such relations should be rejected (Byrne 2006; Raykov and Marcoulides 2006).

There are several positive aspects to utilizing SEM rather than another form of statistical analysis, such as standard Logistic or Ordinary Least Squares (OLS) regressions. First of all, SEM takes on a confirmatory, rather than exploratory, approach to the analysis of data. By determining the pattern of inter-variable relationships, SEM lends itself to the investigation of data for inferential purposes, which allows for hypothesis testing. Second, SEM provides explicit estimates for measurement errors, where other traditional forms of multivariate analyses do not. Third, SEM provides a way of modeling multivariate relationships and for estimating direct and indirect effects among variables. Finally, SEM also allows for the use of dichotomous outcome variables, such as the ones I use for this dissertation (Byrne 2006; Raykov and Marcoulides 2006).

Model Analyses

By using a Structural Equation Modeling approach to estimate my conceptual model, it allows me to test each hypothesized mediating effect as well as assess the effects of multiple mediators simultaneously (MacKinnon 2008). It also allows me to determine how

sociodemographic, personal health, and region control variables affect various variables within the model. Figure 5.1 details my conceptual model listing the operationalized variables that are used for analyses, including sociodemographic, personal health, and regional control variables. The operationalized model shows how all the independent, dependent, and control variables are incorporated into the analysis of my original conceptual model.

After consultation with various experts, I decided on a statistical program known as M*plus* to test my conceptual model. M*plus* allows for a two-level (county- and individual-level in my case) path analysis within an SEM framework. It also includes a technique for analyzing complex survey data, which refers to data obtained by stratification, clustering, or sampling with an unequal chance of selection, such as the BRFSS data I used, as it is clustered by county within metropolitan areas. I also employed a technique in M*plus* that allows for a two-level path analysis with continuous and dichotomous independent and dependent variables, as my conceptual model includes both of these types of variables (Muthen and Muthen 2007). Additionally, M*plus* also allows for testing of direct and indirect effects with a hierarchical linear modeling technique.

In order to account for the complexity of my conceptual model, which includes a hierarchical data structure with both binary and continuous variables, clustering of data, and sampling with an unequal chance of selection, I employ the COMPLEX analysis feature in M*plus*. This analysis feature adjusts standard errors for non-independence by using a Taylor expansion of the Huber-White sandwich estimator. I also employ the weighted least square (WLSMV) parameter estimator, which uses a full diagonal weight matrix with standard errors and mean- and variance-adjusted chi-square test statistics in order to estimate my two-level regression equations. By employing the COMPLEX analysis feature and using the WLSMV

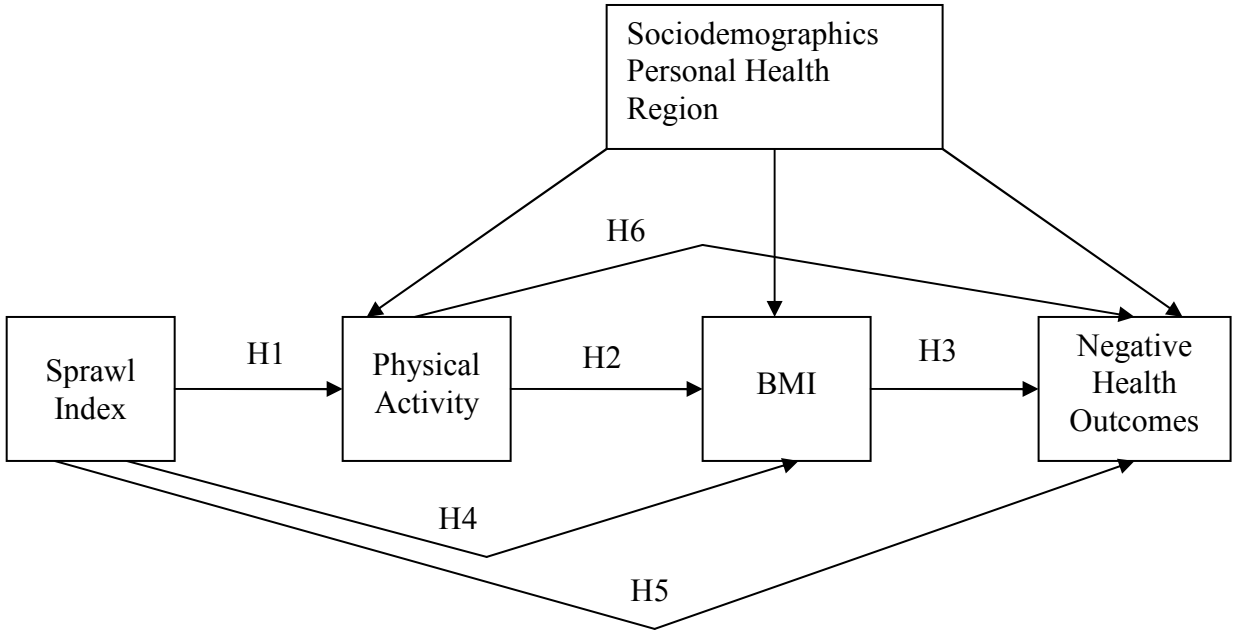

Figure 5.1 Analytical Model

parameter estimator, there is no need to use sample weights, as M*plus* takes into account any under- and/or overrepresentation of variables, such as in this dataset where income, gender, and race/ethnicity categories reveal some issues with representation (Muthen and Muthen 2007; Muthen and Satorra 1995; Perkins et al. 2010).

Even though SEM allows for the testing of all the variables in my model simultaneously, it does not make sense to analyze the entire model all at once because there will not be a "baseline" measurement to compare the associations each variable has on one another, especially when dealing with mediating variables. Therefore, I began with a baseline model before I tested my entire model at once. It should be noted why this method is important when dealing with mediating variables.

In general, a variable may function as a mediator if it accounts for any of the relationship between independent and dependent variables. Mediators explain how events take on

significance and speak to how or why certain effects occur. To help visually model how mediators affect independent and dependent variable relationships, the diagram in Figure 5.2 is referred to (Baron and Kenny 1986; Judd and Kenny 1981).

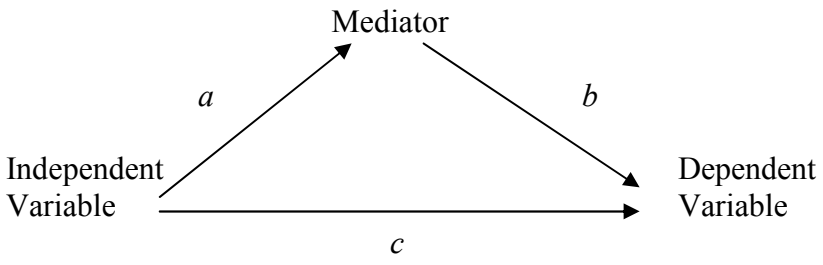

Figure 5.2 Independent, Dependent, and Mediating Variables
Source: Baron and Kenny (1986)

The causal chain depicted in Figure 5.2 involves a mediating variable – a variable that is both a dependent and an independent variable, with a mediating relationship from one variable to another (paths *a* and *b*). It consists of two causal paths feeding into a single outcome variable – the direct impact of the independent variable (path *c*) and the impact of the mediator (path *b*). There is also a path from the independent variable to the mediator (path *a*). If a variable functions as a mediator it will meet the following conditions: (1) variations in levels of the independent variable significantly account for variations in the mediating variable (path *a*); (2) variations in the mediator significantly account for variations in the dependent variable (path *b*); and (3) when paths *a* and *b* are controlled for, a previously significant relation between the independent and dependent variables drops in strength of association (path *c*). When this relationship drops to zero, then there is strong evidence for a single dominant mediating variable within the relationship between the independent and dependent variables. If the relationship

drops in strength of association but does not drop to zero, then partial mediation is indicated (Baron and Kenny 1986; Judd and Kenny 1981).

This latter situation, a relationship between an independent and dependent variable becoming less significant or non-significant, is the reason why it is important to test a model such as mine in different steps, beginning with a baseline model in which to compare additional models with mediating variables. This method allows me to have a comparative baseline in which to analyze my results. Below are the steps in which I analyze my conceptual model.

Baseline Model. Model one serves as the baseline model. It follows path H5, determining the strength and statistical significance of the relationship between sprawl and health outcomes with no other independent/mediating variables considered, but with the specified control variables in place. This first model provides a means in which to compare additional path models. When I build on this baseline model I am able to determine better just how the mediating variables in my conceptual model affect my final dependent variables (those of health outcomes). It is important to establish a baseline model in order to measure goodness of fit between it and the final model in order to determine if it is a better or worse fit for the variables included. If a statistically significant relationship is determined, then whether it (or how much of it) is "direct" or "indirect" can't be known until the mediating variables are included in the analysis, which leads to the second analytical model.

Model #2. The second, and final model, tests my entire conceptual model with all hypothesized paths, including all mediating and control variables, at once. This model is compared to the baseline model to determine goodness of fit and level of improvement.

By testing my conceptual model in two steps (models 1 and 2), I am able to determine which mediating variables have strong, significant effects on the relationship between sprawl and

health outcomes. An SEM approach to analyzing my conceptual model allows me to test each hypothesized mediating effect, as well as to assess the effects of multiple mediators along with sociodemographic, personal health, and regional control variables, simultaneously (MacKinnon 2008).

Similarities and Difference between this Study and Ewing et al. Study

This study is intended to be an extension of and improvement upon the Ewing et al. (2003) and McCann and Ewing (2003) studies. My research is differentiated from these studies in several ways. The Ewing et al. and McCann and Ewing studies focus on HLM regressions to determine relationships between the different variables utilized. My method of analysis did employ HLM regressions, but it was completed within an SEM framework analyzing a path model. This analytical framework allowed me to look at direct and indirect effects on the various outcome and mediating variables. Ewing et al. diagram a model towards the end of their article linking the variables in their research, and this diagram is very similar to my conceptual model. However, I believe my study improves on their study by actually testing a model that depicts the relationship between all the variables involved.

Other ways in which my research differs from Ewing et al. (2003) is the fact that I utilize total minutes of physical activity per week for each individual, whereas Ewing et al. use minutes walked per month, whether or not individuals received any physical activity within the last month, and whether or not individuals received the recommended amount of physical activity within the last month. I employ additional outcome variables than that of Ewing et al. In addition to diabetes, coronary heart disease, and hypertension, I also look at the outcome variables of high cholesterol, heart attack, and stroke. Finally, although I am using the same

sample data source as Ewing et al. and McCann and Ewing, they use BRFSS data from 1998 to 2000, whereas I utilized 2003 data.

Limitations

My study is not without limitations. These pertain to issues related to measurement and to using cross-sectional and secondary data. The first measurement issue concerns the BMI variable I employed in order to determine weight gain. While the BRFSS data utilized in this study may be fairly reliable, the validity of height and weight reported by participants has been questioned. Critics state that many people tend to report being taller and thinner than they in fact are (Lopez 2004; Bowlin et al. 1993; Jackson et al. 1992). This common misreporting with the BRFSS survey does not mean the data are completely flawed. Every survey has shortcomings and is plagued with the fact that some people will not be honest (Rowland 1990). However, I have confidence that the individual height and weight reported in the BRFSS dataset, as well as the calculated BMI, provide a respectable reflection of the population overall. In fact, measures of BRFSS data have been determined to have high reliability and validity, including height and weight (Nelson et al. 2001; Nelson et al. 2003). Respected researchers such as Ewing et al. (2003) and McCann and Ewing (2003) see fit to utilize BMI in their research.

The other measurement limitation encountered in my dissertation is the sprawl index variable. Even though the county-level sprawl index provides a finer level of detail in studying sprawl as it relates to health at the individual level, counties in metropolitan areas can vary greatly in urban form just as metropolitan areas do. Just as studies such as Lopez's (2004) do not reflect how sprawl varies within metropolitan areas, my research does not reflect how sprawl varies within a single county. Metropolitan areas are not homogenous in their development patterns across their geographical span, and neither are counties, with the exception of New

York, NY (Manhattan). Even though I controlled for the way metropolitan areas vary from inner city to suburban locales by utilizing a county-level sprawl index, there is still the issue of how counties differ from place to place.

Another limitation with my study is the fact that the data represent respondents at only one point in time; therefore, there is no way of measuring whether or not respondents became overweight or obese while residing in the metropolitan area or county in which they were surveyed. It may be the case that some became overweight while residing in a dense metropolitan area or county then moved to a sprawling metropolitan area or county before participating in the survey, or vice versa. There is no way of determining the length of time an individuals has lived in their current county.

The final limitation to my study is the fact that it utilized secondary data. Secondary data sources are readily available and easily accessible, but they impose limitations, especially as it relates to the formulation of problems and concepts. This limitation is due to the fact that measurement can only be based on existing data (Baunach 2001; Sund 2003). As with my research project, I was forced to formulate my hypotheses and conceptual and analytical models based on data that already existed. This does not reduce the validity or importance of my study, but it does limit it in some ways. Since I obtained some statistically significant results, I am now able to move forward with new, more in-depth studies with data specifically geared towards testing additional concepts and hypotheses. Therefore, even though there are limitations to using secondary data, secondary data sources can produce well-crafted, thoughtful, important research, as long as the limitations imposed are clearly understood.

Findings

The following two chapters discuss the findings from my analytical models, including how well my analyses support or do not support my original hypotheses. The two chapters are organized by analytical model and discuss how each model affects the various outcome variables included in this study. For the second model, which includes indirect effects, those indirect effects and how the mediating variables affect the model as a whole are discussed. Each results chapter also includes a discussion of how well the variables in each model fit the model as a whole by looking at various fit statistics.

I organize my results chapters this way because the main part of my dissertation concerns the conceptual model that I developed. It makes sense to organize the following chapters according to the way I conduct analyses of my conceptual model in the two different analytical models, rather than to organize it by final outcome variable, especially since my conceptual model and corresponding two analytical models include various mediating variables. I organize my discussion of health outcome variables by control variables, instead of by dependent variables. This way the reader is able to see how certain variables, such as age, education, and income, affect a multitude of outcome variables like the six included in this dissertation.

I then conclude my dissertation with a discussion of how well my analytical models fit, or do not fit, my original hypotheses. It then moves on to discuss the associations between the various variables in my conceptual model, followed by a discussion of which analytical model fits my conceptual framework the best. I also propose a new conceptual model from the results of my analyses. Limitations of my study, contributions my research provides to the literature, implications my research will have at an applied level, as well as possibilities for future research, are also discussed in the final chapter.

CHAPTER 6

SPRAWL AND NEGATIVE HEALTH OUTCOMES

Health outcomes can be the result of several key factors. To determine what variables most affect certain health outcomes and in order to determine mediating effects set forth in my conceptual model, different paths were analyzed in different steps with various outcome variables, two of which act as mediating variables – physical activity and BMI. This and the following chapter outlines the different paths and statistical models and focus on the variables most statistically significant in determining health outcomes for diabetes, heart attacks, heart disease, high cholesterol, hypertension, and stroke. The following chapter also focuses on those variables most significant in mediating the effect of sprawl on these health outcomes.

As stated previously, there are two analytical models. The baseline model examined in this chapter measures the strength and statistical significance of the relationship between sprawl and health outcomes with no mediating variables considered, just with the specific control variables in place. If the relationship in this baseline model turns out to be statistically significant, then whether it (or how much of it) is a direct or indirect effect cannot be determined until the mediating variables are included in the second analysis. This is where the next model comes into play. The second model, discussed in Chapter 7, looks at how sprawl affects health outcomes with both mediating variables – physical activity and BMI – and with all hypothesized paths analyzed simultaneously. The second model will be explored based on comparison to the baseline model.

The baseline model will now be discussed. A pictorial description of this path model is depicted in Figure 6.1. The following sections discuss how well the independent and control

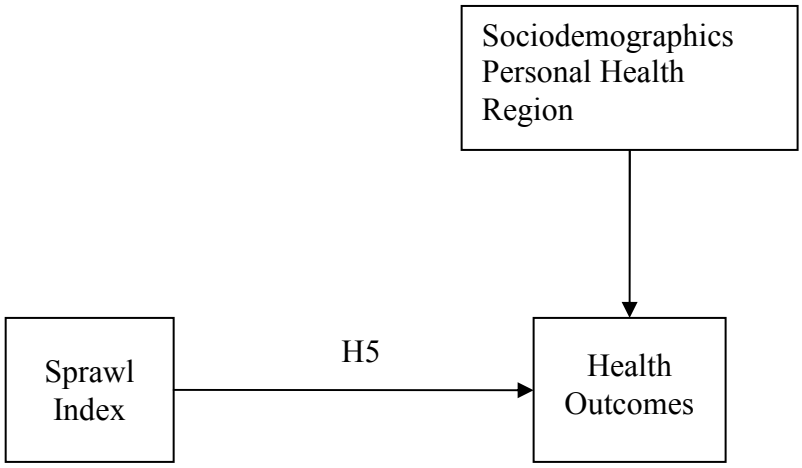

Figure 6.1 Baseline Analytical Path Model

variables set forth in this model predict each of my dependent health outcome variables, which include diabetes, heart attacks, heart disease, high cholesterol, hypertension, and stroke.

Prediction of Negative Health Outcomes with Sprawl

As my fifth hypothesis states**,** sprawl is positively related to negative health outcomes, and as sprawl increases, so does the chance for individual-level negative health outcomes, such as diabetes, heart attacks, heart disease, high cholesterol, hypertension, and stroke. This baseline analytical model is tested to look at the zero-order correlation between sprawl and negative health outcomes. As sprawl increases I expect a positive relationship between sprawl and health problems to be revealed, especially when controlling for other variables such as sociodemographics, personal health, and region. Preliminarily testing of the effect of sprawl on negative health outcomes by means of cross-tabulations reveled highly significant results, with p-values (Pearson's Chi Square) equal to 0.000 on all dependent variables, except for stroke where the crosstab returns a p-value of 0.001. When examining correlations (see Table 6.1),

sprawl has a significant correlation with heart attacks (0.017), and hypertension (0.010). These tests suggest there is a very weak, significant association between sprawl and negative health outcomes, but is it a direct effect?

It is quite possible no direct effect exists between sprawl and negative health outcomes, but to be certain I test this hypothesis in order to determine how much variance in the individual-level variables of health outcomes can be explained by the macro-level variable of sprawl, while controlling for other factors. Additionally, this model gives me a baseline to determine if variables such as physical activity and BMI are in fact mediating the effect of sprawl on health outcomes. As it turns out, results from analyses show there is almost no correlation between sprawl and health outcomes. This is somewhat surprising considering what other studies have shown.

Table 6.1 also lists the unstandardized coefficients for sprawl's zero-order relationship on negative health outcomes, without any control variables. As you can see, four health outcome variables show statistically significant results when it comes to sprawl as an independent variable – heart attack, heart disease, hypertension, and stroke. In fact, sprawl has a highly significant, albeit very tiny, effect for individuals who have had a heart attack ($p < 0.001$). Even though sprawl returns statistically significant results on four of the health outcome variables, it may not necessarily hold true when control variables are added to the statistical models. Additionally, R-Square statistics, which is the percentage of linear variance in the dependent variable that is explained by the independent variables, are not very impressive when predicting health outcomes with sprawl as the only independent variable. All fit statistics, including R-Square, CFI, and RMSEA, will be discussed in greater detail later in this chapter.

Table 6.1

Zero-order Relationship between Sprawl and Negative Health Outcomes

Dependent Variable	Correlations	Unstandardized Coefficient	CFI	RMSEA	R-Square	N
Diabetes	-0.001	-0.001	1.000	0.000	0.000	101,714
Heart Attack	0.017**	0.012***	1.000	0.000	0.002	36,403
Heart Disease	0.010	0.005*	1.000	0.000	0.000	36,297
High Cholesterol	0.002	0.001	1.000	0.000	0.000	83,000
Hypertension	0.010**	0.005*	1.000	0.000	0.000	101,639
Stroke	0.010	0.007*	1.000	0.000	0.001	36,401

*p < 0.050, **p < 0.010, ***p < 0.001

Other factors are important in predicting health outcomes; therefore, it is important to test more variables. Aside from testing a sprawl index as an independent variable to see the effects among these health outcomes, I also employed sociodemographic, personal health, and regional control variables in hierarchical regression models in order to assess the relative influence of sprawl on these negative health outcomes. Table 6.2 presents the results of the hierarchical regressions for the baseline model with sociodemographic, personal health, and regional control variables. As Tables 6.1 and 6.2 indicate, sprawl's effects on health outcomes change when adding control variables.

Initially sprawl does not have a significant effect on high cholesterol when it is the only independent variable. However, with the addition of control variables, sprawl becomes a significant predictor of this variable. According to the baseline model (Table 6.2), each one point increase in the sprawl index as utilized in these regressions (which is the original McCann and Ewing sprawl index reversed and divided by 10) decreases the log odds of someone having high cholesterol by 0.005. That translates to the odds of someone having high cholesterol changing by a factor of 0.995 ($e^b = 2.7183^{-0.005}$) with each additional one point increase in the sprawl index as analyzed, controlling for all the other variables in the model. In other words, the

Table 6.2

Baseline Model SEM Regression Results for Negative Health Outcomes

	Diabetes	Heart Attack	Heart Disease	High Cholesterol	Hypertension	Stroke
Sprawl	0.004	0.001	-0.002	-0.005**	0.005**	0.006
Age	0.025*** †	0.031*** †	0.033*** †	0.021*** †	0.035*** †	0.024*** †
Education	-0.069***	-0.053*	-0.029	-0.054***	-0.073***	-0.045
Income	-0.076***	-0.080***	-0.062***	-0.018***	-0.043***	-0.106***
Gender	-0.151***	-0.365***	-0.243***	-0.125***	-0.180***	-0.194**
Black	0.383***	-0.078	-0.087	-0.113***	0.365***	0.083
Asian	0.073	-0.410**	-0.152	-0.096**	-0.063	-0.215
Native Hawaiian/Pacific Islander	0.401	-0.396	-0.004	-0.116	-0.006	0.309
American Indian	0.352***	0.157	0.460**	0.122*	0.108	0.317
Multi-Racial	0.220***	0.230*	-0.014	-0.010	0.234***	0.366**
Other	0.031	0.321*	0.049	-0.126	-0.034	0.051
Hispanic	0.109**	-0.155	0.125	-0.130***	-0.080**	0.014
Widowed	-0.171***	-0.209**	-0.222**	-0.161***	0.003	-0.000
Separated/Divorced	-0.039	-0.109	-0.123*	-0.005	0.057**	0.140*
Never Married	-0.048	-0.251***	-0.255***	-0.044*	0.045*	-0.117
Married * Gender	-0.070*	-0.187*	-0.236***	-0.098***	-0.017	0.016
Increased Physical Activity	-0.042**	-0.066	-0.006	-0.051***	-0.057***	-0.099*
South	0.035	0.150**	0.127*	0.025	0.048**	0.180*
Midwest	0.026	0.073	0.060	-0.004	0.014	0.079
West	0.013	-0.068	-0.158	-0.030	-0.030	0.062
N	68,402	24,817	24,771	57,275	68,369	24,814
R-Square	0.185	0.276	0.261	0.114	0.274	0.211
CFI	1.000	1.000	1.000	1.000	1.000	1.000
RMSEA	0.000	0.000	0.000	0.000	0.000	0.000

*p < 0.050, **p < 0.010, ***p < 0.001

† = moderate variable association (standardized regression coefficient absolute value is between 0.3 and 0.6)

odds of an individual in the sample of having high cholesterol decreases by approximately 0.5% (e^b -1 = $2.7183^{-0.005}$ – 1) for each one point increase in the sprawl index. That means individuals in the most sprawling county in this study, Yadkin, NC, are 14.15% ([282.9 x 0.005] / 10) (keeping in mind that this formula is divided by 10 to keep it factually related to the way the sprawl index was utilized in analyses) less likely to have high cholesterol than those in the least sprawling county, New York, NY (Manhattan). This result is the opposite of what was expected.

In contrast, when predicting hypertension, each one point increase in the sprawl index increases the log odds of someone having high blood pressure by 0.005. That translates to the odds of someone having high blood pressure changing by a factor of 1.01 (e^b = $2.7183^{0.005}$) with each additional one point increase in the sprawl index, controlling for all the other variables in the model. In other words, the odds of an individual in the sample having high blood pressure increases approximately 0.5% (e^b -1 = $2.7183^{0.005}$ – 1) for each one point increase in the sprawl index. That means individuals in the most sprawling county in this study, Yadkin, NC, are 14.15% ([282.9 x .005] / 10) more likely to have hypertension than those in the least sprawling county, New York, NY (Manhattan).

Initially, when predicting heart attacks, sprawl returns highly significant, but small results (Table 6.1). However, this significance disappears when controlling for other factors (Table 6.2). This is most likely due to the fact that certain control variables are better predictors of heart attacks than sprawl is. This could mean sprawl has an effect on certain aspects of the lives of individuals residing in sprawling areas, such as their propensity to exercise and/or their overall weight. This is a topic that is covered in detail in the next chapter. For the other two health outcome variables, heart disease and stroke, sprawl also begins with significant results, but loses its significance when control variables are added, suggesting there might be mediating effects.

When it comes to the strength of sprawl as an independent variable, the results indicate a very weak association with the outcome variables. Standardized regression coefficients (shown in Appendix C) for sprawl remain below 0.020, indicating almost no significant association (Healy 2006; Menard 1995). Actually, there are no variables within this analytical model that return standardized regression coefficients that rank within the strong category for association (0.7 to 1). In fact, only one variable (age) returns a moderate relationship with the outcome variables, which leads me to also interpret the relative strength for the variables in this analytical model. Even in terms of relative strength, however, sprawl continually falls towards the bottom half of all variables in terms of association with the outcome variables, except for high cholesterol, where it ranks towards the middle in terms of relative strength, suggesting a moderate relationship with this outcome variable.

Age, Education, Income, and Gender as Predictors of Negative Health Outcomes

When adding demographic, personal health, and regional control variables to create my baseline model (Table 6.2), there are quite a few variables that return significant results in predicting health outcomes. Age is significant on all six health outcome variables, indicating that for every year increase in age, individuals within the sample have a greater chance of becoming diabetic, having a heart attack, developing heart disease, having high cholesterol, developing hypertension, and having a stroke. This is not a surprising result, as studies have shown that negative health outcomes tend to increase with age (McCann and Ewing 2003; Ewing et al. 2003; Egede and Zheng 2002; Frank and Engelke 2001; Giles-Corti and Donovan 2002).

According to the regression models (Table 6.2), each one year increase in age increases the log odds of someone having any one of these health conditions between 0.021 and 0.035,

depending on which specific outcome is observed. That translates to the odds of someone having any one of these health conditions changing by a factor of 1.021 to 1.036 with each additional one year increase in age, controlling for all the other variables in the model. In other words, the odds of an individual in the sample having one of these health outcomes increase between 2.12% and 3.56% for each one year increase in age.

Age is also the variable that returns standardized coefficients ranking it as moderate in terms of its strength of association with the outcome variables. Age consistently ranks as the strongest variable across all regressions in this analytical model when it comes to standardized regression coefficients, indicating it is the strongest variable in the prediction of these six health outcomes (Healy 2006; Menard 1995).

Education has a significant effect on diabetes, heart attacks, high cholesterol, and hypertension. With more education, an individual's chance of having diabetes, a heart attack, high cholesterol, or hypertension decreases. This finding coincides with previous research that shows those with higher educational attainment are more likely to be healthier and suffer less from negative health outcomes (McCann and Ewing 2003; Ewing et al. 2003; Giles-Corti and Donovan 2002; Ford et al. 1991; Jeffery et al. 1991).

According to the baseline regression model, each one categorical increase in education (such as going from high school graduate to some college or technical school) decreases the log odds of someone having diabetes, heart attack, high cholesterol, or hypertension between 0.053 and 0.073, depending on which specific outcome is observed. That translates to the odds of someone having any one of these health conditions changing by a factor of 0.948 to 0.930 with each additional categorical increase in education, controlling for all other variables in the model.

In other words, the odds of an individual in the sample having one of these health outcomes decrease between 5.16% and 7.04% for each additional categorical increase in education.

Even though education does not return standardized regression coefficients ranking it in the strong or even moderate category for association with the dependent variables, it usually ranks in the top one-third as far as strength goes, except for its association with heart attack and heart disease, where it ranks towards the middle in relative strength (Healy 2006; Menard 1995).

Income is significant on all six health outcome variables. In all cases, individuals with higher incomes have less chance of being afflicted with these negative health outcomes. According to the regression models, for each one categorical increase in income (such as going from less than $10,000 per year to $10,000 to $14,999 per year) decreases the log odds of someone having any one of these health outcomes between 0.018 and 0.106, depending on which specific outcome is observed. That translates to the odds of someone having any one of these health conditions changing by a factor of 0.982 to 0.899 with each additional categorical increase in income, controlling for all the other variables in the model. In other words, the odds of an individual in the sample having one of these health outcomes decrease between 1.78% and 10.06% for each one categorical increase in income. Even though income standardized regression coefficients do not make it above 0.150, it still ranks in the top one-third of all variables in every regression model, making it one of the strongest variables in this analytical model as far as its association with the outcome variables in terms of relative strength (Healy 2006; Menard 1995).

Income seems to have a greater effect on health outcomes than does education. Income may have a greater effect than education due to the fact that those with higher incomes have better access to medical attention than those with lower incomes. This is not surprising since

research continually shows a clear association between negative health outcomes and low income. This is due in part to the fact that low income people have less access to health care coverage, as well as less access to quality healthcare (Day 2006; Strunk and Reschovsky 2004; Subramanian and Kawachi 2006; Gorin and Moniz 2004; Eberstadt and Satel 2004).

Not surprisingly, gender has a significant effect on all seven health outcome variables. Compared to males, females have less chance of developing diabetes, having a heart attack, heart disease, high cholesterol, hypertension, or stroke. According to the regression models, being female decreases the log odds of having one of these negative health outcomes between 0.125 and 0.365, depending on which specific outcome is observed. That translates to the odds of a female having any one of these health conditions changing by a factor of 0.882 to 0.694, controlling for all the other variables in the model. In other words, the odds of a female in the sample having one of these health conditions decrease between 11.75% and 30.58%.

From the percentage changes listed above, it is not surprising that gender is one of the top three variables in relative strength in its association with the outcome variables. Even though gender does not return standardized regression coefficients ranking it in the strong or even moderate category for association with the dependent variables, relatively it is one of the stronger variables in this analytical model (Healy 2006; Menard 1995).

Race and Ethnicity as Predictors of Negative Health Outcomes

When it comes to race and ethnicity, the results are not surprising, as previous research shows race is a significant predictor for many health outcomes (Table 6.2) (Day 2006; McCann and Ewing 2003; Egede and Zheng 2002; Mokdad et al. 1999). Compared to Whites, those who are Black have a greater risk for developing diabetes (log odds of 0.383, change factor of 1.467,

an increase of 46.67%) and having hypertension (log odds of 0.365, change factor of 1.441, an increase of 44.05%). However, Blacks are less likely to have high cholesterol (log odds of -0.113, change factor of 0.893, a decrease of 10.68%).

Compared to Whites, Asians are significantly less likely to have a heart attack (log odds of -0.410, change factor of 0.664, a decrease of 33.63%) or high cholesterol (log odds of -0.096, change factor of 0.908, a decrease of 9.15%). American Indians have a greater chance of having diabetes (log odds of 0.352, change factor of 1.422, an increase of 42.19%), heart disease (log odds of 0.460, change factor of 1.584, an increase of 58.41%), and high cholesterol (log odds of 0.122, change factor of 1.130, an increase of 12.98%).

Compared to Whites, those who are Multi-racial have a greater chance of having diabetes (log odds of 0.220, change factor of 1.246, an increase of 24.61%), heart attacks (log odds of 0.230, change factor of 1.259, an increase of 25.86%), hypertension (log odds of 0.234, change factor of 1.264, an increase of 26.36%), and stroke (log odds of 0.366, change factor of 1.442, an increase of 44.20%). Individuals of the Other racial category are more likely to have a heart attack (log odds of 0.321, change factor of 1.379, an increase of 37.85%).

Compared to Whites, Hispanics have a greater chance of developing diabetes (log odds of 0.109, change factor of 1.115, an increase of 11.52%), but are less likely to have high cholesterol (log odds of -0.130, change factor of 0.878, a decrease of 12.19%) or hypertension (log odds of -0.080, change factor of 0.923, a decrease of 7.69%). Native Hawaiians and Pacific Islanders are not statistically more likely to have any of these six health outcomes compared to those who are White.

The strength of association to each outcome variables varies by racial and ethnic category, as well as outcome variable. Standardized regression coefficients for race consistently

run 0.102 or much lower, making this set of dummy variables weak as far as standardized strength is concerned. Relative to all the other variables in the models, race and ethnicity is still a rather weak control variable. Age, income, education, and gender are much stronger indicators for these negative health outcomes than is race and ethnicity (Healy 2006; Menard 1995).

Marital Status as Predictors of Negative Health Outcomes

It is not surprising that marital status plays a significant role in the prediction of health outcomes (Table 6.2) (Umberson 1992; Gove 1973; Stroebe and Stroebe 1983). Compared to those who are married, the reference category for this set of dummy variables, those who are widowed are less likely to have diabetes (log odds of -0.171, change factor of 0.843, a decrease of 15.72%), a heart attack (log odds of -0.209, change factor of 0.811, a decrease of 18.86%), heart disease (log odds of -0.222, change factor of 0.801, a decrease of 19.91%), or high cholesterol (log odds of -0.161, change factor of 0.851, a decrease of 14.87%).

Those who are separated or divorced are less likely to have a heart attack (log odds of -0.123, change factor of 0.884, a decrease of 11.57%), but have a higher chance of having hypertension (log odds of 0.057, change factor of 1.059, an increase of 5.87%) and stroke (log odds of 0.140, change factor of 1.150, an increase of 15.03%). Those who have never married are less likely to have a heart attack (log odds of -0.251, change factor of 0.778, a decrease of 22.20%), heart disease (log odds of -0.255, change factor of 0.775, a decrease of 22.51%), or high cholesterol (log odds of -0.044, change factor of 0.957, a decrease of 4.30%), but are more likely to have hypertension (log odds of 0.045, change factor of 1.046, an increase of 4.6%).

The marriage/gender interaction variable, which focuses on the relationship between marital status and health outcomes by gender, shows that married females are less likely to have

diabetes (log odds of -0.070, change factor of 0.932, a decrease of 6.76%), heart attacks (log odds of -0.187, change factor of 0.829, a decrease of 17.06%), heart disease (log odds of -0.236, change factor of 0.790, a decrease of 21.02%), and high cholesterol (log odds of -0.098, change factor of 0.907, a decrease of 9.34%). Therefore, the effect of being married is statistically different for females.

Standardized regression coefficients for marital status consistently run 0.081 or much lower, making this set of dummy variables weak as far as standardized strength is concerned. The strength of marital status as a predictor variable for these health outcomes also runs the gamut in relative strength, much like that of race and ethnicity. Relative to all the other variables in the models, marital status is still a rather weak control variable. Again, age, income, education, and gender are much stronger indicators for these negative health outcomes than is marital status (Healy 2006; Menard 1995).

Personal Health as Predictors of Negative Health Outcomes

Individuals who consciously increase their physical activity in order to lose or maintain weight have less chance of having diabetes, high cholesterol, hypertension, or stroke. This finding is to be expected because physical activity has been shown to have beneficial health outcomes (Perdue 2004; McCann and Ewing 2003; Ewing et al.2003; Ewing et al. 2002a).

According to the regression models (Table 6.2), individuals who consciously increase their physical activity in order to lose or maintain weight decrease their log odds of having diabetes, high cholesterol, hypertension, and stroke between 0.042 and 0.099, depending on which specific outcome is observed. That translates to the odds of an individual who consciously increased their physical activity to lose or maintain weight having any one of these

four health conditions changing by a factor of 0.959 to 0.906, controlling for all the other variables in the model. In other words, the odds of an individual who consciously increase their physical activity decrease the chance of having one of these health conditions between 4.11% and 9.43%.

When it comes to the strength of this personal health control variable, it does not return very strong associations with the outcome variables. In fact, the standardized regression coefficients for this variable remain below 0.040, indicating a weak association (Healy 2006; Menard 1995). Even in terms of relative strength, increase in physical activity continually falls towards the bottom two-thirds of all variables in terms of association with each outcome variable.

Region as a Predictor of Negative Health Outcomes

When it comes to region, only the southern part of the nation stands out as having significantly different health outcomes (Table 6.2). Individuals who reside in the South are more likely to have heart attacks, heart disease, hypertension, and stroke than those in the Northeast, the reference category for this set of dummy variables. According to the regression models, individuals living in the South increase their log odds of having heart attacks, heart disease, hypertension, and stroke between 0.048 and 0.180, depending on which specific outcome is observed. That translates to the odds of 1.049 to 1.197. In other words, the odds that Southerners will have one of these five health conditions is 4.92% to 19.72% more likely, controlling for all the other variables in the model.

When it comes to the strength of region as a control variable, it does not return very strong associations with the outcome variables. In fact, standardized regression coefficients for

this set of dummy variables remain at or below 0.080, indicating a weak association (Healy 2006; Menard 1995). Even in terms of relative strength, region usually falls towards the bottom two-thirds of all variables in terms of association with the outcome variables. The few exceptions to this are when the variable representing Southerners falls in the top third of all variables in regressions where it returns significance results, which include heart attack, heart disease, hypertension, and stroke. It is obvious other control variables, such as age, education, income, and gender, are stronger predictors of the health outcomes included in this research project.

Fit Statistics

R-square, also called Multiple Correlation or the Coefficient of Multiple Determination, is an important fit statistic when using regression models to make accurate predictions concerning independent and dependent variables. It is the percentage of linear variation in the dependent variable that is explained by the independent variables. In other words, it is the proportionate reduction in error when estimating the dependent variables, knowing the independent variables. It reflects the number of errors made when using the regression model to estimate the value of the outcome variable in relation to the total errors made when using only the dependent variable's mean as the basis for estimating all cases. Ideally, explaining most, if not all, of the original variability, is preferred when looking at R-square. The closer to 100% variance explained the better. Of course, an increase in R-square is preferred when adding additional variables or changing the structure of a model, such as I have done for this research project (Garson 2010; StatSoft 2010).

As mentioned earlier, the R-square statistics in the zero-order relationship with sprawl as the only independent variable are very low for all of the health outcome variables (Table 6.1). However, as Table 6.2 shows, when adding control variables the variance explained does increase. For the outcome variable of diabetes, the variance explained increases to 18.5%, for heart attack it increases to 27.6%, for heart disease it increases to 26.1%, for high cholesterol it increases to 11.4%, for hypertension it increases to 27.4%, and for stroke it increases to 21.1%, making for significant improvements in the variance explained.

All of these R-Square statistics are a vast improvement over the ones obtained from the zero-order model with sprawl as the only independent variable, which means these other control variables are explaining more in the prediction of these negative health outcomes than sprawl alone. Such an increase in the variance explained for each health outcome points to variables that fit the models better and account for the multi-dimensional nature of predicting health outcomes. However, model fit cannot be determined by looking at variance alone. In addition to assessing the R-square statistic, I also look at other fit statistics in order to determine model fit.

Many researchers look to the chi-square statistic in order to determine model fit. Even though the chi-square statistic is a global test of a model's ability to reproduce the sample variance and covariance matrix, it is sensitive to sample size. Since my sample sizes are so large, it is important to use fit statistics that are not sensitive to such issues with size. Since there is no one statistic universally accepted as an index of model adequacy, Wheaton (1987) suggests using multiple fit statistics in order to determine model adequacy. Therefore, the understanding of the results from my models relies on the interpretation of not only the R-square statistic, but also two other model fit statistics. For this research project, I employ Root Mean Square Error of

Approximation (RMSEA) and Confirmatory Fit Index (CFI) in order to determine model fit (Table 6.2).

RMSEA measures the amount of model discrepancy per degree of freedom. When RMSEA equals zero, it indicates a perfect fit, values less than 0.05 are considered an indication of good fit, whereas values between 0.05 and 0.08 indicate a reasonable fit. Brown and Cudeck do not recommend using models with RMSEA values greater than 0.10 (Hu and Bentler 1999; Marcoulides and Schumacker 2001; Rigdon 1996; Brown and Cudeck 1993; Doll et al. 2004). CFI estimates each non-centrality parameter by the difference between its *t*-statistic and corresponding degrees of freedom, with values greater than 0.90 indicating a good fit and 1.000 indicating a perfect fit (Hu and Bentler 1999; Marcoulides and Schumacker 2001; Rigdon 1996; Brown and Cudeck 1993; Doll et al. 2004; Tasmin and Woods 2008).

Both fit indexes are conceptually linked to the non-central chi-square distribution. Even though CFI has been used in applied research and is appropriate for use in confirmatory contexts such as this research project, RMSEA is gaining more attention and credibility as a fit index (Hu and Bentler 1999; Marcoulides and Schumacker 2001; Rigdon 1996). I think it is important to incorporate both indexes as a test of model fit and to take both into consideration, as well as the R-square statistic, when determining the overall adequacy of each statistical model. All six statistical models in this baseline model return values of 1.000 for CFI and 0.000 for RMSEA, both of which state an excellent model fit.

Summary

To recap, for the baseline model sprawl has almost no correlation with health outcome variables and is very weak in its association with the two variables it returns a significance effect

on (high cholesterol and hypertension). Control variables such as age, education, income, gender, and marital status are better predictors of health outcomes. Additionally, increasing physical activity in order to control weight and the effect of residing in the South are also significant predictors for certain outcome variables, however neither of these two variables are the strongest predictors of health outcomes. Age, education, income, and gender are the strongest variable relatively speaking when it comes to health outcomes.

CHAPTER 7

SPRAWL AND NEGATIVE HEALTH OUTCOMES WITH PHYSICAL ACTIVITY AND BMI AS MEDIATING VARIABLES INCLUDING ALL HYPOTHESIZED PATHS

The previous chapter looked at the baseline model, examining the strength, statistical significance, and effect of the relationship between sprawl and health outcomes with no mediating variables, only specific control variables. This chapter looks at the relationship between sprawl and health outcomes with both physical activity and BMI as mediating variables, as well as the specified control variables, and compares results to the baseline model. This mediating model is my second and final analytical model, which follows all hypothesized paths. A pictorial description of this path model is depicted in Figure 7.1.

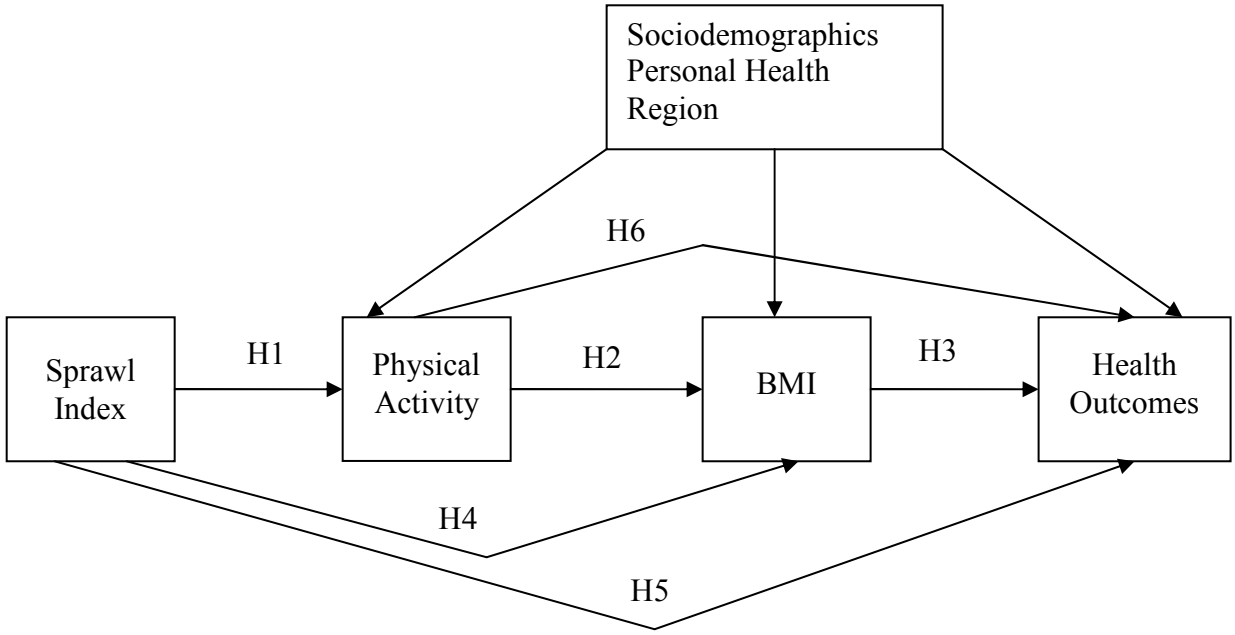

Figure 7.1 Analytical Path Model Two

Prediction of Physical Activity

Besides physical activity being a mediating variable between sprawl and health outcomes and between sprawl and BMI in this analytical model, physical activity is also a dependent variable with sprawl and sociodemographic, personal health, and regional control variables having a direct relationship to it. In order to see just how sprawl affects physical activity, Table 7.1 presents the results of the statistical regressions for this analytical model, including the outcome variable of physical activity.

When predicting physical activity, sprawl has a highly significant effect (Table 7.1). According to the model, for each one point increase on the sprawl scale, individuals in the sample increase their weekly physical activity by 0.009 hours. That translates to individuals in the most sprawling county in this study, Yadkin, NC, getting 0.25 (282.9 x .009 / 10) additional hours (or 15 additional minutes) of physical activity per week than individuals in the least sprawling county in this study, New York, NY (Manhattan). This is not the result expected. It was hypothesized that sprawl has a negative relationship to physical activity, and that as sprawl increases, individual levels of physical activity decrease. However, this hierarchical regression model points to the fact that sprawl has a positive relationship to physical activity.

Even though sprawl is highly significant in the prediction of the amount of physical activity individuals get per week, it returns a rather weak effect. Furthermore, it is not the only significant variable. When adding demographic variables, age, education, income, gender, and race are all highly significant when predicting levels of physical activity. For every year increase in age, individuals obtain less physical activity. So the older an individual becomes the less physical activity he/she gets each week. As for education, the more education an individual has

Table 7.1

Model 2 SEM Regression Results for Negative Health Outcomes

	Physical Activity	BMI	Diabetes	Heart Attack	Heart Disease	High Cholesterol	Hypertension	Stroke
Sprawl	0.009*	0.051***	0.002	0.001	-0.003	-0.006***	0.003	0.005
Physical Activity	------	-0.214***	-0.050***	-0.014	-0.026**	-0.018***	-0.014***	-0.041***
BMI	------	------	0.057***	0.024***	0.025***	0.028***	0.056***	0.016***
Age	-0.006***	0.019***	0.023*** †	0.030*** †	0.032*** †	0.020*** †	0.034*** †	0.024*** †
Education	-0.054***	-0.378***	-0.043***	-0.042	-0.018	-0.041***	-0.046***	-0.038
Income	0.034***	-0.113***	-0.062***	-0.074***	-0.055***	-0.011**	-0.030***	-0.100***
Gender	-0.276***	-0.766***	-0.105***	-0.341***	-0.218***	-0.100***	-0.124***	-0.197**
Black	-0.086***	2.163***	0.240***	-0.139**	-0.150**	-0.182***	0.227***	0.039
Asian	-0.186**	-2.109***	0.192***	-0.365**	-0.095	-0.038	0.054	-0.186
Native Hawaiian/Pacific Islander	0.018	1.006	0.403	-0.300	-0.008	-0.136	-0.024	0.335*
American Indian	0.099	0.510*	0.312***	0.121	0.436*	0.114	0.067	0.301
Multi-Racial	0.170***	0.868***	0.163**	0.221	-0.030	-0.036	0.178***	0.382***
Other	0.060	0.152	0.017	0.314*	0.040	-0.126	-0.047	0.057
Hispanic	-0.064*	0.496***	0.086*	-0.174	0.111	-0.142***	-0.102***	0.010
Widowed	-0.030	-1.294***	-0.081**	-0.077	-0.078	-0.131***	0.001	-0.137*
Separated/Divorced	-0.023	0.267**	0.024	0.105	0.118*	-0.003	-0.068**	-0.146*
Never Married	-0.020	-0.672***	0.003	-0.138*	-0.133	-0.033	0.001	-0.250**
Married * Gender	-0.065***	-0.661***	-0.035	-0.177*	-0.227***	-0.081***	0.018	0.027
Increased Physical Activity	0.461***	-0.379***	0.004	-0.050	0.014	-0.031*	-0.029**	-0.079
South	-0.008	0.017	0.030	0.145*	0.122*	0.022	0.043*	0.173*
Midwest	-0.048	0.413***	-0.007	0.058	0.041	-0.020	-0.017	0.063
West	0.102***	0.058	-0.014	-0.076	-0.168	-0.032	-0.034	0.069
Indirect Effect of Sprawl through Physical Activity	------	-0.001*	-0.0005*	0.0001	-0.0001*	-0.0001*	-0.0001*	-0.0004*
Indirect Effect of Sprawl through BMI	------	------	0.003***	0.001***	0.001***	0.001***	0.003***	0.001***
Indirect Effect of Sprawl through Phys Act and BMI	------	------	-0.0001*	-0.0001*	-0.0001*	-0.0001*	-0.0001*	-0.0001*
Indirect Effect of Physical Activity through BMI	------	------	-0.013***	-0.006***	-0.006***	-0.007***	-0.013***	-0.004***
N	101,819	101,819	68,454	68,262	68,262	68,399	68,455	68,262
R-Square	0.042	0.077	0.267	0.293	0.283	0.139	0.340	0.223
CFI	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
RMSEA	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

*p < 0.050, **p < 0.010, ***p < 0.001

† = moderate variable association (standardized regression coefficient absolute value is between 0.3 and 0.6)

the less physical activity they receive as well. When it comes to income, individuals with higher incomes tend to have higher levels of physical activity.

Compared to their male counterparts, females get less physical activity. The model also shows that compared to Whites, Blacks, Asians, and Hispanics get less physical activity per week. However, those who are Multi-racial get more weekly physical activity than do Whites. Even though marital status does not return any statistically significant results, the gender and marriage interaction effect actually shows statistical significance, indicating that married females tend to get less physical activity per week.

Individuals who consciously increase their physical activity in order to maintain or lose weight do in fact get more physical activity per week than those who do not increase their physical activity in order to control their weight, which is to be expected because these variables are theoretically highly correlated. Regional control variables show that individuals residing in the West get more physical activity per week compared to those who reside in the Northeast, the reference category for this set of dummy variables.

Standardized regression coefficients (Appendix C) for the physical activity hierarchical regression model indicate a weak association between the independent and outcome variable. Relatively speaking though, the variable that measures if someone increases their physical activity in order to control their weight is the strongest variable in the model, followed by gender, age, income, and education, which were four of the strongest variables shown to be associated with health outcomes in the baseline model. The R-Square statistic for this regression model is 0.042. Even though the variance being explained is rather low at 4.2%, CFI and RMSEA values are 1.000 and 0.000 respectively, indicating good model fit (Hu and Bentler 1999; Marcoulides and Schumacker 2001; Rigdon 1996).

Prediction of Body Mass Index (BMI)

Besides BMI acting as a mediating variable between sprawl, physical activity, and negative health outcomes in this analytical model, it is also a dependent variable with sprawl, physical activity, and sociodemographic, personal health, and regional control variables, having a direct relationship to it. Table 7.1 presents the results of the statistical regressions for this analytical model, including the outcome variable of BMI.

Many variables have a significant effect on BMI, including sprawl and physical activity, both of which are highly significant. According to the regression, for each one point increase in the sprawl scale, individuals in the sample have an increase in their BMI of 0.051. Therefore, comparing the least sprawling county in this study, New York, NY (Manhattan), to the most sprawling county in this study, Yadkin, NC, individuals would expect to have a BMI that is 1.44 points higher, which is a rather small effect overall.

As far as physical activity is concerned, for each one hour increase in the amount of physical activity per week, individuals within the sample decrease their BMI by 0.214. This is, of course, an expected result, as conditions such as obesity have reached epidemic levels in the United States and physical inactivity is now implicated as one of the causes of such conditions. An increasing body of evidence suggests that moderate forms of regular physical activity, such as walking, can have beneficial effects in maintaining a healthy weight (McCann and Ewing 2003; Ewing et al. 2003).

Age, education, income, gender, race, and marital status all turn out to be highly significant as well. For every one year increase in age for individuals within the sample, their BMI is expected to increase by 0.019. So the older an individual becomes, the higher their BMI tends to be. As for education, the more education an individual has the lower their BMI tends to be, with BMI decreasing by 0.378 for each higher level category for this ordinal variable. When it comes to

income, the higher an individual's income, the lower their BMI tends to be, by 0.113 for each higher level category for this ordinal variable.

Compared to their male counterparts, females tend to have a lower BMI. The model also shows that those who are Black, American Indian, Multi-racial, and Hispanic have higher BMI's compared to Whites, although those who are Asian tend to have lower BMI's. Compared to those who are married, widows and individuals who have never married have lower BMI's and separated and divorced individuals are likely to have a slightly higher BMI. The model also shows that married females tend to have lower BMI's.

Individuals who consciously increase their physical activity in order to control their weight have lower BMI's than those who do not increase their physical activity in order to control their weight, which is to be expected. Regional control variables reveal that individuals who reside in the Midwest are more likely to have a higher BMI than individuals in the Northeast.

Standardized regression coefficients for the BMI hierarchical regression model indicate a weak association between the independent and outcome variables. Relatively speaking though, the strongest variable in the prediction of BMI is the racial category of Black, followed by gender, education, income, and amount of physical activity. Age, which is one of the strongest variables in predicting other outcome variables in this study ranks in the middle in its association with BMI. The R-Square statistic is 0.077, and even though the variance being explained is rather low at 7.7%, CFI and RMSEA values are 0.994 and 0.025 respectively, indicating good model fit (Hu and Bentler 1999; Marcoulides and Schumacker 2001; Rigdon 1996).

Physical Activity and BMI as Independent Variables

Table 7.1 details the results of the statistical regressions for the health outcome variables of diabetes, heart attack, heart disease, high cholesterol, hypertension, and stroke with sprawl, physical activity, BMI, and sociodemographic, personal health, and regional control variables. As expected, this second and final analytical model, with all variables and hypothesized paths accounted for, indicates that sprawl has an indirect, rather than direct, effect on negative health outcomes. This indirect effect is detailed later in this chapter. This analytical model also shows that physical activity and BMI both have a statistically significant effect on health outcomes, with BMI having a rather strong effect.

As can be seen in Table 7.1, physical activity is highly significant in the prediction of health outcomes, except for heart attacks. The more regular physical activity people engage in, the less likely they are to have five of these six negative health outcomes. Each one hour per week increase in physical activity decreases the log odds of an individual's chance of developing these negative health outcomes between 0.014 and 0.050, according to which health outcome is examined. That translates to the odds of someone having one of these five health outcomes changing by a factor of 0.986 to 0.951 with each additional one hour increase in weekly physical activity, controlling for all the other variables. In other words, the odds of an individual in the sample having any one of these five negative health outcomes decreases between 1.39% and 4.88% for each one hour increase in weekly physical activity. This finding is to be expected because physical activity has been shown to have beneficial health outcomes (Perdue 2004; McCann and Ewing 2003; Ewing et al. 2003; Ewing et al. 2002a).

BMI is also highly significant in the prediction of negative health outcomes. The higher a person's BMI, the more likely he or she is to have all six of these negative health outcomes. Each

one point increase in BMI increases the log odds an individual's chance of developing such negative health outcomes between 0.016 and 0.057, according to which model is examined. That translates to the odds of someone having any of these six health outcomes changing by a factor of 1.016 to 1.059 with each additional one point increase in BMI, controlling for all the other variables in the model. In other words, the odds of an individual in the sample having one of these six negative health outcomes increases between 1.61% and 5.87% for each one point increase in BMI.

It should be mentioned that even though both physical activity and BMI are both highly significant in the prediction negative health outcomes, standardized regression coefficients reveal that BMI is a much stronger predictor of the health outcomes included in this study than is physical activity.

Strongest Variables in the Prediction of Negative Health Outcomes

When it comes to diabetes, the relative strength of the independent and control variables remains fairly unchanged, with age, BMI, Black, Asian, and American Indian as the top five strongest predictors (Table 7.1). The variance explained increases quite a bit from 18.5% in the baseline model to 26.7% in the second model. In the second analytical model CFI is 1.000 and RMSEA is 0.000, both indicating a good fit (Hu and Bentler 1999; Marcoulides and Schumacker 2001; Rigdon 1996). Taking both CFI and RMSEA values and the R-square statistic into consideration, the model fit for the second analytical regression in the prediction of diabetes is very good, being a better fit than the baseline model.

For heart attack, the relative strength of the independent and control variables remains fairly unchanged from the baseline model as well. Age, gender, income, BMI, and the gender-marriage interaction effect are the five strongest predictors. The variance explained increases slightly from

27.6% in the baseline analytical model to 29.3% in the second analytical model. In the second analytical model CFI is 1.000 and RMSEA is 0.000, both indicating a good fit (Hu and Bentler 1999; Marcoulides and Schumacker 2001; Rigdon 1996). Taking both CFI and RMSEA values and the R-square statistic into consideration, the model fit for the second analytical regression in the prediction of heart attacks is very good, being a better fit than the baseline model.

The relative strength of the independent and control variables in the prediction of heart disease also remains fairly unchanged, with age, BMI, income, gender-marriage interaction effect, and gender being the top five strongest predictors. The variance explained increases slightly from 26.1% in the baseline model to 28.3% in the second model. In the second analytical model CFI is 1.000 and RMSEA is 0.000, both indicating a good fit (Hu and Bentler 1999; Marcoulides and Schumacker 2001; Rigdon 1996). Taking both CFI and RMSEA values and the R-square statistic into consideration, the model fit for the second analytical regression in the prediction of heart disease is very good, being a better fit than the baseline model.

The relative strength of the independent and control variables in the prediction of high cholesterol also remains fairly unchanged, with age, BMI, Black, gender, and education as the five strongest predictors. The variance explained increases slightly from 11.4% in the baseline analytical model to 13.9% in the second analytical model. In the second analytical model CFI is 1.000 and RMSEA is 0.000, both indicating a good fit (Hu and Bentler 1999; Marcoulides and Schumacker 2001; Rigdon 1996). Taking both CFI and RMSEA values and the R-square statistic into consideration, the model fit for the second analytical regression in the prediction of high cholesterol is good, being a better fit than the baseline model.

For hypertension, the relative strength of the independent and control variables reveal that age, BMI, Black, income, and gender are the top five strongest variables in its prediction. The

variance explained increases slightly from 27.4% in the baseline analytical model to 34.0% in the second analytical model. In the second analytical model CFI is 1.000 and RMSEA is 0.000, both indicating a good fit (Hu and Bentler 1999; Marcoulides and Schumacker 2001; Rigdon 1996). Taking both CFI and RMSEA values and the R-square statistic into consideration, the model fit for the second analytical regression in the prediction of hypertension is very good, being a better fit over the baseline model.

The relative strength of the variables in the prediction of stroke also remains fairly unchanged, with age, income, gender, gender-marriage interaction effect, and BMI being the top five strongest predictors. The variance explained increases slightly from 21.1% in the baseline analytical model to 22.3% in the second analytical model. In the second analytical model CFI is 1.000 and RMSEA is 0.000, both indicating a good fit (Hu and Bentler 1999; Marcoulides and Schumacker 2001; Rigdon 1996). Taking both CFI and RMSEA values and the R-square statistic into consideration, the model fit for the second analytical regression in the prediction of stoke is very good, being a better fit than the baseline model.

As it can be seen, the addition of BMI to the analytical model shows statistically significant results in the prediction of negative health outcomes; moreover BMI ranks as one of the strongest predictors for all six negative health outcomes (Table 7.1). Even though physical activity returns highly statistically significant results, it does not rank as one of the strongest variables in the prediction of these health outcomes. From examining the fit statistics it can be clearly seen that the variance explained in all statistical models increased from the baseline model (indicating a better fit) and CFI and RMSEA values fall within respectable ranges. Therefore, this second analytical model is explaining more in the prediction of health outcomes than did the baseline model (McCann and Ewing 2003; Ewing, Schmid, Killingsworth, Zlot, and Raudenbush 2003).

Changes in Prediction of Negative Health Outcomes from Previous Models

There are a few changes in the prediction of negative health outcomes from the baseline model (Tables 6.2 and 7.1). In the prediction of diabetes changes in significance include: 1) The effect of being Asian is no longer statistically significant; 2) The effect of being separated/divorced and never married are now statistically significant; and 3) The effect of living in the Midwest is also now statistically significant. None of these variables, with the exception of Asian, are very strong predictors of diabetes, relatively speaking.

In the prediction of heart attacks changes in significance include: 1) Education is no longer statistically significant; 2) The effect of being Black is now statistically significant; 3) The effect of being Multi-racial is no longer statistically significant; and 4) The effect of being widowed is also no longer statistically significant. Even though the effect of being Black is now statistically significant, it does not rank as one of the strongest variables in the prediction of heart attacks.

In the prediction of heart disease the major changes in significance include: 1) The effect of being Black is now statistically significant; and 2) The effect of being widowed and never married are no longer statistically significant. Even though the effect of being Black is now statistically significant, it does not rank as one of the strongest predictor variables for heart disease.

In the prediction of high cholesterol the major changes in significance include: 1) The effect of being Asian or American Indian is no longer statistically significant; and 2) The effect of never married is also no longer statistically significant. It should be noted that Sprawl's direct effect remains statistically significant in the prediction of high cholesterol. This is the only negative health outcome variable that sprawl returns a statistically significant, albeit very tiny, direct effect in this analysis.

In the prediction of hypertension the only change in significance is the effect of never married is no longer statistically significant. In the prediction of stroke, the major changes in significance include: 1) The effect of being Native Hawaiian/Pacific Islander is now statistically significant; 2) The effect of being widowed and never married are now statistically significant; and 3) The effect of increasing physical activity in order to control weight is no longer statistically significant. Even though the effects of being Native Hawaiian/Pacific Islander, widowed, or never married are now statistically significant, none of them rank as any of the strongest variables in the prediction of stroke.

Indirect Effect of Sprawl through Physical Activity on Negative Health Outcomes, Paths H1 and H6

In this analytical model both physical activity and BMI are mediating variables in various paths and determine indirect effects sprawl may have on negative health outcomes (Table 7.1). Not only does physical activity and BMI mediate the relationship between sprawl and negative health outcomes, physical activity also mediates the relationship between sprawl and BMI. Additionally, BMI mediates the relationship between physical activity and negative health outcomes.

The first indirect effect in this analytical model is sprawl through physical activity on negative health outcomes and this effect is significant for all health outcome variables, except for heart attacks. This significance indicates that physical activity does have a mediating effect between sprawl and negative health outcomes. Even though there is a significant indirect effect, it is a very small effect. This indirect effect states that each one point increase in the sprawl index decreases the log odds of someone having any one of these five negative health outcomes between 0.0001 and 0.0005. That translates to the odds of someone having such a health outcome changing

by a factor of 0.9999 to 0.9995 with each additional one point increase in the sprawl index, controlling for all the other variables in the model. In other words, the odds of an individual in the sample of having diabetes, heart attack, heart disease, high cholesterol, hypertension, or stroke decreases between 0.01% and 0.05% for each one point increase in the sprawl index. Therefore; going from the least sprawling county included in this study, New York, NY (Manhattan) to the most sprawling county included in this study, Yadkin, NC, individuals are between 0.28% and 1.41% less likely to have one of these six negative health outcomes when mediated by physical activity. This indirect effect was expected to be positive, but turned out to be negative instead.

Indirect Effect of Sprawl through BMI on Negative Health Outcomes, Paths H4 and H3

The indirect effect of sprawl through BMI on negative health outcomes in this analytical model is highly significant for all health outcomes (Table 7.1). This significance indicates that BMI has a mediating effect between sprawl and negative health outcomes. Even though there is a significant indirect effect, it is a rather small effect. This indirect effect states that each one point increase in the sprawl index increases the log odds of someone having any one of these six negative health outcomes by 0.001 to 0.003, depending on which model is examined. That translates to the odds of someone having such a health outcome changing by a factor of 1.001 to 1.003 with each additional one point increase in the sprawl index, controlling for all the other variables in the model. In other words, the odds of an individual in the sample of having diabetes, heart attack, heart disease, high cholesterol, hypertension, or stroke increases between 0.10% and 0.30% for each one point increase in the sprawl index. Therefore; going from the least sprawling county included in this study, New York, NY (Manhattan) to the most sprawling county included in this study, Yadkin, NC, individuals are between 2.8% and 8.5% more likely to have one of these six negative health

outcomes when mediated by BMI. This indirect effect is positive, whereas the previous indirect effect of sprawl on health outcomes via physical activity was negative.

Indirect Effect of Sprawl through Physical Activity and BMI on Negative Health Outcomes, Paths H1, H2, and H3

The indirect effect of sprawl through physical activity and BMI in this analytical model is significant for all health outcome variables (Table 7.1). This significance indicates that sprawl's effect on health outcomes is mediated by both physical activity and BMI. This indirect effect states that each one point increase in the sprawl index decreases the log odds of someone having any one of these negative health outcomes by 0.0001 (the coefficient is the same for all health outcomes). That translates to the odds of someone having such a health outcome changing by a factor of 0.9999 with each additional one point increase in the sprawl scale, controlling for all the other variables in the model. In other words, the odds of an individual in the sample of having diabetes, heart attack, heart disease, high cholesterol, hypertension, or stroke decreases by 0.01% for each one point increase in the sprawl index. Therefore; going from the least sprawling county included in this study, New York, NY (Manhattan) to the most sprawling county included in this study, Yadkin, NC, individuals are 0.28% less likely to have one of these six negative health outcomes when mediated by both physical activity and BMI. Obviously, at less than half of one percent this is an extremely weak indirect effect. In comparison, the indirect effect of sprawl through BMI is stronger in its effect on negative health outcomes.

Indirect Effect of Physical Activity through BMI on Negative Health Outcomes, Paths H2 and H3

This indirect effect is highly significant for all health outcome variables (Table 7.1). This significance indicates that BMI has a mediating effect between physical activity and health outcomes. Even though there is a significant indirect effect, it is a rather small effect. This indirect effect states that each one hour increase in weekly physical activity decreases the log odds of someone having any one of these six negative health outcomes between 0.004 and 0.013, depending on which model is examined. That translates to the odds of someone having such a health outcome changing by a factor of 0.996 to 0.987 with each additional one hour increase in amount of weekly physical activity, controlling for all the other variables in the model. In other words, the odds of an individual in the sample of having diabetes, heart attack, heart disease, high cholesterol, hypertension, or stroke decreases between 0.40% and 1.29% for each one hour increase in weekly amount of physical activity.

Indirect Effect of Sprawl through Physical Activity on BMI, Paths H1 and H2

The final indirect effect for this analytical model is the mediating effect physical activity has between sprawl and BMI (Table 7.1). This indirect effect is highly significant, indicating that physical activity does have a mediating effect between these two variables. This indirect effect states that each one point increase in the sprawl index decreases BMI by 0.001. Therefore, going from the least sprawling county included in this study, New York, NY (Manhattan) to the most sprawling county included in this study, Yadkin, NC, individual BMI tends to be 0.028 points lower, indicating a extremely small indirect effect.

CHAPTER 8

DISCUSSION AND CONCLUSION

This concluding chapter of my dissertation begins with a discussion of how well my analytical models fit, or do not fit, my original hypotheses. It then moves on to discuss the associations between the various variables in my conceptual model followed by a discussion of which analytical model returns the best results. I also propose a new conceptual model from the results of my analyses. Limitations of my study, contributions my research provides to the literature, implications my research has at an applied level, and possibilities for future research are also discussed in this final chapter.

Validity of Hypotheses

Taking direction from Ewing et al. (2003) and my conceptual model, I hypothesized associations between sprawl, levels of physical activity, body weight, and negative health outcomes. Following is a discussion of my original hypotheses and how well the analyses of data support, or do not support, each hypothesis.

Hypothesis #1

My first hypothesis states that sprawl is negatively related to physical activity and that as sprawl increases, individual levels of physical activity decrease. I formulated this hypothesis due to the fact that previous research shows environmental form does have a relationship to one's propensity to engage in physical activity and suggests that people living in sprawling areas miss out on significant health benefits, such as walking to the store, to work, or other places as part of a daily routine (McCann and Ewing 2003). This lack of beneficial regular physical activity may be

explained by patterns of streets within neighborhoods, such as those found in many suburban subdivisions (Lopez 2004).

However, my analyses do not support this hypothesis. Not only does sprawl return a significant effect on the prediction of physical activity, it shows there is a positive relationship instead of a negative one I predicted. According to the models (Table 7.1), for each one point increase on the sprawl scale, individuals actually increase their amount of weekly physical activity by 0.009 hours per week. Therefore, going from the least sprawling county included in this study, New York, NY (Manhattan), to the most sprawling county included in this study, Yadkin, NC, individuals would expect to have 0.25 additional hours, or 15 additional minutes, of physical activity per week. Even though there is a significant, positive effect; it is a rather weak effect overall.

Hypothesis #2

My second hypothesis states that physical activity is negatively related to weight gain and that as individual levels of physical activity decrease, individual weight increases. I formulated this hypothesis due to the fact that previous research shows that environmental form has a strong relationship to one's propensity to engage in physical activity. Conditions such as obesity, diabetes, and hypertension have reached epidemic levels in the United States and physical inactivity is now implicated as one of the causes of such conditions. An increasing body of evidence suggests that moderate forms of regular physical activity, such as walking, can have beneficial effects in maintaining a healthy weight (McCann and Ewing 2003; Ewing et al. 2003). Unfortunately, the majority of Americans report not obtaining enough exercise to meet the recommended weekly minimum, and many Americans remain completely inactive during their leisure time (McCann and Ewing 2003).

In order to test this hypothesis, I utilized BMI data from the BRFSS when conducting SEM analyses. In the prediction of BMI, regressions show that physical activity has a highly significant effect (Table 7.1). According to the regressions, for each one hour increase in the amount of physical activity per week, individuals within the sample decrease their BMI by 0.214 points. Therefore, the models show a significant, negative relationship between physical activity and BMI, which is the expected result.

Hypothesis #3

My third hypothesis states that weight is positively related to health outcomes and that as individual weight increases, so does the chance for negative health outcomes. I formulated this hypothesis because health experts have found that most Americans are overweight and conditions such as diabetes and heart disease have reached epidemic levels as a consequence of such excessive weight (McCann and Ewing 2003; Ewing et al. 2003). Such connections between body weight and health conditions have long been documented, as has an increase in the risk of mortality coinciding with the severity of obesity (Calle et al. 1999).

As with the testing of my second hypothesis, I utilized BMI data from the BRFSS when conducting SEM analyses. The regression models show that BMI is highly significant in the prediction of health outcomes (Table 7.1). The higher an individual's BMI, the more likely they are to have negative health outcomes. Each one point increase in BMI increases the log odds of someone having any one of these seven negative health outcomes between 0.016 and 0.058, depending on which outcome variable is examined. That translates to the odds of someone having such a negative health outcome changing by a factor of 1.016 to 1.060 with each additional one point increase in BMI, controlling for all the other variables in the model. In other words, the odds of an individual in the sample having diabetes, heart attack, heart disease, high cholesterol,

hypertension, or stroke increases between 1.61% and 5.97% for each one point increase in BMI. BMI as an independent variable is significant in all health outcome statistical models and shows a strong, positive relationship, meaning the higher an individual's BMI, the more likely they are of developing a negative health outcome.

Hypothesis #4

My fourth hypothesis states that sprawl is positively related to weight and that as sprawl increases so does individual-level weight. I formulated this hypothesis in order to determine if there is a direct effect between sprawl and weight. Results show that when predicting BMI, sprawl shows a significant effect. According to the analyses (Table 7.1), for each one point increase on the sprawl scale, individuals in the sample have an increase in their BMI by 0.051 points. Therefore, going from the least sprawling county included in this study, New York, NY (Manhattan), to the most sprawling county included in this study, Yadkin, NC, individuals would expect to have their BMI increase by 1.44 points. Although sprawl does return a significant, positive relationship with BMI, this relationship is fairly weak.

Hypothesis #5

My fifth hypothesis states that sprawl is positively related to individual-level negative health outcomes and that as sprawl increases, so does the chance for negative health outcomes, such as diabetes, heart attack, heart disease, high cholesterol, hypertension, and stroke. Based on existing research, I hypothesized a positive zero-order correlation between sprawl and negative health outcomes. As sprawl increases I expect a positive relationship between sprawl and health problems to be revealed, especially when controlling for other variables such as sociodemographics, personal health, and region. However, with the addition of physical activity and BMI as mediating variables,

I hypothesized that the relationship between sprawl and negative health outcomes will shrink and become statistically insignificant.

As the results clearly show, sprawl is in fact mediated by physical activity and BMI (Table 7.1). This finding indicates a mediating effect between sprawl and health outcomes since the relationship between the independent variable (sprawl) and dependent variables (all health outcomes except high cholesterol) becomes less significant or non-significant (Baron and Kenny 1986) as control and mediating variables are added. This drop in significance is evident in comparing my baseline model (Table 6.2) and my second analytical model (Table 7.1).

When looking at sprawl's indirect effect on health outcomes its effect is negative, for all health outcome variables. This is due to the fact that sprawl has a positive relationship to physical activity (as sprawl increases, so does individual-level physical activity) and physical activity has a negative relationship to BMI (as physical activity increases, BMI decreases) and BMI has a positive relationship to health outcomes (as BMI increases, so does the chance of developing negative health outcomes), indicating an overall negative relationship between sprawl and health outcomes. This indirect effect is statistically significant, but very weak, across all regression models.

Hypothesis #6

My sixth hypothesis states that physical activity is negatively related to health outcomes and that as individual levels of physical activity decrease, the chance for negative health outcomes increase. As shown by my analyses (Table 7.1), physical activity is highly significant in the prediction of all but one health outcome included in this research project. The more regular physical activity they engage in, the less likely they are to have diabetes, heart disease, high cholesterol, hypertension, or stroke. Each one hour per week increase in physical activity decreases the log odds of an individual's chance of developing these negative health outcomes between 0.014

and 0.063, depending to which outcome is examined. That translates to the odds of someone having one of these five negative health outcomes changing by a factor of 0.986 to 0.939 with each additional one hour increase in weekly physical activity, controlling for all the other variables in the model. In other words, the odds of an individual in the sample having any one of these five negative health outcomes decreases between 1.39% and 6.11% for each one hour increase in weekly physical activity. Physical activity as an independent variable shows a statistically significant, negative, but rather weak, relationship to health outcomes, meaning the more physical activity an individual receives each week, the less likely they are to develop one of these five negative health outcomes.

As it can be seen, out of my original six hypotheses, I can accept four of them as valid. These four hypotheses are numbers two, three, four, and six. The two hypotheses that are not supported by analyses are hypotheses one and five. Hypothesis one, which states that sprawl is negatively related to physical activity, is not support because results clearly show that sprawl has a *positive*, although weak, relationship to physical activity. Therefore, hypothesis one should be rejected.

Hypothesis five, which states that sprawl is positively related to health problems, is not supported by analyses either. Even though it was hypothesized that sprawl has an indirect, rather than direct, effect on negative health problems, this hypothesis also states that this relationship is positive. Results show that sprawl actually has a *negative* indirect effect on health outcomes, meaning that via proxy physical activity and BMI, sprawl has a significant, although weak, effect on decreasing negative health outcomes for individuals within the sample. However, when mediated by BMI alone, sprawl appears to be linked to increased health problems. It is sprawl's mediating effect that has the most influence on health outcomes.

Associations between Variables in the Conceptual Model

The usefulness of using an SEM framework to test my conceptual model is found in its ability to both test existing theories and reveal trends that can lead to developing new theories (Byrne 2006; Raykov and Marcoulides 2006). The findings that my analyses have revealed may indicate the discovery of a new trend (Table 7.1). The surprising finding that sprawl has a positive direct effect on physical activity and a negative indirect effect on BMI could be due to the fact that people in sprawling areas are becoming more aware of the epidemic of obesity in American society. As a result they may be consciously increasing their amounts of physical activity in order to combat a sedentary lifestyle caused by a reliance on automobiles for mobility.

When examining correlations related to this issue, results show that amount of weekly physical activity is significantly, but very weakly, correlated with sprawl (0.010) and increased physical activity in order to control weight (0.257). From these figures it is apparent there is almost no correlation between these variables. However, there is no statistically significant correlation between sprawl and increased physical activity in order to control weight. Additionally, when examining BMI statistics, correlations reveal a significant effect between BMI and sprawl (0.018) and between BMI and increasing physical activity in order to control weight (-0.045).

Another explanation could be that individuals living in suburban areas are taking advantage of the many activities available to them. The suburbs have more space for amenities such as parks, green spaces, golf courses, tennis courts, and pools. Such amenities are usually easily accessible to suburbanites, and it is quite possible that people in sprawling suburban areas get more physical activity because they utilize these types of amenities to exercise. Additionally, many suburban housing developments have tennis courts, pools, and homes with yards. Suburban family living

also emphasizes organized sports activities, such as baseball and soccer, for children, as well as adults.

Luckily more and more people are realizing the benefits of regular physical activity and municipalities are encouraging the development of parks and green spaces. However, such development of parks and green spaces need not be limited to suburban locations. Even though cities have limited available space, some have developed creative alternatives. In Atlanta for example, there is a movement to convert abandoned railroad lines into green space with a light rail system. This light rail system would provide an additional mode of transportation for the citizens of Atlanta, but would also provide dedicated green space for active use, such as walking, jogging, and bicycling. If the availability of active public spaces is the key to increasing individuals' regular physical activity, then it is important to look for ways to increase such public spaces in cities. Many scholars believe such active outdoor spaces are important to the health of individuals and there is a movement in the U. S. to document the growing use of health impact assessment methods, which will help planners, policy makers, and developers consider the health consequences of the decisions they make. Such knowledge gained could provide for more parks and green spaces to help increase levels of physical activity of all American citizens (Dannenberg et al. 2008).

Findings that physical activity has an inverse indirect effect on health outcomes and when mediated by BMI was expected (Table 7.1). As levels of physical activity increase, BMI decreases and in turn negative health outcomes decrease as well. It can be seen from this, and previous analyses in other studies, that physical activity is an effective way to reduce the risk for negative health outcomes, such as diabetes, heart attack, heart disease, hypertension, high cholesterol, and stroke. My findings show that when mediated by BMI, physical activity has a highly significant

indirect effect on health outcomes, although physical activity's direct effect is somewhat stronger than its indirect effect.

Best Fitting Analytical Model

Two results indicate that the best fitting analytical model in this research study is the second model (Table 7.1). First, by analyzing my conceptual model in different steps, I was able to have a "baseline" measurement to compare the effect each variable has on one another. This approach was important because of the anticipated role of certain mediating variables. This analytic "road map" allowed me to uncover the indirect (mediating) effects of physical activity and BMI on negative health outcomes. Although sprawl was found to have a significant relationship with certain negative health outcomes in the baseline model, the second model revealed that there is in fact an indirect effect between sprawl and negative health outcomes when mediated by physical activity and BMI, and that the only direct effect between sprawl and a negative health outcome is that of high cholesterol.

The second reason I attest the second analytical model as being the best fitting model relates to the better R-square statistics and CFI and RMSEA values. Such good fit statistics give validity to the statistical models. R-square statistics improved for the outcome variables and CFI and RMSEA values remained in the range of good model fit in all statistical models. Therefore, my second analytical model is the best fitting model for this research project. Even though the second analytical model is the best fitting model, it is necessary to address the positive association of sprawl with physical activity and health outcomes and propose a newer, better fitting conceptual model.

Proposal of a New Conceptual Model

In my conceptual framework, I posed some questions as to the basis of my conceptual model. These questions include: 1) Is there a significant association between sprawl and physical activity? 2) If so, how might sprawl affect physical activity? 3) Do individuals residing in less sprawling areas get significantly more or less physical activity on average than those in more sprawling areas? and 4) Does physical activity have a significant association with health outcomes by way of individual-level weight?

The analyses I conducted not only help answer these questions, but they also help craft a new conceptual model, one that fits the data and results better. The results of my analyses do show there is a significant association between sprawl and physical activity (Table 7.1). However, that association is positive and not negative as originally hypothesized. Additionally, analyses show that physical activity does have a significant direct, as well as indirect (via BMI), association with health outcomes.

In my original conceptual model, hypothesized relationships between variables such as physical activity, have a direct, as well as indirect, effect on health outcomes, and sprawl had a direct, as well as indirect, effect on both weight and health outcomes. However, my analyses show that sprawl has more of an indirect, rather than direct, effect on health outcomes. Therefore, Figure 8.1 shows a new conceptual model I have crafted from the results of my analyses.

In my new conceptual model, I have changed the variables and hypothesized paths slightly. "Less physical activity" is simply "physical activity;" "weight gain" is "weight;" and "negative health outcomes" is "disease." I made these changes in order to add directional signs (+/-) to indicate how variables affect one another directly, as determined from the results of my analyses. I removed the original path H5, which depicted a direct relationship between sprawl and negative

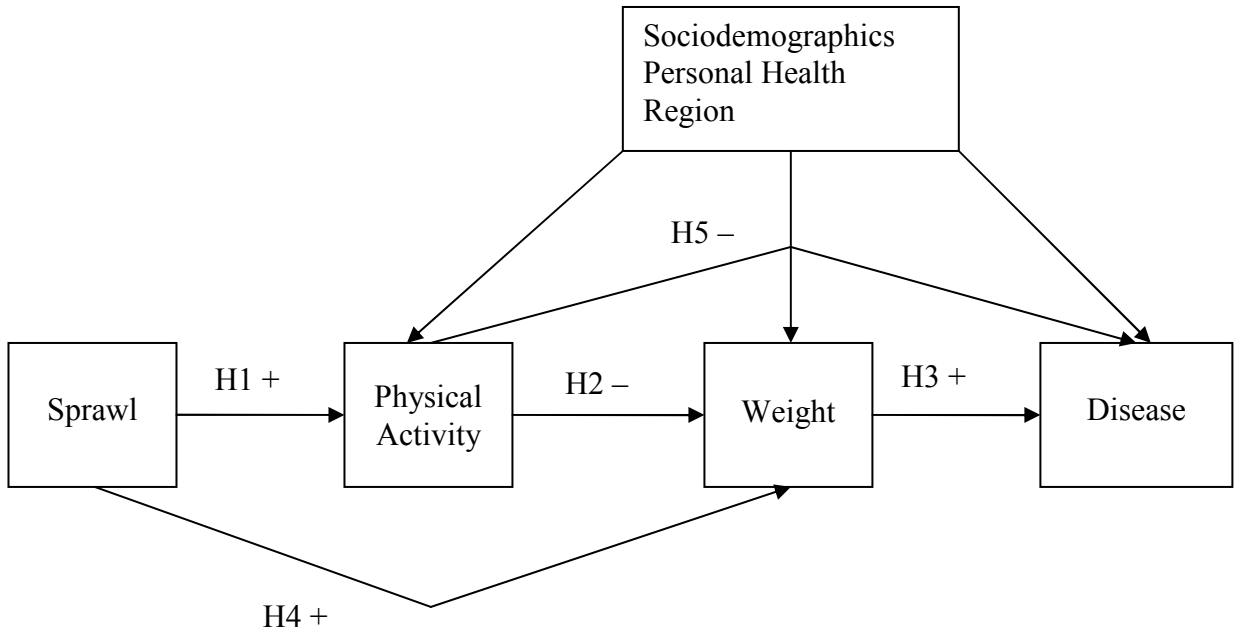

Figure 8.1 New Conceptual Model

health outcomes. I retained the direct path between physical activity and negative health outcomes because results show a statistically significant relationship between these variables. This path is now path H5 instead of path H6 as depicted in my original conceptual model.

In my new conceptual model, the direct path between sprawl and physical activity (H1) has positive directional signage because my analyses points to the fact that sprawl has a positive effect on individual levels of physical activity. As sprawl increases, so do amounts of physical activity. Additionally, results from my analyses show that physical activity has a direct negative effect on weight (H2). As physical activity increases, weight tends to decrease. Physical activity has a similar effect on disease (H5). As physical activity increases, disease decreases. Weight has a positive effect on disease (H3). As weight decreases, disease deceases as well.

The final hypothesized path in my new conceptual model depicts the positive relationship between sprawl and weight (H4), meaning that as sprawl increases, so does weight. This path seems counterintuitive to sprawl's negative indirect effect on weight. However, results show that even though individuals in sprawling areas tend to engage in more physical activity than those in more compact locations, they also tend to have a higher BMI, suggesting that those in suburban areas who do not participate in regular physical activity may have negative health outcomes due to a higher than recommended BMI. Thus, it is important to be ever vigilant in promoting the benefits of exercise.

Contributions to the Literature

My dissertation will contribute to the literature on sprawl and health outcomes in a number of ways. First, guided by existing literature, specifically Ewing et al. (2003), I developed and tested a conceptual path model linking sprawl, physical activity, weight gain/BMI, and several key negative health outcomes. Such a model has not been developed and tested before and my findings make an important contribution to the growing body of knowledge about the connection of these variables.

Among the contributions is the knowledge that after controlling for a wide range of social and economic demographic variables, sprawl has a positive effect on physical activity, a finding contrary to findings of other studies. Existing literature maintains that the design of communities influence health by encouraging or discouraging routine physical activity and that people living in sprawling areas tend not to engage in regular health promoting activity, such as walking, in their daily lives (McCann and Ewing 2003). Neighborhood street patterns, such as those found in many suburban subdivisions, affect automobile use and walking. Metropolitan areas with high levels of

urban sprawl tend to have higher per capita vehicle miles traveled daily, even after controlling for factors such as income, size of metropolitan area, and location within the nation. This suggests that people in high-sprawl areas drive more, quite possibly at the expense of daily physical activity (Lopez 2004). However, results from my analyses show that individuals of similar race, age, education, socioeconomic status, etc. residing in more sprawling areas actually get more physical activity than individuals in less sprawling, more compact locales. So, while individuals in more sprawling areas may be driving more due to the design of their communities, which are usually not very pedestrian friendly environments, they are likely driving to the places such as gyms, parks, soccer fields, and pools, just to name a few. It may be suburban life makes available multiple recreation options, which leads to increased activity. It may also be the fact that tree-lined suburban streets with longer blocks may be more pleasant to walk than are city streets with shorter blocks where one often has to stop at major intersections before crossing the street.

My findings also show that there are significant relationships in two mediating variables (physical activity and BMI) between sprawl and negative health outcomes. Both variables mediate this relationship together, as well as independently. Additionally, BMI mediates the relationship between physical activity and negative health outcomes. The relationship between these variables has not been tested previously using an SEM framework as I have done.

Implications at an Applied Level

Education is, and always has been, an important element in modern society. Education allows individuals, as well as communities, to improve and better their lives. Research plays an important role in education by providing important information that will benefit the lives of many people. My dissertation contains a body of knowledge that will lead to the understanding of how

the place where people reside can affect health outcomes. My findings indicate that living in a more dense versus sprawling locale can affect the amount of physical activity individuals engage in, which affects weight, and ultimately several health conditions that are growing in prevalence in American society. It is important for people to know that no matter where they live, whether in the city or the suburbs, physical activity is important and it is important that everyone get enough exercise to maximize their health outcomes. It has long been known that physical activity is a significant factor in individual weight/BMI and health outcomes. Now it is becoming more apparent just how sprawl affects physical activity and weight/BMI and ultimately, as uncovered in this research project via indirect paths, health outcomes.

Possibilities for Future Research

It was important to find out which associations between sprawl, physical activity, weight/BMI, and negative health outcomes exist in order to proceed with additional, more refined research to test the causal relationships in my conceptual model. It was also important to find out the relative strength and magnitude of significant relationships. Now that I have uncovered significant relationships between the variables in my conceptual model, it is important to discuss other research opportunities in regards to this type of study.

Numerous possibilities for future research exist. Potential studies include an examination of the physical activity variable in two separate components – moderate physical activity and vigorous physical activity – in order to determine how sprawl affects these separate components and how each component affects weight and health outcomes. These variables could be tested within a framework similar to that used in this study. The conceptual model also might be modified to include both moderate and vigorous physical activity as separate variables within one model.

Another significant potential avenue of study would be to test my new conceptual model with newly collected data that would include other variables deemed relevant to health outcomes, including family health history. A more refined sprawl index and migration patterns of metropolitan residents would also be beneficial in testing my new conceptual model. These additional variables would provide a greater understanding of how sprawl affects physical activity, weight, and health outcomes. Additionally, other health outcome variables, such as depression, Chronic Fatigue Syndrome, dementia, renal disease, and cancer can be tested as well.

Another potential study would compare city dwellers, suburbanites, and people that live in "new urban/live, work, play" centers to determine how they differ with respect to sprawl, physical activity, weight, and health outcomes. Collecting my own data will allow me to know whether individuals reside in suburban areas, the city, or new urban centers. A study such as this would make a significant contribution to the understanding of how sprawl affects physical activity, weight, and health outcomes, especially due to the fact that new urban centers are increasingly common and are a hot topic in development community arenas, as well as in the literature.

A study that looks at how suburbanites utilize available amenities, such as parks, green spaces, tennis courts, and sports fields would be an interesting project. It also would be intriguing to compare how suburbanites and city-dwellers differ in their use of public activity spaces, especially as it relates to the amount of physical activity they each receive on a regular basis.

A final study I would like to pursue would examine sprawl and its effects on adults aged 65 and older. Since I have been working in gerontological research for the past several years, I have become very interested and concerned about the health and well-being of the elderly population. Sprawl may not affect older persons the same way it affects younger individuals and a study focusing on individuals 65 years of age and older would provide this information. By collecting my

own data, I can include persons living in a range of housing types, including traditional retirement communities, naturally occurring retirement communities, care facilities such as assisted living, and non-age segregated communities.

When designing future studies dealing with sprawl, physical activity, weight, and health outcomes, it will be important to develop a research design that would be appropriate for multiple studies. This strategy would allow me to conduct multiple research projects. By obtaining all the information I need from one survey, I will be able to conduct multiple research projects that focus on sprawl, physical activity, weight, and health outcomes, and it would have the potential to provide important new information.

Conclusion

As urban sprawl continues to increase throughout the nation, it is important to understand possible consequences of this phenomenon. Although many would consider urbanization to be a form of progress, it can also result in unintended outcomes. Previous studies suggest that sprawl and the low-density, automobile-dependent community design it encourages affect levels of physical activity and that sprawl may have an overarching influence on health. As seen from this research project, sprawl does affect levels of physical activity; however, this effect is positive, not negative as previous research indicates (MacLennan 2004; McCann and Ewing 2003; Ewing et al. 2002a).

A number of arguments exist concerning sprawl and its association with health outcomes, including those that support the existence of definite associations and those that reject such claims (McCann and Ewing 2003; Kelly-Schwartz et al. 2004; Ewing et al. 2003; Lopez 2004; Giles-Corti and Donovan 2002; O'Toole 2002a, b). When beginning this dissertation, it was my intent to find

out what significant associations existed between the variables I proposed in my conceptual model. My research project examined the conceptual model I put forth and systematically looked for possible associations between sprawl, physical activity, weight, and negative health outcomes. By looking at urban sprawl at a multilevel path approach, my dissertation adds additional layers to the analysis of sprawl and health outcomes, as well as evaluates associations through various "paths" instead of simple hierarchical linear regression models.

When predicting physical activity, sprawl was shown to have a significant but small effect. According to the second analytical model, individuals in the most sprawling county in this study, Yadkin, NC, get 15 additional minutes of physical activity per week than do individuals in Manhattan. Even though sprawl is significant in the prediction of the amount of physical activity individuals get per week, it returns a rather weak effect. The effect of increasing one's level of physical activity in order to control weight is a much stronger predictor of amount of weekly physical activity, followed by gender, age, income, and education.

Scholarly studies have shown that sprawl and its component factors (dependence on automobiles, less physical activity, etc.) are found to be greater predictors of health than demographic control variables (Ewing et al. 2002a). Reasoning for this finding is that community design either encourages or discourages routine physical activity, such as walking to the store, to work, or other such places as a part of their daily routine. Other studies have shown that when it comes to whether or not people get regular exercise in their leisure time, such as running, working out, gardening, etc., the degree of sprawl seems to have very little influence, as people in both sprawling and compact areas are equally likely to report they exercise in some fashion (McCann and Ewing 2003). The results of my study show that sprawl does have a significant, albeit very

weak, positive effect on amounts of weekly physical activity. It also shows that sociodemographic and personal health control variables are stronger predictors of physical activity than is sprawl.

Many variables have a significant effect on BMI, including sprawl, which turns out to be highly significant. According to the second analytical model, individuals in Manhattan tend to have a BMI that is 1.44 points higher than those residing in Yadkin, NC. Even though sprawl is highly significant in determining BMI, it has a weak effect. Stronger predictors of BMI include the effect of being Black, followed by gender, education, income, and amount of physical activity. It is not surprising that physical activity is an important predictor of BMI. Results show that for each one hour increase in the amount of physical activity per week, individuals within the sample decrease their BMI by 0.214. It is well documented that moderate forms of regular physical activity, such as walking, can have beneficial effects in maintaining a healthy weight (McCann and Ewing 2003; Ewing et al. 2003).

Based on existing research, I hypothesized a positive zero-order correlation between sprawl and negative health outcomes and that as physical activity and BMI were added as mediating variables, the relationship between sprawl and negative health outcomes would shrink and become statistically insignificant. As the results clearly show, sprawl is in fact mediated by physical activity and BMI. This finding indicates a mediating effect between sprawl and health outcomes since the relationship between the independent variable (sprawl) and dependent variables (health outcomes) becomes less significant or non-significant (Baron and Kenny 1986) as control and mediating variables are added. This drop in significance between the first and second analytical model is the reason why it was important to test my model in different steps.

When looking at sprawl's indirect effect on health outcomes its effect is negative, for all health outcome variables. This is due to the fact that sprawl has a positive relationship to physical

activity (as sprawl increases, so does individual-level physical activity), physical activity has a negative relationship to BMI (as physical activity increases, BMI decreases), and BMI has a positive relationship to health outcomes (as BMI increases, so does the chance of developing negative health outcomes), indicating an overall negative relationship between sprawl and health outcomes. This indirect effect is significant, but very weak, across all regression models.

By employing my conceptual model and analyzing it within an SEM framework, I am able to help fill the need for additional empirical measures to determine relationships between urban form and health. By addressing issues of sprawl and how it relates to negative health outcomes, my dissertation adds to the growing knowledge about sprawl and its associations with health, including physical activity and BMI. It is important now, more than ever, to understand just how the choices society makes in planning, regulating, and zoning the American landscape impact its citizens. The better these issues are understood, the better choices Americans can make in how to maintain their best health in the ever sprawling American landscape.

REFERENCES

Baron, Reuben M., and David A. Kenny. 1986. "The Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations." *Journal of Personality and Social Psychology* 51: 1173-1182.

Baunach, Dawn Michelle. 2001. "Gender Inequality in Childhood: Toward a Life Course Perspective." *Gender Issues* 19: 61-86.

Bianchini, Franca, Rudolf Kaaks, and Harri Vainio. 2002. "Weight Control and Physical Activity in Cancer Prevention." *The International Association for the Study of Obesity* 3: 5 -8.

Blair, S. N., M. J. LaMonte, M. Z. Nichaman. 2004. "The Evolution of Physical Activity Recommendations: How much is Enough?" *American Journal of Clinical Nutrition* 79: 913S-920S.

Bowlin, S. J., B. D. Morrill, A. N. Nafziger, P. L. Jenkins, C. Lewis, and T. A. Pearson. 1993. "Validity of Cardiovascular Disease Risk Factors Assessed by Telephone Survey: The Behavioral Risk Factor Survey." *Journal of Clinical Epidemiology* 46: 561-571.

Brown, M. W., and R. Cudeck. 1993. "Alternative Ways of Assessing Model Fit." Pp. 136-162 in *Testing Structural Equation Models*, edited by K. A. Bollen and J. S. Long. Newbury Park, CA: Sage.

Brueckner, Jan K., and David A. Fansler. 1983. "The Economics of Urban Sprawl: Theory and Evidence on the Spatial Sizes of Cities." *The Review of Economics and Statistics* 65: 479-482.

Bruegmann, Robert. 2005. *Sprawl, A Compact History.* Chicago, IL: The University of Chicago Press.

Byrne, Barbara M. 2006. *Structural Equation Modeling with EQS: Basic Concepts, Applications, and Programming.* Routledge.

Calle, Eugenia E., Michael J. Thun, Jennifer M. Petrelli, Carmen Rodriguez, and Clark W. Heath. 1999. "Body-Mass Index and Mortality in a Prospective Cohort of U.S. Adults." The New England Journal of Medicine 341: 1097-1105.

Center for Disease Control and Prevention (CDC). 2003. *Behavioral Risk Factor Surveillance System Survey Data.* Atlanta, GA: US Department of Health and Human Services.

Dannenberg, , Andrew L., Rajiv Bhatia, Brian L. Cole, Sarah K. Heaton, Jason D. Feldman, and Candace D. Rutt. 2008. "Use of Health Impact Assessment in the U.S." *American Journal of Preventive Medicine* 34: 241-256.

Day, Kristen. 2006. "Active Living and Social Justice: Planning for Physical Activity in Low-Income, Black, and Latino Communities." *Journal of the American Planning Association* 72: 88-99.

Doll, William J., Xiaodong Deng, T. S. Raghunathan, Gholamreza Torkzadeh, and Weidong Xia. 2004. "The Meaning and Measurement of User Satisfaction: A Multigroup Invariance Analysis of the End-User Computing Satisfaction Instrument." *Journal of Management Information Systems* 21: 227-262.

Dresel, Alexandra, Joseph A. Kuhn, and Todd M. McCarty. 2004. "Laparoscopic Roux-en-Y Gastric Bypass in Morbidly Obese and Super Morbidly Obese Patients." *The American Journal of Surgery* 187: 230-232.

Eberstadt, Nicholas, and Sally Satel. 2004. "Health, Inequality, and the Scholars." *Public Interest* Fall 2004: 100-118.

Egede, Leonard E. and Deyi Zheng. 2002. "Modifiable Cardiovascular Risk Factors in Adults with Diabetes, Prevalence and Missed Opportunities for Physician Counseling." *Arch Intern Med* 162: 427-433.

Ewing, Reid, Tom Schmid, Richard Killingsworth, Amy Zlot, and Stephen Raudenbush. 2003. "Relationship Between Urban Sprawl and Physical Activity, Obesity, and Morbidity." *American Journal of Health Promotion* 18: 47-57.

Ewing, Reid, Rolf Pendall, and Don Chen. 2002a. *Measuring Sprawl and Its Impact*. Washington, DC: Smart Growth America.

Ewing, Reid, Rolf Pendall, and Don Chen. 2002b. *Measuring Sprawl and Its Impact, Volume 1*. Washington, DC: Smart Growth America.

Ezzati, Majid, Stephen Vander Hoorn, Carlene M. M. Lawes, Rachel Leach, W. Phillip T. James, Alan D. Lopez, Anthony Rodgers, and Christopher J. L. Murray. 2005. "Rethinking the "Diseases of the Affluence" Paradigm: Global Patterns of Nutritional Risks in Relation to Economic Development." *PLOS Medicine* 2: 404-412.

Flegal, KM, MD Carroll, RJ Kuczmarski, and CL Johnson. 1998. "Overweight and Obesity in the United States: Prevalence and Trends, 1960-1994." *International Journal of Obesity* 22: 39-47.

Ford, Earl S., Robert K. Merritt, Gregory W. Heath, Kenneth E. Powell, Richard A. Washburn, Andrea Kriska, and Gwendolyn Haile. 1991. "Physical Activity Behaviors in Lower and Higher Socioeconomic Status Populations." *American Journal of Epidemiology* 133: 1246-1256.

Freeman, Lance. 2001. "The Effects of Sprawl on Neighborhood Social Ties." *Journal of the American Planning Association* 67: 69-77.

Frank, Lawrence D., and Peter O. Engelke. 2001. "The Built Environment and Human Patterns: Exploring the Impacts of Urban Form on Public Health." *Journal of Planning Literature* 16: 202 -218.

Fulton, William, Rolf Pendall, Mai Nguyen, and Alicia Harrison. 2001. *Who Sprawls Most: How Growth Patterns Differ Across the United States.* Washington, DC: Center on Urban and Metropolitan Policy, The Brookings Institution.

Galdas, Paul M., Francine Cheater, and Paul Marshall. 2005. "Men and Health Help-Seeking Behaviour: Literature Review." *Journal of Advanced Nursing* 49: 616-623.

Galster, George, Royce Hanson, Michael R. Ratcliffe, Harold Wolman, Stephen Coleman, and Jason Freihage. 2001. "Wrestling Sprawl to the Ground: Defining and Measuring an Elusive Concept." *Housing Policy Debate* 12: 681-717.

Garson, G. David. 2010. "Multiple Regression." *North Carolina State University.* Retrieved on June 18, 2010 from http://faculty.chass.ncsu.edu/garson/PA765/regress.htm#concepts.

Giles-Corti, Billie, and Robert J. Donovan. 2002. "The Relative Influence of Individual, Social and Physical Environment Determinants of Physical Activity." *Social Science & Medicine* 54: 1793-1812.

Glaeser, Edward and Matthew Kahn. 2003. "Sprawl and Urban Growth." *National Bureau of Economic Research Working Paper Series.* Working Paper 9733. Cambridge, MA.

Gorin, Stephen, and Cynthia Moniz. 2004. "Will the United States Ever have Universal Health Care?" *Health & Social Work* 29: 340-344.

Gove, Walter R. 1973. "Sex, Marital Status and Mortality." *American Journal of Sociology* 79: 45-67.

Guo, Shumei Sun, Wei Wu, William Cameron Chumlea, and Alex F. Roche. 2002. "Predicting Overweight and Obesity in Adulthood from Body Mass Index Values in Childhood and Adolescence." *American Journal of Clinical Nutrition* 76: 653-658.

Healy, Lawrence M. 2006. "Logistic Regression: An Overview." *Eastern Michigan University, College of Technology.* Retrieved on May 18, 2011 from http://www.emich.edu/ia/pdf/phdresearch/Logistic%20Regression,%20Larry%20Healy%20-%20%20COT-711%20-%20%20-%20An%20Overview.pdf

Hu, Li-tze, and Peter M. Bentler. 1999. "Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria Versus New Alternatives." *Structural Equation Modeling* 6: 1-55.

Jackson, C., D. E. Jatulis, and S. P. Fortmann. 1992. "The Behavioral Risk Factor Survey and the Stanford Five-City Project Survey: A Comparison of Cardiovascular Risk Behavior Estimates." *American Journal of Public Health* 82: 412-416.

Jaret, Charles, Ravi Ghadge, Leslie Williams Reid, and Robert M. Adelman. 2009. "Measurements of Suburban Sprawl: An Evaluation." *City & Community* 8: 65-84.

Jeffery, R.W., S.A. French, J.L. Forster, and V.M. Spry. 1991. "Socioeconomic Status Differences in Health Behaviors Related to Obesity: The Healthy Worker Project." *International Journal of Obesity* 15: 689-696.

Judd, C. M. and Kenny D. A. 1981. "Process Analysis: Estimating Mediation in Treatment Evaluations." *Evaluation Review* 5: 602-619.

Kelly-Schwartz, Alexia C., Jean Stockard, Scott Doyle, and Marc 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: 184-196.

Krieger, Nancy. 1994. "Epidemiology and the Web of Causation: Has Anyone Seen the Spider?" *Society and Medicine* 39: 887-903.

Kuczmarski, Robert J., Katherine M. Flegal, Stephen M. Campbell, and Clifford L. Johnson. 1994. "Increasing Prevalence of Overweight Among US Adults: The National Health and Nutrition Examination Surveys, 1960 – 1991." *Journal of the American Medical Association* 272: 205-211.

Lindstrom, Matthew J., and Hugh Bartling. 2003. *Suburban Sprawl: Culture, Theory, and Politics.* Rowman & Littlefield.

Lopez, Russ. 2004. "Urban Sprawl and Risk for Being Overweight or Obese." *American Journal of Public Health* 94: 1574-1579.

Lopez, Russ, and H. Patricia Hynes. 2003. "Sprawl in the 1990s: Measurement, Distribution, and Trends." *Urban Affairs Review* 38: 325-355.

MacKinnon, David P. 2008. *Introduction to Statistical Mediation Analysis.* CRC Press.

MacLennan, Carol. 2004. "Smart Growth for Community Development." *Journal of Law, Medicine, & Ethics, Winter 2004 Supplement* 32: 28-30.

Marcoulides, George A., and Randall E. Schumacker. 2001. *New Developments and Techniques in Structural Equation Modeling.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Massey, Douglas S. and Nancy A. Denton. 1988. "The Dimensions of Residential Segregation." *Social Forces* 67: 281-315.

McCann, Barbara A., and Reid Ewing. 2003. "Measuring the Health Effects of Sprawl, A National Analysis of Physical Activity, Obesity, and Chronic Disease." *Smart Growth America, Surface Transportation Policy Project*, September 2003.

McNeill, Lorna Haughton, Matthew W. Kreuter, and S. V. Subramanian. 2006. "Social Environment and Physical Activity: A Review of Concepts and Evidence." *Social Science & Medicine* 63: 1011-1022.

Menard, Scott. 1995. "Applied Logistic Regression." *Sage Publications Series: Quantitative Applications in the Social Sciences, No 106*.

Mieszkowski, Peter, and Edwin S. Mills. 1993. "The Causes of Metropolitan Suburbanization." *Journal of Economic Perspectives* 7: 135-147.

Mokdad, Ali H., Mary K. Serdula, William H. Dietz, Barbara A. Bowman, James S. Marks, and Jeffrey P. Koplan. 1999. "The Spread of the Obesity Epidemic in the United States, 1991 – 1998." *Journal of the American Medical Association* 282: 1519-1522.

Mokdad, Ali H., Barbara A. Bowman, Earl S. Ford, Frank Vinicor, James S. Marks, and Jeffrey P. Koplan. 2001. "The Continuing Epidemics of Obesity and Diabetes in the United States." *Journal of the American Medical Association* 286: 1195-1200.

Muthen, L. K., and B. O. Muthen. 2007. M*plus User's Guide. Fifth Edition.* Los Angeles, CA: Muthen & Muthen.

Muthen, Bengt O., and Albert Satorra. 1995. "Complex Sample Data in Structural Equation Modeling." *Sociological Methodology* 59: 216-316.

National Institute of Health. 1998. *Clinical Guidelines on the Identification, Evaluation, and Treatment of Overweight and Obesity in Adults.* Bethesda, MD: Department of Health and Human Services, NIH, National Heart, Lung, and Blood Institute.

Nelson, D. E., D. Holtzman, J. Bolen, C. A. Stanwyck, and K. A. Mack. 2001. "Reliability and Validity of Measures from the Behavioral Risk Factor Surveillance System (BRFSS). *Social and Preventive Medicine* 46 Supplement 1: S03-S42.

Nelson, David E., Eve Powell-Griner, Machell Town, and Mary Grace Kovar. 2003. "A Comparison of National Estimates from the National Health Interview Survey and the Behavioral Risk Factor Surveillance System." *American Journal of Public Health* 93: 1335-1341.

O'Brien, Rosaleen, Kate Hunt, and Graham Hart. 2005. "'It's Caveman Stuff, but that is to a Certain Extent How Guys Still Operate': Men's Accounts of Masculinity and Help Seeking." *Social Science & Medicine* 61: 503-516.

O'Toole, Randal. 2002a. "Special Interests Run with Faulty Obesity Data." *Cascade Commentary*.

O'Toole, Randal. 2002b. "The Myth of the Fat Suburbanites." *The Thoreau Institute.* Retrieved November 17, 2008 from http://www.ti.org/vaupdate25.html.

Paeratakul, S., J. C. Lovejoy, D. H. Ryan, and G. A. Bray. 2002. "The Relation of Gender, Race, and Socioeconomic Status to Obesity and Obesity Comorbidities in a Sample of US Adults." *International Journal of Obesity* 26: 1205-1210.

Perdue, Wendy Collins. 2004. "Smart Growth for Community Development." *Journal of Law, Medicine, & Ethics, Winter 2004 Supplement* 32: 27-28.

Perkins, Molly M., Robert M. Adelman, Carolyn Furlow, W. Mark Sweatman, and Jim Baird. 2010. "Staff Turnover in Assisted Living: A Multilevel Analysis." Pp. 195-219 in *Frontline Workers in Assisted Living*, edited by Mary M. Ball, Molly M. Perkins, Carole Hollingsworth, and Candace L. Kemp. Baltimore: The Johns Hopkins University Press.

Raykov, Tenko, and George A. Marcoulides. 2006. *A First Course in Structural Equation Modeling.* Routledge.

Razin, Eran. 1998. "Policies to Control Urban Sprawl: Planning Regulations or Changes in the 'Rules of the Game'?" *Urban Studies* 35: 321-340.

Rigdon, Edward E. 1996. "CFI and RMSEA: A Comparison of Two Fit Indexes for Structural Equation Modeling." *Structural Equation Modeling: A Multidisciplinary Journal* 3: 369-379.

Rowland, Michael L. 1990. "Self-Reported Weight and Height." *American Journal of Clinical Nutrition* 52: 1125-1133.

Saris, W. H. M., S. N. Blair, M. A. van Baak, S. B. Eaton, P. S. W. Davies, L. DiPietro, M. Fogelholm, A. Rissanen, D. Schoeller, B. Swinburn, A. Tremblay, K. R. Westerterp, and H. Wyatt. 2003. "How Much Physical Activity is Enough to Prevent Unhealthy Weight Gain? Outcome of the IASO 1st Stock Conference and Consensus Statement." *The International Association for the Study of Obesity* 4: 101-114.

StatSoft. 2010. *StatSoft Electronic Statistics Textbook.* Retrieved June 18, 2010 from http://www.statsoft.com/textbook/multiple-regression/#index.

Stroebe, Margaret S. and Wolfgang Stroebe. 1983. "Who Suffers More? Sex Differences in Health Risks of the Widowed." *Psychological Bulletin* 93: 279-301.

Strunk, Bradley C., and James D. Reschovsky. 2004. *Trends in U.S. Health Insurance Coverage, 2001 – 2003.* Center for Studying Health System Change (Tracking Report No. 9). Retrieved June 27, 2007, from http://www.hschange.org/CONTENT/694/.

Subramanian, S.V., and Ichiro Kawachi. 2006. "Being Well and Doing Well: On the Importance of Income for Health." *International Journal of Social Welfare* 15: S13-S22.

Sund, Reijo. 2003. "Utilisation of Administrative Registers Using Scientific Knowledge Discovery." *Intelligent Data Analysis* 7: 501-519.

Tasmin, Rosmaini, and Peter Woods. 2008. "Knowledge Management and Innovation in Peninsular Malaysia." *Knowledge Management International Conference.* Paper.

Timms, Duncan W. G. 1971. *The Urban Mosaic: Towards a Theory of Residential Differentiation.* Cambridge University Press.

Umberson, Debra. 1992. "Gender, Marital Status and the Social Control of Health Behavior." *Social Science & Medicine* 34: 907-917.

U.S. Census Bureau. 2000. *Census 2000 Region Classifications.* Retrieved May 21, 2009 from http://www.census.gov.

U.S. Department of Transportation. 1999. *National Transportation Statistics.* U.S. Government Printing Office.

U.S. Department of Transportation. 2004. *2001 National Household Travel Survey.* U.S. Government Printing Office.

Wassmer, Robert W. 2002. "Fiscalisation of Land Use, Urban Growth Boundaries and Non-central Retail Sprawl in the Western United States." *Urban Studies* 39: 1307-1327.

Wheaton, B. 1987. "Assessment of Fit in Overidentified Models with Latent Variables." *Sociological Methods and Research* 16: 118-154.

Wolman, Harold, George Galster, Royce Hanson, Michael Ratcliffe, Kimberly Furdell, and Andrea Sarzynski. 2005. "The fundamental Challenge in Measuring Sprawl: Which Land Should Be Considered?" *The Professional Geographer* 57: 94-105.

APPENDIX A

LIST OF COUNTY-LEVEL SPRAWL INDEX, FIPS, AND REGION CLASSIFICATION

From: "Measuring the Health Effects of Sprawl" McCann and Ewing (2003)
and U. S. Census (2000)

Counties listed in bold italic were not included in analyses due to the fact that they were not included in the BRFSS database.

State/County	FIPS	Metropolitan Area	Original Sprawl Score
ALABAMA (South)			
Baldwin	01003	Mobile, AL	83.16
Jefferson	01073	Birmingham, AL	108.45
Mobile	01097	Mobile, AL	98.85
Shelby	01117	Birmingham, AL	87.16
St. Clair	01115	Birmingham, AL	83.76
ARIZONA (West)			
Maricopa	04013	Phoenix-Mesa, AZ	111.51
Pima	04019	Tucson, AZ	101.73
ARKANSAS (South)			
Crittenden	05035	Memphis, TN-AR-MS	94.07
Faulkner	05045	Little Rock-North Little Rock, AR	83.45
Lonoke	05085	Little Rock-North Little Rock, AR	81.22
Pulaski	05119	Little Rock-North Little Rock, AR	108.04
Saline	05125	Little Rock-North Little Rock, AR	82.00
CALIFORNIA (West)			
Alameda	06001	Oakland, CA	136.64
Contra Costa	06013	Oakland, CA	115.77
El Dorado	***06017***	***Sacramento, CA***	***85.67***
Fresno	06019	Fresno, CA	98.02
Kern	06029	Bakersfield, CA	95.07
Los Angeles	06037	Los Angeles-Long Beach, CA	141.74
Marin	06041	San Francisco, CA	111.80
Napa	***06055***	***San Francisco-Oakland-San Jose, CA***	***107.01***
Orange	06059	Orange County, CA	131.74
Placer	***06061***	***Sacramento, CA***	***95.58***
Riverside	06065	Riverside-San Bernardino, CA	101.34
Sacramento	06067	Sacramento, CA	116.35
San Bernardino	06071	Riverside-San Bernardino, CA	100.49

State/County	FIPS	Metropolitan Area	Sprawl Score
San Diego	06073	San Diego, CA	119.73
San Francisco	06075	San Francisco, CA	209.27
San Joaquin	06077	Stockton-Lodi, CA	110.94
San Mateo	06081	San Francisco, CA	132.09
Santa Barbara	06083	Santa Barbara-Santa Maria-Lompoc, CA	115.84
Santa Clara	06085	San Jose, CA	127.28
Santa Cruz	***06087***	***San Francisco-Oakland-San Jose, CA***	***111.63***
Solano	***06095***	***San Francisco-Oakland-San Jose, CA***	***110.41***
Sonoma	06097	San Francisco-Oakland-San Jose, CA	101.84
Stanislaus	06099	Modesto, CA	109.91
Ventura	06111	Los Angeles-Riverside-Orange County, CA	112.72
Yolo	***06113***	***Sacramento-Yolo, CA***	***105.60***
COLORADO (West)			
Adams	08001	Denver, CO	125.56
Arapahoe	08005	Denver, CO	114.56
Boulder	08013	Denver-Boulder-Greeley, CO	108.15
Douglas	08035	Denver, CO	94.83
El Paso	08041	Colorado Springs, CO	105.29
Jefferson	08059	Denver, CO	112.59
CONNECTICUT (Northeast)			
Fairfield	09001	New York-Northern New Jersey-Long Island, NY-NJ-CT-PA	107.34
Hartford	09003	Hartford, CT	102.75
Middlesex	09007	Hartford, CT	90.28
New Haven	09009	New York-Northern New Jersey-Long Island, NY-NJ-CT-PA	107.10
Tolland	09013	Hartford, CT	81.77
DELAWARE (South)			
New Castle		Philadelphia-Wilmington-Atlantic City, PA-NJ-DE-MD	114.83
FLORIDA (South)			
Brevard	12009	Melbourne-Titusville-Palm Bay, FL	104.77
Broward	12011	Fort Lauderdale, FL	127.01
Clay	12019	Jacksonville, FL	87.51
Dade	12086	Miami, FL	136.17
Duval	12031	Jacksonville, FL	114.74
Hernando	12053	Tampa-St Petersburg-Clearwater, FL	94.97
Hillsborough	12057	Tampa-St Petersburg-Clearwater, FL	114.31
Nassau	***12089***	***Jacksonville, FL***	***80.31***

State/County	FIPS	Metropolitan Area	Sprawl Score
Orange	12095	Orlando, FL	115.05
Osceola	12097	Orlando, FL	105.55
Palm Beach	12099	West Palm Beach-Boca Raton, FL	110.38
Pasco	12101	Tampa-St Petersburg-Clearwater, FL	110.22
Pinellas	12103	Tampa-St Petersburg-Clearwater, FL	126.64
Polk	12105	Lakeland-Winter Haven, FL	105.24
Seminole	12117	Orlando, FL	112.13
St. Johns	***12109***	***Jacksonville, FL***	***99.33***
Volusia	12127	Daytona Beach, FL	104.77
GEORGIA (South)			
Barrow	***13013***	***Atlanta, GA***	***77.67***
Butts	***13035***	***Atlanta, GA***	***78.71***
Catoosa	***13047***	***Chattanooga, TN-GA***	***85.61***
Cherokee	13057	Atlanta, GA	85.22
Clayton	13063	Atlanta, GA	99.61
Cobb	13067	Atlanta, GA	101.01
Columbia	13073	Augusta-Aiken, GA-SC	87.30
Coweta	***13077***	***Atlanta, GA***	***80.87***
Dade	***13083***	***Chattanooga, TN-GA***	***77.61***
De Kalb	13089	Atlanta, GA	103.94
Douglas	***13097***	***Atlanta, GA***	***80.29***
Fayette	***13113***	***Atlanta, GA***	***75.74***
Forsyth	13117	Atlanta, GA	72.04
Fulton	13121	Atlanta, GA	105.46
Gwinnett	13135	Atlanta, GA	93.76
Henry	13151	Atlanta, GA	74.13
McDuffie	***13189***	***Augusta-Aiken, GA-SC***	***78.00***
Newton	***13217***	***Atlanta, GA***	***79.44***
Paulding	***13223***	***Atlanta, GA***	***82.10***
Richmond	13245	Augusta-Aiken, GA-SC	102.47
Rockdale	***13247***	***Atlanta, GA***	***82.82***
Spalding	***13255***	***Atlanta, GA***	***85.19***
Walker	***13295***	***Chattanooga, TN-GA***	***81.26***
Walton	13297	Atlanta, GA	69.61
HAWAII (West)			
Honolulu	15003	Honolulu, HI	126.76
ILLINOIS (Midwest)			
Clinton	***17027***	***St Louis, MO-IL***	***86.69***
Cook	17031	Chicago, IL	150.15
Du Page	17043	Chicago, IL	114.64
Grundy	***17063***	***Chicago, IL***	***87.74***

State/County	FIPS	Metropolitan Area	Sprawl Score
Kane	17089	Chicago, IL	108.53
Kendall	***17093***	***Chicago, IL***	***90.37***
Lake	17097	Chicago, IL	108.92
Madison	17119	St Louis, MO-IL	102.62
McHenry	17111	Chicago, IL	100.08
Monroe	***17133***	***St Louis, MO-IL***	***85.64***
St. Clair	17163	St Louis, MO-IL	104.41
Will	17197	Chicago, IL	98.81
INDIANA (Midwest)			
Allen	18003	Fort Wayne, IN	97.44
Boone	18011	Indianapolis, IN	78.10
Clark	18019	Louisville, KY-IN	102.63
De Kalb	***18033***	***Fort Wayne, IN***	***80.34***
Dearborn	***18029***	***Cincinnati, OH-KY-IN***	***84.42***
Floyd	18043	Louisville, KY-IN	97.71
Hamilton	18057	Indianapolis, IN	93.90
Hancock	***18059***	***Indianapolis, IN***	***82.44***
Harrison	***18061***	***Louisville, KY-IN***	***74.37***
Hendricks	18063	Indianapolis, IN	85.56
Johnson	18081	Indianapolis, IN	96.44
Lake	18089	Chicago-Gary-Kenosha, IL-IN-WI	110.99
Marion	18097	Indianapolis, IN	113.13
Morgan	18109	Indianapolis, IN	88.06
Porter	18127	Chicago-Gary-Kenosha, IL-IN-WI	94.09
Shelby	***18145***	***Indianapolis, IN***	***88.18***
Whitley	***18183***	***Fort Wayne, IN***	***72.38***
IOWA (Midwest)			
Dallas	19049	Des Moines, IA	81.68
Polk	19153	Des Moines, IA	105.34
Pottawattamie	19155	Omaha, NE-IA	93.73
Warren	19181	Des Moines, IA	79.86
KANSAS (Midwest)			
Butler	20015	Wichita, KS	79.81
Harvey	20079	Wichita, KS	73.89
Johnson	20091	Kansas City, MO-KS	103.50
Leavenworth	20103	Kansas City, MO-KS	94.46
Miami	20121	Kansas City, MO-KS	71.03
Sedgwick	20173	Wichita, KS	106.18
Wyandotte	20209	Kansas City, MO-KS	111.25

State/County	FIPS	Metropolitan Area	Sprawl Score
KENTUCKY (South)			
Boone	21015	Cincinnati, OH-KY-IN	92.11
Bullitt	***21029***	***Louisville, KY-IN***	***86.26***
Campbell	21037	Cincinnati, OH-KY-IN	109.57
Jefferson	21111	Louisville, KY-IN	110.08
Kenton	21117	Cincinnati, OH-KY-IN	108.82
Oldham	***21185***	***Louisville, KY-IN***	***84.27***
LOUISIANA (South)			
Ascension	22005	Baton Rouge, LA	87.25
East Baton Rouge	22033	Baton Rouge, LA	105.67
Jefferson	22051	New Orleans, LA	124.23
Livingston	22063	Baton Rouge, LA	82.76
Orleans	22071	New Orleans, LA	149.47
St. Bernard	22087	New Orleans, LA	113.59
St. Charles	***22089***	***New Orleans, LA***	***91.16***
St. John the Baptist	***22095***	***New Orleans, LA***	***98.13***
St. Tammany	22103	New Orleans, LA	96.48
West Baton Rouge	***22121***	***Baton Rouge, LA***	***91.68***
MARYLAND (South)			
Anne Arundel	24003	Baltimore, MD	107.75
Baltimore	24005	Baltimore, MD	107.02
Baltimore city	24510	Baltimore, MD	162.76
Calvert	24009	Washington, DC-MD-VA-WV	90.84
Carroll	24013	Baltimore, MD	81.92
Cecil	24015	Philadelphia-Wilmington-Atlantic City, PA-NJ-DE-MD	86.87
Charles	24017	Washington, DC-MD-VA-WV	89.72
Frederick	24021	Washington, DC-MD-VA-WV	87.09
Harford	24025	Baltimore, MD	92.47
Howard	24027	Baltimore, MD	93.65
Montgomery	24031	Washington, DC-MD-VA-WV	112.70
Prince George's	24033	Washington, DC-MD-VA-WV	112.42
Queen Anne's	24035	Baltimore, MD	77.24
MASSACHUSETTS (Northeast)			
Bristol	25005	Boston, MA-NH	113.62
Essex	25009	Boston, MA-NH	118.56
Hampden	25013	Springfield, MA NECMA	108.58
Hampshire	25015	Springfield, MA NECMA	87.77
Middlesex	25017	Boston, MA-NH	121.56
Norfolk	25021	Boston, MA-NH	113.64
Plymouth	25023	Boston, MA-NH	100.26

State/County	FIPS	Metropolitan Area	Sprawl Score
Suffolk	25025	Boston, MA-NH	179.37
Worcester	25027	Boston, MA-NH	99.44
MICHIGAN (Midwest)			
Bay	26017	Saginaw-Bay City-Midland, MI	95.10
Clinton	***26037***	***Lansing-East Lansing, MI***	***66.63***
Eaton	***26045***	***Lansing-East Lansing, MI***	***77.36***
Genesee	26049	Detroit-Ann Arbor-Flint, MI	99.04
Ingham	26065	Lansing-East Lansing, MI	103.26
Kent	26081	Grand Rapids-Muskegon-Holland, MI	98.44
Lapeer	***26087***	***Detroit, MI***	***71.56***
Livingston	26093	Detroit-Ann Arbor-Flint, MI	82.72
Macomb	26099	Detroit, MI	107.27
Midland	***26111***	***Saginaw-Bay City-Midland, MI***	***82.25***
Monroe	26115	Detroit, MI	83.45
Oakland	26125	Detroit, MI	105.71
Ottawa	26139	Grand Rapids-Muskegon-Holland, MI	87.36
Saginaw	26145	Saginaw-Bay City-Midland, MI	96.22
St.Clair	26147	Detroit, MI	88.30
Washtenaw	26161	Detroit-Ann Arbor-Flint, MI	99.27
Wayne	26163	Detroit, MI	123.22
MINNESOTA (Midwest)			
Anoka	27003	Minneapolis-St Paul, MN-WI	95.92
Carver	27019	Minneapolis-St Paul, MN-WI	85.66
Chisago	***27025***	***Minneapolis-St Paul, MN-WI***	***79.39***
Dakota	27037	Minneapolis-St Paul, MN-WI	98.09
Hennepin	27053	Minneapolis-St Paul, MN-WI	119.74
Isanti	***27059***	***Minneapolis-St Paul, MN-WI***	***70.12***
Ramsey	27123	Minneapolis-St Paul, MN-WI	123.09
Scott	***27139***	***Minneapolis-St Paul, MN-WI***	***90.36***
Washington	27163	Minneapolis-St Paul, MN-WI	96.80
Wright	27171	Minneapolis-St Paul, MN-WI	79.85
MISSISSIPPI (South)			
De Soto	28033	Memphis, TN-AR-MS	82.01
Hinds	28049	Jackson, MS	97.30
Madison	28089	Jackson, MS	80.16
Rankin	28121	Jackson, MS	81.66
MISSOURI (Midwest)			
Cass	***29037***	***Kansas City, MO-KS***	***83.70***
Clay	29047	Kansas City, MO-KS	98.39
Franklin	***29071***	***St Louis, MO-IL***	***83.89***

State/County	FIPS	Metropolitan Area	Sprawl Score
Jackson	29095	Kansas City, MO-KS	113.43
Jefferson	29099	St Louis, MO-IL	93.97
Lafayette	***29107***	***Kansas City, MO-KS***	***85.00***
Platte	***29165***	***Kansas City, MO-KS***	***90.26***
Ray	***29177***	***Kansas City, MO-KS***	***74.60***
St. Charles	29183	St Louis, MO-IL	108.45
St. Louis	29189	St Louis, MO-IL	118.14
NEBRASKA (Midwest)			
Douglas	31055	Omaha, NE-IA	118.10
Sarpy	31153	Omaha, NE-IA	101.76
Washington	31177	Omaha, NE-IA	76.51
NEVADA (West)			
Clark	32003	Las Vegas, NV-AZ	114.46
NEW JERSEY (Northeast)			
Bergen	34003	Bergen-Passaic, NJ	130.41
Burlington	34005	Philadelphia, PA-NJ	101.58
Camden	34007	Philadelphia, PA-NJ	123.76
Cumberland	34011	Philadelphia-Wilmington-Atlantic City, PA-NJ-DE-MD	94.16
Essex	34013	Newark, NJ	152.13
Gloucester	34015	Philadelphia, PA-NJ	101.53
Hudson	34017	New York-Northern New Jersey-Long Island, NY-NJ-CT-PA	190.06
Hunterdon	34019	Middlesex-Somerset-Hunterdon, NJ	81.28
Mercer	34021	New York-Northern New Jersey-Long Island, NY-NJ-CT-PA	116.27
Middlesex	34023	Middlesex-Somerset-Hunterdon, NJ	121.71
Monmouth	34025	Monmouth-Ocean, NJ	110.86
Morris	34027	Newark, NJ	101.20
Ocean	34029	Monmouth-Ocean, NJ	112.14
Passaic	34031	Bergen-Passaic, NJ	140.41
Salem	34033	Philadelphia, PA-NJ	89.71
Somerset	34035	Middlesex-Somerset-Hunterdon, NJ	97.01
Sussex	34037	Newark, NJ	91.77
Union	34039	Newark, NJ	136.13
Warren	34041	Newark, NJ	96.66
NEW MEXICO (West)			
Bernalillo	35001	Albuquerque, NM	112.10

State/County	FIPS	Metropolitan Area	Sprawl Score
NEW YORK (Northeast)			
Albany	36001	Albany-Schenectady-Troy, NY	105.13
Bronx	36005	New York, NY	250.72
Erie	36029	Buffalo-Niagara Falls, NY	106.81
Kings	36047	New York, NY	263.65
Livingston	***36051***	***Rochester, NY***	***76.56***
Madison	***36053***	***Syracuse, NY***	***75.57***
Monroe	36055	Rochester, NY	103.62
Montgomery	***36057***	***Albany-Schenectady-Troy, NY***	***89.71***
Nassau	36059	Nassau-Suffolk, NY	136.56
New York	36061	New York, NY	352.07
Niagara	36063	Buffalo-Niagara Falls, NY	98.52
Onondaga	36067	Syracuse, NY	101.73
Ontario	36069	Rochester, NY	79.85
Orange	36071	New York-Northern New Jersey-Long Island, NY-NJ-CT-PA	98.10
Orleans	***36073***	***Rochester, NY***	***79.66***
Oswego	***36075***	***Syracuse, NY***	***83.93***
Putnam	***36079***	***New York, NY***	***92.69***
Queens	36081	New York, NY	218.90
Rensselaer	36083	Albany-Schenectady-Troy, NY	99.04
Richmond	36085	New York, NY	162.89
Rockland	36087	New York, NY	110.19
Saratoga	36091	Albany-Schenectady-Troy, NY	88.90
Schenectady	36093	Albany-Schenectady-Troy, NY	108.87
Suffolk	36103	Nassau-Suffolk, NY	109.88
Wayne	***36117***	***Rochester, NY***	***74.63***
Westchester	36119	New York, NY	128.37
NORTH CAROLINA (South)			
Cabarrus	37025	Charlotte-Gastonia-Rock Hill, NC-SC	89.47
Davidson	37057	Greensboro-Winston-Salem-High Point, NC	85.42
Davie	***37059***	***Greensboro-Winston-Salem-High Point, NC***	***70.99***
Durham	37063	Raleigh-Durham-Chapel Hill, NC	99.12
Forsyth	37067	Greensboro-Winston-Salem-High Point, NC	96.58
Franklin	37069	Raleigh-Durham-Chapel Hill, NC	76.50
Gaston	37071	Charlotte-Gastonia-Rock Hill, NC-SC	93.06
Guilford	37081	Greensboro-Winston-Salem-High Point, NC	97.26
Lincoln	***37109***	***Charlotte-Gastonia-Rock Hill, NC-SC***	***78.56***
Mecklenburg	37119	Charlotte-Gastonia-Rock Hill, NC-SC	96.82
Orange	37135	Raleigh-Durham-Chapel Hill, NC	86.21
Randolph	37151	Greensboro-Winston-Salem-High Point, NC	77.32
Rowan	37159	Charlotte-Gastonia-Rock Hill, NC-SC	87.22
Stokes	37169	Greensboro-Winston-Salem-High Point, NC	71.26

State/County	FIPS	Metropolitan Area	Sprawl Score
Union	***37179***	***Charlotte-Gastonia-Rock Hill, NC-SC***	***75.93***
Wake	37183	Raleigh-Durham-Chapel Hill, NC	95.89
Yadkin	37197	Greensboro-Winston-Salem-High Point, NC	69.17
OHIO (Midwest)			
Butler	39017	Cincinnati-Hamilton, OH-KY-IN	102.29
Carroll	***39019***	***Canton-Massillon, OH***	***79.10***
Clark	***39023***	***Dayton-Springfield, OH***	***96.10***
Clermont	39025	Cincinnati, OH-KY-IN	86.90
Cuyahoga	39035	Cleveland-Lorain-Elyria, OH	115.84
Delaware	***39041***	***Columbus, OH***	***81.99***
Fairfield	***39045***	***Columbus, OH***	***85.77***
Franklin	39049	Columbus, OH	116.72
Fulton	***39051***	***Toledo, OH***	***66.83***
Geauga	***39055***	***Cleveland-Lorain-Elyria, OH***	***63.12***
Greene	***39057***	***Dayton-Springfield, OH***	***91.03***
Hamilton	39061	Cincinnati, OH-KY-IN	112.45
Lake	***39085***	***Cleveland-Lorain-Elyria, OH***	***96.84***
Licking	***39089***	***Columbus, OH***	***84.56***
Lorain	39093	Cleveland-Lorain-Elyria, OH	94.50
Lucas	39095	Toledo, OH	111.48
Madison	***39097***	***Columbus, OH***	***83.00***
Mahoning	39099	Youngstown-Warren, OH	98.13
Medina	***39103***	***Cleveland-Lorain-Elyria, OH***	***76.59***
Miami	***39109***	***Dayton-Springfield, OH***	***86.81***
Montgomery	39113	Dayton-Springfield, OH	108.47
Pickaway	***39129***	***Columbus, OH***	***84.74***
Portage	***39133***	***Cleveland-Akron, OH***	***83.97***
Stark	39151	Canton-Massillon, OH	106.62
Summit	39153	Cleveland-Akron, OH	106.62
Trumbull	***39155***	***Youngstown-Warren, OH***	***93.59***
Warren	***39165***	***Cincinnati, OH-KY-IN***	***89.95***
Wood	***39173***	***Toledo, OH***	***84.24***
OKLAHOMA (South)			
Canadian	40017	Oklahoma City, OK	81.11
Cleveland	40027	Oklahoma City, OK	95.07
Creek	40037	Tulsa, OK	91.30
Logan	40083	Oklahoma City, OK	80.83
McClain	40087	Oklahoma City, OK	79.97
Oklahoma	40109	Oklahoma City, OK	106.31
Osage	40113	Tulsa, OK	98.63
Pottawatomie	40125	Oklahoma City, OK	88.26
Rogers	40131	Tulsa, OK	87.03

State/County	FIPS	Metropolitan Area	Sprawl Score
Tulsa	40143	Tulsa, OK	108.64
Wagoner	40145	Tulsa, OK	88.89
OREGON (West)			
Clackamas	41005	Portland-Vancouver, OR-WA	98.45
Multnomah	41051	Portland-Vancouver, OR-WA	131.41
Washington	41067	Portland-Vancouver, OR-WA	108.29
Yamhill	41071	Portland-Vancouver, OR-WA	98.23
PENNSYLVANIA (Northeast)			
Allegheny	42003	Pittsburgh, PA	120.99
Beaver	42007	Pittsburgh, PA	105.52
Bucks	42017	Philadelphia, PA-NJ	100.15
Carbon	***42025***	***Allentown-Bethlehem-Easton, PA***	***93.99***
Chester	42029	Philadelphia, PA-NJ	89.84
Columbia	***42037***	***Scranton—Wilkes-Barre—Hazleton, PA***	***92.46***
Cumberland	42041	Harrisburg-Lebanon-Carlisle, PA	97.64
Dauphin	42043	Harrisburg-Lebanon-Carlisle, PA	113.77
Delaware	42045	Philadelphia, PA-NJ	125.34
Fayette	***42051***	***Pittsburgh, PA***	***98.66***
Lackawanna	42069	Scranton—Wilkes-Barre—Hazleton, PA	111.01
Lancaster	42071	Lancaster, PA	94.09
Lebanon	***42075***	***Harrisburg-Lebanon-Carlisle, PA***	***102.33***
Lehigh	42077	Allentown-Bethlehem-Easton, PA	119.67
Luzerne	42079	Scranton—Wilkes-Barre—Hazleton, PA	107.09
Montgomery	42091	Philadelphia, PA-NJ	107.06
Northampton	42095	Allentown-Bethlehem-Easton, PA	110.65
Perry	***42099***	***Harrisburg-Lebanon-Carlisle, PA***	***82.91***
Philadelphia	42101	Philadelphia, PA-NJ	187.78
Washington	42125	Pittsburgh, PA	100.95
Westmoreland	42129	Pittsburgh, PA	100.53
Wyoming	***42131***	***Scranton—Wilkes-Barre—Hazleton, PA***	***78.64***
York	42133	York, PA	94.78
RHODE ISLAND (Northeast)			
Bristol	44001	Providence-Fall River-Warwick, RI-MA	118.66
Kent	44003	Providence-Fall River-Warwick, RI-MA	115.99
Providence	44007	Providence-Fall River-Warwick, RI-MA	130.56
Washington	44009	Providence-Fall River-Warwick, RI-MA	92.45
SOUTH CAROLINA (South)			
Aiken	45003	Augusta-Aiken, GA-SC	86.39
Berkeley	45015	Charleston-North Charleston, SC	90.12
Charleston	45019	Charleston-North Charleston, SC	110.28

State/County	FIPS	Metropolitan Area	Sprawl Score
Dorchester	45035	Charleston-North Charleston, SC	87.82
Greenville	45045	Greenville-Spartanburg-Anderson, SC	94.35
Lexington	45063	Columbia, SC	86.41
Pickens	45077	Greenville-Spartanburg-Anderson, SC	83.78
Richland	45079	Columbia, SC	101.86
Spartanburg	45083	Greenville-Spartanburg-Anderson, SC	86.73
York	45091	Charlotte-Gastonia-Rock Hill, NC-SC	84.11
TENNESSEE (South)			
Anderson	***47001***	***Knoxville, TN***	***90.20***
Blount	47009	Knoxville, TN	89.51
Carter	***47019***	***Johnson City-Kingsport-Bristol, TN-VA***	***97.93***
Cheatham	***47021***	***Nashville, TN***	***74.75***
Davidson	47037	Nashville, TN	101.17
Dickson	***47043***	***Nashville, TN***	***80.92***
Hamilton	47065	Chattanooga, TN-GA	99.83
Hawkins	***47073***	***Johnson City-Kingsport-Bristol, TN-VA***	***86.81***
Knox	47093	Knoxville, TN	99.34
Robertson	***47147***	***Nashville, TN***	***77.32***
Rutherford	47149	Nashville, TN	85.34
Sevier	***47155***	***Knoxville, TN***	***88.16***
Shelby	47157	Memphis, TN-AR-MS	103.98
Sullivan	47163	Johnson City-Kingsport-Bristol, TN-VA	93.28
Sumner	47165	Nashville, TN	87.09
Tipton	***47167***	***Memphis, TN-AR-MS***	***77.54***
Unicoi	***47171***	***Johnson City-Kingsport-Bristol, TN-VA***	***104.18***
Union	***47173***	***Knoxville, TN***	***84.39***
Washington	47179	Johnson City-Kingsport-Bristol, TN-VA	92.36
Williamson	47187	Nashville, TN	83.12
Wilson	***47189***	***Nashville, TN***	***78.67***
TEXAS (South)			
Bexar	48029	San Antonio, TX	112.72
Brazoria	48039	Houston-Galveston-Brazoria, TX	96.02
Collin	48085	Dallas, TX	101.00
Comal	***48091***	***San Antonio, TX***	***92.67***
Dallas	48113	Dallas, TX	114.55
Denton	48121	Dallas, TX	98.68
El Paso	48141	El Paso, TX	110.26
Ellis	***48139***	***Dallas, TX***	***88.64***
Fort Bend	48157	Houston, TX	100.63
Galveston	48167	Houston-Galveston-Brazoria, TX	109.98
Guadalupe	***48187***	***San Antonio, TX***	***91.01***
Harris	48201	Houston, TX	113.25

State/County	FIPS	Metropolitan Area	Sprawl Score
Hays	***48209***	***Austin-San Marcos, TX***	***88.93***
Hidalgo	48215	McAllen-Edinburg-Mission, TX	100.30
Johnson	***48251***	***Fort Worth-Arlington, TX***	***89.94***
Kaufman	***48257***	***Dallas, TX***	***88.42***
Liberty	***48291***	***Houston, TX***	***85.00***
Montgomery	48339	Houston, TX	88.10
Parker	***48367***	***Fort Worth-Arlington, TX***	***80.94***
Rockwall	***48397***	***Dallas, TX***	***90.98***
Tarrant	48439	Fort Worth-Arlington, TX	110.62
Travis	48453	Austin-San Marcos, TX	106.79
Waller	***48473***	***Houston, TX***	***94.45***
Williamson	48491	Austin-San Marcos, TX	98.61
UTAH (West)			
Davis	49011	Salt Lake City-Ogden, UT	107.27
Salt Lake	49035	Salt Lake City-Ogden, UT	114.43
Weber	49057	Salt Lake City-Ogden, UT	106.07
VIRGINIA (South)			
Chesapeake city	51550	Norfolk-Virginia Beach-Newport News, VA-NC	103.17
Chesterfield	51041	Richmond-Petersburg, VA	93.89
Dinwiddie	***51053***	***Richmond-Petersburg, VA***	***72.45***
Fairfax	51059	Washington, DC-MD-VA-WV	117.81
Gloucester	***51073***	***Norfolk-Virginia Beach-Newport News, VA-NC***	***82.82***
Goochland	***51075***	***Richmond-Petersburg, VA***	***67.59***
Hanover	51085	Richmond-Petersburg, VA	74.97
Henrico	51087	Richmond-Petersburg, VA	100.73
James city	***51095***	***Norfolk-Virginia Beach-*** Newport News, VA-NC	90.41
Loudoun	51107	Washington, DC-MD-VA-WV	94.57
New Kent	***51127***	***Richmond-Petersburg, VA***	***76.49***
Norfolk city	51710	Norfolk-Virginia Beach-Newport News, VA-NC	131.92
Portsmouth city	51740	Norfolk-Virginia Beach-Newport News, VA-NC	124.93
Powhatan	***51145***	***Richmond-Petersburg, VA***	***72.48***
Prince George	***51149***	***Richmond-Petersburg, VA***	***105.81***
Prince William	51153	Washington, DC-MD-VA-WV	99.98
Richmond city	51760	Richmond-Petersburg, VA	127.18
Scott	***51169***	***Johnson City-Kingsport-Bristol, TN-VA***	***89.47***
Stafford	51179	Washington, DC-MD-VA-WV	87.90

State/County	FIPS	Metropolitan Area	Sprawl Score
Suffolk city	51800	Norfolk-Virginia Beach-Newport News, VA-NC	89.16
Virginia Beach city	51810	Norfolk-Virginia Beach-Newport News, VA-NC	113.91
Washington	***51191***	***Johnson City-Kingsport-Bristol, TN-VA***	***94.38***
York	***51199***	***Norfolk-Virginia Beach-Newport News, VA-NC***	***110.62***
WASHINGTON (West)			
Clark	53011	Portland-Vancouver, OR-WA	103.44
King	53033	Seattle-Bellevue-Everett, WA	118.01
Pierce	53053	Seattle-Tacoma-Bremerton, WA	107.59
Snohomish	53061	Seattle-Bellevue-Everett, WA	100.73
WISCONSIN (Midwest)			
Dane	55025	Madison, WI	102.46
Kenosha	55059	Chicago-Gary-Kenosha, IL-IN-WI	106.16
Milwaukee	55079	Milwaukee-Waukesha, WI	132.54
Ozaukee	***55089***	***Milwaukee-Waukesha, WI***	***88.43***
Racine	55101	Milwaukee-Racine, WI	103.10
St. Croix	***55109***	***Minneapolis-St Paul, MN-WI***	***76.22***
Washington	55131	Milwaukee-Waukesha, WI	80.75
Waukesha	55133	Milwaukee-Waukesha, WI	90.44

APPENDIX B

COUNTY SPRAWL INDEX VARIABLES

From: "Relationship Between Urban Sprawl and Physical Activity, Obesity, and Morbidity" Ewing et al. (2003) p. 51 and "Measuring the Health Effects of Sprawl" McCann and Ewing (2003) p. 10

Factor	Variable	Source
Residential Density	Gross population density in persons per square mile in county	U. S. Census (2000)
	Percentage of county population living at densities less than 1,500 person per square mile (low suburban density)	U. S. Census (2000)
	Percentage of county population living at densities greater than 12,500 persons per square mile (urban density that begins to be transit supportive)	U. S. Census (2000)
	Net population density of urban lands in county	USDA Natural Resources Inventory (2000)
Connectivity of the Street Network	County's average block size in square miles	Census TIGER files (2000)
	Percentage of small blocks (≤ 0.01 square mile) in county	Census TIGER files (2000)

APPENDIX C

STANDARDIZED REGRESSION COEFFICIENTS

Baseline Model SEM Standardized Regression Coefficients						
	Diabetes	**Heart Attack**	**Heart Disease**	**High Cholesterol**	**Hypertension**	**Stroke**
Sprawl	0.011	0.002	-0.008	-0.014	0.013	0.020
Physical Activity	------	------	------	------	------	------
BMI	------	------	------	------	------	------
Age	0.356	0.408	0.436	0.301	0.475	0.332
Education	-0.063	-0.045	-0.025	-0.050	-0.063	-0.040
Income	-0.140	-0.137	-0.107	-0.034	-0.074	-0.190
Gender	-0.066	-0.150	-0.101	-0.057	-0.074	-0.084
Black	0.102	-0.022	-0.025	-0.031	0.092	0.025
Asian	0.011	-0.066	-0.025	-0.014	-0.009	-0.036
Native Hawaiian/Pacific Islander	0.022	-0.023	0.000	-0.006	0.000	0.019
American Indian	0.027	0.009	0.027	0.009	0.008	0.019
Multi-Racial	0.024	0.025	-0.002	-0.001	0.024	0.042
Other	0.002	0.026	0.004	-0.010	-0.002	0.004
Hispanic	0.026	-0.027	0.022	-0.029	-0.018	0.003
Widowed	-0.042	-0.048	-0.052	-0.044	0.001	-0.002
Separated/Divorced	-0.013	-0.034	-0.039	-0.002	0.018	0.046
Never Married	-0.016	-0.081	-0.084	-0.014	0.014	-0.039
Married * Gender	-0.029	-0.075	-0.095	-0.043	-0.007	0.007
Increased Physical Activity	-0.017	-0.025	-0.002	-0.022	-0.022	-0.040
South	0.015	0.064	0.054	0.011	0.019	0.080
Midwest	0.009	0.023	0.019	-0.002	0.005	0.026
West	-0.005	-0.018	-0.043	-0.012	-0.011	0.017

Model 2 SEM Standardized Regression Coefficients

	Physical Activity	BMI	Diabetes	Heart Attack	Heart Disease	High Cholesterol	Hypertension	Stroke
Sprawl	0.016	0.025	0.005	-0.001	-0.008	-0.017	0.007	0.014
Physical Activity	------	-0.064	-0.069	-0.018	-0.033	-0.026	-0.018	-0.055
BMI	------	------	0.286	0.116	0.121	0.147	0.265	0.078
Age	-0.064	0.052	0.335	0.406	0.433	0.301	0.459	0.330
Education	-0.036	-0.095	-0.039	-0.036	-0.016	-0.039	-0.040	-0.034
Income	0.045	-0.080	-0.113	-0.127	-0.095	-0.021	-0.051	-0.180
Gender	-0.088	-0.099	-0.046	-0.140	-0.090	-0.046	-0.051	-0.084
Black	-0.017	0.128	0.064	-0.035	-0.038	-0.050	0.057	0.010
Asian	-0.020	-0.066	0.028	-0.051	-0.013	-0.006	0.007	-0.027
Native Hawaiian/Pacific Islander	0.001	0.003	0.022	-0.015	0.000	-0.008	-0.001	0.018
American Indian	0.006	0.012	0.024	0.009	0.032	0.009	0.005	0.023
Multi-Racial	0.014	0.025	0.018	0.023	-0.003	-0.004	0.019	0.041
Other	0.003	0.004	0.001	0.022	0.003	-0.010	-0.003	0.004
Hispanic	-0.011	0.016	0.020	-0.038	0.025	-0.035	-0.022	0.002
Widowed	-0.005	-0.047	-0.020	-0.018	-0.018	-0.034	0.000	-0.033
Separated/Divorced	-0.007	0.019	0.011	0.044	0.050	-0.001	-0.029	-0.064
Never Married	-0.005	-0.015	0.001	-0.044	-0.043	-0.012	0.000	-0.083
Married * Gender	-0.020	-0.055	-0.015	-0.070	-0.091	-0.035	0.007	0.011
Increased Physical Activity	0.138	-0.024	0.002	-0.019	0.006	-0.013	-0.011	-0.032
South	-0.002	0.008	0.012	0.056	0.048	0.009	0.017	0.070
Midwest	-0.012	0.036	-0.002	0.019	0.013	-0.007	-0.005	0.021
West	0.028	0.010	-0.005	-0.027	-0.059	-0.012	-0.012	0.025